PRINCIPLES OF MEDICAL BIOLOGY

Edited by

E. EDWARD BITTAR

NEVILLE BITTAR

CELL CHEMISTRY AND PHYSIOLOGY: PART IV

Cell Chemistry and Physiology: Part IV

PRINCIPLES OF MEDICAL BIOLOGY
A Multi-Volume Work, Volume 4

Editors: **E. EDWARD BITTAR,** *Department of Physiology, University of Wisconsin, Madison*
NEVILLE BITTAR, *Department of Medicine, University of Wisconsin, Madison*

This Page Intentionally Left Blank

Principles of Medical Biology
A Multi-Volume Work

Edited by **E. Edward Bittar,** *Department of Physiology, University of Wisconsin, Madison and*
Neville Bittar, *Department of Medicine University of Wisconsin, Madison*

This work provides:

* A holistic treatment of the main medical disciplines. The basic sciences including most of the achievements in cell and molecular biology have been blended with pathology and clinical medicine. Thus, a special feature is that departmental barriers have been overcome.

* The subject matter covered in preclinical and clinical courses has been reduced by almost one-third without sacrificing any of the essentials of a sound medical education. This information base thus represents an integrated core curriculum.

* The movement towards reform in medical teaching calls for the adoption of an integrated core curriculum involving small-group teaching and the recognition of the student as an active learner.

* There are increasing indications that the traditional education system in which the teacher plays the role of expert and the student that of a passive learner is undergoing reform in many medical schools. The trend can only grow.

* Medical biology as the new profession has the power to simplify the problem of reductionism.

* Over 700 internationally acclaimed medical scientists, pathologists, clinical investigators, clinicians and bioethicists are participants in this undertaking.

This Page Intentionally Left Blank

Cell Chemistry and Physiology: Part IV

Edited by **E. EDWARD BITTAR**
Department of Physiology
University of Wisconsin
Madison, Wisconsin

NEVILLE BITTAR
Department of Medicine
University of Wisconsin
Madison, Wisconsin


JAI PRESS INC.

Greenwich, Connecticut *London, England*

Library of Congress Cataloging-in-Publication Data

Cell chemistry and physiology / edited by E. Edward Bittar, Neville Bittar.
p. cm.—(Principles of medical biology ; v. 4)
Includes index.
ISBN 1-55938-807-6
1. Cytochemistry. 2. Cell physiology. I. Bittar, E. Edward.
II. Bittar, Neville. III. Series.
[DNLM: 1. Cells—chemistry. 2. Cells—physiology. QH 581.2 C392 1996]
QH611.C4214 1996
574.87—cd20
for Library of Congress 94-37215
CIP

55 Old Post Road, No. 2
Greenwich, Connecticut 06836

JAI PRESS LTD.
The Courtyard
29 High Street
Hampton Hill, Middlesex TW12 1PD
England

ISBN: 1-55938-808-0
Library of Congress Catalog No.: 95-33561

Manufactured in the United States of America

CONTENTS

LIST OF CONTRIBUTORS

Louise V.B. Anderson
Regional Neurosciences Centre
Newcastle General Hospital
Newcastle upon Tyne, England

Lloyd Barr
Department of Physiology and Biophysics
University of Illinois at Urbana
Urbana, Illinois

Kate M.D. Bushby
Department of Human Genetics
Newcastle General Hospital
Newcastle upon Tyne, England

Marie-France Carlier
Laboratoire d'Enzymologie
Centre National de la Reserche
Scientifique
Cedex, France

Primal de Lanerolle
Department of Physiology and Biophysics
University of Illinois at Chicago
Chicago, Illinois

John B. Harris
Regional Neurosciences Centre
Newcastle General Hospital
Newcastle upon Tyne, England

Eric Hultman
Department of Clinical Chemistry
Huddinge University Hospital
Huddinge, Sweden

Calvin Johnson
Department of Anesthesiology
Hutzel Hospital
Wayne State University
Detroit, Michigan

Margaret A. Johnson
Regional Neurosciences Centre
Newcastle General Hospital
Newcastle upon Tyne, England

Donald A. Jurivich
Department of Medicine
Northwestern University
Chicago, Illinois

Sujatha Kotamraju
Department of Anesthesiology
Hutzel Hospital
Wayne State University
Detroit, Michigan

Donald D. Lorimer
Department of Physiology and Biophysics
University of Illinois at Chicago
Chicago, Illinois

Sutherland Maciver
Department of Biochemistry
University of Edinburgh
Edinburgh, Scotland

S.K. Malhotra
Departments of Zoology and Pathology
University of Alberta
Edmonton, Alberta, Canada

Peter Mazur
Biology Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee

Dominique Pantaloni
Laboratoire d'Enzymologie
Centre National de la Reserche
Scientifique
Cedex, France

Michelle Peckham
Division of Biomolecular Sciences
Biophysics Section
King's College
London, England

H.S.A. Sherratt
Division of Clinical Neuroscience
The Medical School
Newcastle upon Tyne, England

T.K. Shnitka
Departments of Zoology and Pathology
University of Alberta
Edmonton, Alberta, Canada

James A. Southard
Department of Surgery
University of Wisconsin
Madison, Wisconsin

Lawrence L. Spriet
School of Human Biology
University of Guelph
Guelph, Ontario, Canada

Mary L. Taub
Department of Biochemistry
State University of New York at Buffalo
Buffalo, New York

D.M. Turnbull
Division of Clinical Neuroscience
The Medical School
Newcastle upon Tyne, England

Alan Weeds
MRC Laboratory of Molecular Biology
University of Cambridge
Cambridge, England

This Page Intentionally Left Blank

PREFACE

This volume is intended to complete the Cell Chemistry and Physiology module. It is about how the traditional boundaries of cell chemistry and physiology are being erased by molecular biology. We do not think it necessary to elaborate on this theme, particularly since the body of core knowledge found in this volume brings us a stage closer to answering the question, "what makes cell biology into a new discipline?"

The first part of the volume deals with the chemistry of actin and myosin and is followed by chapters on cell motility, ATP synthesis in muscle, and contraction in smooth and skeletal muscle. Here the reader is immediately made aware of the contributions molecular biology is making to our understanding of the molecular mechanisms underlying muscle contraction. It is perhaps enough to point out that Huxley's concept of the cross-bridge cycle and generation of force can now be explained in molecular terms. Topics such as muscle fatigue and muscle disorders, as well as malignant hyperthermia are bound to arouse active learning in the student and set the stage for problem-based learning.

Most medical students look askance at thermobiology. We think this is a mistake; hence, we have included a section dealing with this subject. This brings us to the chapter on the heat shock response which at the very outset makes clear that many stressors besides heat are known to result in heat shock gene expression. Many of the heat shock proteins occur in unstressed cells and some of them behave as chaperones. These proteins also reach high levels in a wide range of diseases

including neurodegenerative disorders. Whether certain diseases are the result of mutations in the heat shock genes is not yet known. As will be appreciated, much of the work done in this field involves the use of cultured cells. Animal cells in culture is the subject of the last chapter.

As editors we have to thank the authors who have kindly contributed chapters to this volume. We also thank Ms. Lauren Manjoney and staff members of JAI Press for their courtesy and skill.

E. EDWARD BITTAR
NEVILLE BITTAR

Chapter 1

The Cytoskeleton—Microtubules and Microfilaments: A Biological Perspective

S.K. MALHOTRA and T.K. SHNITKA

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 1–41.

ISBN: 1-55938-808-0

INTRODUCTION

General Comments

The cytoplasm of all eukaryotic cells contains a cytoskeletal framework that serves a multitude of dynamic functions exemplified by the control of cell shape, the internal positioning and movement of organelles, and the capacity of the cell to move and undergo division.

The major types of cytoskeletal filaments are 7-nm-thick microfilaments, 25-nm-thick microtubules, and 10-nm-thick intermediate filaments (IFs). These are respectively composed of actin, tubulin, and a variety of interrelated sparsely soluble fibrous proteins termed intermediate filament proteins. In addition, thick myosin filaments are present in large numbers in skeletal and heart muscle cells and in small numbers in many other types of eukaryotic cells.

The cytoskeleton also contains different accessory proteins, which, in accordance with their affinities and functions, are designated as microtubule-associated proteins (MAPs), actin-binding proteins (ABPs), intermediate-filament-associated proteins (IFAPs), and myosin-binding proteins. This chapter is focused on those parts of the cytoskeleton that are composed of microfilaments and microtubules and their associated proteins. The subject of intermediate filaments is dealt with in detail in Volume 2.

A Brief History of Cytoskeletal Studies

Historically, the notion that the cytoplasm contains a cytoskeleton is both old and new (Porter, 1984; Bershadsky and Vasiliev, 1988). The mitotic spindle was the first dynamic fibrillary structure to be discovered in the cytoplasm; it was seen with the light microscope during studies of cell division by Walter Flemming in 1879. Flemming coined the term mitosis from the Greek root "mitos," meaning thread. Perhaps by extension he suggested that the protoplasm also consists of two components, namely, a fibrillar network and interfibrillary substance. Subsequently, these sightings were largely discounted as artifacts produced by coagulant fixatives. During the early part of the 20th century, experiments on living cells (reviewed by Porter, 1984), including mechanical displacement of the nucleus with

a needle, attraction with an electromagnet of iron particles introduced into the cytoplasm, and the dislocation by centrifugation of pigment granules in the cytoplasm of egg cells, showed that the cytoplasm has the properties of a viscoelastic gel. At the time, these interesting observations could not be pursued because of the lack of suitable methods. However, during this era, improvements in the compound light microscope and the development of special stains led to the discovery of abundant, highly ordered arrays of cytoplasmic filaments in the sarcomeres of skeletal muscle, the terminal web of intestinal epithelial cells, cytokeratin filaments in epidermal epithelial cells, neurofilaments in neurons, and glial fibers in astrocytes.

Studies on muscle contraction carried out between 1930 and 1960 heralded the modern era of research on cytoskeletal structures. Actin and myosin were identified as the major contractile proteins of muscle, and detailed electron microscopic studies on sarcomeres by H.E. Huxley and associates in the 1950s produced the concept of the "sliding filament model," which remains the keystone to an understanding of the molecular mechanisms responsible for cytoskeletal motility.

Powerful new technologies and unifying biological and molecular concepts have led to remarkable advances in our knowledge of the cytoskeleton over the past two decades. The electron microscope (EM) is 100 times better than the light microscope in resolving detail and magnifying images. The application of this instrument to biological materials in the 1950s permitted for the first time the direct visualization of the three main types of cytoskeletal fibers, but these were still thought to be specialized structures occurring only in certain types of differentiated cells. Indeed, the cytoplasm of the "generalized cell" then was thought to be composed of a gel-like cytosol in which various membrane-bound organelles, ribosomes, and the nucleus floated free. Keith Porter (1966) and his associates, and others, established the universality of microtubules in the general scheme of the cell by employing glutaraldehyde fixation and high-resolution electron microscopy. Other notable EM discoveries were the cytoskeletal apparatus of the intestinal brush border, the association of cytokeratin filaments (tonofibrils) with desmosomes, the assembly of microtubules to form mitotic spindle fibers, and the uniformity in the pattern of arrangement of microtubules within cilia and flagella from protozoa to primates. Also, a plasma-membrane-associated network of filaments composed of spectrin was discovered in mammalian red blood cells. Currently, live-cell imaging using fluorescent cytoskeletal probes with fluorescent sensors of the local physiological microenvironment in the cytoplasm ('molecular biosensors') are providing new information on how local biochemical changes modulate cytoskeletal structure and function (Giuliano and Taylor, 1995).

In this chapter we describe the distribution, assembly, and interaction of microfilaments and microtubules and their functional roles in cell movement and in the maintenance of the spatial organization of the cytoplasm. Also, the relative roles

of motor proteins and polymerization/depolymerization dynamics in generating force from the cytoskeleton and performing mechanical work will be discussed.

MICROTUBULES

Microtubules are universally present in eukaryotes from protozoa to the cells of higher animals and plants (Porter, 1966; Hardham and Gunning, 1978; Lloyd, 1987), but they are absent in mammalian erythrocytes and in prokaryotes. Microtubules participate in a number of cellular functions including the maintenance of cell shape and polarity, mitosis, cytokinesis, the positioning of organelles, intracellular transport to specific domains, axoplasmic transport, and cell locomotion. The diversity of microtubule functions suggests that not all microtubules are identical and that different classes of microtubules are present in different cell types or are localized in distinct domains in the same cell type (Ginzburg et al., 1989).

Structure and Composition of Microtubules

Microtubules (MT) are the largest of the cytoskeletal filaments with an outer diameter of about 25 nm, a wall thickness of about 5 nm, and a central lumen measuring about 15 nm. They consist of tubulin and associated proteins. Vertebrate brain tissue is a rich source of extractable tubulin because of the large numbers of microtubules that are present in axons and dendrites. Tubulin obtained from such a natural source is a heterodimer of 100 kD composed of α-tubulin and β-tubulin. Brain α-tubulin is a globular polypeptide that contains 451 amino acid residues, whereas β-tubulin, which is somewhat shorter, is made up of 445 amino acid residues. These two molecular species of tubulin share in common 40% of their amino acid residues.

Tubulins arose very early during the course of evolution of unicellular eukaryotes and provide the machinery for the equipartitioning of chromosomes in mitosis, cell locomotion, and the maintenance of cell shape. The primordial genes that coded for tubulins likely were few in number. As metazoan evolution progressed, natural selection processes conserved multiple and mutant tubulin genes in response to the requirements for differentiated cell types (Sullivan, 1988).

Tubulin Gene Families in Vertebrates

In-situ hybridization studies with DNA probes have revealed the existence in various vertebrate species of multiple small multigene families of α- and β-tubulin sequences that are not linked to each other in the genome (Sullivan, 1988). The occurrence of heterogeneous multigene families for α- and β-tubulins allows for the differential regulation of these proteins under multiple independent programs of gene expression during growth and differentiation and in the cell cycle (Sullivan, 1988). Cleveland and Sullivan (1985) have postulated that at the cellular level, the

expression of tubulin genes can be regulated by (1) the selective expression of some gene families and not others, (2) integration of the synthesis of α- and β-tubulins, and (3) establishment of the proper levels of α- and β-tubulin synthesis during development and differentiation and in the cell cycle.

Gene control occurs at the levels of both transcription and mRNA stability. Some tubulin genes are constitutively expressed, whereas other tubulin genes are tissue or cell-type specific in their expression. Experimentally, it has been shown that cultured mammalian cells can autoregulate tubulin gene expression through a feedback control mechanism that is based on the concentration of unpolymerized tubulin subunits in the cytoplasm (Cleveland and Sullivan, 1985). As yet, no information is available to explain the tissue- or cell-specific programs of gene expression.

The proportion of α- and β-tubulin isoforms varies in neurons in different brain regions, and a single neuron may express a number of different tubulin isoforms. The "multitubulin hypothesis" suggests that microtubules composed of different tubulin isoforms have different functional roles (Sullivan, 1988).

Assembly of Microtubules

When tubulin heterodimers are assembled into microtubules, they form linear "protofilaments" with the β-tubulin subunit of one tubulin molecule linking covalently with the α-subunit of the next. Direct examination by electron microscopy of tannic acid-treated specimens has shown that microtubules in neurons and the A-microtubules of cilia and flagella have 13 protofilaments arranged side to side to form a cylinder around what appears to be an empty lumen.

Microtubules may form stable structures such as the axoneme of cilia and flagella or transient structures such as mitotic spindle fibers (Dustin, 1980). Microtubules are assembled from the soluble tubulin dimer pool in the cytosol by polymerization and unidirectional elongation. The process is complex and is regulated by many factors including ion concentrations, microtubule-associated proteins (which have been studied in *in vitro* systems), and by microtubule-organizing centers (MTOCs) of diverse morphologies (Joshi, 1994), which orchestrate microtubules into parallel or radial arrays in living cells (Bershadsky and Vasiliev, 1988; Amos, 1995).

Reconstitution Experiments *in Vitro*

Microtubules can be reconstituted *in vitro* at 37 °C from a solution that contains a "physiological mixture" of brain tubulin, MAPs, small amounts of guanosine 5′-triphosphate (GTP), magnesium ions, and the calcium-chelating agent EGTA [ethylene glycol-bis(2-aminoethyl ether) N, N′-tetraacetic acid]. Tubulin assembly is inhibited by low temperature and by the presence of calcium ions.

The polymerization of tubulin follows a "growth curve" (sigmoidal in shape). Presumably the lag phase corresponds to the formation of nucleation centers and the growth phase to the elongation of microtubules. Electron microscopic studies indicate that the nucleation centers are comprised of protofilaments that associate laterally and assemble in groups of 13 into cylinders. Elongation occurs by the addition of tubulin dimer molecules at the ends of fragments of microtubules and is similar to the formation of actin filaments. The tubulin dimer binds two molecules of GTP. The attachment of this dimer to the end of the microtubule produces the hydrolysis of the bound GTP to GDP. Although the polymerization process does not require energy, the hydrolysis of bound exchangeable GTP in the wall of the microtubule does significantly speed up the rate of polymerization. In the presence of high concentrations of GTP-tubulin, "GTP caps" at the ends of microtubules are numerous, and elongation proceeds rapidly. In contrast, when GTP-tubulin concentrations are low, uncapped microtubules begin to polymerize. At high concentrations of GTP-tubulin, the rate of elongation of the microtubule at its plus end is considerably higher than at its minus end, which explains the effect of polarity on elongation. Treadmilling can occur at certain intermediate concentrations of GTP-tubulin that favor the addition of tubulin subunits at one end of the microtubule and their concurrent loss at the other end (Bershadsky and Vasiliev, 1988). Changes in the concentration of Mg^{2+} can play a similar role in microtubule treadmilling (see Herrmann, 1989).

Microtubule-Associated Proteins

Microtubule-associated proteins bind to microtubules *in vivo* and subserve a number of functions including the promotion of microtubule assembly and bundling, chemomechanical force generation, and the attachment of microtubules to transport vesicles and organelles (Olmsted, 1986). Tubulin purified from brain tissue by repeated polymerization-depolymerization contains up to 20% MAPs. The latter can be dissociated from tubulin by ion-exchange chromatography. The MAPs from brain can be resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Two principal categories of MAPs have been identified in neurons:

1. Fibrous MAPs, including MAP2 (270 kD) and tau (5—8 kD), which serve to regulate the assembly and bundling of microtubules (Olmsted, 1986; Gibbons, 1988; Lewis et al., 1989; Mandelkow and Mandelkow, 1995).
2. Force-producing MAPs, including kinesin (110—134 kD), dynein (1200—1900 kD), and dynamin (100 kD). These are mechanochemical ATPases that are involved in microtubule-dependent movement and transport processes (Olmsted, 1986; Gibbons, 1988; Paschal et al., 1989). There is also

evidence that brain dynein (Amos, 1989) and dynamin (Shpetner and Vallee, 1989) cross-bridge microtubules into bundles.

The MAPs vary in number and relative abundance in different cell types: MAP2 is principally found in dendrites, whereas tau is restricted to axons. This selective distribution of MAP2 molecules is the result of subcellular sorting of its messenger RNA (Lewis et al., 1989). Currently, there is interest in the observation that tau is a component of the neurofibrillary tangles of Alzheimer's disease (Goedert et al., 1989).

The microtubule-associated proteins MAP2 and tau both have two separate functional regions (Lewis et al., 1989). One is the microtubule-binding site, which nucleates microtubule assembly and controls the rate of elongation (by slowing the rate of assembly). The second functional domain shared by MAP2 and tau is a short C-terminal α-helical sequence that can cross-link microtubules into bundles by self-interaction. This domain has some of the properties of a "leucine zipper." Likely it is responsible for the organization of microtubules into dense stable parallel arrays in axons and dendrites (Lewis et al., 1989).

The MAPs that regulate the stability of microtubules can themselves be reversibly regulated by phosphorylation and dephosphorylation. The phosphorylation of MAP2 is catalyzed by kinases, such as calcium-calmodulin cAMP-dependent kinase. This lessens the ability of MAP2 to bind to tubulin and thereby decreases its ability to stabilize microtubules. Aoki and Siekevitz (1988) have suggested that alterations in MAP2 produced by phosphorylation and dephosphorylation influence the developmental plasticity of the brain. This concept is based on light-dark adaptation effects on the visual cortex of the cat. Phosphorylation of MAP2 (and many other proteins) is dependent on cAMP, which is the well-known second messenger in cells that mediates intracellular events initiated at the cell surface. The latter can be triggered by hormones, neurotransmitters, or light (in light-sensitive cells). In certain locations, such as in the axoneme, microtubules can undergo another type of more permanent stabilization through slow "maturation" as a result of two types of post-translational modifications in their α-tubulin subunits. Detyrosination and acetylation primarily affect preformed microtubules, whereas tyrosination and deacetylation occur in the soluble tubulin pool, thereby producing a dynamic asymmetry in the dimer-polymer equilibrium for tubulin. Detyrosination involves the enzymatic removal of the carboxyl-terminal tyrosine residue of α-tubulin. Acetylation, mediated by tubulin acetyltransferase (TAT), attaches an acetyl group to a specific lysine of α-tubulin. These posttranslational changes also may modify the assembly dynamics of microtubules by altering the binding properties of α-tubulin to MAPs (Cleveland and Sullivan, 1985; Sullivan, 1988; Mandelkow and Mandelkow, 1995).

The force-producing MAPs (kinesin, dynein, and dynamin) function as energy-transducing ATPases to provide the motive force for cilia and flagella by means of

a sliding filament mechanism and for anterograde and retrograde axonal transport and the directed transport of Golgi vesicles, presumably by means of a "walking mechanism." (Further details concerning these processes are given in subsequent sections of this chapter.)

Both dynein and MAP2 interact with microtubules at the same binding sites, namely, the C termini of α- and β-tubulin. Also, MAP2 inhibits the microtubule-activated ATPase of dynein and prevents microtubule gliding on dynein-coated glass coverslips. Thus, MAP2 and other fibrous MAPs may be regulators of microtubule-based motility *in vivo* (Paschal et al., 1989).

Functions of Microtubule-Organizing Centers

Microtubule-organizing centers (MTOCs) serve to nucleate, stabilize, and organize the arrangement of microtubules, both *in vivo* and *in vitro* (Brinkley, 1985; Joshi, 1994). Two large morphological groups of MTOCs are identified:

1. MTOCs containing highly regular, parallel arrays of microtubules, as seen in centrioles and basal bodies.
2. MTOCs with a ring-like, spherical structure consisting of amorphous electron-dense material.

The MTOCs of both types initiate the rapid nucleation of microtubules *in vitro*, even at low concentrations of pure tubulin. All centrosome-nucleated microtubules grow with their plus ends facing out. Centrosomes nucleate an asteroid arrangement of microtubules that radiates in all directions. Basal bodies initiate the parallel array of microtubules found in the axoneme. The proximal structures of basal bodies and centrioles are identical. In some species, there is an interconversion of centrioles and basal bodies, and in other species basal bodies arise *de novo*. From molecular genetic studies on a basal body mutant of *Chlamydomonas reinhardtii* (a single-celled green alga), Hall, Ramanis, and Luck (1989) have established that in this species the basal body contains a unique six- to nine-megabase chromosome. The foregoing, like mitochondrial and chloroplast genomes, carries genes that are concerned with the formation of the organelle, and these replicate and combine at different times from the nuclear genome. Also, the centriole/basal body chromosome is not autonomous, inasmuch as the genes that control the synthesis of tubulin are nuclear, as are the genes that affect the assembly and number of basal bodies (Goodenough, 1989).

Morphology and Functions of Special Microtubular Structures

Parallel arrays of microtubules are found in the axoneme of cilia and flagella of eukaryotic cells, and these are of constant pattern throughout the phylogenetic scale.

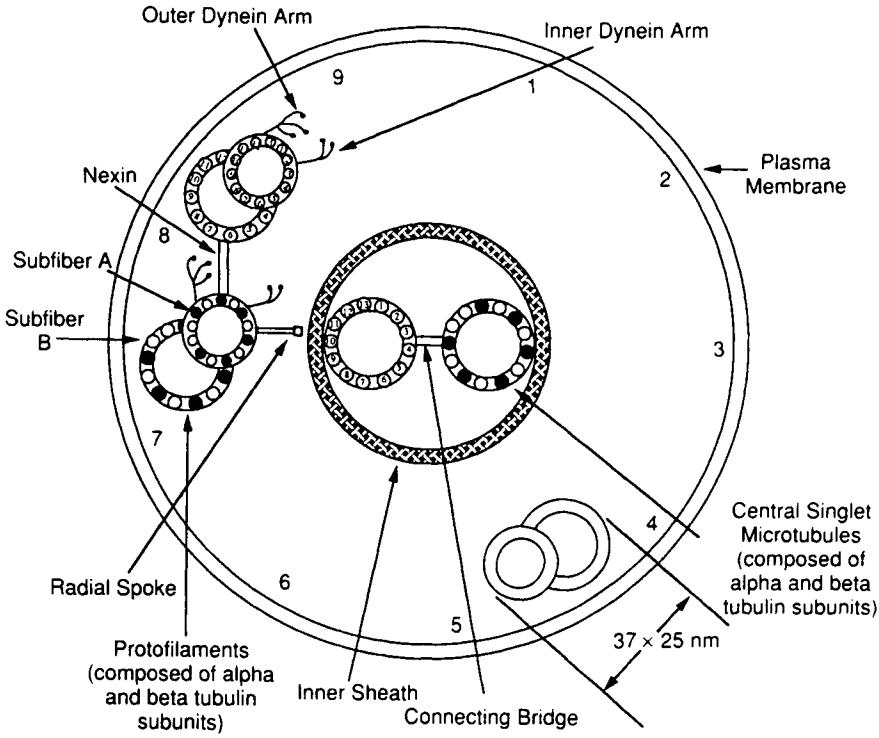


Figure 1. Cross-sectional view of an idealized cilium. For clarity only three outer doublet pairs of microtubules are shown. (This diagram is constructed from data previously published; for example, see Alberts et al., 1988.)

In mammals, ciliated cells line the respiratory air passages, the fallopian tubes, and the ventricles of the brain. The cilia beat in a coordinated manner in waves that propel fluids, suspended cells, and small particles along a surface. The motility of the sperm cell is provided by a single flagellum.

The axoneme consists of a cylinder of nine outer doublets of fused microtubules and a pair of discrete central microtubules (commonly referred to as the 9 + 2 arrangement of microtubules). The outer doublets each consist of a complete A-microtubule and an incomplete B-microtubule, the deficiency in the wall of the latter being made up by a sharing of wall material with the former. The tip of the axoneme contains the plus ends of all of the constituent microtubules. Two curved sidearms, composed of the MAP protein dynein, are attached at regular intervals to the A-microtubules of each fused outer doublet (Figures 1 and 2).

Other lateral structures that are bound to the microtubules of cilia and flagella are filamentous interdoublet links composed of nexin, radial spokes directed to the central pair of microtubules, and a central sheath that is attached to the walls of the central pair (Figure 1).

Axoneme Movement Is Produced by a Sliding Microtubule Mechanism

Dynein sidearms interact with the walls of B-microtubules of adjacent doublets by means of a sliding-filament mechanism to produce ciliary movement. The process is energized by ATP hydrolysis. Movement of the cilium occurs in two stages, termed the power stroke and the recovery stroke.

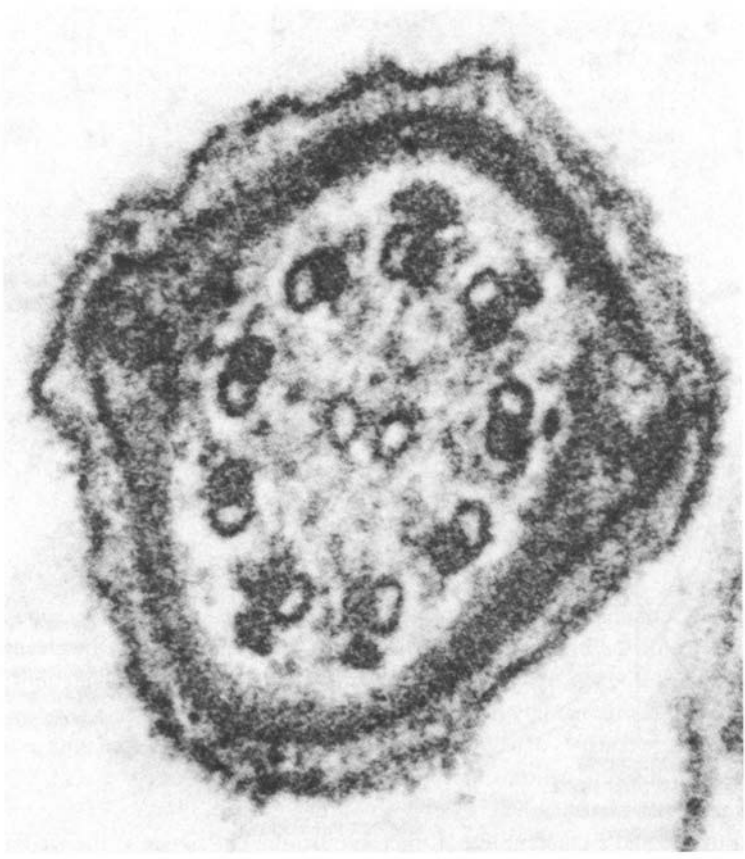


Figure 2. Electron micrograph of cross section of flagellum of mouse sperm, taken near the tip. The axoneme contains nine outer pairs of doublet microtubules and two central singlet microtubules. Several dynein arms and the fibrous sheath of the sperm are also shown.

An isolated flagellum will continue to bend actively, indicating that this function is linked to its intrinsic structure. Treatment of cilia from the protozoan *Tetrahymena* with the proteolytic enzyme trypsin selectively dissolves the nexin links and radial spokes but leaves unaffected the microtubules and dynein arms. If such a preparation is treated with a small amount of ATP, the loosened microtubule doublets slide against each other and through longitudinal overlap, extend for a distance that is up to nine times the original length of the cilium (Warner and Mitchell, 1981).

Functional studies indicate that the dynein side arms projecting from a doublet microtubule interact with their counterparts on an adjacent doublet microtubule and that ATP binding and hydrolysis by dynein ATPase activity slides the dynein heads along from the plus end toward the minus end of the microtubule. The sliding movement of the microtubules is converted to a bending movement of the axoneme by the constraining action of the radial spokes that are attached to the inner sheath and the central pair of single microtubules (Alberts et al., 1989; Herrmann, 1989). Also, localized differences in sliding rates between outer doublet microtubules contribute to axonemal bending. Intiation and propagation of bending appear to require different outer-arm dynein heavy chain isoforms that are located in specific regions of the axoneme (Asai and Brokaw, 1993).

Another example of parallel arrays of microtubules but with a much looser pattern is provided by axons, showing the involvement of such arrays in axoplasmic transport. This is illustrated in Figure 3.

Radial arrays are best exemplified by mitotic half-spindles, which have a central MTOC, the centrosome. The centrosome consists of two centrioles (which are homologous with the basal body) surrounded by dense pericentriolar material (Kalt and Schliwa, 1993). In plant cells, the MTOC of the mitotic spindle consists of dense material only without centrioles. The plus ends of microtubules of the mitotic spindle are directed toward the equator of the cell. Some are free, and others attach to kinetochores on chromatids (see Figure 4).

Submembranous microtubules are often present in parallel bundles beneath the plasma membrane in the cells of higher plants, particularly during cell wall formation (Hardham and Gunning, 1978). Circular submembranous bundles of microtubules are a feature of bird erythrocytes and mammalian blood platelets, where they maintain the discoid shape of these structures (Dustin, 1980).

Genetic Mutations Affecting Cilia and Flagella

The biflagellate unicellular green alga *Chlamydomonas reinhardtii* is prone to spontaneous mutations that produce deficiencies in flagellar proteins and MT assembly, substructure, and function. Viable mutants that are either nonmotile or slow moving can be isolated and analyzed biochemically and morphologically, thereby establishing structure-function correlations. Electron microscopic analysis

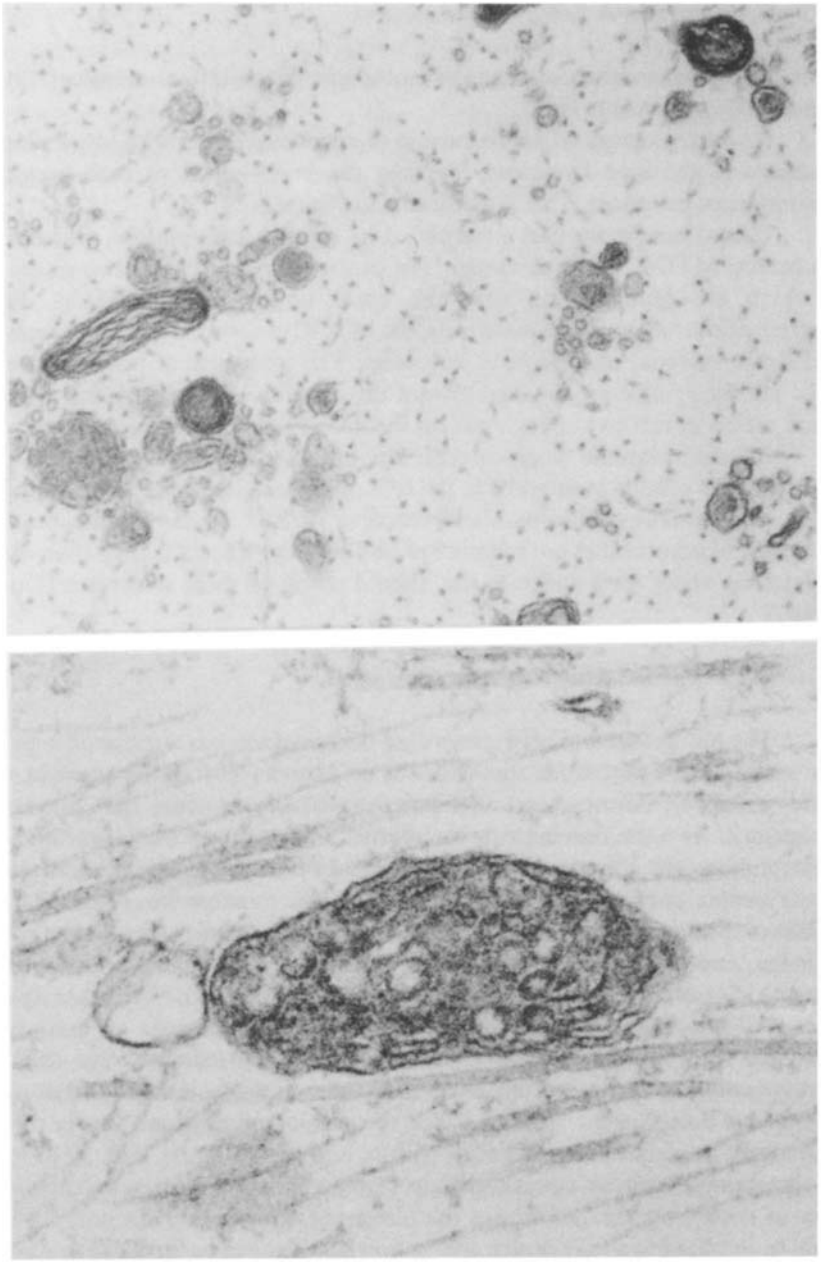


Figure 3. Electron micrographs of myelinated axons of Xenopus laevis. *Upper figure*: Cross section of axon showing microtubules in groups in association with membrane-bound organelles. *Lower figure*: Longitudinal section of axon showing neurofilaments and microtubules in close proximity to membrane-bound organelles. (Courtesy of Dr. R. Smith.)

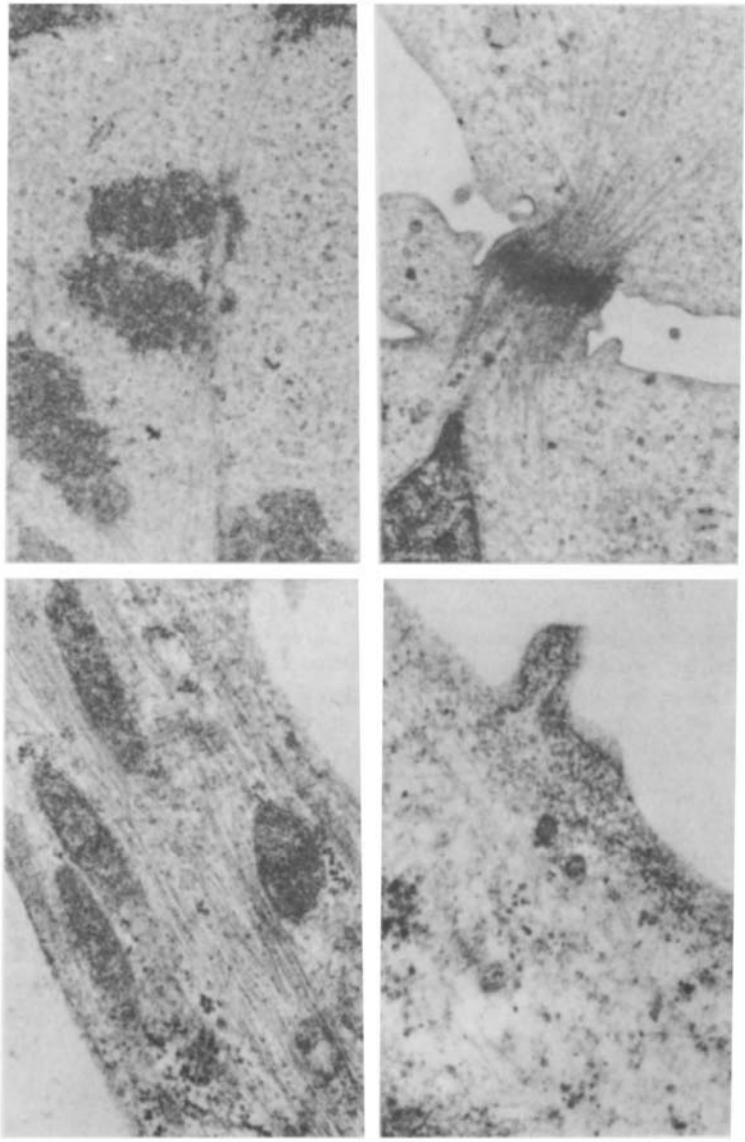


Figure 4. Electron micrographs of cultured rat glioma cells. *Upper left*: Kinetochore microtubules and chromatids of metaphase plate. *Upper right*: Dividing cell in late telophase. Cytokinesis is almost complete. The two daughter cells are still joined by a narrow band of cytoplasm that contains a midbody and a region of overlap of interdigitating polar microtubules. *Lower left*: Interphase cell showing a cytoplasmic process with parallel bundles of microtubules. *Lower right*: Interphase cell showing the actin-rich cortical cytoskeleton and a few microtubules. (From Bhatnagar and Malhotra, unpublished.)

of flagella from nonmotile mutants has revealed three different patterns of abnormality, namely, loss of the dynein arms, lack of radial spokes, and absence of both the central pair of microtubules and the inner sheath. Some slow moving mutants lack either the outer or the inner dynein arms. Two-dimensional polyacrylamide gel electrophoresis of axonemal proteins from wild-type (normal) flagella resolved approximately 200 different polypeptides as well as α- and β-tubulins. Mutant analysis showed that some structural defects are related to a lack of specific polypeptides. For example, paralyzed mutants without radial spokes and spoke heads lack 17 specific polypeptides, whereas mutants deficient only in spoke heads are missing only six of these polypeptides. From the foregoing, presumably six polypeptides can be localized to spoke heads, and 11 polypeptides can be localized to spokes. Similarly, the inner and outer dynein arms have been shown to be composed of different protein subspecies (Huang, 1986).

Very rarely, parallel genetic defects in axonemes can occur in man. The best-known of these is the immotile cilia syndrome described by Afzelius (1985), in which nonmotile respiratory cilia and sperm flagella lack dynein arms. Such individuals suffer from male sterility and from recurrent bronchitis and chronic sinusitis because of inadequate clearance of mucus from the respiratory tract. Also, about one-half of these patients have situs inversus, which is a mirror-image reversal in the position of the thoracic and abdominal organs. The entire complex of these abnormalities is known as Kartagener's syndrome. Electron microscopic examination of cilia in biopsies of respiratory epithelium is used as a laboratory test for the diagnosis of the immotile cilia syndrome.

Role of Microtubules in Cytokinesis and in Plant Cell Wall Formation

In dividing plant cells, microtubules are found in at least four arrays. Thus, during most of interphase, the cortical microtubule array is transversely oriented to the major cellular axis. The preprophase band of microtubules appears prior to mitosis. During metaphase, the spindle apparatus is typically anastral. After mitosis, the formation of a new primary cell wall between two daughter cells depends on the directed transport of Golgi-derived vesicles containing cell wall precursors along bundles of actin microfilaments coaligned with microtubules to a disk-like zone located in the plane of the former metaphase equator, termed the phragmoplast. The vesicles fuse with one another to form the cell plate, which expands circumferentially until it merges with the mother cell wall, comprising the middle lamella (Cyr, 1994; Goddard et al., 1994).

As plant cells grow, they deposit new layers of cellulose external to the plasma membrane by exocytosis. The newest regions, which are laid down successively in three layers next to the plasma membrane, are termed the secondary cell wall. Because the latter varies in its chemical composition and structure at different locations around the cell, Golgi-derived vesicles must be guided by the cytoskeleton

in a regulated fashion. During the differentiation of xylem cells to form vessels in young growing plants, peripheral bands of microtubules alternating with sheets of endoplasmic reticulum define future regions of cell wall thickening through the deposition of cellulose (Pickett-Heaps, 1967; Lloyd, 1987; Alberts et al., 1989). Mechanistic details about how microtubules in plant cells acquire spatial information are unavailable at present (Cyr, 1994).

Axonal (Axoplasmic) Transport

Nerve cells are of large size. Much of their cytoplasm is in the axon. For example, axons of the sciatic nerve in man can measure up to a meter in length. The synthetic activity of the nerve cell predominantly occurs in the cell body (perikaryon), which is crowded with endoplasmic reticulum, ribosomes, and mitochondria. The most abundant proteins of the axon are those that comprise the subunits of microtubules, neurofilaments (a category of intermediate filament), and actin filaments. These proteins are synthesized in the cell body and move down the axon at a rate of 1 to 5 mm per day by a process termed slow axonal transport (Hollenbeck, 1989; Coy and Howard, 1994). Other materials that are required at the axonal synapse for neurotransmission (such as neuropeptides and lipids) are carried in Golgi-derived vesicles along axonal microtubules (Figure 3) by fast axonal transport (anterograde transport) at speeds of up to 400 mm per day. Fast axonal transport is also required for the addition of new plasma membrane and cytoplasm at the tips of growing axons during development (see subsequent description of the 'neuronal growth cone' in this chapter). Fast anterograde transport (driven by kinesin) is in balance with fast retrograde transport (driven by cytoplasmic dynein), which moves membrane material in vesicles as well as aging mitochondria back to the cell body at one-half the speed (i.e., 200–300 mm/day) of fast anterograde transport (Alberts et al., 1989).

Neuroanatomists have taken advantage of the phenomenon of fast retrograde transport to locate remote nerve cell bodies in the CNS of an experimental animal that are connected to an identified axonal fiber tract whose origin is uncertain. The tracer material [purified horseradish peroxidase (HRP) enzyme] is injected in the region of the axon terminals, where it is taken up by endocytosis and then is carried by retrograde axonal transport over a period of several hours to days back to the nerve cell body. The animal is sacrificed, and the enzyme tracer is localized by staining thin sections of the brain for peroxidase activity.

Experimentally, the giant axon of the squid has proven to be an invaluable model system for studying the mechanisms of fast axoplasmic transport (see Allen, 1987). It was discovered through the use of video-enhanced light microscopy that the cytoplasm that could be extruded from an isolated giant axon displayed bidirectional movement of organelles along fine filamentous tracks. The latter were identified as single microtubules by correlative studies employing immunofluores-

cence microscopy and electron microscopy. Fast axoplasmic transport involved the binding of vesicles to microtubules and required ATP as an energy source. A nonhydrolyzable analogue of ATP, 5′-adenylylimidodiphosphate (AMP-PNP), caused organelles to become immobilized on microtubules in extruded squid axoplasm (Lasek and Brady, 1985), and this observation led to the development of a purification scheme for the identification of a large protein complex with ATPase activity (Vale et al., 1985a). This MAP was named kinesin. In an *in vitro* motility assay (originally designed for myosin-mediated motility), kinesin induced in the presence of ATP a "plus end"-directed (anterograde) movement of carboxylated latex beads along purified microtubules or a gliding of microtubules on a carpet of motor molecules coated on glass (Vale et al., 1985b). Using a quantitative method based on the foregoing system for measuring the movement produced *in vitro* by individual kinesin molecules, Howard, Hudspeth, and Vale (1989) have shown that a single kinesin molecule can move a microtubule for several micrometers.

The native kinesin molecule is composed of two 110- to 134-kD polypeptides that form two heads (now called heavy chains) and two 60- to 80-kD polypeptides (now called light chains). To explain the mechanoenzyme properties of kinesin, this molecular complex must possess a binding site for microtubules, a binding site for ATP, and an attachment site for the membranes of vesicles or organelles that are undergoing axoplasmic transport (Schliwa, 1989). The fact that kinesin has been identified in a variety of organisms and cell types suggests that it plays a role in different microtubule-based motor systems, including the formation of the endoplasmic reticulum network (Hoyt, 1994).

Recently, another high-molecular-weight MAP has been purified from mammalian brain that is responsible for the fast retrograde transport of vesicles along microtubules from the axon terminus back to the cell body of the neuron. This globular two-headed protein complex (1200 to 1900 kD) was first called MT-associated protein 1C (MAP 1C) but now has been renamed cytoplasmic dynein (in recognition of its homology with axoneme dynein) (Vallee et al., 1988; Holzbaur and Vallee, 1994; Schroer, 1994). Purified cytoplasmic dynein can produce an ATP-dependent sliding of taxol-stabilized purified microtubules on glass, and the direction of travel over microtubules is toward their minus end (Paschal et al., 1987).

Dynein, kinesin, and myosin are motor proteins with ATPase activity that convert the chemical bond energy released by ATP hydrolysis into mechanical work. Each motor molecule reacts cyclically with a polymerized cytoskeletal filament in this chemomechanical transduction process. The motor protein first binds to the filament and then undergoes a conformational change that produces an increment of movement, known as the "power stroke." The motor protein then releases its hold on the filament before reattaching at a new site to begin another cycle. Events in the mechanical cycle are believed to depend on intermediate steps in the ATPase cycle. Cytoplasmic dynein and kinesin "walk" (albeit in opposite

directions) on microtubules composed of tubulin; myosin moves on microfilaments of actin (Stryer, 1988; Howard et al., 1989).

Even though dynein, kinesin, and myosin serve similar ATPase-dependent chemomechanical functions and have structural similarities, they do not appear to be related to each other in molecular terms. Their similarity lies in the overall shape of the molecule, which is composed of a pair of globular heads that bind microtubules and a fan-shaped tail piece (not present in myosin) that is suspected to carry the attachment site for membranous vesicles and other cytoplasmic components transported by MT. The cytoplasmic and axonemal dyneins are similar in structure (Hirokawa et al., 1989; Holzbaur and Vallee, 1994). Current studies on mutant phenotypes are likely to lead to a better understanding of the cellular roles of molecular motor proteins and their mechanisms of action (Endow and Titus, 1992).

Microtubules Determine the Intracellular Location, Shapes, and Dynamics of Membrane-Bound Organelles

Microtubules are nucleated during interphase from the perinuclear microtubule-organizing center (MTOC) and extend to the cell periphery in the form of an extensive radiating network. Organelles of the central membrane system and their transport components differentially distribute within this microtubule network, with the endoplasmic reticulum (ER) and early endosomes selectively distributed toward the plus (fast growing) ends of the microtubules toward the cell periphery, and the Golgi, late endosomes and lysosomal membranes concentrated near the minus (slow growing) ends of the microtubules close to the nucleus. In this way, organelles of the central membrane system utilize microtubules both for the maintenance of their normal spatial distribution and for the directed transport of their protein and lipid secretory products along specific membrane traffic routes through the cytoplasm. Current research is focusing on organelle-microtubule interactions, the roles of microtubules and microtubule motors in the membrane trafficking pathways between organelles, and the mechanisms that regulate organelle mobility (see Cole and Lippincott-Schwartz, 1995). Because of space limitations, only the role of microtubules in fashioning and maintaining the ER and the Golgi complex are considered here.

The ER has a reticular morphology which provides a large surface area, which presumably is required for the synthesis and transport of proteins and lipids and for the storage of calcium. The ER is associated with microtubules, and the two are highly interdependent structures. Terasaki et al. (1986) found that when microtubules in the cell are depolymerized by colchicine, the ER network slowly retracts toward the center of the cell. If the microtubules are repolymerized, the ER network is restored to its original morphology, thereby suggesting that the MTs participate in the formation and maintenance of the ER.

In vitro, a tubulovesicular network analogous to the ER can form from membranous aggregates in the presence of MTs and ATP. This network, like the ER, can show branching, fusion, ring closure, and sliding. Dabora and Sheetz (1988) have termed this process microtubule-dependent tethering and suggest that a motor enzyme such as kinesin attaches to sites on the membrane aggregates and then binds to and moves along a stationary microtubule, thereby providing the force to draw out tubular branches. A new membrane branch forms when another active motor interacts with an intersecting microtubule. Lee et al. (1989) have confirmed these details by directly examining the reconstruction of the ER in cultured mammalian cells that were first exposed to the microtubule-disrupting drug nocodazole and then were allowed to recover in drug-free medium. Thus, the developing ER is pulled out from the cell center to the periphery on tracks of microtubules. While one end of the ER remains anchored to the outer nuclear membrane, the other end is stretched toward the plus ends of microtubules, which is the direction of kinesin-induced movement. Membranes of Golgi cisternae appear to move in the opposite direction, as if propelled along microtubules by cytoplasmic dynein. Microtubule depolymerization induced with nocodazole in cultured bovine kidney cells results in the reversible fragmentation and dispersal of the Golgi complex. However, during recovery, there is some dissociation between Golgi complex distribution and microtubule status, thereby suggesting that the reorganization of the Golgi complex is an active process that is not solely determined by microtubule binding (Turner and Tartakoff, 1989).

Role of the Cytoskeleton in Cell Division: Formation of the Mitotic Spindle, Mitosis, and Cytokinesis

Cell division involves a programmed, multistage reorganization of the cytoskeleton, and the latter plays a pivotal role in the entire process (McIntosh and Koonce, 1989; Wadsworth, 1993). In cell division, nuclear division (mitosis) is followed by cytoplasmic division (cytokinesis). These events will already be familiar to the reader from general biology courses. Of the three types of cytoskeletal fibers, microtubules are the major players in mitosis. The sequence of their involvement is as follows. Two polar microtubule-organizing centers (MTOCs), formed from duplicated centrioles (composed of nine triplet microtubules), migrate to the region of the spindle poles and become surrounded by a halo of amorphous material. Interphase cytoplasmic microtubules undergo depolymerization. Two half-spindles grow from the two MTOCs. During early prophase, kinetochores, which are specialized trilaminar plate-like structures (Verma, 1990), form in the centrometric regions of chromatids, and each becomes attached to an existing spindle fiber (Mitchison, 1988, 1990). Centromeric DNA directs the formation of the kinetochore (Mitchison and Hyman, 1988). Half-spindles consist of polar fibers that extend from the two poles of the spindle toward the equator, kinetochore fibers

that are attached to the centromeres of chromatids and extend toward the spindle poles, and astral fibers that radiate outward from the MTOCs at the poles. The dividing cells of higher plants lack centrioles and astral fibers, and the mechanism of induction of their spindle fibers is not well understood. By metaphase, the two half-spindles have become integrated to form a complete spindle. The bivalent chromosomes of the metaphase plate are immobilized at the equator by opposite traction forces exerted through the kinetochore fibers. Colchicine arrests cells undergoing mitosis in metaphase because it blocks the polymerization of spindle microtubules (Darnell et al., 1986; Alberts et al., 1989).

During anaphase, the two sets of daughter chromosomes separate and move toward the poles (anaphase A), and the poles move away from each other with an accompanying three- to fourfold elongation of the pole-to-pole distance by polar fibers but a shortening of the kinetochore fibers in anaphase B.

The microtubule-based mechanisms that produce chromosome movement are still under investigation. The subject was critically reviewed in 1982 by Pickett-Heaps and colleagues and in 1986 by Darnell et al. Recent photobleaching studies of cells injected with fluorescent tubulin (cited by Gorbsky and Borisy, 1989) indicate that in metaphase, kinetochore fiber microtubules undergo rapid polymerization and depolymerization. However, the exact sites of net tubulin loss and addition remain uncertain. Two theories have been proposed (Ault and Rieder, 1994). One is the "sliding collar hypothesis" in which depolymerization and ATP-driven sequential lateral interactions between the kinetochore and the kinetochore microtubule move the chromosome poleward. In this model, the kinetochore microtubule depolymerizes at the kinetochore but remains stationary with respect to the pole (McIntosh and McDonald, 1989). The second theory is the "treadmilling model," which suggests that the kinetochore microtubule moves poleward by the addition of tubulin subunits at the kinetochore (plus end) and loss at the pole (minus end) (Mitchison, 1989).

During anaphase B, a "sliding microtubule model" could be relevant to both the movement of chromosomes and the separation of the poles of the mitotic spindle (McIntosh and McDonald, 1989). Spindle MT-MT sliding might be driven by dynamin or by other as yet unidentified ATP-dependent motor molecules. Dynamin (a 100-kD polypeptide) is a newly discovered mechanochemical enzyme that exhibits both nucleotide-sensitive binding to microtubules and microtubule-enhanced ATPase activity. Also dynamin forms periodic cross-bridges between microtubules, thereby resulting in highly organized bundles of microtubules. The latter fragment and elongate in the presence of ATP. The postulated function of dynamin is to bridge microtubles and generate force between them (Shpetner and Vallee, 1989). Experimentally, the migration of the half-spindles is inhibited by vanadate but not by cytochalasin or phalloidin, thereby eliminating an actin-myosin-mediated process.

During telophase, a midbody forms between the separating daughter cells (see Figure 4), and disassembly of the remaining spindle microtubules and kinetochores takes place.

Cytokinesis (cell division) in animal cells involves the progressive formation in telophase of a furrow between the two daughter cells in the equator of the mitotic spindle. Immunofluorescent staining of the cortical cytoplasm at the site of the contraction ring reveals an abundance of actin as well as myosin, α-actinin, and filamin (Fishkind and Wang, 1995). Cytokinesis is highly sensitive to actin-myosin inhibitors such as cytochalasin and phalloidin.

Extracellular factors that induce mitosis are termed mitogens. Some of these, such as epidermal growth factor (EGF) and fibroblast growth factor (FGF) are now being characterized.

Maturation-promoting factor (MPF) appears to be the key intracellular regulator of mitosis and meiosis in eukaryotic cells. The MPF purified from starfish and frog eggs is a protein kinase that phosphorylates histone H1. The MPF activity rises as cells enter mitosis and falls as they leave mitosis to enter interphase. On the other hand, cyclin, a prominent 56-kD protein that was first discovered in the initial cell cycles of dividing sea urchin eggs, accumulates during interphase and disappears during mitosis. There is now evidence to suggest that cyclin accumulation leads to the activation of MPF, which in turn leads to the destruction of cyclin and the inactivation of MPP (Murray and Kirschner, 1989). The abrupt transition from metaphase to anaphase is thought to be initiated by the dephosphorylation of histone H1 molecules, lamins, and other proteins. Concurrently, there is a release of Ca^{2+} into the cytosol from spindle-associated vesicles (Alberts et al., 1989; Darnell et al., 1986). The regulation of mitosis is highly relevant to an understanding of embryological development (ontogeny), tissue regeneration, and cancer (neoplasia).

In terms of evolutionary biology, the complex mitotic process of higher animals and plants has evolved through a progression of steps from simple prokaryotic fission sequences. In prokaryotic cells, the two copies of replicated chromosomes become attached to specialized regions of the cell membrane and are separated by the slow intrusion of the membrane between them. In many primitive eukaryotes, the nuclear membrane participates in a similar process and remains intact; the spindle microtubules are extranuclear but may indent the nuclear membrane to form parallel channels. In yeasts and diatoms, the nuclear membrane also remains intact, an intranuclear polar spindle forms and attaches at each pole to the nuclear envelope, and a single kinetochore microtubule moves each chromosome to a pole. In the cells of higher animals and plants, the mitotic spindle starts to form outside of the nucleus, the nuclear envelope breaks down, and the spindle microtubules are captured by chromosomes (Kubai, 1975; Heath, 1980; Alberts et al., 1989).

Drug Effects on Microtubules

Several groups of drugs that bind to tubulin at different sites interfere with its polymerization into microtubules. These drugs are of experimental and clinical importance (Bershadsky and Vasiliev, 1988). For example, colchicine, an alkaloid derived from the meadow saffron plant (*Colchicum autumnale* or *Colchicum speciosum*), is the oldest and most widely studied of these drugs. It forms a molecular complex with tubulin in the cytosol pool and prevents its polymerization into microtubules. Other substances such as colcemid, podophyllotoxin, and nocodazole bind to the tubulin molecule at the same site as colchicine and produce a similar effect, albeit with some kinetic differences. Mature ciliary microtubules are resistant to colchicine, whereas those of the mitotic spindle are very sensitive. Colchicine and colcemid block cell division in metaphase and are widely used in cytogenetic studies of cultured cells to enhance the yield of metaphase plate chromosomes.

Vincristine and vinblastine (vinca alkaloids) comprise another class of drugs that inhibit the polymerization of microtubules but do so by binding to the tubulin molecule at a site different from the colchicine site. Cultured cells exposed to high concentrations of vinca alkaloids develop intracytoplasmic paracrystalline aggregates of tubulin. These drugs are employed clinically in cancer chemotherapy to inhibit the growth of tumors composed of rapidly dividing cells.

Another drug is taxol, which is extracted from the bark of the Pacific yew tree, *Taxus brevijolia*. Unlike colchicine and the vinca alkaloids, taxol binds tightly to microtubules and stabilizes them against depolymerization by Ca^{2+}. It also enhances the rate and yield of microtubule assembly, thereby decreasing the amount of soluble tubulin in the cytosol pool. Again, the overall effect of taxol is to arrest dividing cells in mitosis. Taxol is used in cancer chemotherapy.

MICROFILAMENTS: ACTIN FILAMENTS

Structure and Composition

The principal molecular constituent of thin filaments is actin. Actin has been highly conserved during the course of evolution and is present in all eukaryotes, including single-celled organisms such as yeasts. Actin was first extracted and purified from skeletal muscle, where it forms the thin filaments of sarcomeres. It also is the main contractile protein of smooth muscle. Refined techniques for the detection of small amounts of actin (e.g., immunofluorescence microscopy, gel electrophoresis, and EM cytochemistry) subsequently confirmed the presence of actin in a great variety of nonmuscle cells. Muscle and nonmuscle actins are encoded by different genes and are isoforms.

Actin isolated from skeletal muscle in the presence of ATP and Ca^{2+} and minimal concentrations of Na^{+} and K^{+}, depolymerizes into globular polypeptide subunits (42 kD) consisting of 375 amino acids (known as G-actin). Polymerization into actin filaments (F-actin) occurs *in vitro* in solutions containing Mg^{2+} and high concentrations of Na^{+} and K^{+}. Actin from diverse sources can copolymerize into filaments. Actin filaments elongate by the addition of G-actin subunits at their ends. The G-actin monomer is tightly bound to one molecule of ATP, which slowly hydrolyzes to ADP after polymerization. Although the process of actin polymerization does not require energy, ATP hydrolysis does permit actin filaments to "treadmill" by lengthening at one end through the incorporation of actin-ATP subunits and concurrently shortening at the opposite end through a release of actin-ADP monomers (Small, 1995). As already mentioned, microtubule treadmills utilize GTP. An "exchange diffusion model" has been proposed as an alternative mechanism for the growth of actin filaments, and both models are discussed by Herrmann (1989).

Molecular models and electron micrographs indicate that the actin filament is a 6 to 10 nm wide (commonly referred to as 7 nm), thread-like, tightly wound helical polar structure with two nonidentical ends. The polarity of actin filaments can be established by incubating actin filaments with a specific fragment of myosin (S1), which decorates the actin filament in an arrowhead fashion. It is now customary to speak of the "pointed end" and the "barbed end" of the actin filament, based on the configuration of the attached S1 fragments. With this terminology used to describe the polymerization of actin filaments, net association takes place at the "barbed end," and dissociation occurs at the "pointed end."

Nonmuscle Actin-Binding Proteins

More than 50 proteins have been discovered in the cytosol of nonmuscle cells that bind to actin and affect the assembly and disassembly of actin filaments or the cross-linking of actin filaments with each other, with other filamentous components of the cytoskeleton, or with the plasma membrane. Collectively, these are known as actin-binding proteins (ABPs). Their mechanisms of actions are complex and are subject to regulation by specific binding affinities to actin and other molecules, cooperation or competition with other ABPs, local changes in the concentrations of ions in the cytosol, and physical forces (Way and Weeds, 1990). Classifications of ABPs have been proposed that are based on their site of binding to actin and on their molecular structure and function (Pollard and Cooper, 1986; Herrmann, 1989; Pollard et al., 1994). These include the following:

1. Proteins that bind to actin monomers and inhibit polymerization are designated as profilins (12–15 kD) (Sun et al., 1995). In addition to functioning as an actin-monomer-sequestering protein, profilin binds at least three other

ligands, and may play a role in membrane-cytoskeleton communication (Machesky and Pollard, 1993).

2. Barbed-end-capping proteins (gelsolin and villin, 95 kD) attach to this specific end of the actin filament and inhibit the further addition of actin molecules.
3. Pointed-end-capping proteins are acumentin (65 kD), spectrin (220–260 kD), and β-actinin (37 kD). They also regulate the length of actin filaments.
4. Proteins that cross-link actin filaments bind to their sides to produce bundles or three-dimensional networks (Otto, 1994). In microvilli, approximately 20 actin filaments of the core are cross-linked by villin (95 kD) and fimbrin (68 kD) in helical array to form a compact bundle (Figure 5). Filamin (2 × 250 kD) induces the formation of an actin network with gel formation. By immunofluorescence microscopy, this ABP is found in the ruffled, motile edge of cultured cells, where only actin filaments are abundant.
5. Proteins that link actin filaments to the plasma membrane deserve special comment. In the erythrocyte, spectrins (220–260 kD) attach the sides of actin filaments to the inner face of the plasma membrane through a linker protein (ankyrin) to form a membrane-cytoskeletal complex. Nonerythroid cell types, including neurons, exocrine cells, and polarized epithelial cells, contain homologues of ankyrin and spectrin (referred to as fodrin) in specific membrane domains (Nelson and Veshnock, 1987). At the tip of the microvillus, the barbed ends of actin filaments of the core are connected to a dense plaque of unidentified material on the inner face of the plasma membrane by bridges of a 110 kD calmodulin (calcium-binding protein) complex (Mooseker, 1985) (see Figure 5).
6. Filament-severing proteins bind to actin subunits in the interior of filaments, causing filament breakage. Gelsolin (90 kD) expresses this activity in the presence of Ca^{2+}. Fragmin and severin (45 kD) are Ca^{2+}-dependent severing and capping proteins.

Actin-binding proteins (gelsolin, villin, fragmin, and severin) have been shown to have similar regions of amino acid sequence, and these regions also resemble the amino acid sequence of actins. Thus, actin-binding proteins (ABPs) may serve their role(s) by mimicking certain regions of actin that are involved in the formation of actin filaments by actin-actin interactions among actin monomers (Tellam et al., 1989).

Drugs Affecting Actin Polymerization

The cytochalasins are a group of metabolites produced by certain fungi (e.g., *Helminthosporium dermatoideum*) that inhibit actin-dependent types of cell movement such as leukocyte locomotion, phagocytosis, cytokinesis, the retraction of

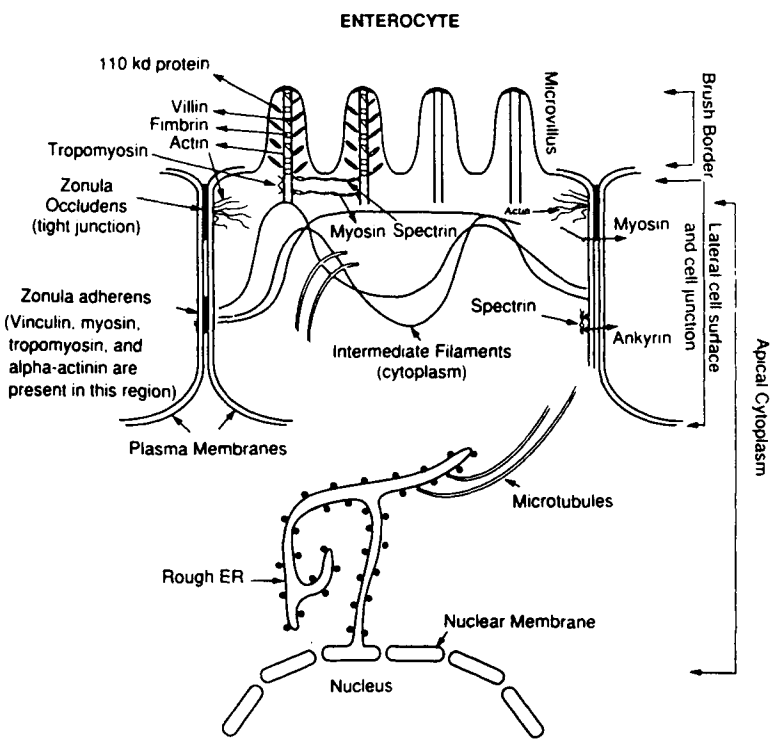


Figure 5. Diagrammatic representation of the cytoskeleton in the apical region of the intestinal epithelial cell (enterocyte). (This diagram is from data previously published; for example, see Mooseker, 1985.)

blood platelets, the surface ruffling of fibroblasts, and the outgrowth of neurites. Cytochalasins interfere with the assembly of actin filaments by capping their fast-growing barbed ends. Electron micrographs of cytochalasin-treated cells reveal a paucity of microfilaments. Such findings indicate the importance of the assembly and disassembly of labile actin filaments in producing cortical cytoplasmic flow. Cytochalasins do not affect the contraction of skeletal muscle because the actin (thin) filaments of sarcomeres are stable structures.

Phalloidin is a highly poisonous oligopeptide produced by the mushroom *Amanita phalloides*. Phalloidin stabilizes actin filaments by binding to them stoichiometrically and inhibiting their depolymerization. Fluorochrome-labeled phalloidin has been used to specifically stain actin filaments in cultured cells. Phalloidin does not cross the plasma membrane as readily as the cytochalasins. When microinjected into cultured cells, it blocks their ability to migrate, again underscoring the significance of actin-based assembly-disassembly mechanisms in cell movement.

Patterns of Arrangement of Actin Filaments in Animal Cells

Actin is plentiful in the cytoplasm of many animal cells, comprising 5% or more of the total protein present. Four principal patterns of arrangement can be recognized: (1) three-dimensional networks of filaments, (2) bundles of parallel filaments with the same polarity, (3) submembranous actin-spectrin (fodrin) networks, and (4) bundles of parallel filaments with alternating polarities.

The diversity of these subcellular actin structures is remarkable and appears to be determined by the interactions of many actin-binding proteins (ABPs) as well as by changes in the concentrations of intracellular signaling molecules such as Ca^{2+} and cAMP, by small GTP-binding proteins, and by signals arising from mechanical stress. Approximately 50% of the actin molecules in most animal cells are unpolymerized subunits in the cytosolic pool and exist in a state of dynamic equilibrium with labile F-actin filamentous structures (i.e., new structures are formed while existing structures are renewed) (Hall, 1994).

Three-Dimensional Networks

Three-dimensional networks are made up of a web of filaments that intersect at various distances and angles. Cross-linking ABPs such as filamin form flexible links between actin filaments at points of contact. Complex F-actin networks are present in the submembranous cortex (superficial plasmagel layer) of many types of animal cells. Actin filaments attach by means of associated proteins to the inner face of the plasma membrane and thereby provide a resilient framework for the delicate overlying lipid bilayer. Also, the gel-like cortical layer has both viscous and elastic properties, resisting rapid deformation but yielding to slow deformation. Local contractions of the cell cortex, together with local changes in the state of gelation and solation of actin, provide the basis for cell locomotion and migration, blood platelet activation, cytokinesis, phagocytosis, lymphocyte capping, and neuronal growth cone formation (Bray and White, 1988). In all of the foregoing movements, cytoplasmic flow is driven by gradients of tension in the actin-rich cortex, which pull regions of relaxation toward regions of contraction. The details vary with the particular function and cell type.

In migrating *Dictyostelium* (slime mold) amebae, nonfilamentous myosin I occurs at the leading edges of lamellipodia, whereas filamentous (high-molecular-weight) myosin is concentrated in the posterior of these cells. It has been suggested that actomyosin I may assist in the forward extension of a locomoting cell and that the contraction of actinomyosin II in the trailing end of the cell squeezes the cytoplasm forward (Fukui et al., 1989). However, mutant strains of *Dictyostelium* lacking myosin II or α-actinin (an actin cross-linking protein) or severin (an actin filament-fragmenting protein) are capable of relatively normal motility and chemotaxis. Therefore, it appears that cell locomotion depends on actin and the combined activities of actin-binding

proteins as a group and that within this group there is considerable overlapping redundancy, thereby providing a "fail-safe" system (Bray and Vasiliev, 1989).

Leukocyte Migration

Neutrophilic granulocytes and macrophages migrate into injured or infected tissues in response to inflammatory chemotactic stimuli in man and higher vertebrates (Devreotes, 1988). Direct observations indicate that the locomotion of these leukocytes is accomplished by the extension from the cell surface of undulating sheets of cytoplasm called "lamellipodia." The lamellipodia at the leading edge of the cell adhere to the substratum, while others move back with a wave-like motion termed "ruffling." Cytoplasm flows into the lead lamellipodium, carrying the cell forward. The pear-shaped tail of the cell or "uropod" periodically detaches and is pulled into the cell body. Although the giant ameba (*Amoeba proteus*) provided the archetype for such descriptions of cellular movement (hence the term "ameboid movement"), the details of the process differ somewhat from those encountered in the neutrophilic granulocyte.

Lamellipodia contain a meshwork of actin filaments that are anchored to focal contacts on the plasma membrane through actin-binding proteins. Focal cytoplasmic protrusive activity likely is determined at the outset by the location of uncapped barbed ends of actin filaments, which nucleate the growth of actin filaments. Free barbed ends probably result from the removal from filament ends of capping proteins like gelsolin and aginactin, and/or from the *de novo* assembly of actin oligomers (Condeelis, 1993; Lee et al., 1993). The actin-binding drug cytochalasin paralyzes such movement in neutrophilic granulocytes. The continuance of cell migration in the absence of protein synthesis suggests that the cortical motor components that move to the uropod are being continually recycled by cytoplasmic streaming to the leading edge of the cell.

Most of the available evidence indicates that the actin-rich cortical cytoskeleton controls cell locomotion. An alternative membrane-flow model has been used to suggest that plasma membrane is delivered continually by exocytosis at the leading edge of the cell to push the cell forward (Singer and Kupfer, 1986). Simultaneously, mobile proteins within the cell are carried backward, and plasma membrane material is recycled by endocytosis. More recently, Sheetz et al. (1989) have questioned the membrane-flow model of cell locomotion because of discordant findings from direct nanometer-level measurements of the movements of membrane glycoproteins tagged with gold particles.

Fibroblast Locomotion

Fibroblast locomotion has been studied in tissue culture systems and is unexpectedly complex. Initially when fibroblasts are plated, they are rounded and

migrate freely. Subsequently they become flattened and move jerkily, because cortical flow is modulated by attachment of the undersurface of the cell to the substratum by adhesion plaques (Burridge et al., 1988). Locomotion begins with an extension of lamellipodia from the leading edge of the cell, so that the cell begins to move forward. At the same time adhesion plaques anchor the undersurface of the cell to the substratum. Ruffling of the cortex develops in the tail region. The tail end of the cell remains temporarily attached to the substratum and becomes attenuated to form a retraction fiber, which then detaches and is drawn into the body of the cell (Bray and White, 1988; Lee et al., 1993). *In vivo*, fibroblasts are stationary cells except during wound healing. Normally, they adhere to collagen fibrils and maintain traction on these.

Stress fibers are parallel bundles of actin filaments that develop in the cytoplasm of fibroblasts from the cortical actin network in response to mechanical tension. These often bind to the plasma membrane at focal contacts and, through transmembrane linker glycoproteins, to the extracellular matrix. Thus, actin filaments of stress fibers indirectly join to the inner face of the plasma membrane through molecular assemblies of attachment proteins, which include an actin-capping protein, α-actinin, vinculin, and talin (Small, 1988).

Fibronectin receptor is a two-chain glycoprotein of the integrin family that serves as a transmembrane linker by binding to talin on the cytoplasmic side and to fibronectin on the external side of the membrane. The pull exerted by stress fibers on attached structures may be produced by bipolar assemblies of nonmuscle myosin molecules producing a sliding of actin filaments of opposite polarity.

Platelets

Blood platelets are key players in the blood-clotting mechanism. These tiny fragments of cytoplasm are shed into the circulation from the surface of megakaryocytes located in the bone marrow. When the lining of a blood vessel is injured, "activated" platelets release clotting factors, adhere to each other and to damaged surfaces, and send out numerous filopodia. The shape changes that occur in activated platelets are the result of actin polymerization. Before activation, there are no microfilaments because profilin binds to G-actin and prevents its polymerization. After activation, profilin dissociates from G-actin, and bundles and networks of F-actin filaments rapidly appear within the platelet.

Cytokinesis

Cytokinesis begins with astral relaxation of the cell cortex, perhaps triggered by the mitotic spindle, followed by the accumulation in a circumferential equatorial band of actin filaments and associated myosin molecules to form a "*contractile*

ring." The latter constricts the cytoplasm, producing a circumferential "*cleavage furrow*."

Phagocytosis

Phagocytosis is a process whereby solid particles are ingested by leukocytes (neutrophilic granulocytes and macrophages) or by protozoa. Initially, tight binding occurs between membrane receptors and ligands, respectively located on the surfaces of the plasma membrane and the solid particle. Waves of lamellipodia extend from the cell surface to engulf the particle. Beneath the contact site, cortical actin filaments become organized into radiating bundles. Within several minutes the edges of the lamellipodia fuse, and the particle is invaginated within a membrane-bound phagocytosis vesicle. The latter fuses with a lysosome to give rise to a phagosome, wherein digestion of the particle occurs through the action of lysosomal acid hydrolases.

Lymphocyte Capping

Lymphocyte capping involves a concerted movement of integral membrane proteins and lipid molecules by the actin-rich membrane-associated cortical cytoskeleton. When lymphocytes are exposed to polyclonal antibodies that bind to certain surface antigens, these antibodies collect into a number of small "patches" on the cell surface. Then, with further incubation at 37 °C for several minutes, the patches aggregate into a single cap, usually in the region of the posterior uropod of the lymphocyte. Capping requires metabolic energy, whereas patching does not. Capping is associated with an accumulation of actin and actin-binding proteins in the subjacent cortical cytoplasm of the lymphocyte. The process is inhibited by cytochalasin.

Neuronal Growth Cone

Neuronal growth cone formation involves the motile expansion of cytoplasm at the tip of a growing neurite (axon or dendrite) and can be observed both in developing embryos and in neurons growing in tissue culture. Different types of nerve cells show differences in the morphology and size of their growth cones, which range from a simple filopodium to a vast expansion of filopodia and lamellipodia (see Bray and Hollenbeck, 1988). Those growth cones that are able to reach a target are transformed into synapses, and others undergo degeneration. Growth cones are rich in the three principal cytoskeletal fibers—microfilaments, microtubules, and neurofilaments—and their various associated proteins (Bentley and O'Connor, 1994). Actin-containing lamellipodia and filopodia at the apex of the growth cone undergo ruffling movements similar to those observed in neutro-

philic granulocytes. These are coupled to an elongation of the neurite as a result of the transport on tracks of microtubules of membrane components and other building materials. The three steps that are recognized in the formation of a new segment of neurite are protrusion, engorgement, and consolidation (Goldberg and Burmeister, 1989).

Various external and internal factors, including ion concentrations, cell-to-cell interactions, and specific messenger molecules, initiate and regulate growth cone formation and neurite migration (see Bray and Hollenbeck, 1988). For example, GAP-43 is a membrane protein that induces cells to extend processes (filopodia). This protein is normally present in the plasma membrane of neural cells and is particularly concentrated at the growth cones of neurites. Recently the gene coding GAP-43 was introduced into nonneuronal cells, and the resulting transfected cells showed the presence of GAP-43 in their plasma membrane and also formed extensive filopodia (Zuber et al., 1989). The specific extracellular adhesive molecules fibronectin and laminin also are known to enhance neurite growth (Rauvala et al., 1989).

Bundles of parallel actin filaments with uniform polarity. The microvilli of intestinal epithelial cells (enterocytes) are packed with actin filaments that are attached to the overlying plasma membrane through a complex composed of a 110-kD protein and calmodulin. The actin filaments are attached to each other through fimbrin (68 kD) and villin (95 kD). The actin bundles that emerge out of the roots of microvilli disperse horizontally to form a filamentous complex, the terminal web, in which several cytoskeletal proteins, spectrin (fodrin), myosin, actinin, and tropomyosin are present. Actin in the terminal web also forms a peripheral ring, which is associated with the plasma membrane on the lateral surfaces of the enterocyte (see Figure 5, p. 24).

A striking example of the controlled rapid formation of bundles of parallel actin filaments with uniform polarity is observed in the acrosomal reaction of sea urchin and sea cucumber sperm as a prelude to fertilization. The acrosomal vesicle (acrosome) lies beneath the plasma membrane at the tip of the sperm. The periacrosomal region (located between the acrosome and the nucleus of the sperm) contains a high concentration of G-actin complexed to profilin. The acrosomal reaction is triggered when the head of the sea urchin sperm binds by virtue of specific receptors to the jelly coat of the egg. A chain reaction ensues. The acrosome is expelled by exocytosis and releases proteases, which digest the jelly coat and vitelline layer of the egg, thereby exposing the egg plasma membrane. Prolifin dissociates from G-actin, resulting in a rapid polymerization of parallel bundles of actin filaments, which in turn cause the protrusion of the acrosomal process. The membrane of the latter, which is derived from the acrosomal membrane, fuses with the egg plasma membrane, thereby providing a cytoplasmic channel for the entry of the sperm nucleus into the egg (Tilney and Inoue, 1982; Tilney et al., 1983; see also Darnell et al., 1986).

Submembranous actin-spectrin network. The nature of the molecular interactions between the cytoskeletal complex and the plasma membrane is beginning to be revealed in the mammalian red blood cell. The major submembranous cytoskeletal protein is spectrin, which is composed of two subunits (α and β). The two subunits are associated with each other to form heterodimers. These form a multiprotein complex with ankyrin and actin. Ankyrin is also complexed with a transmembrane protein (band 3), which is an anion transporter in the plasma membrane. Actin also interacts with the plasma membrane through another cytoskeletal protein called band 4.1 (82 kD), which also binds to spectrin as well as to glycophorin in the plasma membrane. Several other proteins, e.g., tropomyosin, calmodulin, and protein 4.9, interact with the spectrin-actin-protein 4.1 complex (Figure 6). Glycophorin (26 kD) is one of the two major transmembrane proteins in the plasma membrane of the red blood cell, the other being band 3 protein. Band 3 protein also binds several other proteins, such as protein 4.1, glycolytic enzymes, and hemoglobin (see Davies and Lux, 1989; Bennett and Gilligan, 1993).

High-resolution electron microscopy of erythrocyte membrane skeletons has yielded images of a regular lattice-like structure with five or six spectrin molecules attached to short actin filaments, 30 to 50 nm in length, to form a sheet of five- and six-sided polygons (Bennett, 1989). The interactions between various proteins in the membrane and underneath the plasma membrane provide red blood cells with the flexibility to undergo changes in shape during their passage through narrow capillaries (Davies and Lux, 1989).

Inherited deficiencies or structural abnormalities in erythrocyte membrane skeletal proteins result in abnormally fragile or misshapen red blood cells, causing fragmentation, accumulation of red cells in the spleen, and mild to severe hemolytic anemias (Delaunay and Dhermy, 1993). Hereditary spherocytosis (HS) in man is the best-known example. However, there is considerable genetic and clinical heterogeneity in HS because of mutations in many different genes. About 80% of families affected by HS exhibit an autosomal dominant pattern of transmission of the disorder (HS-2), with variable degrees of penetrance from family to family. A 15% to 40% quantitative deficiency in β-spectrin and impaired binding of β-spectrin to protein 4.1 are found in cases with mild anemia. Recessively inherited spherocytosis (HS-1), observed in about 20% of families with HS, is clinically more severe than the dominant form (Bennett, 1989). This is associated with a 25% to 70% quantitative deficiency in α-spectrin.

Ankyrin deficiency, amounting to a reduction of approximately 50%, has been found in some patients with the dominant type of HS. A primary defect in ankyrin almost certainly reduces the assembly of spectrin into the membrane skeleton of the red blood cell (Davies and Lux, 1989).

Proteins closely related to spectrin and ankyrin and many of the associated proteins first identified in erythrocytes also are prevalent in most vertebrate tissues. In these other tissues, spectrin and ankyrin and their associated proteins interact with proteins that may not be expressed in erythrocytes, including ion channel

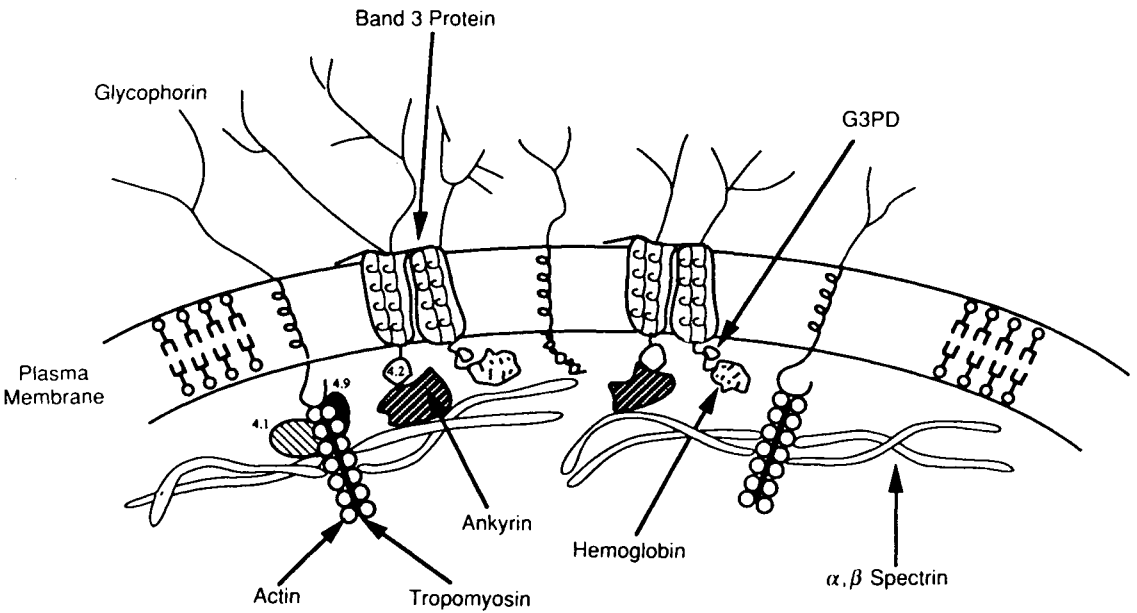


Figure 6. Diagrammatic representation of the red blood cell cytoskeletal-plasma membrane complex. Spectrin is made up of many homologous triple-helical segments joined by nonhelical regions (Speicher and Marchesi, 1984). Spectrin and actin require accessory proteins to form a membrane-associated network. (This diagram is constructed from data previously published; for example, see Stryer, 1988; Davies and Lux, 1989; Bennett and Gilligan, 1993).

proteins, cell adhesion molecules, tubulin and intermediate filaments (Bennett and Gilligan, 1993). One function of spectrin and ankyrin-based structures in differentiated cells in tissues involves the organization of specialized domains of the plasma membrane. Thus, in a culture system of Madin-Darby canine kidney (M-DCK) epithelial cells, Nelson and Hammerton (1989) have shown that ankyrin/fodrin is complexed to the well-known Na^+, K^+-ATPase. In these cells, Na^+, K^+-ATPase is preferentially localized in the basolateral plasma membrane, and such interactions between cytoskeletal proteins and proteins in the plasma membrane serve to maintain regional specializations of the plasma membrane. Sites of cell-to-cell contact in epithelial tissues are the best characterized example of the involvement of membrane skeletal proteins in the organization of a membrane domain. Thus, isolated junctional complexes, like membrane skeletons, contain four proteins that are stable components of this structure, namely, ankyrin/fodrin, actin, protein 4.1, and protein 4.9 (based on mobility on SDS gels). *In vitro* studies suggest that tropomyosin, tropomyosin-binding protein, and adducin also are linked to junctional complexes, but as yet these have not been localized *in situ* (Burridge et al., 1988; Bennett, 1989; Nelson, 1992).

Bundles of actin filaments with alternate polarities. Such bundles of filaments are best exemplified by the myofibrils of sarcomeres of skeletal muscle cells and cardiac muscle cells, which are highly organized stable structures (Figure 7). The sarcomeres are periodically repeating units consisting of two arrays of parallel actin filaments with opposite polarities. The barbed ends of the actin filaments of each array are attached by means of α-actinin (an actin-binding protein) to transverse partitions termed Z disks. The pointed ends of the actin filaments are oriented to face the midline of the sarcomere. The central region of the sarcomere contains thick myosin filaments, which interdigitate with and slide along the actin thin filaments during muscle contraction, so that the sarcomere shortens. The heads of the bipolar myosin thick filaments form cross-bridges with the actin thin filaments. The contractile force is generated in sarcomeres by the cyclic formation and dissociation of complexes between actin and the heads of myosin, producing a ratchet-like power stroke. The hydrolysis of ATP, generated by oxidative phosphorylation in muscle mitochondria, drives the cyclic association and dissociation of actin and myosin. Myosin has Mg^{2+}-ATPase activity that is enhanced when it binds to F-actin. Actomyosin then complexes with ATP, which causes the hydrolysis of ATP and the dissociation of actin and myosin, enabling the cycle to be repeated.

Muscle contraction is initiated by a signal from a motor nerve. This triggers an action potential, which is propagated along the muscle plasma membrane to the T-tubule system and the sarcotubular reticulum, where a sudden large electrically excited release of Ca^{2+} into the cytosol occurs. Accessory proteins closely associated with actin (troponins T, I, and C) together with tropomyosin mediate the Ca^{2+}-dependent motor command within the sarcomere. Other accessory proteins (titin, nebulin, myomesin, etc.) serve to provide the myofibril with both stability

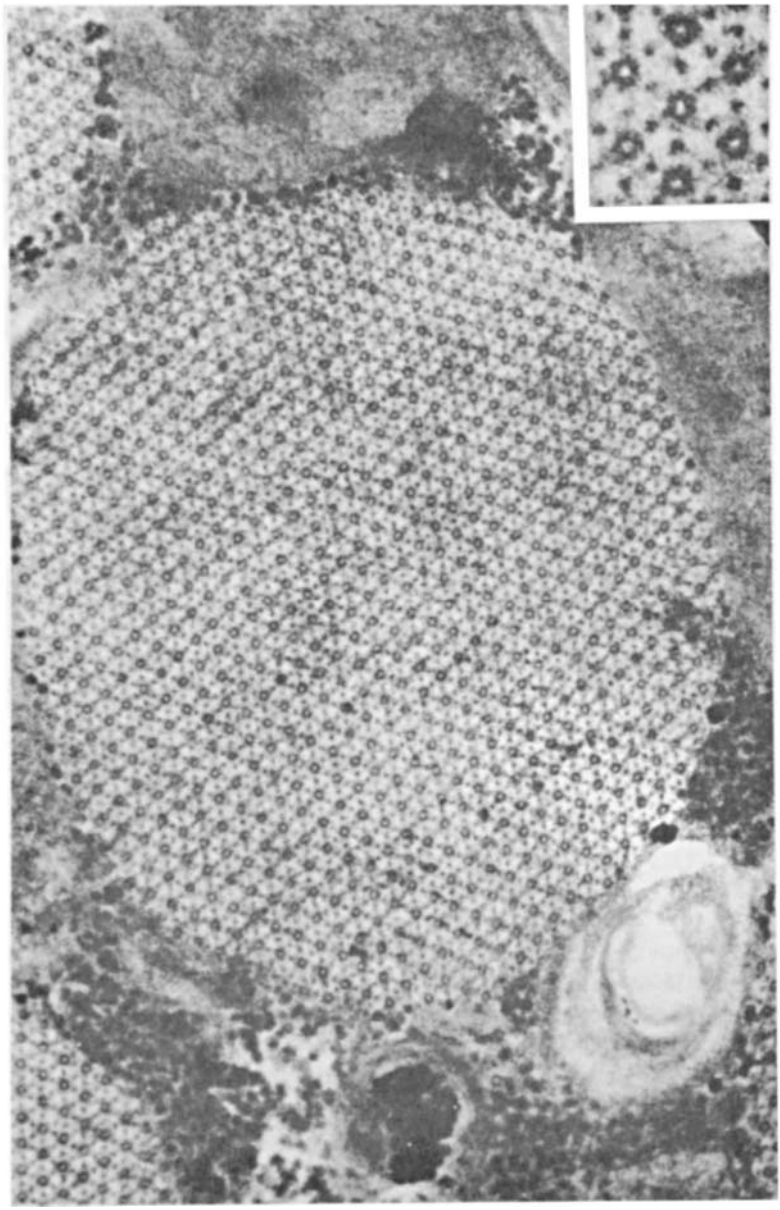


Figure 7. Electron micrograph of cross section of *Drosophila* leg muscle to show the hexagonal lattice arrangement of thick (myosin) and thin (actin) filaments. The thick filaments appear hollow in this insect muscle (*see insert*).

and elasticity. For details concerning the molecular biochemistry of muscle contraction see Alberts et al. (1989), Herrmann (1989) and Part IV of this modeule. Myosin motors in nonmuscle cells are reviewed by Spudich (1989).

THE MICROTRABECULAR LATTICE

The question has been raised as to whether the three principal types of filaments of the cytoskeleton and the intracellular membranes can account for all of the observed phenomena in the cytoplasm, or whether there is an additional order of compartmentalization and structural organization. Porter and his associates (Wolosewick and Porter, 1979; Porter, 1984) have presented evidence for the existence of a dynamic meshwork of fibers, 5–10 nm in diameter, based on high-resolution/high-voltage electron microscopy and the use of specimen preparation methods that minimize artifacts. However, there is still a lack of general agreement about the existence of the microtrabecular lattice. In support of the concept are observations that the lattice can be found in cells prepared for electron microscopy by different methods and that it undergoes regular alterations in response to various treatments. Other investigators have suggested that the microtrabecular lattice is an artifact produced by the coagulation of soluble proteins in the cytosol during fixation to form fibrous deposits (Ris, 1985). Several hundred proteins can be fractionated from the cytosol, but none of these appears to be a specific marker for the microtrabecular lattice (see Herrmann, 1989). Functionally, a microtrabecular lattice, if it exists, could provide a large intracellular surface for the binding of water and small molecules and for the selective adsorption and release of macromolecules (Gershon et al., 1985).

CONCLUDING REMARKS ON THE ORGANIZATION AND POLYFUNCTIONALITY OF THE CYTOSKELETON

Thus far, microtubules and actin filaments and their associated proteins have been discussed to advantage as independent cytoskeletal components. In actual fact, all of the components of the cytoskeleton (including intermediate filaments) are precisely integrated with one another (Langford, 1995), as well as with various cytoplasmic organelles, the nuclear membrane, the plasma membrane, and the extracellular matrix. In its totality the cytoskeleton subserves many coordinated and regulated functions in the cell:

1. Cell migration and cytoplasmic movement involve predominantly actin filaments in the locomotion of neutrophilic granulocytes, both actin filaments and microtubules in the elongation of neuronal growth cones and migration of neurites, and both actin and myosin in cytokinesis and the contraction of skeletal and cardiac muscle.

2. The cytoskeleton undergoes extensive reorganization during mitosis, and is responsible for the equipartition of a diploid set of chromosomes to each daughter cell (McIntosh and Koonce, 1989; Wadsworth, 1993).
3. The cytoskeleton is involved in the maintenance of cell shape and cytoplasmic processes (e.g., microvilli). In polarized epithelial cells, distinct cytocortical cytoskeletal complexes are associated with the apical and basal-lateral domains of the plasma membrane (Rodriguez-Boulan and Nelson, 1989; Mays et al., 1994).
4. The cytoskeleton is involved in the movement and positioning of cytoplasmic organelles (Cole and Lippincott-Schwartz, 1995).
5. Components of the cytoplasmic cytoskeleton (particularly microtubules) appear to provide specific "tracks" for the sorting, targeting, and transport of vesicles from the Golgi apparatus to specific plasma membrane domains. In polarized epithelial cells, the tracks from the Golgi apparatus to the apical or basal-lateral regions of the plasma membrane may be different, and these differences may be recognized by domain-specific receptors in the membrane of the transport vesicle, or the recognition event may occur between the vesicle and the specific plasma membrane domain (Rodriguez-Boulan and Nelson, 1989). Similarly, the cytoskeleton directs and regulates primary and secondary cell wall formation in plants (Tiwari et al., 1984; Lloyd, 1988).
6. The membrane cytoskeleton (inclusive of the submembranous actin-spectrin network) may function in the establishment and maintenance of restricted domains of specific proteins on the plasma membrane of polarized epithelial cells (Rodriguez-Boulan and Nelson, 1989; Mays et al., 1994).
7. The cytoskeleton is involved in anterograde and retrograde axoplasmic transport (Hollenbeck, 1989; Coy and Howard, 1994).
8. The cytoskeleton enables cultured fibroblasts to respond to the physical characteristics of a solid surface by the formation of adhesion plaques and stress fibers (Burridge et al., 1988).
9. The organization of the cytoskeleton is responsive to cell-to-cell signaling, which can cross cell membranes. This capability is important for the maintenance of histological structure in adult tissues and for embryogenesis (Fleming and Johnson, 1988). For example, the folding of sheets of epithelia, which is a fundamental process in early animal development, may involve the contraction of apical bundles of actin filaments and myosin in the adhesion belt of a single cell and then the propagation of tension to neighboring cells to initiate similar contractions by them (see Alberts et al., 1988).
10. A family of transmembrane proteins referred to as integrins is capable of interactions with extracellular ligands as well as with the cytoskeleton. These have been implicated in cell-to-cell interactions through the extracel-

lular matrix. When antibodies to integrins are used to block the binding with their extracellular ligands, cell migration and the outgrowth of neurites are interfered with (reviewed by Fessler and Fessler, 1989).

11. It is of interest that proteins termed "motility factors" (55–70 kD) are secreted by fetal cells and some tumor cells. These proteins act as autocrine factors and stimulate rapid movement by these cells. Motility factors induce the formation of cell processes that are packed with actin filaments and have an increased number of receptors for the matrix proteins laminin and fibronectin. The latter enhance the ability of the cells to bind to the extracellular matrix. Thus, it is likely that motility factors influence the organization of the cytoskeleton through changes taking place at the cell surface (reviewed by Warn and Dowrick, 1989).

Only fragmentary information is available concerning the factors that regulate the expression of the multigene families for cytoskeletal proteins and their associated proteins and the posttranslational modification and integration of these gene products to subserve complex mechanisms of specialized cell functions. The formation of adhesion plaques, cell locomotion, and phagocytosis are examples that readily come to mind. Ben-Ze'ev (1986) has discussed experimental evidence derived from mammalian cell cultures that suggests that changes in cell shape and cytoplasmic organization, acting through the cytoskeleton, can serve as major regulators of growth, gene expression, and cellular differentiation. How events taking place at the cell surface and within the cytoskeletal framework can influence changes at the level of the nucleus and gene regulation is still a matter for speculation. In a general discussion concerning the mechanisms by which ligand- and cell-specific transcriptional responses may be generated by extracellular signals, Herschman (1989) has suggested the following possibilities:

1. The induction of expression by a variety of extracellular signals of a relatively small set of common primary response genes.
2. Quantitative differences in expression of primary response genes to different extracellular signals.
3. Restrictive cell-type-specific expression of primary response genes.
4. Restrictive cell-type-specific expression of secondary response genes.
5. Posttranslational modifications of primary response gene products.

The possible combinations generated by these mechanisms could be sufficient to account for ligand-specific and cell-specific biological responses, notwithstanding the limited number of second messengers that are available for the transcription of the primary response genes (Herschman, 1989). It is apparent that much remains to be done to understand the complexity of the cytoskeleton and its interactions within cells and across cell membranes.

ACKNOWLEDGMENT

We thank Mr. George Braybrook and Mr. David Hildebrandt for the preparation of the diagrams, Ms. A. Cowan for the typing of the manuscript, and Mr. R. Bhatnagar for his invaluable help in the preparation of this manuscript.

REFERENCES

In order to provide students with a manageable list of references we have resorted to the use of citations from textbooks and reviews. We therefore apologize to the large number of original investigators whose publications are not cited. Some recent original papers have been included to provide current information.

Afzelius, B. (1985). The immotile-cilia syndrome: A microtubule-associated defect. CRC Crit. Rev. Biochem. 19, 63–87.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J.D. (1989). Molecular Biology of the Cell (2nd ed.). Garland, New York.

Allen, R.D. (1987). The microtubule as an intracellular engine. Sci. Am. 256, 42–49.

Amos, L.A. (1989). Brain dynein crossbridges microtubules into bundles. J. Cell Sci. 93, 19–28.

Amos, L.A. (1995). The microtubule lattice—20 years on. Trends Cell Biol. 5, 48–51.

Aoki, C. & Siekevitz, P. (1988). Plasticity in brain development. Sci. Am. 259, 56–64.

Asai, D.J. & Brokaw, C.J. (1993). Dynein heavy chain isoforms and axonemal motility. Trends Cell Biol. 3, 398–402.

Ault, A. & Rieder, C.L. (1994). Centrosome and kinetochore movement during mitosis. Curr. Opin. Cell Biol. 6, 41–49.

Bennett, V. (1989). The spectrin-actin junction of erythrocyte membrane skeletons. Biochim. Biophys. Acta 988, 107–121.

Bennett, V. & Gilligan, D.M. (1993). The spectrin-based membrane skeletal and micron-scale organization of the plasma membrane. Ann. Rev. Cell Biol. 9, 27–66.

Bentley, D. & O'Connor, T.P. (1994). Cytoskeletal events in growth cone steering. Curr. Opin. Neurobiol. 4, 43–48.

Ben-Ze'ev, A. (1986). The relationship between cytoplasmic organization, gene expression and morphogenesis. Trends Biochem. Sci. 11, 478–481.

Bershadsky, A.D. & Vasiliev, J.M. (1980). The Cytoskeleton. Plenum, New York.

Bray, D. & Hollenbeck, P.J. (1988). Growth cone motility and guidance. Ann. Rev. Cell Biol. 4, 43–61.

Bray, D. & Vasiliev, J. (1989). Networks from mutants. Nature 338, 203–204.

Bray, D. & White, J.G. (1988). Cortical flow in animal cells. Science 239, 883–888.

Brinkley, B.R. (1985). Microtubule organizing centers. Ann. Rev. Cell Biol. 1, 145–172.

Burridge, K., Fath, K., Kelly, T., Nuckolls, G., & Turner, C. (1988). Focal adhesions: Transmembrane junctions between the extracellular matrix and the cytoskeleton. Ann. Rev. Cell Biol. 4, 487–525.

Cleveland, D.W. & Sullivan, K.F. (1985). Molecular biology and genetics of tubulin. Ann. Rev. Biochem. 54, 331–365.

Cole, N.B. & Lippincott-Schwartz, J. (1995). Organization of organelles and membrane traffic by microtubules. Curr. Opin. Cell Biol. 7, 55–64.

Condeelis, J. (1993). Life at the leading edge: The formation of cell protrusions. Ann. Rev. Cell Biol. 9, 41–44.

Coy, D.L. & Howard, J. (1994). Organelle transport and sorting in axons. Curr. Opin. Neurobiol. 4, 662–667.

Cyr, R.J. (1994). Microtubules in plant morphogenesis: Role of the cortical array. Ann. Rev. Cell Biol. 10, 153–180.

Dabora, S.L. & Sheetz, M.P. (1988). The microtubule-dependent formation of a tubulovesicular network with characteristics of the ER from cultured cell extracts. Cell 54, 27–35.

Darnell, J., Lodish, H., & Baltimore, D. (1986). Molecular Cell Biology. W.H. Freeman, New York.

Davies, K.A. & Lux, S.E. (1989). Hereditary disorders of the red cell membrane skeleton. Trends Genet. 5, 222–227.

Delaunay, J. & Dhermy, D. (1993). Mutations involving the spectrin heterodimer contact site: Clinical expression and alterations in specific function. Sem. Hematol. 30, 21–33.

Devreotes, P.N. (1988). Chemotaxis in eukaryotic cells: A focus on leukocytes and *Dictyostelium*. Ann. Rev. Cell Biol. 4, 649–686.

Dustin, P. (1980). Microtubules. Sci. Am. 243, 67–76.

Endow, S.A. & Titus, M.A. (1992). Genetic approaches to molecular motors. Ann. Rev. Cell Biol. 8, 29–66.

Fessler, J.H. & Fessler, L.I. (1989). Drosophila extracellular matrix. Ann. Rev. Cell Biol. 5, 309–339.

Fishkind, D.J. & Wang, Y-L. (1995). New horizons for cytokinesis. Curr. Opin. Cell Biol. 7, 23–31.

Fleming, T.P. & Johnson, M.H. (1988). From egg to epithelium. Ann. Rev. Cell Biol. 4, 459–485.

Fukui, Y., Lynch, T.J., Brzeska, H., & Korn, E.D. (1989). Myosin 1 is located at the leading edges of locomoting *Dictyostelium* amoebae. Nature 341, 328–331.

Gershon, N.D., Porter, K.R., & Trus, B.L. (1985). The cytoplasmic matrix: Its volume and surface area and the diffusion of molecules through it. Proc. Natl. Acad. Sci. USA 82, 5030–5034.

Gibbons, I.R. (1988). Dynein ATPases as microtubule motors. J. Biol. Chem. 263, 15837–15840.

Ginzburg, I., Behar, L., Littauer, U.Z., & Griffin, W.S.T. (1989). Differential localization of microtubules in cerebellar cells. J. Neurosci. 9, 1961–1967.

Giuliano, K.A. & Taylor, D.L. (1995). Measurement and manipulation of cytoskeletal dynamics in living cells. Curr. Opin. Cell Biol. 7, 4–12.

Goddard, R.H., Wick, S.W., Silflow, C.D., & Snustad, D.P. (1994). Microtubule components of the plant cell cytoskeleton. Plant Physiol. 104, 1–6.

Goedert, M., Spillantini, M.G., Jakes, R., Rutherford, D., & Crowther, R.A. (1989). Multiple isoforms of human microtubule-associated protein tau: Sequences and localization in neurofibrillary tangles of Alzheimer's disease. Neuron 3, 519–526.

Goldberg, D.J. & Burmeister, D.W. (1989). Looking into growth cones. Trends Neurosci. 12, 503–506.

Goodenough, U.W. (1989). Basal body chromosomes. Cell 55, 1–3.

Gorbsky, G.J. & Borisy, G.G. (1989). Microtubules of the kinetochore fiber turn over in metaphase but not in anaphase. J. Cell Biol. 109, 653–662.

Hall, J.L., Ramanis, Z., & Luck, D.J.L. (1989). Basal body/centriolar DNA: Molecular genetic studies in *Chlamydomonas*. Cell 59, 121–132.

Hall, A. (1994). Small GTP-binding proteins and the regulation of the actin cytoskeleton. Ann. Rev. Cell Biol. 10, 31–54.

Hardham, A.R. & Gunning, B.E.S. (1978). Structure of cortical microtubule arrays in plant cells. J. Cell Biol. 77, 14–34.

Heath, I.B. (1980). Variant mitoses in lower eukaryotes: Indicators of the evolution of mitosis? Intl. Rev. Cytol. 64, 1–80.

Herrmann, H. (1989). Cell Biology. Harper & Row, New York.

Herschman, H.R. (1989). Extracellular signals, transcriptional responses and cellular specificity. Trends Biochem. Sci. 14, 455–458.

Hirokawa, N., Pfister, K.K., Yorifuji, H., Wagner, M.C., Brady, S.T., & Bloom, G.S. (1989). Submolecular domains of bovine brain kinesin identified by electron microscopy and monoclonal antibody decoration. Cell 56, 867–878.

Hollenbeck, P.J. (1984). The transport and assembly of the axonal cytoskeleton. J. Cell Biol. 108, 223–227.

Holzbaur, E.L.F. & Vallee, R.B. (1994). Dyneins: Molecular structure and cellular function. Ann. Rev. Cell Biol. 10, 339–372.

Howard, J., Hudspeth, A.J., & Vale, R.D. (1989). Movement of microtubules by single kinesin molecules. Nature 342, 154–158.

Hoyt, M.A. (1994). Cellular roles of kinesin and related proteins. Curr. Opin. Cell Biol. 6, 63–68.

Huang, B. (1986). *Chlamydomonas reinhardtii*. A model system for genetic analysis of flagellar structure and motility. Intl. Rev. Cytol. 99, 181–215.

Huxley, H.E. & Hanson, J. (1960). The molecular basis of contraction in cross-straited muscle. In: Structure and Function of Muscle (Bourne, G.H., ed.), Vol. 1, pp. 183–227. Academic Press, New York.

Joshi, H.C. (1994). Microtubule organizing centers and γ-tubulin. Curr. Opin. Cell Biol. 6, 55–62.

Kalt, A. & Schliwa, M. (1993). Molecular components of the centrosome. Trends Cell Biol. 3, 118–128.

Kubai, D.F. (1975). The evolution of the mitotic spindle. Intl. Rev. Cytol. 43, 167–227.

Langford, G.M. (1995). Actin- and microtubule-dependent organelle motors: Interrelationship between the two motility systems. Curr. Opin. Cell Biol. 7, 82–88.

Lasek, R.J. & Brady, S.T. (1985). Attachment of transported vesicles to microtubules in axoplasm is facilitated by AMP-PNP. Nature 316, 645–647.

Lee, C., Ferguson, M., & Chen, L.B. (1989). Construction of the endoplasmic reticulum. J. Cell Biol. 109, 2045–2055.

Lee, J., Ishihara, A., & Jacobson, K. (1993). How do cells move along surfaces? Trends Cell Biol. 3, 366–370.

Lewis, S.A., Ivanov, I.E., Lee, G.-H., & Cowan, N.J. (1989). Organization of microtubules in dendrites and axons is determined by a short hydrophobic zipper in microtubule-associated proteins MAP2 and tau. Nature 342, 498–505.

Lloyd, C.W. (1987). The plant cytoskeleton: The impact of fluorescence microscopy. Ann. Rev. Plant Physiol. 38, 119–139.

Machesky, L.M. & Pollard, T.D. (1993). Profilin as a potential mediator of membrane-cytoskeleton communication. Trends Cell Biol. 3, 381–385.

McIntosh, J.R. & Koonce, M.P. (1989). Mitosis. Science 246, 622–628.

McIntosh, J.R. & McDonald, K.L. (1989). The mitotic spindle. Sci. Am. 261, 48–56.

Mandelkow, E. & Mandelkow, E.-M. (1995). Microtubules and microtubule-associated proteins. Curr. Opin. Cell Biol. 7, 72–81.

Mays, R.W., Beck, K.A., & Nelson, W.J. (1994). Organization and function of the cytoskeleton in polarized epithelial cells: a component of the protein sorting machinery. Curr. Opin. Cell Biol. 6, 16–24.

Mitchison, T.J. (1988). Microtubule dynamics and kinetochore function in mitosis. Ann. Rev. Cell Biol. 4, 527–549.

Mitchison, T.J. (1989). Polewards microtubule flux in the mitotic spindle: Evidence from photoactivation of fluorescence. J. Cell Biol. 109, 637–652.

Mitchison, T.J. (1990). The kinetochore in captivity. Nature 348, 14–15.

Mitchison, T.J. & Hyman, A. (1988). Kinetochores on the move. Nature 336, 200–201.

Mooseker, M.S. (1985). Organization, chemistry, and assembly of the cytoskeletal apparatus of the intestinal brush border. Ann. Rev. Cell Biol. 1, 209–241.

Murray, A.W. & Kirschner, M.W. (1989). Dominoes and clocks: The union of two views of the cell cycle. Science 246, 614–621.

Nelson, W.J. & Hammerton, R.W. (1989). A membrane-cytoskeletal complex containing Na^+, K^+-ATPase, ankyrin, and fodrin in Madin-Darby canine kidney (MDCK cells): Implications for the biogenesis of epithelial cell polarity. J. Cell Biol. 108, 893–902.

Nelson, W.J. & Veshnock, P.J. (1987). Ankyrin binding to ($Na^+ + K^+$)-ATPase and implications for the organization of membrane domains in polarized cells. Nature 328, 533–536.

Nelson, W.J. (1992). Regulation of cell surface polarity from bacteria to mammals. Science 258, 948–955.

Olmsted, J.B. (1986). Microtubule-associated proteins. Ann. Rev. Cell Biol. 2, 421–457.

Otto, J.J. (1994). Actin-bundling proteins. Curr. Opin. Cell Biol. 6, 105–109.

Paschal, B.M., Shpetner, H.S., & Vallee, R.B. (1987). MAP 1C is a microtubule-activated ATPase which translocates microtubules in vitro and has dynein-like properties. J. Cell Biol. 105, 1273–1282.

Paschal, B.M., Obar, R.A., & Vallee, R.B. (1989). Interaction of brain cytoplasmic dynein and MAP2 with a common sequence at the C terminus of tubulin. Nature 342, 569–572.

Pickett-Heaps, J.D. (1967). The effects of colchicine on the ultrastructure of dividing plant cells, xylem wall differentiation and distribution of cytoplasmic microtubules. Dev. Biol. 15, 206–236.

Pickett-Heaps, J.D., Tippit, D.A., & Porter, K.R. (1982). Rethinking mitosis. Cell 29, 729–744.

Pollard, T.D. & Cooper, J.R. (1986). Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. Ann. Rev. Biochem. 55, 987–1035.

Pollard, T.D., Almo, S., Quirk, S., Vinson, V., & Lattman, E.E. (1994). Structure of actin binding proteins: Insights about function at atomic resolution. Ann. Rev. Cell Biol. 10, 207–249.

Porter, K.R. (1966). Cytoplasmic microtubules and their functions. In: Principles of Biomolecular Organization: Ciba Foundation Symposia (Wolstenholme, G.E.W., & O'Connor, M., eds.), pp. 308–345. Little, Brown, Boston.

Porter, K.R. (1984). The cytomatrix: A short history of its study. J. Cell Biol. 99, 3S–12S.

Rauvala, H., Pihlaskari, R., Laitinin, J., & Merenmies, J. (1989). Extracellular adhesive molecules in neurite growth. Biosci. Rep. 9, 1–12.

Ris, H. (1985). The cytoplasmic filament system in critical point-dried whole mounts and plastic-embedded section. J. Cell Biol. 100, 1474–1487.

Rodriguez-Boulan, E. & Nelson, W.J. (1989). Morphogenesis of the polarized epithelial cell phenotype. Science 245, 718–725.

Schliwa, M. (1989). Head and tail. Cell 56, 719–720.

Schroer, T.A. (1994). Structure, function and regulation of cytoplasmic dynein. Curr. Opin. Cell Biol. 6, 69–73.

Sheetz, M.P., Turney, S., Qian, H., & Elson, E.L. (1989). Nanometre-level analyses demonstrates that lipid flow does not drive membrane glycoprotein movements. Nature 340, 284–288.

Shpetner, H.S. & Vallee, R.B. (1989). Identification of dyamin, a novel mechanochemical enzyme that mediates interactions between microtubules. Cell 59, 421–432.

Singer, S.J. & Kupfer, A. (1986). The directed migration of eukaryotic cells. Ann. Rev. Cell Biol. 2, 337–365.

Small, J.V. (1988). The actin cytoskeleton. Electron Miscrosc. Rev. 1, 155–174.

Small, J.V. (1995). Getting the actin filaments straight: nucleation-release or treadmilling? Trends Cell Biol. 5, 52–55.

Smith, R.S. (1980). The short term accumulation of axonally transported organelles in the region of localized lesions of single myelinated axons. J. Neurocytol. 9, 39–65.

Speicher, D.W. & Marchesi, V.T. (1984). Erythrocyte spectrin is comprised of many homologous triple helical segments. Nature 311, 177–180.

Spudich, J.A. (1989). In pursuit of myosin function. Cell Regul. 1, 1–11.

Stryer, L. (1988). Biochemistry (3rd edn.). W.H. Freeman, New York.

Sullivan, K.F. (1988). Structure and utilization of tubulin isotypes. Ann. Rev. Cell Biol. 4, 687–716.

Sun, H-Q., Kwiatkowska, K., & Yin, H.L. (1995). Actin monomer binding proteins. Curr. Opin. Cell Biol. 7, 102–110.

Tellman, R.L., Morton, D.J., & Clarke, F.M. (1989). A common theme in the amino acid sequences of actin and many actin-binding proteins. Trends Biochem. Sci. 14, 130–133.

Terasaki, M., Chen, L.B., & Fujiwara, K. (1986). Microtubules and the endoplasmic reticulum are highly interdependent structures. J. Cell Biol. 103, 1557–1568.

Tilney, L.G. & Inoue, S. (1982). Acrosomal reaction or *Thyone* sperm. 11. The kinetics and possible mechanism of acrosomal process elongation. J. Cell Biol. 93, 820–827.

Tilney, L.G., Bonder, E.M., Coluccio, L.M., & Moosekar, M.S. (1983). Actin from *Thyone* sperm assembles on only one end of an actin filament: A behavior regulated by profilin. J. Cell Biol. 97, 112–124.

Tiwari, S.C., Wick, S.M., Williamson, R.E., & Gunning, B.E. (1984). Cytoskeleton and integration of cellular function in cells of higher plants. J. Cell Biol. 99, 63S–69S.

Turner, G.R. & Tartakoff, A.M. (1989). The response of the Golgi complex to microtubule alterations: The roles of metabolic energy and membrane traffic in Golgi complex organization. J. Cell Biol. 109, 2081–2088.

Vale, R.D., Reese, T.S., & Sheetz, M.P. (1985a). Identification of a novel force generating protein, kinesin, involved in microtubule-based motility. Cell 42, 39–50.

Vale, R.D., Schnapp, B.J., Mitchison, T., Steuer, E., Reese, T.S., & Sheetz, M.P. (1985b). Different axoplasmic proteins generate movement in opposite directions along microtubules in vitro. Cell 43, 623–632.

Vallee, R.B., Wall, J.S., Paschal, B.M., & Shpetner, H.S. (1988). Microtubule-associated protein 1C from brain is a two-headed, cytosolic dynein. Nature 332, 561–563.

Verma, R.S. (1990). The Genome. VCH Publishers, New York.

Wadsworth, P. (1993). Mitosis: spindle assembly and chromosome motion. Curr. Opin. Cell Biol. 5, 123–128.

Warn, R.M. & Dowrick, P. (1989). Motility factors on the march. Nature 340, 186–187.

Warner, F.D. & Mitchell, D.R. (1981). Polarity of dynein-microtubule interactions *in vitro*: Cross-bridging between parallel and antiparallel microtubules. J. Cell Biol. 89, 35–44.

Way, M. & Weeds, A. (1990). Cytoskeletal ups and downs. Nature 344, 292–294.

Wolosewick, J.J. & Porter, K.R. (1979). Microtrabecular lattice of the cytoplasmic ground substance. Artifact or reality. J. Cell Biol. 82, 114–139.

Zuber, M.X., Goodman, D.W., Karns, L.R., & Fishman, M.C. (1989). The neuronal growth-associated protein GAP-43 induces filopodia in non-neuronal cells. Science 224, 1193–1195.

This Page Intentionally Left Blank

Chapter 2

Actin Polymerization: Regulation by Divalent Metal Ion and Nucleotide Binding, ATP Hydrolysis and Actin Binding Proteins

MARIE-FRANCE CARLIER and
DOMINIQUE PANTALONI

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 43–57.

ISBN: 1-55938-808-0

INTRODUCTION

Actin filaments are major dynamic components of the cytoskeleton of eukaryotic cells. Assembly of filaments from monomeric actin occurs with expenditure of energy, the tightly bound ATP being irreversibly hydrolyzed during polymerization. This dissipation of energy perturbs the laws of reversible helical polymerization defined by Oosawa and Asakura (1975), and affects the dynamics of actin filaments. In 1976, Wegner showed that the free energy of ATP hydrolysis was used in actin assembly to establish an energetic difference (i.e., different critical concentrations for assembly) between the two ends of the filaments. We have additionally shown that ATP hydrolysis destabilizes actin-actin interactions in the filament. The destabilization is linked to the liberation of P_i that follows cleavage of the γ-phosphate. P_i release therefore plays the role of a conformational switch. Because ATP hydrolysis is uncoupled from polymerization, the nucleotide content of the filaments changes during the polymerization process, and filaments grow with a stabilizing cap of terminal ADP-P_i subunits. The fact that the dynamic properties of F-actin are affected by ATP hydrolysis results in a nonlinear dependence of the rate of filament elongation on monomer concentration.

Possible modes of regulation of filament assembly may be anticipated from the basic properties of actin. We have shown that the tightly bound divalent metal ion (Ca^{2+} or Mg^{2+}) interacts with the β- and γ-phosphates of ATP bound to actin, and that the Me-ATP bidentate chelate is bound to G-actin in the Λ configuration. The nature of the bound metal ion affects the conformation of actin, the binding kinetics of ATP and ADP, and the rate of ATP hydrolysis.

In living cells, actin binding proteins modify the assembly properties of actin in a variety of fashions. Some proteins, like filament bundling proteins, or filament cross-linking proteins, affect the spatial organization or polymorphism of filaments, and play a structural role in organizing the architecture of the cytoplasm. Other proteins play an effective role in the dynamics of filament assembly, by affecting the monomer-polymer steady-state; these are barbed end capping proteins which block the fast elongating barbed end, hence increase the critical concentration for F-actin assembly up to that of the pointed end (Bonder et al., 1983; Coué and Korn, 1985; Pollard and Cooper, 1986 for review). Profilin, on the other hand, forms a 1:1 complex with G-actin which can participate in actin assembly at the barbed end (Pollard and Cooper, 1984). Hence, profilin energetically facilitates actin assembly

and decreases the critical concentration at the barbed end in the presence of ATP (Pantaloni and Carlier, 1993). Barbed end capping proteins and profilin therefore govern the steady-state of actin assembly in two opposite ways. By changing the concentration of G-actin in the cell, these proteins in turn affect the concentration of all non-covalent complexes of G-actin with monomer-binding proteins (also called G-actin sequestering or G-actin buffering factors) like thymosin β_4 and its variants. In motile cells, a large pool of actin ($\sim 10^{-4}$ M) is maintained unpolymerized by interaction with these factors. This reservoir is used for rapid actin assembly in response to extracellular stimuli (Cooper, 1991).

In conclusion, the kinetic and steady-state basic parameters for pure actin assembly are finely modulated by accessory factors, themselves connected to signal transduction pathways, to operate the changes in actin assembly that trigger cell shape changes and movement. Finally, the role of G-actin polymerizing proteins may be crucial in defining the patterns of filament assembly. The myosin head (myosin subfragment-1) may be considered as a model actin polymerizing protein that may be the closest model to the short-tailed myosin I family. The mechanism of assembly of decorated filaments from G-actin and myosin subfragment-1 has therefore been examined.

ATP HYDROLYSIS LINKED TO ACTIN POLYMERIZATION PERTURBS THE THERMODYNAMICS OF REVERSIBLE POLYMERIZATION

ATP Hydrolysis Is Linked to the Destabilization of Actin Filaments and Microtubules

Monomeric actin binds ATP very tightly with an association constant K_a of 10^{10} M^{-1} in low ionic strength buffers in the presence of Ca^{2+} ions. A polymerization cycle involves addition of the ATP-monomer to the polymer end, hydrolysis of ATP on the incorporated subunit, liberation of P_i in solution, and dissociation of the ADP-monomer. Exchange of ATP for bound ADP occurs on the monomer only, and precedes its involvement in another polymerization cycle. Therefore, monomer-polymer exchange reactions are linked to the expenditure of energy: exactly one mol of ATP per mol of actin is incorporated into actin filaments. As a result, up to 40% of the ATP consumed in motile cells is used to maintain the dynamic state of actin. Thus, it is important to understand how the free energy of nucleotide hydrolysis is utilized in cytoskeleton assembly.

The critical concentration, that is, the monomer $\Leftrightarrow$ polymer equilibrium dissociation constant for polymerization of ADP-actin, is 25-fold that for polymerization of ATP-actin. However, in both cases the filament is made of F-ADP subunits, and the rate constant for association of ADP-actin to filament ends is only 2.5-fold lower than the rate constant for association of ATP-actin. In the absence of free ATP, the

1:1 ATP-actin complex can polymerize, but the polymer once formed spontaneously depolymerizes. Depolymerization stops when the concentration of ADP-monomer in the medium reaches the value of the critical concentration for polymerization of ADP-actin (for review see Korn et al., 1987; Carlier, 1991).

The above observations are inconsistent with a simple two-state polymerization model within which only two species, ATP-G-actin and ADP-F-actin, coexist in solution.

Thermodynamic and Kinetic Parameters for Reversible Polymerization (Oosawa's Law)

The theory of reversible helical polymerization of proteins has been fully described by Oosawa and Asakura (1975). The following equation describes polymer growth:

$$J(c) = dc/dt = k_+ [P] c - k_- [P] \quad (1)$$

where J(c) is the rate of polymer growth; c and P represent the concentrations of monomer and polymer elongating sites, respectively; k_+ and k_- are the rate constants for monomer association to and dissociation from polymer ends. According to equation (1), k_+ and k_- can easily be derived from the linear dependence of J(c) on c, and the critical concentration $c_c = k_-/k_+$ defined as the monomer concentration at which J(c) = 0.

Thermodynamic and Kinetic Parameters for Mg-ATP-Actin Polymerization

Polymer growth J(c) showed nonlinear monomer concentration dependence in the presence of ATP (Carlier et al., 1984), while in the presence of ADP, the plot of J(c) versus monomer concentration for actin was a straight line, as expected for reversible polymerization. The data imply that newly incorporated subunits dissociate from the filament at a slower rate than internal ADP-subunits; in other words, (a) the effect of nucleotide hydrolysis is to decrease the stability of the polymer by increasing k_-; and (b) nucleotide hydrolysis is uncoupled from polymerization and occurs in a step that follows incorporation of a ATP-subunit in the polymer. Newly incorporated, slowly dissociating, terminal ATP-subunits form a stable cap at the ends of F-actin filaments.

The above results demonstrate that ATP hydrolysis associated with actin and tubulin polymerization acts as a regulatory switch affecting the strength of protein-protein interactions. In this respect, this biological system appears similar to the G-proteins or other regulatory nucleoside triphosphatases. Indeed, one can consider that actin exists in two states: a noninteracting state in which ATP-G-actin does not hydrolyse nucleotide, and an interacting state in which F-actin hydrolyzes ATP in

a single turnover reaction, and nucleotide hydrolysis is linked to a weakening of actin-actin interactions in the polymer lattice.

In order to anticipate possible modes of regulation of cytoskeleton dynamics *in vivo*, it is necessary (a) to identify the kinetic intermediates involved in the polymerization process and to characterize their structural and functional properties; and (b) to define the essential elementary steps in the hydrolysis process.

Kinetic Steps in ATP Hydrolysis on F-Actin

ATP is hydrolyzed in at least two consecutive steps on F-actin, viz. cleavage of the γ-phosphoester bond, followed by P_i release, according to the following scheme:

$$\text{F-ATP} \Leftrightarrow \text{F-ADP-P}_i \Leftrightarrow \text{F-ADP} + P_i \qquad (2)$$

P_i release occurs at a relatively apparent slow rate ($k_{obs} = 0.005\ s^{-1}$), so that the transient intermediate F-ADP-P_i in which P_i is non-covalently bound, has a life time of 2–3 minutes (Carlier and Pantaloni, 1986; Carlier, 1987). While the γ-phosphate cleavage step is irreversible as assessed by ^{18}O exchange studies (Carlier et al., 1987), the release of P_i is reversible. Binding of $H_2PO_4^-$ (K_p 10^{-3}M) causes the stabilization of actin filaments and the rate of filament growth varies linearly with the concentration of actin monomer in the presence of P_i (Carlier and Pantaloni, 1988). Therefore, P_i release appears as the elementary step responsible for the destabilization of actin-actin interactions in the filament.

Probing the Intermediate ADP-P State on F-Actin Using Structural Analogues of P_i: AlF_4^- and BeF_3^-, H_2O

Fluoroaluminate and fluoroberyllate have been proposed to be structural analogues of inorganic phosphate, and have been shown to restore the functional properties of GTP-transducin when added to GDP-transducin (Bigay et al., 1987). These phosphate analogues bind to F-ADP-actin in competition with P_i, but with an affinity three orders of magnitude higher than that for P_i (Combeau and Carlier, 1988, 1989). The F-ADP-BeF_3^- filaments are extremely stable, the rate of dissociation of ADP-BeF_3 subunits from filament ends is very low; actually, even lower than the rate of dissociation of ADP-P_i subunits. In addition, some evidence suggests that the conformation of the F-ADP-BeF_3^- state is different from that of the F-ADP-P_i state. BeF_3^- and AlF_4^- bind to and dissociate from the ADP-polymer at very slow rates. All the above properties of BeF_3^- and AlF_4^- are very similar to those of vanadate with respect to other ATPases, e.g., vanadate binding to ADP-myosin (Goodno, 1979). This has led to the suggestion that BeF_3^- and AlF_4^- could mimic the ADP-P^* transition state, or at least adopt a configuration closer to that of bound ATP than to bound ADP-P_i. Similar results have been obtained for binding

of BeF_3^- and AlF_4^- to the bacterial F_1-ATPase (Dupuis et al., 1989) and to myosin (Phan and Reisler, 1992). Further, experiments should be aimed at understanding the structure of bound ADP-AlF_4^-; for example, by using the superhyperfine coupling of the Mn ESR signal with ^{17}O labeled ADP, or using ^{19}F NMR. Interestingly, AlF_4^- and BeF_3^- do not bind to monomeric G-ADP-actin which is not able to hydrolyze the nucleotide. This observation indicates that the environment of the γ-phosphoester bond of the nucleotide in the monomer and that of the polymerized states of actin are not the same.

The results of experiments obtained with phosphate analogues lead to another additional step in the kinetic scheme for the hydrolysis of ATP or GTP on F-actin or microtubules. This is described as follows:

$$\text{F-ATP} \rightarrow \text{F-ADP-P}^* \Leftrightarrow \text{F-ADP-}\ P_i \Leftrightarrow \text{F-ADP-P}_i \qquad (3)$$

In the above scheme, F-ADP-P* represents the transition state energetically identical to the F-ADP BeF_3^- state. The transition from F-ADP-P* to F-ADP-P_i would be slow and rate limiting for P_i release. In this scheme, which resembles the one proposed for ATP hydrolysis on myosin (e.g., Hibberd and Trentham, 1986), P_i binds to F-ADP in rapid equilibrium, while dissociation of P_i following cleavage of ATP is slow.

Mechanistic Models for ATP Hydrolysis in F-Actin Assembly

ATP may potentially be hydrolyzed in several ways following the incorporation of an ATP-actin subunit into the filament: the rate of ATP hydrolysis may be independent of the nature of the nucleotide bound to neighboring subunits (ATP or ADP), which can be called random hydrolysis, or it may be affected by induced-fit, viz. by the conformation of neighboring subunits, i.e., by the bound nucleotide. An extreme case is the one where hydrolysis occurs at a much faster rate on an ATP-subunit distally adjacent to an ADP-subunit: in this vectorial hydrolysis model, ATP hydrolysis occurs essentially at the ATP cap/ADP core boundary which migrates distally like a zipper at a constant rate. This latter theoretical model appears to adequately account for the data obtained with MgATP-actin (Carlier et al., 1986, 1987) with the additional formation of new ATP/ADP boundaries taking place at a high rate of filament growth, via random hydrolysis in long stretches of rapidly assembled F-ATP-actin. In contrast, the data obtained with Ca-ATP-actin are well described by a model of random hydrolysis of ATP on any F-ATP subunit independently of the nature of the neighboring subunit. Hence, the cap of F-ATP is larger on filaments growing from Ca-ATP-actin than on filaments growing from MgATP-actin, other medium conditions being the same (0.1 M KCl).

Structural Change of F-Actin Linked to ATP Hydrolysis

The results of thermodynamic and kinetic studies carried out on actin and tubulin polymerization indicate that a structural change of the polymer is linked to P_i release. The nature of this change, though not yet known, is a challenging issue. In a recent study combining electron microscopy and image reconstruction from negatively stained F-ADP and F-ADP-BeF_3^- filaments, a structural change localized in subdomain two of the actin subunit has been detected (Orlova and Egelman, 1992). Evidence for different structural states of the filament in the F-ADP, F-ADP-P_i, and F-ADP-BeF_3^- states has also been obtained by a combination of cryoelectromicroscopy and solution low angle X-ray scattering techniques (Lepault et al., in press). This structural change is expected to be less spectacular than in the case of ras p21, because the conformation of the subunit is somewhat constrained in the NTP state by the structure of the polymer itself. The change in the coordination of the divalent metal ion following the release of P_i is likely to trigger this structural change.

In the three-dimensional structure of actin, the environment of the phosphate moiety of the nucleotide appears roughly the same when CaADP or CaATP is bound. This observation argues against two different conformations. The reason why this is so is unclear. However, it must be stressed that the three-dimensional structure is derived from X-ray diffraction of crystals of the DNaseI-actin complex, which, like G-actin, is unable to hydrolyze ATP. The conformation obtained may therefore correspond to G-actin frozen in the G-ATP state independently of the bound nucleotide. Structural studies in conjunction with site-directed mutagenesis experiments should eventually solve this problem.

THE CRITICAL CONCENTRATION INCREASES WITH THE NUMBER OF FILAMENTS IN THE PRESENCE OF ATP

In reversible polymerization, the critical concentration is equal to the equilibrium dissociation constant for polymer formation. This parameter is therefore independent of the number of polymers in solution. Confirmation comes from studying reversible polymerization of ADP-actin: when sonic vibration is applied to a solution of F-ADP-actin filaments at equilibrium with G-ADP monomers, no change is observed in the proportion of G- and F-actin (Carlier et al., 1985). Therefore, the only effect of sonic vibration is to increase the number of filaments without affecting the rates of monomer association to and dissociation from filament ends.

When sonic vibration is applied to a solution of F-actin in the steady state in the presence of ATP, the observed behavior of F-actin is strikingly different: fragmentation is accompanied by a rapid, partial depolymerizaton leading to a new steady-

state (Pantaloni et al., 1984). Further examination of this phenomenon shows that the extent of depolymerization is a function of the regime of fragmentation imposed by the sonication, i.e., of the number of ends maintained in solution. The fragmentation can actually be controlled at will using a time controller attached to the sonicator, which allows application of sonication periodically for short periods (say, 0.5 sec) separated by variable time intervals. The shorter the interval between two sonications, the smaller the average size of the fragments generated, i.e., the higher the number of filaments. Indeed, polymerization under sustained sonication can be understood as a polymerization with constant filament length, as opposed to seeded polymerization which develops with a constant number of filaments. Very simply, filaments are fragmented when their length exceeds a certain size, so that a parameter similar to a "generation time period" can be defined as for bacterial growth. This being so, it can be demonstrated that the polymerization curve under continuous sonication is symmetric with respect to the point of half polymerization, and can be described by the following equation (Carlier et al., 1985):

$$\ln [C_o - C(t)/(C(t) - C_c)] = (k_+/m)(C_o - C_c)(t - t_{1/2}) \quad (4)$$

where C_o, C_c, and $C(t)$ are the total actin concentration, the critical concentration and the monomer concentration at time t, respectively, and k_+ is the rate constant for monomer association with filament ends, and m is the average number of subunits of sonicated filaments (40–60 subunits). Polymerization under sonication of ADP-actin is adequately described by the above equation. In the presence of ADP, the same monomer-polymer equilibrium is reached with or without sonication, and the same critical concentration can be determined over a range of ADP-actin concentrations.

The situation is quite different when actin is polymerized under sonication in the presence of ATP. In this case, the polymerization curve cannot be described by equation (4). At a high actin concentration, overshoot polymerization kinetics are observed, with a maximum and subsequent decrease to a lower stable plateau (Carlier et al., 1985). The final amount of polymer is the same as that obtained when sonication is applied to F-actin that had polymerized spontaneously without sonication. Conversely, when sonication is stopped, repolymerization accompanies the spontaneous length redistribution to a population of less numerous, but longer filaments.

In summary, then, polymerization of ATP-actin under conditions of sonication displays two characteristic deviations from the simple law described by equation (4), which is only valid for reversible polymerization. These are: (a) overshoot polymerization kinetics, and (b) the steady-state amount of polymer formed decreases, or the steady-state monomer concentration increases, with the number of filaments. These two features are the direct consequence of ATP hydrolysis accompanying the polymerization of ATP-actin, as will be explained now.

Because ATP hydrolysis on F-actin takes place with a delay following the incorporation of ATP-subunits, and because in the transient F-ATP state filaments are more stable than in the final F-ADP state, polymerization under conditions of sonication can be complete, within a time short enough for practically all subunits of the filaments to be F-ATP. At a later stage, as P_i is liberated, the F-ADP filament becomes less stable and loses ADP-subunits steadily. The G-ADP-actin liberated in solution is not immediately converted into easily polymerizable G-ATP-actin, because nucleotide exchange on G-actin is relatively slow, and is not able to polymerize by itself unless a high concentration (the critical concentration of ADP-actin) is reached. Therefore, G-ADP-actin accumulates in solution. A steady-state concentration of G-ADP-actin is established when the rate of depolymerization of ADP-actin ($k_-[F]$) is equal to the sum of the rates of disappearance of G-ADP-actin via nucleotide exchange and association to filament ends. $[G\text{-}ADP]_{ss}$ in this scheme is described by the following equation (Pantaloni et al., 1984):

$$[G\text{-}ADP]_{ss} = \frac{k_{23}[F]}{k_{31} + k_{32}[F]} \tag{5}$$

where k_{23} and k_{32} are the rate constants for ADP-actin dissociation from and association to filament ends, k_{31} is the rate constant for nucleotide exchange on G-actin, and [F] the number concentration of filaments. At a very high filament concentration $[G\text{-}ADP]_{ss}$ reaches a higher limit, $[G\text{-}ADP]_{ss, \infty}$ equal to k_{23}/k_{32} which is the critical concentration for polymerization of ADP-actin. This point can also be experimentally verified (Pantaloni et al., 1984).

The fact that the concentration of G-actin at steady-state in the presence of ATP varies with the number of filaments may have some biological significance: indeed, in cells, large pools of G-ADP-actin may accumulate in regions where a large number of short filaments exist. This behavior is the direct consequence of two combined features of actin polymerization; namely, the hydrolysis of ATP, and the relatively slow rate of ATP exchange for ADP on G-actin.

It is likely that this regulation of the concentration of monomeric actin is itself modulated by the action of G-actin binding proteins, which may exhibit different affinities for G-ATP-actin and G-ADP-actin. For instance, $T\beta_4$ has a 50-fold higher affinity for G-ATP actin than for G-ADP-actin; in fact, $T\beta_4$ hardly binds to G-ADP actin ($K_D \sim 100\ \mu M$). Hence, the presence of $T\beta_4$ increases the proportion of ATP-bound G-actin in cells, in the presence of mixtures of ATP and ADP (Carlier et al., 1993). Profilin, another G-actin binding protein, also has a 7–10-fold higher affinity for G-ATP-actin than for G-ADP-actin (Pantaloni and Carlier, 1993), while gelsolin, a barbed end capping protein, has similar affinities for G-ATP- and G-ADP-actin, but its binding kinetics may be different for ATP- and ADP-G-actin (Laham et al., 1993).

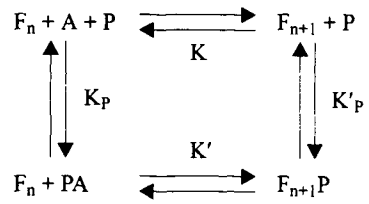
STEREOCHEMISTRY OF NUCLEOTIDE BINDING TO ACTIN AND TUBULIN: ROLE OF DIVALENT METAL ION IN NUCLEOTIDE BINDING AND HYDROLYSIS

Actin binds ATP very tightly in the presence of a divalent metal ion that can be Ca^{2+} or Mg^{2+}. The exchange-inert analogue of Mg-ATP, β,γ-CrATP, can displace both tightly bound nucleotide and divalent metal ion from G-actin. The inference drawn is that the tightly bound metal ion interacts with the β- and γ-phosphate of ATP in the nucleotide site (Valentin-Ranc and Carlier, 1989). The conformation, ability to polymerize, and rate of ATP hydrolysis and nucleotide exchange differ when CaATP or MgATP is bound to actin (see Carlier [1991] for a review, and Kinosian et al. (1993) for most recent results). In particular, the hydrolysis of MgATP is fast, whereas the hydrolysis of CaATP is slow (Carlier et al., 1986) and corresponds to ATP hydrolysis on divalent cation-free actin (Valentin-Ranc and Carlier, 1991). It thus appears that only Mg^{2+} is able to play an effective role in catalysis. CrATP has also been useful for probing the stereochemistry of ATP binding. The metal-ATP chelate is bound in the Λ configuration, which is confirmed by the three-dimensional structure of actin at atomic resolution (Kabsch et al., 1990). CrATP is hydrolyzed on F-actin upon polymerization: the hydrolysis product is Cr-ADP-P_i that remains bound to F-actin and P_i is not released. The resulting F-CrADP-P_i-actin filament shows a high stability, as expected.

After metal-ATP hydrolysis on F-actin. only P_i is released in solution, and the β- monodentate metal-ADP remains bound to F-actin. The F CrADP-P_i filaments are very stable.

HOW PROFILIN CONTROLS MONOMER-POLYMER STEADY-STATE AND PROMOTES ACTIN FILAMENT ASSEMBLY IN THE PRESENCE OF THYMOSYN β4

Profilin is a small (12–15 kDa) protein present in amoebae and animal cells (two profilin species, I and II, exist in *Acanthamoeba* and *Dictyostelium*, but whether only one species exists in vertebrates is not clear). Profilin was the first protein, discovered by Lindberg and his coworkers (Carlsson et al., 1977), that specifically bound to monomeric actin. Profilin was initially thought to be a G-actin sequestering agent that prevented actin assembly. However, the kinetics of filament elongation (Pollard and Cooper, 1984; Pring et al., 1992) could not be accounted for in terms of this simple function alone. Rather, results indicated that the profilin-actin complex actively participated in filament assembly at the barbed ends. That is, in the presence of profilin, a filament of n subunits, F_n, gains one subunit at the barbed end via either one of the two pathways: (a) association of G-actin (A) to the barbed end, and (b) association of the profilin-actin complex (PA) with the barbed end, followed by dissociation of P from this end. This is described by the following scheme:



$$K = \frac{k_-}{k_+}, \qquad K' = \frac{k'_-}{k'_+}$$

$$K_p = \frac{k_{-p}}{k_{+p}}, \qquad K'_p = \frac{k'_p}{k'_{+p}}$$

When operating at equilibrium, thermodynamic detailed balance would require the two elongation pathways to be energetically equivalent, i.e., $K = K'$. K_P/K'_P (detailed balance). This equilibrium model was first thought to account for the effect of profilin at the barbed end (Pring et al., 1992). In the above model, it is implicit that the critical concentration of G-actin, i.e., the value of K, remains unchanged in the presence of profilin. The demonstration (Pantaloni and Carlier, 1993) that the steady-state concentration of G-actin in fact decreases in the presence of profilin means that profilin energetically facilitates actin assembly. The decrease in concentration of G-actin in the steady-state can be accounted for quantitatively by solving the steady-state equation stating that the rate of elongation equals zero at steady-state. This is: $v_e = k_+ [F] . [A] + k'_+[F][PA] - k_-[F] - k'_-[FP]$

The fact that profilin controls the steady state of actin assembly (i.e., regulates the steady-state concentration of G-actin) has an immediate important consequence for the action of profilin in cell motility. This can be stated as follows: the concentration of all non-covalent complexes of G-actin with monomer sequestering agents such as $T\beta_4$ decreases as the concentration of G-actin decreases, in accordance with the law of mass action. The decrease in the pool of G-actin-$T\beta_4$ complex is paralleled by an increase in the pool of F-actin. In other words, profilin promotes F-actin assembly at the barbed ends in the presence of a reservoir of unpolymerized, sequestered G-actin (Pantaloni and Carlier, 1993). In this view, the *in vivo* function of profilin in the rapid assembly of actin (e.g., following neutrophil platelet or neutrophil activation) and chemotactic stimulation of *Dictyostelium*, or in promoting actin assembly leading to propulsion of *Listeria*, can be readily explained. Clearly, the living cell has the power of controlling the monomer-polymer steady state of actin assembly in two opposite ways, namely using either capping proteins or profilin as follows. Profilin causes a decrease in the steady-state concentration of G-actin, which leads to enhanced actin assembly in the presence of a reservoir of unpolymerized actin, whereas capping proteins cause an increase in the steady-state concentration of G-actin, which results in disassembly and an increase in sequestered monomeric actin. When barbed ends are capped, profilin is only a G-actin sequestering protein. Thus the possibility seems real that interaction of these proteins with the signaling pathway may well provide the basis of a link "from signal to pseudopod" (Stossel, 1989) in motility events.

MYOSIN SUBFRAGMENT-1 INDUCED POLYMERIZATION OF G-ACTIN

The myosin head has long been shown to induce, even in low ionic strength buffers, polymerization of G-actin into decorated F-actin-S_{-1} filaments that exhibit the classical arrowhead structure (Miller et al., 1988 and older references therein). However, to date, the molecular mechanism of this polymerization process remains unknown.

In an effort to understand how actin-actin interactions might be affected by the binding of the myosin head, and in order to gain more insight into the nature of the actin-myosin interface, we have investigated the nature of the kinetic actin-myosin intermediates involved in the process of S_1-induced polymerization of G-actin. For this purpose, a variety of fluorescent probes (e.g., pyrene, NBD, AEDANS) have been covalently attached to the C-terminus of G-actin to probe the G-actin-S_1 interaction under conditions of tightest binding, i.e., in the absence of ATP.

Myosin Subfragment-1 Interacts With Two G-Actin Molecules

The change in intensity of pyrenyl-actin fluorescence (Valentin-Ranc et al., 1991), as well as the change in anisotropy of fluorescence of AEDANS-labeled G-actin (Valentin-Ranc and Carlier, 1992) upon addition of increasing amounts of S_1 yielded titration curves incompatible with the formation of the 1:1 G-actin-S_1 complex previously proposed (Chaussepied and Kasprzak, 1989), but only compatible with the formation of a G_2S ternary complex. The S_1A_1 isomer of S_1 showed a higher affinity than S_1A_2 in this complex. It is plausible that the two G-actin molecules present in the G_2S complex have the same orientation with respect to S_1, as the two actin subunits that appear to interact with the myosin head in the rigor state (Milligan et al., 1990), i.e., the two actin monomers which are in contact via longitudinal bonds along the long pitch helix of the actin filament. In this view, one would expect that subdomain-2 of one actin molecule in G_2S is in contact with subdomain-1 of the other G-actin molecule. Conformational changes in subdomain-2 can actually be monitored by limited proteolysis. Using subtilisin, α-chymotrypsin, trypsin, and ArgC protease, we are able to demonstrate that binding of S_1 to G-actin induces the same changes in subdomain-2 as those found in the G → F transition (Fievez and Carlier, 1993).

Covalent crosslinking is traditionally a useful tool to monitor the actin-S_1 interface (Audemard et al., 1988 for review). Covalent crosslinking of G-actin-S_1 complexes using the zero-length crosslinker EDC has revealed that the G-actin-S_1 electrostatic contacts were very similar if not identical to the F-actin-S_1 contacts present in the rigor filament (Combeau et al., 1992). As in F-actin-S_1 complexes, only one G-actin could be crosslinked to S_1. On the other hand, the main difference between G-actin-S_1 and F-actin-S_1 is the proximity of the actin C-terminal cys374 to S_1 in G-actin-S_1, but not in F-actin-S_1. Using pPDM (spanning 10 Å), or by

photoirradiation of benzophenone-G-actin (prepared by reacting benzophenone maleimide with actin) in complex S_1, a 1:1 crosslinked complex of apparent molecular mass 195 kDa in SDS PAGE was obtained. The nature of the aminoacid of S_1 that can be crosslinked to G-actin is currently under investigation. This result shows that upon polymerization of actin, a change occurs in the environment of the C-terminal segment of actin. This is not unexpected since in the F-actin (and F-actin-S_1) state, cys374 can be crosslinked by pPDM to lys191 of the adjacent actin subunit along the short pitch helix (Elzinga and Phelan, 1984).

Oligomers of G-Actin and S_1 Are the Second Intermediates in F-Actin-S_1 Assembly

Light scattering, as well as pyrenyl actin or NBD-actin fluorescence changes, provide a convenient way in which to monitor the process of S_1-induced polymerization of G-actin. We have shown that actin-S_1 oligomers form rapidly (within 5 sec) following the formation of GS and G_2S complexes. Analysis of fluorescence data shows that the actin:S_1 molar ratio is 2:1 in these oligomers. The fluorescence of NBD- G-actin, which is not modified upon formation of GS and G_2S complexes, is increased ~ twofold in the oligomers. The results indicate that oligomers are assembled by condensation of G_2S units, and that new actin-actin interactions, in which hydrophobic contacts are involved, are formed upon oligomer assembly. It is proposed that these actin-actin interactions correspond to the lateral bonds between actin subunits in the filament, along the short pitch helix. Both light scattering and anisotropy of fluorescence measurements indicate that oligomers contain 2–4 G_2S units only. Formation of the decorated filament, in which the actin:S_1 molar ratio is 1:1, requires further endwise condensation of oligomers which results in the creation of new S_1 binding sites, due to the formation of new actin-actin longitudinal bonds. Increased binding of S_1 presumably accompanies the increase in stability of the final F-actin-S_1 product. The kinetic analysis of these elementary steps leading to the decorated filaments is currently underway.

CONCLUSION

The kinetics of F-actin-S_1 assembly from G-actin and S_1 via nucleation of actin filaments, followed by S_1 binding are not observed in a low ionic strength medium. Instead, the mechanism involves condensation of high affinity $(G\text{-actin})_2$ S_1 complexes rapidly preformed in solution. Assembly of F-actin-S_1 in the presence of $S_1 \geq$ G-actin is a quasi-irreversible process. This mechanism is therefore different from that involving the assembly of F-actin filaments, which is characterized by the initial, energetically unfavorable formation of a small number of nuclei representing a minute fraction of the population of actin molecules, followed by endwise elongation from G-actin subunits.

REFERENCES

Bigay, J., Deterre, P., Pfister, C., & Chabre, M. (1987). Fluoride complexes of aluminium or beryllium act on G-proteins as reversibly bound analogues of the γ-phosphate of GTP. EMBO J. 6, 2907–2912.

Bonder, E.M., Fishkind, D.J., & Mooseker, M.S. (1983). Direct measurement of critical concentrations and assembly rate constants at the two ends of an actin filament. Cell 34, 491–501.

Carlier, M.-F. (1987). Measurement of P_i dissociation from actin filaments following ATP hydrolysis using an enzyme-linked assay. Biochem. Biophys. Res. Comm. 143, 1069–1073.

Carlier, M.-F. (1989). Role of nucleotide hydrolysis in the dynamics of actin filaments and microtubules. Intl. Rev. Cytol. 115, 139–170.

Carlier, M.-F. (1991). Actin:protein structure and filament dynamics, J. Biol. Chem. 266, 1–4.

Carlier, M.-F. & Pantaloni, D. (1986). Kinetic evidence for F-ADP-P_i as a major transient in polymerization of ATP-actin. Biochemistry 25, 7789–7793.

Carlier, M.-F. & Pantaloni, D. (1988). Binding of P_i to F-ADP-actin and characterization of the F-ADP-P_i filament. J. Biol. Chem. 263, 817.

Carlier, M.-F., Pantaloni, D., & Korn, E.D. (1984). Evidence for an ATP cap at the ends of actin filaments and its regulation of the F-actin steady state. J. Biol. Chem. 259, 9983–9986.

Carlier, M.-F., Pantaloni, D., & Korn, E.D. (1985). Polymerization of ADP-actin and ATP-actin under sonication and characteristics of the ATP-actin equilibrium polymer. J. Biol. Chem. 260, 6565–6571.

Carlier, M.-F., Pantaloni, D., & Korn, E.D. (1986). The effects of Mg^{2+} at the high-affinity and low-affinity sites on the polymerization of actin and associated ATP hydrolysis. J. Biol. Chem. 261, 10785–10792.

Carlier, M.-F., Pantaloni, D., & Korn, E.D. (1987a). The mechanisms of ATP hydrolysis accompanying the polymerization of Mg-actin and Ca-actin. J. Biol. Chem. 262, 3052–3059.

Carlier, M.-F., Pantaloni, D., Evans, J.A., Lambooy, P.K., Korn, E.D., & Webb, M.R. (1987b). The hydrolysis of ATP that accompanies actin polymerization is essentially irreversible. FEBS Lett. 235, 211–214.

Carlier, M.F., Jean, C., Rieger, K., Lenfant, M., & Pantaloni, D. (1993). Modulation of the interaction between G-actin and thymosin β_4 by the ATP/ADP ratio: Possible implication in the regulation of actin dynamics. Proc. Natl. Acad. Sci. USA 90, 5034–5038.

Carlsson, L., Nyström, L.E., Sundkvist, I., Markey, F., & Lindberg, U. (1977). Actin polymerizability is influenced by profilin, a low molecular weight protein in non-muscle cells. J. Mol. Biol. 115, 465–483.

Chaussepied, P. & Kasprzak, A.A. (1989). Isolation and characterization of the G-actin-myosin head complex. Nature 342, 950–953.

Combeau, C. & Carlier, M.-F. (1988). Probing the mechanism of ATP hydrolysis on F-actin using vanadate and the structural analogs of phosphate BeF_3^- and AlF_4^-. J. Biol. Chem. 263, 17429–17434.

Combeau, C. & Carlier, M.-F. (1989). Characterization of the aluminium and beryllium fluoride species bound to F-actin and microtubules at the site of the γ-phosphate of the nucleotide. J. Biol. Chem. 264, 19017–19022.

Combeau, C., Didry, D., & Carlier, M.-F. (1992). Interaction between G-actin and myosin subfragment-1 probed by covalent crosslinking. J. Biol. Chem. 267, 14038–14046.

Cooper, J.A. (1991). The role of actin polymerization in cell motility. Ann. Rev. Physiology 53, 585–605.

Coué, M. & Korn, E.D. (1985). Interaction of plasma gelsolin with G-actin and F-actin in the presence and absence of Ca^{+2} ions. J. Biol. Chem. 260, 15033–15041.

Dupuis, A., Israel, J.-P., & Vignais, P.V. (1989). Direct identification of the fluoroaluminate and fluoroberyllate species responsible for inhibition of the mitochondrial F_1- ATPase. FEBS Lett. 255, 47–52.

Elzinga, M. & Phelan, J.J. (1984). F-actin is intermolecularly crosslinked by NN′ phenylenedimaleimide through lysine 191 and cysteine 374. Proc. Natl. Acad. Sci. USA 81, 6599–6603.

Fievez, S. & Carlier, M.-F. (1993). Conformational changes in subdomain-2 of G-actin upon polymerization into F-actin and upon binding myosin subfragment-1. FEBS Lett. 316, 186–190.

Goodno, C.C. (1979). Inhibition of myosin ATPase by vanadate ion. Proc. Natl. Acad. Sci. USA 76, 2620–2624.

Hibberd, M.G. & Trentham, D.R. (1986). Relationships between chemical and mechanical events during muscular contraction. Ann. Rev. Biophys. Chem. 15, 119–136.

Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F., & Holmes, K.C. (1990). Atomic structure of the actin: DNase I complex. Nature Lond. 347, 37–43.

Kinosian, H.J., Selden, L.A., Estes, J.E., & Gershman, L.C. (1993). Nucleotide binding to actin:cation dependence of nucleotide dissociation and exchange rates. J. Biol. Chem. 268, 8683–8691.

Korn, E.D., Carlier, M.-F., & Pantaloni, D. (1987). Actin polymerization and ATP hydrolysis. Science. Wash. 238, 638–644.

Laham, L.E., Lamb, J.A., Allen, P.G., & Janmey, P.A. (1993). Selective binding of gelsolin to actin monomers containing ADP. J. Biol. Chem. 268, 14202–14207.

Miller, L., Phillips, M., & Reisler, E. (1988). Polymerization of G-actin by myosin subfragment-1. J. Biol. Chem. 263, 1996–2002.

Milligan, R.A., Whittaker, M., & Safer, D. (1990). Molecular structure of F-actin and location of surface binding sites. Nature 248, 217–221.

Mornet, D., Bertrand, R., Pantel, P., Audemard, E., & Kassab, R. (1981). Structure of the actin-myosin interface. Nature 292, 301–310.

Oosawa, F. & Asakura, S. (1975). Thermodynamics of the polymerization of protein. Academic Press, London.

Orlova, A. & Egelman, E.H. (1992). Structural basis for the destabilization of F-actin by phosphate release following ATP hydrolysis. J. Mol. Biol. 227, 1043–1048.

Pantaloni, D., Carlier, M.-F., Coué, M., Lal, A.A., Brenner, S.L., & Korn, E.D. (1984). The critical concentration of actin in the presence of ATP increases with the number of filaments and approaches the critical concentration of ADP-actin. J. Biol. Chem., 259, 6274–6283.

Pantaloni, D. & Carlier, M.-F. (1993). How profilin promotes actin filament assembly in the presence of thymosin beta 4. Cell, 75, 1007–1014.

Phan, B. & Reisler, E. (1992). Inhibition of myosin ATPase by beryllium fluoride. Biochemistry 31, 4787–4792.

Pollard, T.D. & Cooper, J.A. (1984). Quantitative analysis of the effect of Acanthamoeba profilin on actin filament nucleation and elongation. Biochemistry 23, 6631–6641.

Pollard, T.D. & Cooper, J.A. (1986). Actin and actin binding proteins: A critical evaluation of mechanisms and functions. Ann. Rev. Biochem. 55, 987–1036.

Stossel, T.P. (1989). From signal to pseudopod: How cells control cytoplasmic actin assembly. J. Biol. Chem. 264, 18261–18264.

Stossel, T.P. (1993). On the crawling of animal cells. Science 260, 1086–1094.

Valentin-Ranc, C. & Carlier, M.-F. (1989). Evidence for the direct interaction between tightly bound divalent metal ion and ATP on actin. Binding of the Λ isomers of βγ-bidentate CrATP to actin. J. Biol. Chem. 264, 20871–20879.

Valentin-Ranc, C. & Carlier, M.-F. (1991). Role of ATP-bound divalent metal ion in the conformation and function of actin. J. Biol. Chem. 266, 7668–7674.

Valentin-Ranc, C. & Carlier, M.-F. (1992). Characterization of oligomers as kinetic intermediates in myosin subfragment-1 induced polymerization of G-actin. J. Biol. Chem. 267, 21543–21550.

Valentin-Ranc, C., Combeau, C., Carlier, M.-F., & Pantaloni, D. (1991). Myosin subfragment-1 interacts with two G-actin molecules in the absence of ATP. J. Biol. Chem. 266, 17872–17879.

Wegner, A. (1976). Head to tail polymerization of actin. J. Mol. Biol. 108, 139–150.

This Page Intentionally Left Blank

Chapter 3

Myosins

DONALD D. LORIMER and PRIMAL DE LANEROLLE

INTRODUCTION

The myosins are a superfamily of proteins that have the ability to convert energy released by ATP is hydrolysis into mechanical work. There are many forms of myosin, all of which have ATPase activity and an actin-binding site that is located

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 59–76.

ISBN: 1-55938-808-0

on a globular, head domain that is attached to a variety of tail structures. The myosin family can be divided into at least nine classes but the traditional classification has been based upon whether they form monomers or dimers. The two classes have historically been called myosin-I and myosin-II. The myosins-II are two-headed dimers that are able to form filaments, and are usually associated with skeletal, cardiac, and smooth muscles, although myosin-II is also found in non-muscle cells. Myosins-I are smaller and monomeric. Although myosins-I are incapable of forming filaments, they are able, in some cases, to bind to membranes. Myosin-I was first discovered in the single cell species *Acanthamoeba* but it is now recognized to be widely distributed in nonmuscle cells.

In this chapter we will discuss the various forms of myosin and the roles they play in living systems. We will compare and contrast the function and regulation of myosin activity in different cellular environments. Finally, we will examine the clinical aspects of myosin structure and function.

THE MYOSIN SUPERFAMILY

Since the discovery of myosin more than 50 years ago (Engelhardt and Ljubimova, 1939), at least nine different classes of myosins have been described (Figure 1). All myosins contain one or two heavy chains, and one or more lower molecular weight proteins known as light chains. The amino terminal of the heavy chain is folded into a globular, head domain that is connected to an alpha helical tail domain of variable length (Figure 2). The defining features of any myosin are actin binding and ATP hydrolysis, both of which occur in the head. Light chains bind to the head and neck region of the heavy chain, and the light chains are important in the regulation of the enzymatic properties of the myosin molecule (see under Regulation). All of the myosin heavy chains studied so far have one or more conserved amino acid sequences termed IQ sequences that are thought to be the sites of light chain binding (Cheney and Mooseker, 1992). This consensus sequence (IQxxxRGxxxR, where I is isoleucine, Q glutamine, R arginine and G glycine) has been useful in identifying members of the myosin superfamily. In contrast, analysis of the tail regions have demonstrated much greater variability among the members of the superfamily.

Most of the myosins (III–IX) described in Figure 1 have been identified only by their DNA sequences and their unique biochemical properties have not been determined nor have their physiological roles been examined. We shall therefore focus on the two classes of myosins, myosin-I and -II, that have been studied extensively. The myosin-II group has two major subgroups: the striated muscle myosins and the nonmuscle/smooth muscle myosins. They have similar structural characteristics but dissimilar enzymatic and regulatory properties. The myosin-I group contains at least three, and possibly more, subgroups: (a) the original *Acanthamoeba* myosin-I that has actin and membrane binding sites in its tail region; (b) a second group from *Dictyostelium* that contains a membrane binding domain

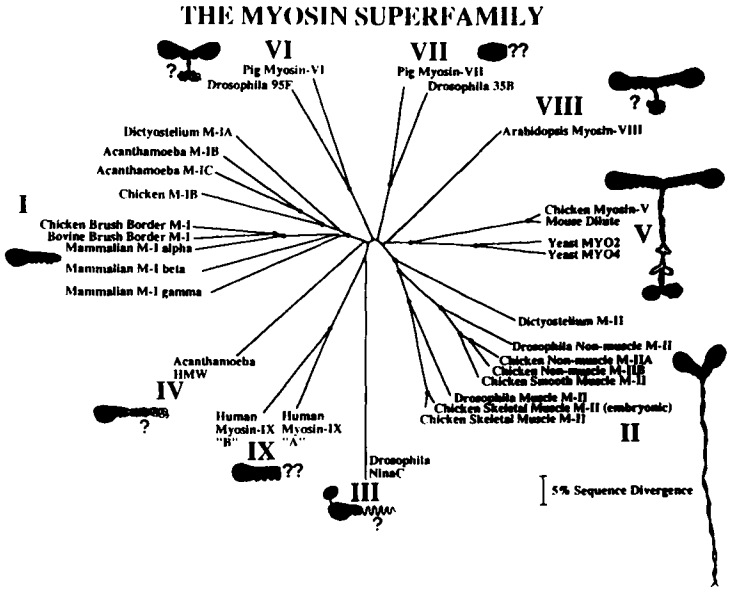


Figure 1. An unrooted phylogenetic tree of the myosins based on the amino acid sequence comparison of their head domains demonstrating the division of the myosin superfamily into nine classes. The lengths of the branches are proportional to the percent of amino acid sequence divergence and a calibration bar for 5% sequence divergence is shown. The different classes of myosins have been numbered using Roman numerals in rough order of their discovery and hypothetical models of the different myosin structures are shown. Question marks indicate either hypothetical or unknown structural features, and only a fraction of the known myosins are shown. (Taken, in modified form, from Cheney et al., 1993).

in its tail; and (c) myosin-I group from metazoans including brush-border myosin-I, mammalian myosin-Ia from mouse brain, and a myosin-I from *Drosophila* (Cheney et al., 1993). The myosin-I group is quite diverse in nature yet appears to coordinate interactions between membranes and the cytoskeleton (Pollard et al., 1991).

MYOSIN-II

Structure

Skeletal muscle myosin-II was first purified in the 1930s and has been extensively studied since (Engelhardt and Ljubimova, 1939). Myosin-II is a dimer composed of two molecules of myosin joined by intertwined, filamentous tails, with each monomer containing two pairs of light chains (Figure 2) (Adelstein and

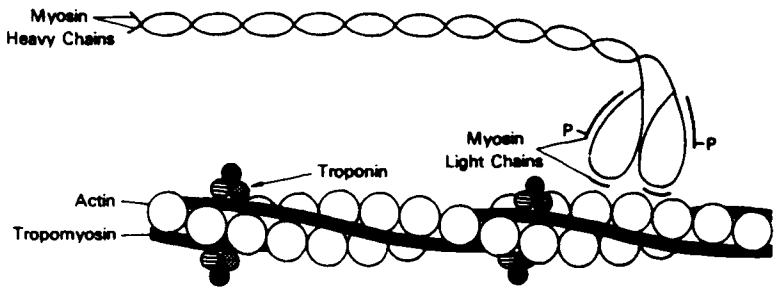


Figure 2. Diagrammatic representation of the myosin-II molecule and the organization of actin into filaments. Myosin-II molecules are long, hexameric proteins consisting of heavy chains and light chains. The heavy chains contain two domains: an α-helical tail domain and a globular head domain. The tail domains are important in the polymerization of myosin into filaments and the head domains contain the actin, ATP, and light chain binding domains. One set of light chains can be phosphorylated and these are identified with a "P". Thin filaments consist mostly of polymerized actin molecules that are woven around tropomyosin molecules at a molar ratio of seven actins per tropomyosin. The troponin complex, consisting of three subunits (see text), is also shown. Stimulation of striated muscle results in an increase in intracellular Ca^{2+} and Ca^{2+} binding to the c-subunit of the troponin molecule (Tn-C). This initiates a conformational change in the troponin molecule and a shift in the location of tropomyosin molecule relative to the actin filament and the myosin heads. In this configuration, actin interacts with myosin, ATP is hydrolyzed, thin filaments move past thick filaments, sarcomeres (see Figure 3) shorten, and muscles contract. Relaxation is initiated when intracellular Ca^{2+} decreases due to the termination of the contractile stimulus. The dissociation of Ca^{2+} from Tn-C reverses the contractile process and tropomyosin molecules return to a location on the thin filaments that sterically inhibit the actin-myosin interaction. Troponin is only found in skeletal and cardiac muscle, but not smooth muscle and nonmuscle thin filaments. A schematic of smooth muscle contraction is presented in Figure 4. (Taken, in modified form, from Adelstein and Eisenberg, 1980).

Eisenberg, 1980). Each myosin monomer consists of an N-terminal globular head domain, containing the sites for actin binding and ATP hydrolysis, attached to a filamentous C-terminal tail. A myosin-II heavy chain monomer has a molecular weight in the range of 190–210,000. Associated with each head are a pair of light chains each with molecular weights ranging from 15–30,000. Thus, a complete myosin-II molecule has a molecular weight of about 450,000.

The myosin-II group can be divided into two groups derived from striated muscle and non-muscle/smooth muscle (Cheney et al., 1993). The striated muscle subgroup can be further divided into the forms found in skeletal and cardiac muscles. The myosins from striated and cardiac muscles have different biochemical

properties related to their physiological roles. Mammalian muscles respond to stimuli with either a short and powerful effort called a fast twitch or a slow, sustained contraction called a slow twitch (Ruegg, 1988). In general, slow twitch muscles receive continuous, repetitive stimulation, contain large numbers of mitochondria and myoglobin, and are not easily fatigued. On the other hand, fast twitch muscles respond quickly to stimulation, contain less mitochondria and myoglobin, and tire easily. Muscles from the slow and fast twitch groups contain myosin isoforms with different biochemical properties that match their muscle type. The contractility of slow and fast twitch muscles differs by a factor of 2–3. Similarly, the myosin isoforms found in these muscles have ATPase activities with similar differences. For example, myosin-II from guinea pig soleus (a slow muscle) hydrolyzes ATP at a rate equal to 50 nmol/min/mg of protein, while the myosin-II ATPase from the vastus muscle hydrolyzes at a rate of 130 nmol/min/mg, or nearly three times faster (Ruegg, 1988).

Nonmuscle/smooth muscle myosins-II are structurally similar to striated muscle myosin-II, but they have slower rates of ATP hydrolysis than do their striated muscle counterparts. Nonmuscle/smooth muscle myosin-II is also regulated differently than striated muscle myosin-II. Nonmuscle myosin-II is divided into the invertebrate and vertebrate branches (Cheney et al., 1993). This group is ubiquitous because it is present in most lower organisms, such as slime molds, amoeba, sea urchins, etc., and in virtually all mammalian nonmuscle cells. Smooth muscle myosin-II is also somewhat heterogeneous in that at least three separate forms of smooth muscle heavy chains, with molecular weights of 196,000, 200,000, and 204,000 have been identified (Kawamoto and Adelstein, 1987). The physiological properties of these separate myosin heavy chains are not yet known.

All myosins contain light chains located near the head domain that play important roles in the regulation of the ATPase activity of myosin-II (Adelstein and Eisenberg, 1980). The naming of the light chains depends on the muscle type. Skeletal muscle light chains are called LC-1, LC-2, and LC-3 based on their electrophoretic mobilities, while the nonmuscle/smooth muscle forms are called the essential and the regulatory light chains. LC-1 and LC-3 are also known as alkali light chains because they can be separated from the heavy chains by brief alkaline treatment. Similarly, LC-2 is known as the DTNB light chain because it can be isolated by treating striated muscle myosin with 5,5′-dithiobis(2-nitrobenzoic acid). The essential light chains are equivalent to LC-1 and LC-3, while the regulatory light chain is equivalent to LC-2. Removal of the essential light chains eliminates ATPase activity. The regulatory light chains are required for proper regulation of ATPase activity as its name implies. Despite their functional similarities, it is important to note that the skeletal, cardiac, smooth muscle and nonmuscle light chains represent the products of separate genes.

The regulatory light chains from vertebrate forms of myosin-II undergo reversible phosphorylation by a calmodulin dependent enzyme called myosin light chain

kinase (Adelstein and Eisenberg, 1980). There has been a great deal of interest in myosin phosphorylation because of the well-known importance of protein phosphorylation reactions in signal transduction processes. Phosphorylation of the regulatory light chain of myosin-II purified from skeletal or cardiac muscle does not seem to be required for activity, and the physiological importance of this reaction in striated muscles remains unclear. The activity of smooth muscle myosin, however, is directly correlated with light chain phosphorylation (de Lanerolle and Paul, 1991), and this important regulatory reaction will be discussed in greater detail below.

The myosin-II molecule has two hinge regions located on the heavy chains (Tan et al., 1992). One is near the head and the other at a somewhat more distal site on the tail. Proteases act at the hinge regions to give subfragments of myosin. Brief proteolysis of muscle myosin-II by chymotrypsin will cut the molecule in the tail, yielding a two headed subfragment called heavy meromyosin (HMM). Further digestion of HMM by an enzyme such as papain cuts at the hinge near the head and gives rise to a single headed molecule called subfragment-1 (S-1). Both HMM and S-1 retain ATPase and actin binding activities. Similarly, fragments of *Dictyostelium* myosin-II analogous to HMM and S-1 have been generated by molecular biological techniques, and have been shown to retain enzyme activity (Manstein et al., 1989). The hinge regions are important to the mechanical function of myosin.

Filament Organization

One can easily identify light and dark bands in electron micrographs of striated muscles (Figure 3) (Eisenberg, 1983). This pattern of light and dark bands, which appear as striations, give skeletal and cardiac muscles their name. In the middle of each light band is a narrow line called the Z disk. The region between two Z disks is called a sarcomere. Analysis of individual sarcomeres reveals that there are two kinds of filaments: thick and thin (Figure 3). Thick filaments are composed of many myosin molecules aligned in a bipolar fashion, and myosin heads protrude from the thick filament surface at an interval of 14 nm.

Smooth muscles, as the name implies, do not contain sarcomeres. In fact, it was initially difficult to demonstrate the presence of thick filaments in smooth muscle, although their presence is now well-established. On the other hand, it is very difficult to demonstrate thick filaments in highly motile cells, such as macrophages and neutrophils, and this may reflect the necessity to rapidly form and redistribute cytoskeletal elements during migration. Thick filaments in smooth muscles appear to be considerably longer than those in striated muscles. They run diagonally in smooth muscle cells and attach to the membrane at structures known as dense bodies. Thus, there is a cork-screw effect when smooth muscles contract (Warshaw et al., 1987).

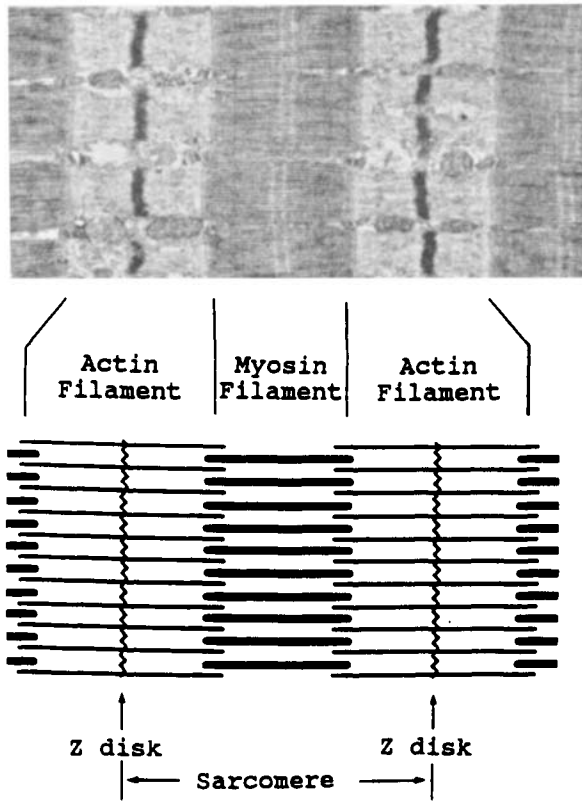


Figure 3. Ultrastructure of striated muscle myosin. The top part of the figure is an electron micrograph of guinea pig soleus muscle (magnification = 15,000×) and the bottom part is a schematic drawing of the salient features. The light regions contain actin (thin) filaments. The dark bands running vertically through the thin filaments are Z disks. The regions between Z disks define the sarcomere, and thin filaments extend in both directions from the Z disks. The dark regions in the center of the sarcomere contain myosin or thick filaments. Thin and thick filaments overlap at their edges, and it is in this region of overlap that ATP hydrolysis and filament sliding take place. The movement of actin filaments towards the center of the sarcomere shortens the length of the sarcomere and, eventually, results in muscle contraction. (The micrograph was taken from Eisenberg, 1983).

Myosin-II phosphorylation is also an important mechanism for regulating myosin assembly in nonmuscle and smooth muscle cells (Korn and Hammer, 1988). For example, myosin-II from *Acanthamoeba* is more soluble when the heavy chain is phosphorylated compared to the unphosphorylated species. Similarly, phosphorylation of the light chains of vertebrate smooth muscle and nonmuscle myosin-II affects filament formation by these myosins. These myosins undergo a

conformational change that is regulated by phosphorylation of the regulatory light chain; this affects their ability to form filaments (Trybus et al., 1982). When unphosphorylated, a conformation in which the tail regions of the heavy chains are bent back towards the head predominates. This is known as the 10S conformation because of its sedimentation characteristics (S stands for Svedberg). Upon phosphorylation of the light chains, the tail extends away from the head and the molecule converts to an extended, or 6S, conformation that is typically associated with the myosin-II molecule (Figure 2). An important aspect of this conformational change is that only the 6S conformation is capable of assembling into filaments. This ability of vertebrate nonmuscle myosin-II to assemble/disassemble is probably a unique feature of these myosins that is related to the necessity in these cells to restructure the myosin filaments during cell motility. In contrast, striated muscle myosin-II does not undergo a similar 10S–6S transition, probably because it rarely redistributes itself, and is more stringently held in place in the sarcomeres.

Thin filaments in striated muscles are composed mainly of actin while the thick filaments are composed of myosin. The thin and thick filaments overlap in the dark region of Figure 3, and it is in this region that actin and myosin interact. Striated muscle has six actin filaments surrounding each myosin filament, and two myosin filaments around each actin filament. Tropomyosin is another important constituent of thin filaments (Figure 2). Tropomyosin is found in most muscle systems and has a molecular weight of 66,000 (Adelstein and Eisenberg, 1980). It is composed of two α-helical subunits coiled around each other. Tropomyosin binds mainly to actin but also interacts with the troponin complex, which plays a critical role in regulating striated muscle contraction. The troponin complex is composed of three proteins: troponin I (Tn-I), troponin T (Tn-T), and troponin C (Tn-C). The importance of these proteins in regulating the actin-myosin interaction and striated muscle contractility is discussed in the next section.

Activity and Regulation

ATP hydrolysis by actin and myosin must be regulated to be physiologically relevant and this is the role of Ca^{2+}. Calcium regulation of the actin-myosin interaction and, hence, ATP hydrolysis and contraction, differs with muscle type. There are two general forms of regulation: actin- and myosin-based. In skeletal muscle, regulation is considered to be thin filament or actin-based because it is mediated through Ca^{2+} binding to Tn-C, while smooth and nonmuscle systems are regulated in a thick filament or myosin-based fashion. Irrespective of muscle type, the general model for muscle contraction relies on the formation of crossbridges between actin and myosin, followed by the activation of the myosin ATPase. During contraction the thick and thin filaments slide past one another when ATP is hydrolyzed (Huxley and Niedergerke, 1954).

Myosins hydrolyze ATP under a variety of physiological and nonphysiological conditions. Physiological concentrations of magnesium inhibit ATP hydrolysis by vertebrate myosin-II (Adelstein and Eisenberg, 1980). Actin dramatically stimulates the rate of ATP hydrolysis by promoting product release from the myosin head. Thus, the physiologically relevant mechanism with respect to muscle contraction is the actin-activated, Mg^{2+}-dependent hydrolysis of ATP by myosin-II. The role of Ca^{2+} is to regulate this interaction.

Calcium regulation of striated muscle contraction is mediated through Tn-C (Figure 2) (Adelstein and Eisenberg, 1980). In the absence of Ca^{2+}, the troponin-tropomyosin complex lies in a groove formed by the actin helix. In this location, the tropomyosin effectively blocks the actin-myosin interaction in a steric manner. The binding of Ca^{2+} to Tn-C affects the interactions of the various troponin subunits and a shift in the location of the tropomyosin molecule relative to the actin groove. This movement of the tropomyosin molecule relieves the steric inhibition of the actin-myosin interaction, and results in the interaction of the myosin head with actin and ATP hydrolysis. This mechanism is called the steric blocking model, and is based on data from numerous biochemical experiments. Relaxation occurs when intracellular Ca^{2+} falls, the conformational changes in the troponin molecule are reversed, and the tropomyosin slips back into the actin groove.

The regulation of smooth muscle and nonmuscle myosin-II is substantially different from the mechanism described above for two important reasons. First, there is no troponin in smooth muscle and nonmuscle cells. Second, although the rate of hydrolysis of ATP by these myosins is low in the presence of physiological concentrations of Mg^{2+}, the addition of actin does not necessarily result in the stimulation of ATP hydrolysis by smooth muscle or nonmuscle myosin-II. These observations suggest the presence of a unique mechanism for Ca^{2+} regulation in smooth and nonmuscle cells, and that these myosins require an activation process before actin can stimulate ATP hydrolysis.

It is now clear that both processes reside in a single reaction (de Lanerolle and Paul, 1991). As discussed previously, the regulatory light chains of vertebrate myosin-II molecules are reversibly phosphorylated. Phosphorylation of a specific serine residue (ser 19) on the phosphorylatable light chains of vertebrate smooth muscle and nonmuscle myosin-II, but not skeletal or cardiac myosin-II, results in actin-activation of ATP hydrolysis. That is, the addition of actin to phosphorylated smooth muscle or nonmuscle myosin-II stimulates the rate of ATP hydrolysis. Ca^{2+} is involved in this process because it regulates the activity of myosin light chain kinase, the enzyme that phosphorylates myosin-II light chain. In order to do so, Ca^{2+} first binds to calmodulin, a ubiquitous calcium binding protein similar to troponin-C and the myosin light chains. The Ca^{2+}-calmodulin complex then binds to myosin light chain kinase, thereby activating the enzyme. Myosin light chain kinase transfers the terminal phosphate from ATP to serine 19 on the regulatory light chain, stimulating the actin activated myosin ATPase. This mechanism of

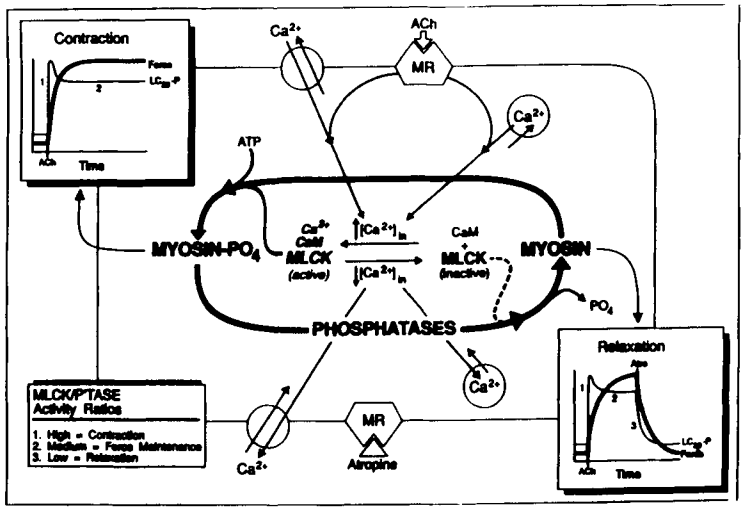


Figure 4. Schematic presentation of the molecular basis of smooth muscle contraction. Stimulation of an excitatory receptor by a neurotransmitter or hormone, in this case the muscarinic receptor (MR) by acetylcholine (Ach), results in an increase in intracellular Ca^{2+} due to the entry of external Ca^{2+}, and the release of Ca^{2+} from internal storage sites. Ca^{2+} binds to calmodulin (CaM) and the Ca^{2+}/calmodulin complex then binds to and activates myosin light chain kinase (MLCK). Phosphorylation of ser 19 of the 20 kDa myosin light chain by MLCK stimulates ATP hydrolysis by actin and myosin, which results in smooth muscle contraction. The predicted temporal relationships between myosin light chain phosphorylation (LC_{20}-P) and force generation is shown in the inset at the top left. Relaxation begins with the cessation of agonist stimulation, in this case due to the displacement of acetylcholine from the muscarinic receptor by atropine. This results in the decrease of intracellular Ca^{2+}, dissociation of Ca^{2+} from calmodulin, inactivation of MLCK due to the dissociation of the Ca^{2+}/calmodulin complex from the enzyme, dephosphorylation of myosin by phosphoprotein phosphatases, and relaxation. The broken line from inactive MLCK indicates that MLCK is passively involved in relaxation. The temporal relationships between light chain dephosphorylation and relaxation are shown in the inset at the right. (Taken from de Lanerolle and Paul, 1991).

activation involves the alteration of the enzyme kinetics of the ATPase. Kinetic experiments have shown that phosphorylation increases the V_{max} of the enzyme, thus increasing the rate at which ATP is hydrolyzed (Korn and Hammer, 1988).

The following series of events is thought to take place when smooth muscles contract (Figure 4) (de Lanerolle and Paul, 1991). A contractile stimulus results in an increase in intracellular Ca^{2+} and Ca^{2+} binding to calmodulin. The Ca^{2+}-calmodulin complex then binds to and activates myosin light chain kinase. The activated enzyme then phosphorylates ser 19 on the regulatory light chain, which leads to actin activation of the myosin and ATP hydrolysis. ATP hydrolysis results

in the movement of filaments past each other and smooth muscle contraction. Relaxation occurs when the system is turned off; that is, when the contractile stimulus is withdrawn. The resulting decrease in intracellular Ca^{2+} leads to a dissociation of Ca^{2+} from calmodulin, dissociation of calmodulin, and inactivation of myosin light chain kinase. When myosin light chain kinase is inactivated, constitutively-active phosphoprotein phosphatases dephosphorylate the regulatory light chain, and turn off ATP hydrolysis. Such a mechanism is also thought to regulate myosin-II in vertebrate nonmuscle cells, and the importance of this mechanism in regulating nonmuscle cell function cannot be overemphasized.

Physiological Properties

Myosin-II is widely distributed in eukaryotic cells and its role in muscle contraction is well-established: ATP hydrolysis following the cyclic interaction of actin with myosin is essential for skeletal, cardiac, and smooth muscles to contract. The importance of striated muscle contraction is obvious and smooth muscle myosin-II is known to play a crucial role in regulating blood flow, air flow in the lungs, digestion, and in childbirth. Equally important, but less obvious, is the importance of myosin-II in nonmuscle cells (Korn and Hammer, 1988). Elegant studies on *Dictyostelium* and other lower eukaryotes have demonstrated that myosin-II functions in many processes such as cytokinesis, cell motility, receptor capping, and cytoskeletal formation. Vertebrate nonmuscle myosin-II is widely distributed and found in virtually all mammalian cells apart from skeletal and cardiac muscle cells. Studies on mammalian cells have demonstrated that myosin-II is involved in receptor capping, cell motility, endothelial cell contractility, epithelial barrier function, and secretion. Thus, myosin-II appears to be involved in mediating a host of cellular functions that have far-reaching physiological and clinical implications.

MYOSIN-I

A new kind of myosin was observed in *Acanthamoeba* in the early 1970s (Pollard and Korn, 1973). It was characterized as having a single heavy chain containing actin binding and ATPase domains in the head but with a small tail that was incapable of forming filaments. Even though the structure was quite different from known myosins, it was labeled a myosin due to its actin binding and ATPase activities. Later, investigators discovered a vertebrate form of this myosin in the plasma membrane of intestinal microvilli (Matsudaira and Burgess, 1979). Evidence has now accumulated that single-headed myosins occur in all eukaryotic cells (Pollard et al., 1991).

Structure

The myosin-I family consists of a group of soluble proteins that are characterized by a single heavy chain (115–140 kDa) and one or two light chains (14–30 kDa) (Pollard et al., 1991). Most of the heavy chain is folded into a globular head that is followed by a very short tail region. The globular heads contain ATP and actin binding sites that are very similar to the comparable regions of myosin-II. There are also some unique myosin-I specific sequences that are well conserved. There is a domain in the head of the *Acanthamoeba* enzyme that contains a phosphorylation site. An amino acid sequence near to this site is conserved in most forms of myosin-I, even though not all myosins-I are phosphorylated at this site.

Myosin-I molecules have several IQ sequences on or near the head and have light chains associated with them (Cheney and Mooseker, 1992; Cheney et al., 1993). Frequently, the light chains appear to be calmodulin molecules and some myosin-I molecules can bind three to four molecules of calmodulin at one time. Brush-border and adrenal myosin-I also bind calmodulin. *Acanthamoeba* myosin-I has a light chain that can be removed, *in vitro*, without adversely affecting the ATPase activity or the heavy chain phosphorylation (Korn and Hammer, 1988). The role of these calmodulin molecules in regulating myosin-I is complex and poorly understood. One possibility is that the calmodulin molecules dissociate from the heavy chains when calcium binds to the calmodulin, thereby imparting greater flexibility to the head of the myosin-I molecules.

Although there is greater diversity in the tail domains, there are at least three regions of high homology that are important. First, they have a stretch of positively charged amino acids that have a combined net charge of +19 to +42 (Pollard et al., 1991). *Acanthamoeba* and brush-border myosin-I both have such a positively-charged sequence, yet the amino acid sequences are quite distinct. Although the role this sequence plays in myosin-I function is unclear, it is speculated to be important due to its high level of conservation. The second domain is rich in glycine, proline, and alanine or glutamate, and is called the GPA or GPQ domain. The GPA/GPQ domain has been implicated in actin binding. The third domain is the SH3 or src homology domain. This 50 amino acid sequence shares homology with conserved sequences on src kinase, spectrin and a yeast actin binding protein. Because this sequence appears in other membrane binding proteins, it has been postulated to be important for membrane binding. Thus, depending on the isoform, these proteins can function as actin-actin or actin-membrane cross-linking proteins.

Regulation

The regulation of myosin-I activity is not well understood. As mentioned above, the heavy chain is phosphorylated in *Acanthamoeba* but the regulatory effect of this process is unclear. Myosin-I from *Acanthamoeba*, *Dictyostelium*, brush-bor-

der, and adrenal cortex have unique light chains associated with them, while brush-border and adrenal myosin-I have several molecules of bound calmodulin apparently acting as light chains. The *Dictyostelium* light chain is phosphorylated and may play a role in the regulation of enzyme activity (Tan et al., 1992). The role of calmodulin binding in brush-border and adrenal myosins-I is probably important in the regulation of the enzymatic activity. As will be recalled, when calmodulin binds Ca^{2+} its affinity for myosin is greatly reduced. The reversible binding of calmodulin to myosin-I results in a point for Ca^{2+} regulation of myosin-I activity.

Physiological Properties

There is a great deal of interest in myosin-I. It has been clear for many years that myosin-II cannot fulfill all the energy transduction functions within cells. For instance, the sliding filament model (Huxley and Niedergerke, 1954), which is critical to our understanding of muscle contraction, accommodates muscle contraction by a mechanism that involves the movement of actin filaments, powered by ATP hydrolysis by myosin-II, towards the center of the cell. In contrast, cell motility involves the *extension* of the leading edge of a cell and, as far as we know, myosin-II is incapable of extending actin filaments at the leading edge (Wilson et al., 1992). This point was underscored by the demonstration that *Dictyostelium* amoeba not having myosin-II are capable of an altered form of motility (De Lozanne and Spudich, 1987).

The presence of an actin binding site or a membrane binding site on the tails of myosin-I molecules has raised the possibility that myosin-I may act as the energy transduction mechanism for lamellapodial extension during cell migration. It is well-known that actin is localized at the leading edges of migrating cells, and the thought is that myosin-I, either by binding to adjacent actin filaments or to the membrane, could power the movement of actin filaments past them. Such a mechanism would be consistent with the sliding filament model but with the filaments sliding in the opposite direction; in other words, away from the center of the cell. The demonstration that myosin-I is localized at the leading edges of migrating cells has provided important support for this hypothesis.

Myosin-I has also been suggested to be involved in a variety of other cellular processes in nonmuscle cells that were initially thought to be mediated by myosin-II. One such process is phagocytosis. *Dictyostelium* mutants devoid of myosin-II feed on bacteria, which means that they are capable of phagocytosis in the absence of myosin-II. In addition, recent studies on macrophages have demonstrated that these cells ingest yeast particles without an increase in myosin light chain phosphorylation (de Lanerolle et al., 1993). These studies have questioned whether myosin-II is required for phagocytosis and suggested that myosin-I may be involved. Again, the localization of myosin-I in phagocytic cups around ingested

particles has provided support for this idea. Other experiments have localized myosin-I in the Golgi apparatus of intestinal epithelial cells, which has led to the suggestion that myosin-I is important in vectorial transport of organelles and membranes in polarized cells.

CLINICAL ASPECTS

Asthma

Asthma is a complex respiratory disorder that involves mast cell degranulation, mucous secretions, and smooth muscle hypertrophy and hyperresponsiveness. Smooth muscle hyperresponsiveness has suggested some defect in the regulation of smooth muscle contractility. Therefore, a number of studies concerning asthma have centered on whether alterations in the regulation of smooth muscle contraction (Figure 4) are responsible for hyperactivity in asthmatic airway smooth muscle.

The essence of the regulatory mechanism depicted in Figure 4 is that changes in myosin light chain kinase activity result in changes in the contractile properties of smooth muscles (de Lanerolle and Paul, 1991). If an increase in myosin light chain kinase activity results in smooth muscle contraction, then a chronic increase in myosin light chain kinase activity and/or improper activation of myosin light chain kinase and/or changes in the enzymatic properties of the myosin molecule could be responsible for airway hyperreactivity seen in asthma.

Several studies have been undertaken to address these issues. Jiang and colleagues have used a model of asthma in which dogs were sensitized to ovalbumin to investigate whether asthma induces changes at the level of light chain phosphorylation (Jiang et al., 1992). They showed a significant increase in light chain phosphorylation in sensitized dogs compared to control animals. Their studies to determine whether the higher level of phosphorylation is a result of an increase in myosin light chain kinase activity or a decrease in phosphatase activity suggested that myosin light chain kinase activity is higher in sensitized dogs than in normal animals, whereas phosphatase activity appeared to be the same in control and test animals. These studies showed for the first time that an increase in myosin light chain kinase activity may be involved in asthma.

Myosin light chain kinase activity could also have an indirect role in airway response. It is known that an increase in cAMP correlates with smooth muscle relaxation, and it has been suggested that asthma may be related to a lack of responsiveness to cAMP (de Lanerolle and Paul, 1991). There are two ways in which cAMP could affect myosin light chain kinase activity. First, it has been postulated that cAMP decreases intracellular calcium, hence, lowering myosin light chain kinase activity. Second, phosphorylation of myosin light chain kinase by cAMP-dependent protein kinase decreases myosin light chain kinase activity (de Lanerolle and Paul, 1991). Both models result in reduced light chain phosphoryla-

tion and less force. Prior stimulation with agents that stimulate cAMP production should result in slower force generation and less total force following stimulation.

The role of cAMP in asthma has been investigated in two ways. First, resting tracheal muscle was pretreated with forskolin, an agent that increases cAMP and relaxes smooth muscles. In this experiment it was shown that myosin light chain kinase is phosphorylated in the resting state, and that phosphorylation was higher with forskolin treatment (de Lanerolle et al., 1984). In addition, Obara and de Lanerolle showed that isoproterenol results in a slower rate of light chain phosphorylation upon stimulation by carbachol. Both of these studies suggest that alteration of myosin light chain kinase activity by phosphorylation of myosin light chain kinase by cAMP-dependent protein kinase has a relaxing effect on smooth muscle.

In conclusion, studies from these two groups suggest that changes in myosin light chain kinase activity could result in the types of changes seen in asthma. Further studies are needed to address the possible role of myosin phosphorylation in asthma. For instance, it is possible that asthmatics may have a higher than normal level of myosin light chain kinase in their smooth muscle, leading to increased light chain phosphorylation. It is also possible that the myosin light chain kinase from these individuals may be kinetically altered to be more active *in vivo*. Modern molecular biology techniques are bound to become very useful in examining this important problem.

Hypertrophic Cardiomyopathy

In a previous section we mentioned the significance of myosin filament structure. In nematodes two forms of myosin-II, myosin A and B, are required for proper filament structure (Epstein, 1988). The two forms of myosin are expressed at the proper time to allow for correct filament assembly. An accessory protein called paramyosin is also required for correct filament assembly. In vertebrate cardiac muscle, there are also two isoforms of myosin-II: α-myosin and β-myosin. The proper ratio of these two proteins is of utmost importance for proper muscle activity. The incorrect synthesis of α- and β-myosins results in a severe cardiac disorder known as hypertrophic cardiomyopathy. Genetic transmission of the disease occurs in about 55% of families. The inherited condition is called familial hypertrophic cardiomyopathy (FHC), and this condition is a leading cause of sudden death in young athletes.

Hypertrophic cardiomyopathy (HCM) is characterized by abnormal left ventricular thickening. The left ventricular septum is the most common site of involvement. Pathologically, the disease is characterized by myocardial fiber disarray. The myocardium may exhibit extensive scarring and disorganization of interstitial and intercellular tissue (Elstein et al., 1992). The severity of HCM depends on the age of the patient, as well as the extent of the disarray. Patients with HCM have variable

histories ranging from asymptomatic to rapidly fatal disease. The annual death rate is 3–5%.

Genetic linkage studies of FHC have implicated the cardiac myosin genes (Elstein et al., 1992). Animal models of the disease suggest that altered expression of the α- and β-cardiac myosin genes plays a role in the development of the disease. Tanagawa et al. (1990) demonstrated that a mutation of the cardiac myosin genes resulting in an α/β hybrid gene occurs in one family with FHC. This suggests that this mutation could cause the disease, although it does not rule out that a closely linked gene is responsible.

Geisterfer-Lowrance et al. (1990) have isolated and sequenced the human cardiac β-myosin gene. Using the sequence information, they found that a missense mutation in exon 13 of the gene is present in a large group of affected individuals. This mutation resulted in the conversion of a highly conserved arginine (Arg-403) to glutamine in all individuals tested. This amino acid is invariant in at least fifteen different myosin genes from human, rat, rabbit, chicken, and *Acanthamoeba*. Man and amoeba diverged around 600 million years ago, and the fact that this amino acid has been conserved for so long indicates its apparent importance in myosin function (Cheney et al., 1993). This mutation occurs in the head region of the molecule but is not in any previously determined functional domain, such as the ATP activity or actin binding domains.

It is interesting to note that the missense mutation cited above occurs only in the β-cardiac gene of affected individuals. The β-cardiac genes are expressed in both cardiac and skeletal muscles, while the α-cardiac gene is expressed in cardiac muscle only. A β-cardiac mutation might also lead to altered skeletal muscle organization and some studies indicate that this may be the case. The involvement of the α-cardiac genes in FHC is only implicated in cases with the α/β hybrid gene product. The fact that this disease is caused by two different types of mutations can be explained by the critical nature of filament formation. Perhaps any mutation of the β-cardiac gene (missense or hybrid) upsets the ratio of α- to β-myosins. Thus, functionally, the two mutations lead to the same result. The complete understanding of the biochemical nature of the mutant proteins will improve our understanding of the disease.

FUTURE DIRECTIONS

The recent explosion in the discovery of new myosin genes has led to the idea that myosins from different classes probably co-exist in cells. This has raised the obvious question as to what functions these myosins subserve within cells. Up to now, only the genes have been cloned for many of the 35 unique myosins. But this is not a question that can be answered solely by cloning; rather, it is absolutely imperative to biochemically characterize these proteins if we are to understand their physiological properties. One way to do this is to express the entire protein or parts of the proteins in bacteria, yeast, or insect cells, and to then purify and characterize

the expressed proteins. The expressed proteins can also be used to generate antibodies that could be very helpful in defining the localization of myosins. Moreover, antibodies could also be very useful in defining the physiological and biochemical properties of myosins. Expressing myosin, however, is a complicated problem because they are multimeric proteins with complex tertiary and quaternary structures.

Two other questions beg answers. First, it is known from studies on *Dictyostelium* and other lower organisms that myosin-II is essential for cytokinesis (Korn and Hammer, 1988). Based on these data, it is assumed that myosin-II is essential for mammalian cells to divide correctly. However, there is no direct evidence from studies on mammalian cells to support such a notion. Given the importance of cell division in normal development and cancer, it is essential to determine whether myosin-II plays a comparable role in mammalian cell division.

Second, it is equally important to understand the role of individual myosins in cells. In light of the presence of multiple myosins within a single cell, it is important to determine whether each of these myosins mediates a separate or unique function. For instance, it has been demonstrated that myosin-IC is involved in the response of *Acanthamoeba* to osmotic stress (Doberstein et al., 1993). Similarly, analysis of the roles of the various myosins is likely to greatly increase our understanding of the biochemical mechanisms that regulate individual cellular responses. Other studies have demonstrated the presence of developmentally regulated myosin light chains in muscle tissues and in lymphocytes. Do myosins containing these light chains have unique functions and do mutations or alterations in their expression affect cellular development or function? The combination of modern molecular biology techniques with physiology and medicine should make it possible to address these questions.

ACKNOWLEDGMENTS

D.D.L. was supported by an NIH training grant (HL 076922). P.deL. is the Florence and Arthur Brock Established Investigator of the Chicago Lung Association. This work was supported, in part, by grants from the National Institutes of Health (HL 35808 and HL 02411) to P.deL.

REFERENCES

Cheney, R.E. & Mooseker, M.S. (1992). Unconventional myosins. Curr. Op. Cell Biol. 4, 27–35.

de Lanerolle, P., Nishikawa, M., Yost, D.A., & Adelstein, R.S. (1984). Increased phosphorylation of myosin light chain kinase after an increase in cyclic AMP in intact smooth muscle. Science 223, 1415–1417.

de Lanerolle, P., Gorgas, G., Li, X., & Schluns, K. (1993). Myosin light chain phosphorylation does not increase during yeast phagocytosis by macrophages. J. Biol. Chem. 268, 16883–16886.

De Lozanne, A. & Spudich, J.A. (1987). Disruption of the *Dictyostelium* myosin heavy chain gene by homologous recombination. Science 236, 1086–1091.

Doberstein, S.K., Baines, I.C., Wiegand, G., Korn, E.D., & Pollard, T.D. (1993). Inhibition of contractile vacuole function *in vivo* by antibodies against myosin-I. Nature 365, 841–843.

Eisenberg, B.R. (1983). Quantitative ultrastructure of mammalian skeletal muscle. In: Handbook of Physiology, Section 10, Skeletal Muscle (Peachey, L.D., Adrian, R.H., & Geiger, S.R., eds.) American Physiological Society, Bethesda, MD.

Elstein, E., Liew, C., & Sole, M.J. (1992). The genetic basis of hypertrophic cardiomyopathy. J. Mol. Cell. Cardiol. 24, 1471–1477.

Engelhardt, W.A. & Ljubimova, M.N. (1939). Myosin and adenosinetriphosphatase. Nature 144, 668–669.

Epstein, H.F. (1988). Modulation of myosin assembly. Bioessays 9, 197–200.

Geisterfer-Lowrance, A.A., Kass, S., Tanigawa, G., Vosberg, H.-P., McKenna, W., Seidman, C.E., & Seidman, J.G. (1990). A molecular basis for familial hypertrophic cardiomyopathy: a β-cardiac myosin heavy chain gene missense mutation. Cell 62, 999–1006.

Huxley, A.F. & Niedergerke, R. (1954). Structural changes in muscle contraction. Interference microscopy of living muscle fibres. Nature 173, 971–973.

Jiang, H., Rao, K., Halayko, A.J., Liu, X., & Stephens, N.L. (1992). Ragweed sensitization-induced increase of myosin light chain kinase content in canine airway smooth muscle. Am. J. Respir. Cell. Mol. Biol. 7, 567–573.

Kawamoto, S. & Adelstein, R.S. (1987). Characterization of myosin heavy chains in cultured aorta smooth muscle cells. A comparative study. J. Biol. Chem. 262, 7282–7288.

Korn, E.D. & Hammer, J.A. (1988). Myosins of nonmuscle cells. Ann. Rev. Biophys. Chem. 17, 23–45.

Manstein, D.J., Ruppel, K.M., & Spudich, J.A. (1989). Expression and characterization of a functional myosin head fragment in *Dictyostelium discoideum*. Science 246, 656–658.

Matsudaira, P.T. & Burgess, D.R. (1979). Identification and organization of the components in the isolated microvillus cytoskeleton. J. Cell. Biol. 83, 667–673.

Obara, K. & de Lanerolle, P. (1989). Isoproterenol attentuates myosin phosphorylation and contraction of tracheal muscle. J. Appl. Physiol. 66, 2017–2022.

Pollard, T.D. & Korn, E.D. (1973). Acanthamoeba myosin: I. Isolation form *Acanthamoeba castellani* of an enzyme similar to muscle myosin. J. Biol. Chem. 248, 4682–4690.

Pollard, T.D., Doberstein, S.K., & Zot, H.G. (1991). Myosin-I. Ann. Rev. Physiol. 53, 653–681.

Ruegg, J.C. (1988). Calcium in Muscle Activation. Springer-Verlag, Berlin.

Tanigawa, G., Jarcho, J.A., Kass, S., Soloman, S.D., Vosberg, H.-P., Seidman, J.G., & Seidman, C.E. (1990). A molecular basis for familial hypertrophic cardiomyopathy: A cardiac α/β myosin heavy chain hybrid gene. Cell 62, 991–998.

Trybus, K.M., Huiatt, T.W., & Lowey, S. (1982). A bent monomeric conformation of myosin from smooth muscle. Proc. Natl. Acad. Sci. USA 79, 6151–6155.

Warshaw, D.M., McBride, W.J., & Work, S.S. (1987). Corkscrew-like shortening in single smooth muscle cells. Science 236, 1457–1459.

Wilson. A.K., Pollenz, R.S., Chisholm, R.L., & de Lanerolle, P. (1992). The role of myosin-I and -II in cell motility. Cancer and Metastasis Rev. 11, 79–91.

RECOMMENDED READINGS

Adelstein, R.S. & Eisenberg, E. (1980). Regulation and kinetics of the actin-myosin-ATP interaction. Ann. Rev. Biochem. 49, 921–956.

Cheney, R.E., Riley, M.A., & Mooseker, M.S. (1993). Phylogenetic analysis of the myosin superfamily. Cell Mot. Cytoskel. 24, 215–223.

de Lanerolle, P. & Paul, R.J. (1991). Myosin phosphorylation/dephosphorylation and regulation of airway smooth muscle contractility. Am. J. Physiol. 261 (Lung Cell. Mol. Physiol. 5), L1–L14.

Tan, J.L., Ravid, S., & Spudich, J.A. (1992). Control of nonmuscle myosins by phosphorylation. Ann. Rev. Bioch. 61, 721–759.

Chapter 4

Cell Motility

SUTHERLAND K. MACIVER and ALAN G. WEEDS

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 77–106.

ISBN: 1-55938-808-0

INTRODUCTION

Cell motility is the expenditure of energy, usually chemical, to bring about cellular movement. Cells indulge in two major forms of motility: the translocation of the cell itself from one point to another (cell locomotion), which is largely the domain of the microfilament system, and the transportation of material within the cell (intracellular motility), for which the microtubule system is mainly responsible. Cell locomotion is a crucial part of vertebrate life. A variety of cell types move during embryogenesis, fibroblasts and keratinocytes move in the process of wound healing, white blood cells constantly rove the body in immune surveillance, and sperm swim to fertilize the egg in reproduction. In addition to these helpful motile functions, cell locomotion is an integral part of metastasis, a process by which malignantly transformed cells travel from their site of origin to distant sites, thereby greatly complicating eradication of the tumor. Despite the enormous interest and work committed to the field of cell locomotion, we remain largely ignorant of the molecular mechanisms involved. To the considerable surprise of many, myosin-II (muscle myosin) has largely been eliminated as being responsible for force production in locomotion. Myosin-II, however, remains likely to play a supporting role in cell motility. Evidence is accumulating in favor of cyclical actin polymerization perhaps in conjunction with myosin-I-type molecular motors for providing the motile force for cell locomotion.

Intracellular motility is also of vital importance in the lives of cells and the organisms they form. Material and organelles are transported within cells along microtubules and microfilaments; an extreme example of this are the axons of nerve cells which transport materials to the synapses where they are secreted—another motile event. Other examples of intracellular motility include phagocytosis, pinocytosis, the separating of chromosomes and cells in cell division, and maintenance of cell polarity.

The widespread nature of cell motility is readily apparent, but the attention of this chapter will be focused on phenomena that relate particularly to human medicine in order to accommodate the medical nature of this compendium.

CELL LOCOMOTION

Elucidation of the mechanisms directing and producing force for the locomotion of vertebrate cells has long been a popular goal in cell biology, partly because of potential medical significance and partly from the intrinsic fascination of observing these motile events, which often seem to be directed by an apparent intelligence. Our present knowledge of the mechanisms involved in the locomotion of cells has been gleaned over several decades from a diverse group of organisms and cell types. The earliest studies concerned themselves largely with amoeboid protozoans, since their rapid locomotory rates made observation possible in the absence of sophisticated microscopy and time-lapse equipment (de Bruyn, 1947; Komnick et al., 1973). With a few notable exceptions, such as the sperm of certain nematodes, we can assume that all cells that crawl do so by the grace of their actin-based cytoskeletons. That cells demonstrate enormous differences in the appearance of their locomotory form and habit probably reflects different emphasis on actin-based structures common to all. Thus much can be learned about motility by comparing properties of one cell type with another.

The Locomotion of Amoeba

The locomotion of the so-called giant amoebae, such as *Amoeba proteus* and *Chaos carolinense*, is traditionally a subject used as an introduction to biology in the classroom. Although it is now recognized that the cytoskeletal organization of these organisms is fundamentally different from vertebrate cells (Bray and White, 1988), many important concepts such as gel-sol transformation were developed from studies with these giant amoebae (de Bruyn, 1947). A host of other amoebae and amoeboid organisms are similar enough to higher cells to be usefully compared. The slime mold *Dictyostelium discoideum* has revealed much of what is now known about the locomotion of all cell types and promises to continue to lead the way. At first glance this organism does not seem a promising candidate to unravel the mechanisms of locomotion of human cells, but on closer inspection the similarities are remarkable. *Dictyostelium* has an actin-based cytoskeleton composed of a group of actin-binding proteins similar to those found in vertebrate cells. In fact, many of these actin-binding proteins were discovered in this organism (and a similar protozoan, *Acanthamoeba castellanii*), before their presence was detected in vertebrate tissues. *Dictyostelium* and *Acanthamoeba* can be viewed as stripped down versions of vertebrate cells whose cytoskeletons are organized almost solely for locomotion, unencumbered by other cytoskeletal functions such as contractile stress fibers, static cell/cell adhesion plaques, and tissue strengthening intermediate filaments. In addition to the basic structural similarities between vertebrate cells and these protozoans, there are similarities in rudimentary behavior such as chemokinesis, chemotaxis, and cytokinesis. *Dictyostelium* is haploid which

facilitates genetic manipulation, such as gene deletion; it is therefore an attractive model for studying the mechanisms involved in cell motility.

The Locomotion of Fibroblastic Cell Types

Fibroblasts have become a commonly studied motile system (Abercrombie, 1980), despite the fact that these cells move at an extremely slow rate of 1 μm/min. Fibroblasts in culture adopt a typical flattened, locomotory morphology with a broad leading edge devoid of internal organelles that is usually the area of greatest motile activity. This activity is not often fast enough to be easily appreciated by direct microscopic observation, but with time lapse recordings this leading edge can be seen to ruffle (Abercrombie, 1980), i.e., very small veils of cytoplasmic extensions rise and fall back to be reabsorbed slightly further towards the centrally positioned nucleus. Many fibroblasts have their broadest extent at, or close to, the leading edge and taper to a point at the terminating trailing filopodia. The sides of fibroblasts are often concave structures supported under the plasma-membrane by thick actin bundles. In order to produce a locomotory force, cells must be able to develop traction against the structure on which they are crawling (Figure 1A). Fibroblasts seem to be stretched out by adhesive contacts close to the leading lamella and at the termini of trailing filopodia. Detachment of these filopodia (which are typically under tension) causes a biphasic retraction into the cell body and a subsequent increase in the rate of lamellar protrusion at the leading edge (Chen, 1981). Two types of adhesive contacts can be recognized in fibroblasts, focal contacts and associated contacts. The focal contact is a small area usually elongated radially with respect to the cell center, which is in molecular contact with the extracellular matrix. On the cytoplasmic side, focal contacts are associated with a dense plaque of organized cytoskeletal proteins and often form the termini of actin rich stress fibers (see the paragraph on the organization of actin in cells). The focal contact and its associated stress fibers are not thought to be important in cell locomotion since there is an inverse relationship between the speed of a particular cell and the number of focal contacts (Couchman and Rees, 1979); indeed fibroblast migration can be switched off and on by factors that concomitantly produce or destroy focal contacts (Dunlevy and Couchman, 1993). Faster cell types such as leukocytes and keratinocytes produce neither focal contacts nor stress fibers. The associated contacts are large areas held about 30–50 nm from the substratum which often contain focal contacts within their domain. Associated contacts are believed to contribute substantially to total cell adhesion despite the fact that molecular contact is not made directly with the substratum (Abercrombie, 1980).

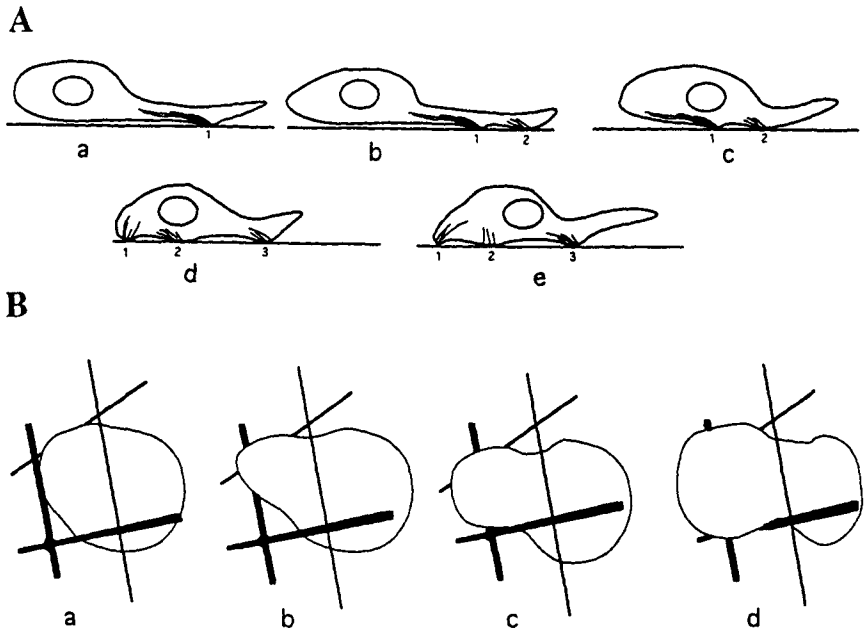


Figure 1. Mechanisms for traction in cell locomotion. **A:** Adhesion mediated traction. The cell moving from left to right, makes an adhesion at position 1 (a). The cell continues to expand the lamellae and forms a second adhesion at position 2 (b), and so on until the original adhesion (1) is left at the rear of the cell (d). This now redundant adhesion weakens and breaks, which results in increased lamellar protrusion. **B:** Adhesion-free traction. Certain leukocytes are able to locomote through matrices of materials to which they are incapable of adhering. This is accomplished by expanding a pseudopod through a constriction in the matrix (b), then expanding the pseudopod (c). A constriction ring forms at the base of this pseudopod which then travels down the cell body, driving the cell forwards (d). Although contact is made no adhesion between the cell and the matrix is necessary for locomotion in this situation.

The Locomotion of Leukocytes

Leukocytes are among the fastest crawling cells in the vertebrate body, travelling throughout the body passively via the blood stream and through tissues by active locomotion. On planar substrates there seems little difference between a locomoting fibroblast and a leukocyte, apart that is, from the huge increase in speed! Puzzlingly, some leukocytes are unable to adhere to biological surfaces and are incapable of locomotion on planar surfaces, yet these same cells are able to invade three-dimensional matrices rapidly (Brown, 1982; Schmalsteig et al., 1986; Lackie and Wilkin-

son, 1984). Locomotion without adhesion in this manner is thought to be possible by the expansion of pseudopods through constrictions (Brown, 1982; Haston and Shields, 1984) (Figure 1B) in a series of movements that have been described as "fist jamming," by analogy to a rather desperate method employed by mountaineers to obtain a hold in a rock crevice. (Lackie, 1986). Leukocytes perform movements which could, according to this hypothesis, produce locomotion in a three-dimensional gel even in suspension (Keller and Cottier, 1981). Constriction rings, described as early as 1931 (Lewis, 1931), can be seen forming as a fresh pseudopod is extended from the cell front. These constriction rings (or contraction waves) then move rearward with respect to the cell body but remain stationary with respect to the coverslip of adherent cells (Haston et al., 1982; Haston and Shields, 1984). In three-dimensional matrices the movement of these constriction rings down the cell body may drive the cell forwards by becoming engaged in constrictions in the matrix (Figure 1B). This mechanism would provide an adhesion-free method of locomotion in some circumstances while allowing normal adhesion-dependent crawling locomotion in others.

The Behavior of Locomoting Cells

Cells communicate with each other in order to form and maintain the vertebrate body. Cells must "know" when and where to move, and what to do once they get there. This is accomplished by sending and receiving chemical, tactile, and perhaps even electrical signals (Cooper and Schliwa, 1986). Cells can be observed responding to at least some of these environmental cues in tissue culture, and also responding to each other. The response of cells to these factors is known as cell behavior.

Contact Inhibition of Locomotion

In culture, untransformed vertebrate cells can be observed moving at random over the culture dish surface. When locomoting cells collide, a remarkable behavior occurs. The cells do not crawl over each other or push each other out of the way, but the actively spreading lamellae freeze as they touch another cell, and a new area of the lamella becomes motile. This behavior is known as contact inhibition of locomotion (Abercrombie and Heaysman, 1954). Contact inhibition is important in the formation and maintenance of body tissues. Many tissues are composed of a sheet of cells overlaying the extracellular matrix. In the event of a wound appearing in this sheet, surrounding cells respond to the space by moving into it, stopping only when they touch other cells. Malignantly transformed cells lose the ability to contact inhibit and so invade and penetrate healthy tissues (Guelstein et al., 1973; Abercrombie, 1979). The majority of lymphocytes are typically immotile when

isolated from blood and the few that move do not show contact inhibition. However, lymphocytes show a peculiar negative form of contact inhibition when they contact fibroblastic cells, i.e., contact stimulation of locomotion (Arencibia et al., 1987).

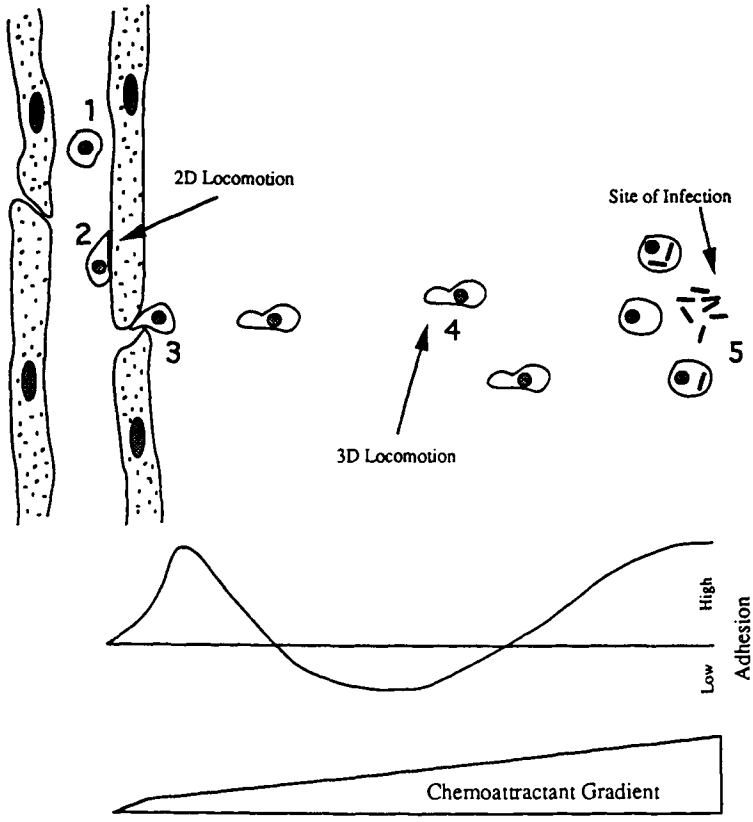


Figure 2. (1) Neutrophils circulating passively in blood capillary. (2) Chemoattractants may be detected by the circulating neutrophils, by the endothelial cells lining the lumen, or both in order that the neutrophils become adhesive. This adhesion is mediated by selectins, a group of cell surface proteins. Neutrophils roll on the surface of the endothelial cells and then actively locomote seeking out spaces between the endothelial cells. (3) The adhesive neutrophils begin to squeeze between endothelial cells. (4) Cells move through the extracellular matrix towards the site of infection. Here adhesion is low and may not be necessary for locomotion. (5) At the site of infection, neutrophils become trapped by increased adhesion where they phagocytose bacteria and liberate the contents of their granules. After Lackie (1982, 1986).

Chemotaxis

The direction of cell locomotion can be modulated by chemoattractants, small molecules, peptides, lipids, or proteins that form concentration gradients detectable by cells. The neutrophil leukocyte has been the subject of intense study with regard to its chemotactic ability (Lackie, 1982; Wilkinson, 1987). These cells are at the forefront of the body's defence against a range of infective agents such as bacteria (Figure 2). Their ability to congregate at localized sites can be appreciated by the study of infected wounds, where pus forms from neutrophils killed after engulfing bacteria and debris. Neutrophils respond chemotactically to a huge variety of molecules. Some of these are produced by the infecting cells themselves, others by other host cells in a coordinated attack on the invading organism (Figure 2). A strong chemotactic response is elicited by peptides such as fMet-Leu-Phe (bacterial cells usually produce proteins with formylated initiating methionine residues, eukaryotic cells do not). Neutrophils also respond to lipopolysaccharides, components of the bacterial cell wall. In addition to homing in on bacteria by direct signals it is possible that neutrophils close to the site relay chemotactic signals to attract more distant neutrophils (Ford-Hutchinson et al., 1980).

Chemokinesis

Cells respond to some extracellular factors such as leukotriene B (Ford-Hutchinson et al., 1980) by increasing the locomotion rate in an undirected manner as opposed to chemotaxis. This mechanism, known as chemokinesis, is likely on purely statistical grounds to result in cells accumulating at the site of origin of this stimulus (Wilkinson, 1987). The differentiation of factors that are chemotactic from chemokinetic responses can be difficult, but this has been greatly facilitated using the "Boyden chamber" (Lackie, 1986).

Haptotaxis

Many cell types are extremely selective about the types of surface on which they will spread or to which they will adhere. Such surfaces guide cells during embryogenesis and determine the fate of metastasizing cells. Cells accumulate in areas of optimal adhesion by passive adhesive entrapment or they may respond directionally to the adhesive gradient (Lackie, 1986). The latter is known as haptotaxis. A peculiar mixture of chemotaxis and haptotaxis has recently been described where a gradient of chemoattractant has been immobilized on a surface, thereby altering adhesion directly by chemoattractant recognition (Rot, 1993). Haptotaxis may be an additional mechanism to accumulate leukocytes at inflammatory foci (Figure 2).

Other Motility Factors

A factor known as scatter factor has been characterized which causes the break up and stimulates motility of epithelial cell clumps (Stoker et al., 1987). This factor is identical to hepatocyte growth factor and increases the rate of locomotion of several other cell types. Motility factors elaborated from tumor cells are considered to play an important role in metastasis (see later). Guidance of cells by the physical topography of the substratum is another factor that profoundly affects the behavior of cells.

THE ROLE OF THE CYTOSKELETON IN CELL LOCOMOTION

A three-dimensional meshwork of proteinaceous filaments of various sizes fills the space between the organelles of all eukaryotic cell types. This material is known collectively as the cytoskeleton, but despite the static property implied by this name, the cytoskeleton is plastic and dynamic. Not only must the cytoplasm move and modify its shape when a cell changes its position or shape, but the cytoskeleton itself causes these movements. In addition to motility, the cytoskeleton plays a role in metabolism. Several glycolytic enzymes are known to be associated with actin filaments, possibly to concentrate substrate and enzymes locally . Many mRNA species appear to be bound by filaments, especially in egg cells where they may be immobilized in distinct regions thereby becoming concentrated in defined tissues upon subsequent cell divisions.

The Microtubule-Based Cytoskeleton

Many cell types have a considerable microtubule-based cytoskeleton. Although available evidence strongly favors a role in intracellular transport and motility, direct involvement in cell locomotion seems unlikely (Schliwa and Höner, 1993). The microtubule organizing center (MTOC) is often seen to occupy a position directly in front of the nucleus in locomoting cells; because of this it has been suggested that microtubules perform a role in the polarization of cells. However, it has also been suggested that the MTOC is merely attempting to maintain its position in the cell centroid (Euteneuer and Schliwa, 1992). Although cells with a highly developed microtubule network, such as fibroblasts and newt eosinophils, become depolarized and lose speed and directionality upon disruption of microtubules, it is likely that microtubule disruption causes concomitant changes in the microfilament system. Cells that are naturally devoid of microtubules, (e.g., *Naegleria*) are capable of efficient and rapid locomotion. A most vivid illustration of this point is provided by fish keratinocytes, among the fastest crawling vertebrate cell type at 1 μm/sec. (Euteneuer and Schliwa, 1984). Fragments of these cells spontaneously break free from cell clusters and move without detectable microtubules. Despite

the above evidence which would argue against a direct involvement in locomotion, microtubules may well play a supporting role in the fine tuning of cell locomotion perhaps by transporting membrane towards the leading edge (see later), or being involved in contact inhibition of locomotion (Nagasaki et al., 1992).

The Intermediate Filament-Based Cytoskeleton

Intermediate filaments form a meshwork throughout tissue culture cells from cell/cell adhesions to a filament rich zone around the nucleus. The role of this network is probably to give physical integrity to tissues. No motor protein has yet been discovered that uses intermediate filaments as a railway, in the way that kinesin and myosin do for microtubules and microfilaments, respectively. The meshwork formed by intermediate filaments is extremely tough, and to some extent dynamic (Vikstrom et al., 1989), but is not thought to contribute directly to the generation of force for locomotion. Intermediate filaments are excluded from actively protruding sites, such as ruffling membranes and cell fragments derived from the lamella of keratinocytes are able to locomote in the apparent absence of this class of filament (Euteneuer and Schliwa, 1984). Moreover, the extent of interconnection between intermediate filaments has been correlated with a decrease in the rate of locomotion of aging skin fibroblasts (E. Wang, 1985). Several studies have produced evidence for a correlation between the expression of a particular type of intermediate filament, keratin, and the acquisition of metastatic potential. A recent study indicates that cell migration and invasion are enhanced by the expression of keratin in mouse L cells. The reason for this increased locomotion is not clear (Chu et al., 1993; Schliwa and Höner, 1993).

The Microfilament-Based Cytoskeleton

The discovery of actin (Hatano and Oosawa, 1966), followed promptly by the discovery of myosin (Adelman and Taylor, 1969) in nonmuscle cells (the slime mold *Physarum*), naturally led workers to formulate hypotheses paralleling those of muscle contraction (Huxley, 1973). This connection was made virtually irresistible when it was discovered that fibroblasts contained minisarcomeres, periodic arrangements of actin and myosin filaments known as stress fibers, which were demonstrated to be contractile when removed from cells by laser microsurgery (Isenberg et al., 1976). However, these structures are now known to be involved in contraction of the cell, e.g., to provide tension for clot retraction and wound healing, rather than being engines for locomotion. Indeed, there appears to be an inverse relationship between the locomotory speed of a particular cell and the number of stress fibers it contains, as discussed above. Time lapse recordings of locomoting fibroblasts show that the leading edge of the cell is its most motile area and consequently most likely to contain the force generating machinery. The word ruffling is most often used to describe this activity. Actin is in a state of constant

polymerization and depolymerization in cells, generally, but especially at the leading edge of the expanding lamellae where it is present at a particularly high concentration. Many factors, such as components of the cell's signaling pathways acting either directly or indirectly on an array of actin binding proteins (Maciver, 1995), are thought to influence this cycle.

The Organization of Microfilaments in Cells

Cell types vary in the arrangement of their microfilaments. In fibroblasts a hierarchy of increasing organization generally exists from the leading lamella to the perinucleus. The leading lamella is dominated by actin filaments arranged in orthogonal arrays and crosslinked by a variety of actin-binding proteins. Also in this region, small bundles of filaments may exist, oriented perpendicularly with respect to the leading edge, often associated with projections from the lamellae. Further rearwards and close to the dorsal surface especially, the actin filaments tend to be more bundled, but these bundles are likely to be parallel to one another increasing in size toward the nucleus and forming arcs (Heath, 1983). In the same region, but towards the ventral surface, large bundles oriented radially from the cell center terminate at focal contacts (see above). The focal contact is a dense plaque containing a large number of proteins that connect molecules physically attached to the extracellular matrix, through the membrane, to the actin bundles within cells. Finally, at the nucleus, thick actin bundles are arranged in a dense "cage" around the nucleus. This hierarchy of organization is reflected by the age of the F-actin from which these structures are fashioned. Actin polymerizes at the leading lamella and the resulting filamentous actin slowly depolymerizes, aggregating into structures as it does so. *In vitro* studies with isolated actin and actin-binding proteins have shown time-dependent increases in order resulting in some highly ordered structures (Maciver, 1991).

MICROFILAMENT DYNAMICS AND CELL LOCOMOTION

Cyclical actin polymerization and depolymerization seems to be pivotal in cell locomotion (Cooper, 1991; Lee et al., 1993; Zigmond, 1993). Polymerization takes place at the edge of cells, especially the leading edge (Y.L. Wang, 1985; Smith, 1988); the site of disassembly is less clear, but depolymerization probably takes place everywhere but the leading edge (Zigmond, 1993; Fechheimer and Zigmond, 1993). *In vitro*, the rates of actin assembly onto the two ends of preformed filaments differ. The barbed end is favored over the pointed end. This situation also seems to hold for actin within cells: filaments polymerizing at the intracellular face of the plasma membrane are oriented with their pointed ends facing the nucleus. As a result of the different assembly rates at the two ends, monomers are likely to add at the barbed end and travel down the filament to fall off at the pointed end, a phenomenon termed treadmilling (Figure 3A). Treadmilling has been postulated to drive lamellae forwards, but the rate of treadmilling is limited by the rate of

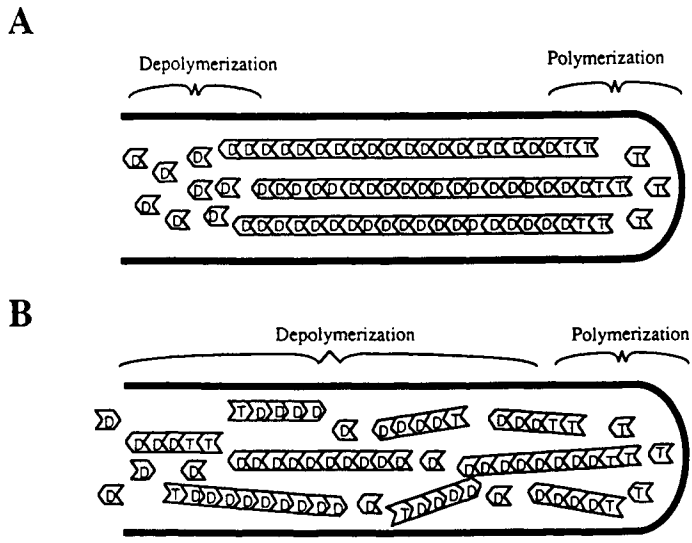


Figure 3. Actin polymerization and cell locomotion. **A:** Simple actin filament treadmilling. In this model actin polymerizes at the leading edge of the locomoting cell, at the barbed ends of filaments. Some distance away from the leading edge these filaments are depolymerizing, releasing monomers for another cycle of polymerization. The difference in behavior of actin at the barbed and pointed ends exists because of the markedly different rates of monomer addition and loss. A weakness of this hypothesis is that the rates measured *in vitro* are not high enough to accommodate the level of lamellar protrusions seen in some cell types. **B:** The nucleation-release model (Theriot and Mitchison, 1992). This model is more complex than the previous although certain key elements have been retained. The lamellae are considered as a meshwork of short actin filaments nucleated and then released from the leading edge. Subunits flux through the lamellae as they are incorporated into filaments which are themselves pushed backwards with respect to the leading edge by the formation of newer filaments. This model is supported by ultrastructural studies.

depolymerization, which at 0.8 monomers per second would permit an extension of 0.13 μm per minute (Theriot and Mitchison, 1991). This is much too slow to account for the rapid rates of locomotion achieved by many cell types. An alternative but related hypothesis has been suggested to account for the flux of actin seen arising from the leading edge. The nucleation-release model (Theriot and Mitchison, 1992) (Figure 3B) considers the lamellae to consist of a moving meshwork of actin filaments rather than long parallel filaments. Actin polymerization can occur only when sufficient actin monomers become available, thus the rate of depolymerization may again be the overall limiting step. Indeed, a tentative link between the rate of depolymerization and the rate of cell locomotion has been made (Zigmond, 1993). The mechanism by which overexpression of gelsolin (a protein

that severs microfilaments) produces an increase in the rate of locomotion (Cunningham et al., 1991) may result directly from an increase in actin flux.

Theoretically at least, it is possible that the energy available from the polymerization of actin could be used to produce lamellar protrusions (Peskin et al., 1993). Certainly it has been found that pure unpolymerized actin within membrane vesicles is able to distort the membrane when the actin polymerizes (Cortese et al., 1989). The acrosomal process of sea cucumber (*Thyone*) sperm has long been cited as an example of actin polymerization driving cellular extension. At rest the actin is held in a non-polymerized form by a profilin like protein, but upon an as yet unknown signal the actin/profilin complex dissociates allowing the actin to polymerize in an almost explosive manner (Tilney and Inoue, 1982). Actin monomers spontaneously polymerize into microfilaments (F-actin) under ionic conditions found in cells. Over 98% of the actin in most cell types is expected to exist in polymerized form, yet the observed amount is closer to 50%. Actin-binding proteins such as profilin have been suggested to be responsible for this, but recent evidence points to a polypeptide of about 5 kDa known as thymosin β4 (Safer, 1992) as the most abundant monomer sequestering protein. Other actin binding proteins thought to sequester monomers include the ADF/cofilin group. Regulation of the dissociation of these complexes is the key to the increased polymerization observed when cells such as leukocytes and *Dictyostelium* are stimulated to move. However, this is not the whole story because actin polymerization occurs only after nuclei (believed to consist of three actin subunits in a particular configuration) have formed. The rate of formation of this trimeric nucleus is very slow, but can be bypassed by certain actin filament capping proteins that nucleate polymerization (reviewed in Weeds and Maciver, 1993). It has been suggested that actin monomers are held at discrete foci located immediately behind the leading edge, prior to their release and polymerization (Cao et al., 1993; Fechheimer and Zigmond, 1993).

Sites of Lamellar Protrusion May Be Determined by the Nucleation of Actin Polymerization

In order to understand the processes controlling actin polymerization it is crucial to determine the limiting steps operating within cells. To this end, experiments have been conducted in which actin nuclei (small fragments of F-actin onto which endogenous cellular actin may polymerize) were microinjected into living cells. Two similar studies have produced exactly the opposite conclusion. Sanders and Wang (1990) concluded that the exogenous nuclei failed to induce detectable polymerization, whereas Handel et al. (1990) found that their nuclei caused dramatic nucleation of endogenous actin. The second group of authors used covalently cross-linked filamentous nuclei which, while permitting actin to polymerize, could not reanneal with each other to extinguish their nucleating activity. Handel and colleagues suggest that this is the likely explanation for these differences. We

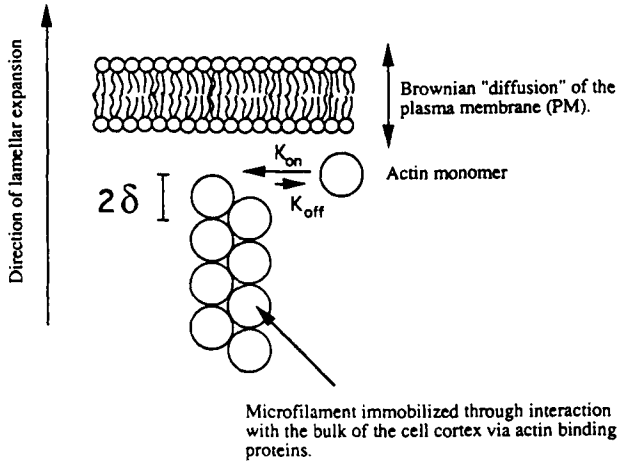


Figure 4. The Brownian ratchet model of lamellar protrusion (Peskin et al., 1993). According to this hypothesis, the distance between the plasma membrane (PM) and the filament end fluctuates randomly. At a point in time when the PM is most distant from the filament end, a new monomer is able to add on. Consequently, the PM is no longer able to return to its former position since the filament is now longer. The filament cannot be pushed backwards by the returning PM as it is locked into the mass of the cell cortex by actin binding proteins. In this way, the PM is permitted to diffuse only in an outward direction. The maximum force which a single filament can exert (the stalling force) is related to the thermal energy of the actin monomer by kinetic theory according to the following equation:

$$\text{Maximum force} = \frac{k_B T}{\delta} \ln\left(\frac{k_{on}}{k_{off}}\right)$$

k_B = Boltzmans constant; T = Absolute temperature; δ = Radius of actin monomer, k = Rate constant

believe that these experiments tell us that localized actin polymerization in a cell can be initiated merely by the supply of nucleation sites and that cells gain polarity by an imbalance in their distribution or nuclei. This makes potentially pivotal those proteins with the ability to nucleate polymerization (Weeds and Maciver, 1993).

Recently, a model for how actin polymerization alone might drive filopodial protrusion was presented (Peskin et al., 1993). The "Brownian Ratchet" model states that a protrusive force is exerted during polymerization of actin by utilizing the energy of Brownian motion in which the plasma-membrane is permitted to diffuse in only one direction by the growing actin filament (Figure 4). The maximum force produced by a single actin filament according to this model is 9 pN (six times that produced by a single myosin S1). A striking feature of this model

is that it requires the barbed end of the filaments to be free, i.e., not capped by other proteins. Consequently the Brownian ratchet model is not easy to reconcile with mechanisms outlined in the preceding paragraph, which describe the importance of nucleated assembly in cell locomotion. Perhaps we can fuse the Brownian ratchet model and the nucleation release models by assuming that other proteins localized at the leading edge of locomoting cells provide privileged conditions for the assembly of actin. Such conditions might be provided by activation of the PIP_2 cycle (Lassing and Lindberg, 1988). Recent evidence (Redmond and Zigmond, 1993) demonstrates that barbed ends are freely available at the edge of fMet-Leu-Phe stimulated neutrophil leukocytes (see Chemotaxis above).

Listeria, an Unlikely Key to Understanding Cellular Actin Nucleation and Possibly Cell Locomotion?

Recently, it has come to light that certain bacteria belonging to the genus *Listeria* are capable of invading cells and use the cellular actin to propel themselves through the cytoplasm of one cell and into neighboring cells (Tilney and Tilney, 1993). The speed at which the bacteria are driven is cell specific varying between 0.02 and 1.5 μm/sec. Propulsion is thought to be achieved by the production of an actin-nucleating activity expressed on the surface of the bacteria. The mechanism is not yet understood, but appears to involve cellular components such as profilin. *Listeria*-induced actin polymerization may be considered as a simplified leading lamella and this system promises to be a useful tool to unravel the role of nucleation in cell locomotion.

The Role of Myosins in Cell Locomotion

Myosins are a diverse group of motor proteins which produce movement along actin filaments in conjunction with ATP hydrolysis. The early discovery of myosin-II homologs in nonmuscle cells led, irresistibly, to theories based on the sliding filament model of skeletal muscle contraction to explain cell locomotion. Several techniques have been employed to disrupt myosin-II function in cells to establish its role in motility. Antibodies to myosin-II have been microinjected into cells, but the results of these experiments have been equivocal. In one study the injected antibodies appeared to increase the motility of epithelial cells and caused a general disruption of their cytoskeletons (Höner et al., 1988). In another study, microinjection of antibodies to myosin-II caused a slight decrease in the locomotory rates of *Acanthamoeba* (Sinard and Pollard, 1989). The disruption of myosin-II heavy chain expression in *Dictyostelium* has provided a spectacularly clear and surprising result. The gene has been disrupted by two methods. The amoebae have been transfected with a vector producing anti-sense mRNA. This binds to, and thereby inactivates, the mRNA encoding the myosin-II heavy chain (Knecht and Loomis, 1987). At the

same time, DeLozanne and Spudich (1987) were able to truncate the single gene encoding the myosin-II heavy chain by homologous recombination made possible by the fact that *Dictyostelium* is haploid. Both methods produced cells that were essentially devoid of functional myosin-II and established that these cells were still able to move (at reduced rates), but could not carry out cytokinesis in the normal fashion. The nagging doubt that both techniques may retain residual myosin-II function was later clarified by the isolation of mutants in which the entire myosin heavy chain gene had been deleted from the genome (Manstein et al., 1989). These myosin-II null mutants had the same phenotype; motility was reduced but still possible. All these experiments indicate that myosin-II is not absolutely required for the locomotion of *Dictyostelium*, but possibly fine tunes the process.

It is most unlikely that the sole functions of mysoin-II in nonmuscle cells are to provide the contractile force to bisect cells during cytokinesis and for the contractility of stress fibers. Myosin-II is present in a variety of cell types at moderate concentrations in tissues such as brain, which are almost totally non-mitotic and do

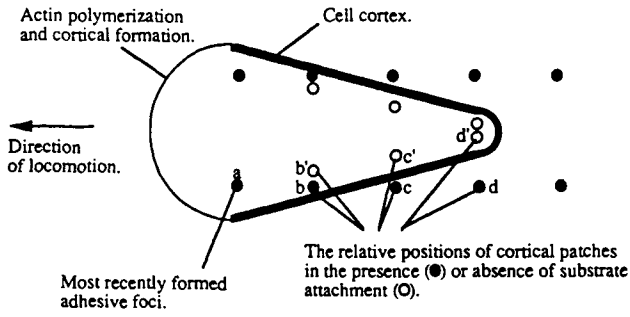


Figure 5. A possible mechanism for the detachment of redundant adhesion in locomoting cells. Actin polymerizes at the leading edge of locomoting cells to produce a layer of micro-filaments arranged in a meshwork with a host of actin binding proteins. Although excluded from the leading pseudopod, myosin-II is able to infiltrate the cortex where it forms thick bipolar filaments. Myosin-II mediated contraction of the cortex (which results in cortical tension absent from myosin-II null cells (Pasternak et al., 1989), and the simultaneous depolymerization of microfilaments leads to a reduction in volume to produce a vaguely cone-shaped cell. Strong, punctate adhesions are produced at the anterior of locomoting cells (a). Soon after these adhesion plaques are formed the cortical tension increases which would result in the sliding of any particular adhesion to a more central point with respect to the cell center if sliding were possible (b-b', c-c' and d-d'). As adhesive foci travel (relative to the cell) to the posterior, so the tension between them and their neighbors increases until one by one they break. (After Maciver, 1987).

not have stress fibers. A possible clue as to other functions comes from the observation that myosin-II null *Dictyostelium* cells lack cortical tension throughout the cell cycle. It has been suggested that the role of myosin-II filaments in interphase cells of this type may be to facilitate the removal of redundant adhesions toward the rear of the locomoting cell (Maciver, 1987) (Figure 5). Cells that gain traction by the usual adhesive mechanisms (Figure 1A) must make adhesions at the anterior and break the same adhesions later at the rear of the cell, where they would otherwise impede further locomotion. Cells from fibroblasts to soil amoebae make strong adhesions to the substrate at specific punctate foci (King and Maciver, 1987). Proteins directly involved with the adhesion plaque are linked to the general cell cortex on the cytoplasmic side of these foci; hence myosin-II mediated cortical contraction also generates tension between these foci, increasing as the distance between the foci increases at the rear of the cell. Evidence for this hypothesis comes from work with amoebae (Bhowmick, 1967) in which myosin-II filaments have been found arranged in the predicted orientation. More recent evidence has come from studies with myosin-II null *Dictyostelium*. Null cells were much less able than wild type to detach adhesions towards the rear when challenged with sticky surfaces (Jay et al., 1993). These null cells are known not to be capable of generating cortical tension (Pasternack et al., 1989). Like actin, myosin-II assembles at the leading edge then slowly disassembles towards the rear of the locomoting cell (Kolega and Taylor, 1993). However, myosin-II assembles at the base of the advancing lamella, not at the leading edge, so it is largely excluded from the lamella itself (De Biasio et al., 1988; Kolega and Taylor, 1993). Consistent with myosin-II having a role in adhesion detachment, it is reported that "assembly competent" myosin-II was "high in the rear of some cells but not in others. Myosin-II concentration and assembly may vary in the rear of the cell depending on the stage of the tail contraction" (Kolega and Taylor, 1993).

Dictyostelium expresses genes encoding at least five distinct myosin-I type proteins, a class of single headed myosins that does not form filaments (Pollard et al., 1991). Some of these genes (known as myoA to myoE) have also been tested by gene disruption. Cells lacking myoA and cells lacking myoB have a very similar phenotype. These cells are slower, pseudopod formation is less frequent, and the cells turn more frequently than control wild type cells (Titus et al., 1993). MyoB is localized at the leading edge of *Dictyostelium* consistent with a direct role in cell locomotion (Fukui et al., 1989). The similarity in phenotypes of myoA$^-$ and myoB$^-$ mutants might be due to both proteins performing a similar task and are thus able to compensate to some extent for each other's absence. A double mutant would have to be created to test this. Other myosin-I isotypes have been found to be localized to the leading edge in *Dictyostelium, Acanthamoeba* (Baines et al., 1992; Yonemura and Pollard, 1992), and vertebrate cells (Wagner et al., 1992).

The Role of Actin-Binding Proteins in Cell Locomotion

It is often very difficult to identify proteins that are specifically and directly involved in activities as complex as cell locomotion. A glycolytic enzyme, for example, might be expected to reduce the rate of cell locomotion simply as a result of diminished general metabolism. Nevertheless, we would predict that alterations in the amount of a single actin-binding protein might result in altered locomotion. To some extent this is what has been found (Table 1), yet disruption of the expression of actin-binding proteins has produced some very surprising results. The case of myosin-II has already been discussed.

The number of actin-binding protein types is large (reviewed in Maciver, 1994) and many of these can be expected to influence actin polymerization. There have been suggestions of great redundancy among cytoskeletal proteins, i.e., a particular protein has a function that is shared with, or overlaps that of, another protein. This may be at least partially true, but presently available techniques may be inadequate to detect the subtle differences that altered actin-binding protein expression causes. In some instances alteration of the expression level has been shown to be deleterious. Thus, reducing the amount of vinculin in fibroblasts increases the rate of cell locomotion by reducing focal contact area and thereby releasing the cells from supra-optimal adhesion. In melanoma cell lines lacking filamin, a reduction in locomotion results apparently from failure to crosslink the actin filaments needed to produce normal pseudopods.

Table 1. The Effect of Altered Concentration of Actin-Binding Proteins on Locomotion

Protein	*Cell Type/Organism*	*Expression Level*	*Effect on Cell Locomotion*	*Reference*
Myosin II	*Dictyostelium*	ablated	little	DeLozanne and Spudich, 1987
Myosin IA	*Dictyostelium*	ablated	decreased	Titus et al., 1993
Myosin IB	*Dictyostelium*	ablated	decreased	Wessels et al., 1989
Severin	*Dictyostelium*	ablated	no apparent change	André et al., 1989
Gelsolin	Vert. Fibroblast	increased 2×	increased rate	Cunningham et al., 1991
ABP 120	*Dictyostelium*	ablated	decreased rate	Cox et al., 1992
Filamin	Melanoma	naturally ablated	decreased rate	Cunningham et al., 1992
α-actinin	*Dictyostelium*	ablated	no apparent change	Wallraff et al., 1986
Villin	Vert. Fibroblast	increased	no apparent change	Franck et al., 1990
Vinculin	Vert. Fibroblast	decreased	increased rate	Luis Roderíguez Fernández et al., 1993

The Transduction of Extracellular Motility Signals to the Cytoskeleton

Extracellular signals of a chemical nature are generally received at the cell surface by binding to a specific receptor, usually a transmembrane protein. Binding of this extracellular signal induces a change in the intracellular domain of the receptor which affects the properties associated with the receptor, such as kinase activity or G-protein binding. The signal transduction cascade produces a series of intracellular messengers such as calcium release, polyphosphoinositide hydrolysis, or protein phosphorylation.

It has often been observed that cells respond to growth factors by increased motility, both of stimulated locomotion and ruffling (O'Neill et al., 1985; Wilkinson, 1986). Growth factors mediate their effects through a variety of different intracellular signaling pathways and these have been suggested to modulate motility differentially. For example, growth factors that utilize the phosphoinositide pathway are found to increase motility (O'Neill et al., 1985; Maciver, 1987; Newell et al., 1987; Lassing and Lindberg, 1988), while those that increase cAMP levels depress motility (Katzin and Gershman, 1984; O'Neil et al., 1985) and severely alter cell shape and cytoskeletal organization (Edwards et al., 1993). Clues are beginning to emerge concerning mechanisms connecting extracellular signals to the actin-based cytoskeleton. Two small GTP-binding proteins are known to be involved in modulating actin dynamics. One of these, *rho*, regulates the assembly of focal contacts and stress fibers in response to various growth factors (Ridley and Hall, 1992), and another, *rac*, regulates growth factor stimulated membrane ruffling (Ridley et al., 1992). The integrins are a group of integral membrane proteins that bind both to the extracellular matrix on the outside of the cell and to cytoskeletal components such as α-actinin and talin on the cytoplasmic side of the plasma membrane. Clustering of integrins at focal contacts (or artificially by antibodies) results in the phosphorylation of a host of cytoskeletal proteins on tyrosine residues (Gingell and Owens, 1992), concomitant with actin polymerization (Löfgren et al., 1993).

Lipid Flow and Cell Locomotion

There is a substantial weight of evidence for the cytoskeleton being responsible for the force production and control of cell locomotion. This view has not yet been accepted unanimously. However, an alternative hypothesis continues to be argued which states that membrane cycling is the motive force driving cell locomotion (Bretscher, 1987). One of the predictions of the membrane flow hypothesis is that there should be a discernible flow of lipid from the front to the rear of the cell. Lipid flow has proven very difficult to study, because of the lack of suitable methods to label single lipid molecules and the heterogenous behavior of membrane-associated proteins. The observation that particles were transported rearward when they bound

to the dorsal surfaces of lamellae of locomoting fibroblasts was seen as evidence for the lipid flow hypothesis (Abercrombie, 1980). Other cell types display quite different behavior. Particles deposited on the large lamellae of the flattened amoeba *Vannella* move rapidly forwards (Hülsmann and Haberey, 1973). Evidence against this hypothesis has come from experiments in which the behavior of 40 nm gold particles on macrophage dorsal surfaces showed the absence of lipid flow (Sheetz et al., 1989). Although many feel the matter has finally been put to rest (Heath and Holifield, 1991), the argument continues (Bretscher, 1989).

The Role of Cell Locomotion in Metastasis

A common feature of solid malignant tumors is the invasion of local tissues. This often follows lines of weakness between different layers of cells and gives rise to the alleged crab shape after which cancer is named. Enzymatic breakdown of the extracellular matrix is unquestionably important in the invasive process, but the tumor cells must also move to form this shape. This could be accomplished either by cell multiplication and/or active locomotion. Increased cell locomotion is often associated with highly metastatic cells. Indeed it is generally thought that locomotion is important both in the spread of cells close to the original malignancy and at distant sites when seed cells have been transported via the blood stream or lymphatics (Sträuli and Weiss, 1977). A wide variety of motility factors that stimulate the locomotion of tumor cells have been characterized. Among these is tumor autocrine motility factor, a 64 kDa protein produced by sarcoma and melanoma cell lines that enhances locomotion of these cells (VanRoy and Mareel, 1992). The association between metastasis and cell locomotion may not be absolute, however, because many growth factors are also stimulators of cell locomotion (O'Neill et al., 1985; Wilkinson, 1986; Lassing and Linberg, 1988). Thus, it can be argued that increased locomotion is merely a by-product of locally produced growth factor activity. Whether increased locomotion is a cause or consequence of metastasis (Grimstad, 1987; Taylor et al., 1993), it seems probable that this property will be a profitable target for anti-metastatic chemotherapy.

INTRACELLULAR MOTILITY

Cells of many types not only move themselves from one point to another, but can be seen transporting material such as organelles around the cell body. Some cells such as the algae *Nitella* (which are incapable of locomotion because of the confinement of a rigid cell wall) are in a state of constant motility known as cytoplasmic streaming. Organelle movement is required by these very large cells in order to exchange gases and nutrients. Most cell types display some level of polarity, in the form of physical shape and the arrangement of organelles. Many protozoan cells exhibit a quite fantastic degree of polarity (Sleigh, 1989). Such order costs the cell energy; organelles must be both directed to particular areas and

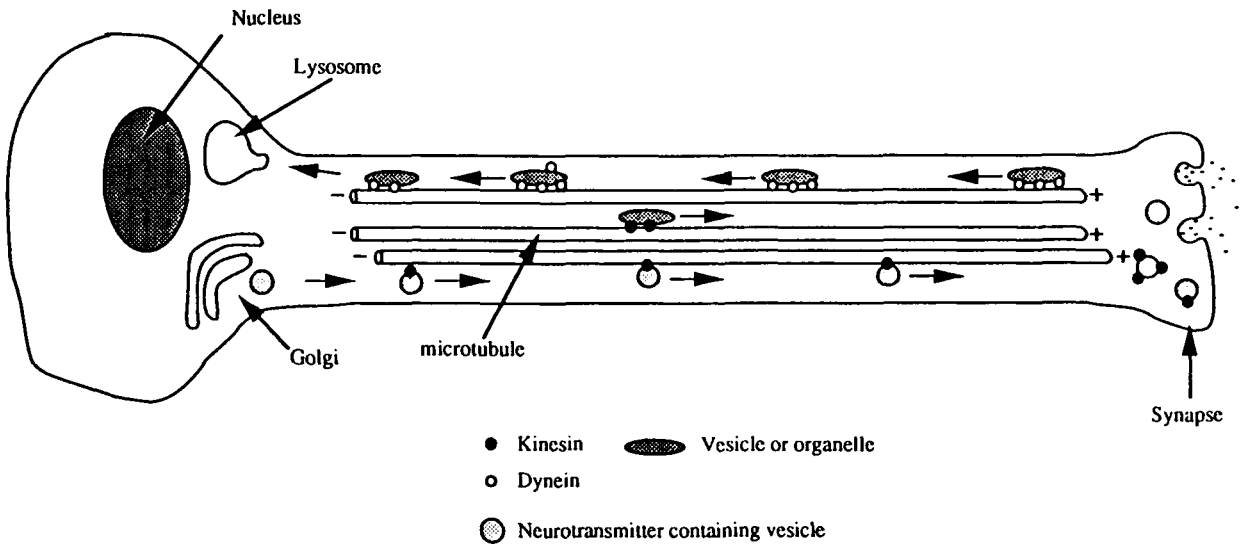


Figure 6. Transport of material along the nerve axon. Materials such as neurotransmitter peptides are synthesized in the cell body and sequestered in vesicles at the Golgi. Vesicles are then transported down the axon towards the synapse by kinesin motors. Other materials are transported from the synapse to the cell body by dynein motors.

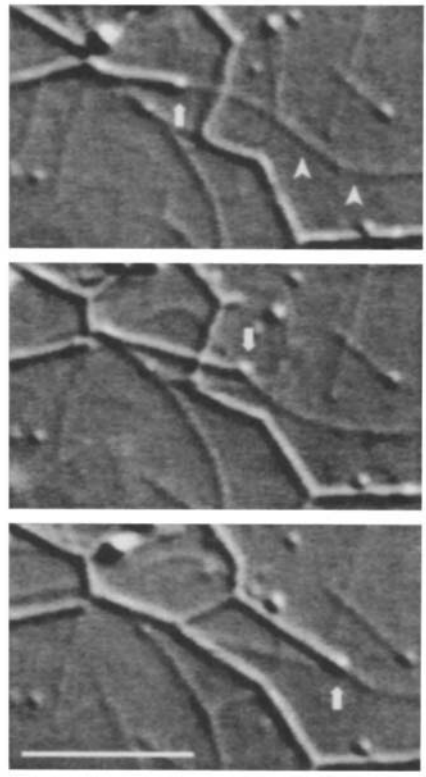


Figure 7. Video-enhanced DIC microscopy of rat liver Golgi apparatus membrane networks moving along microtubules using *Xenopus* egg microtubule motors (Allan and Vale, 1994). Top panel: membrane extension with a bulbous terminus (arrow) attached to a microtubule (arrow heads). Middle panel: same field two seconds later. The membrane has advanced about 3 μm along the microtubule (arrow). Bottom panel: membrane has now advanced further along the microtubule (arrow). Bar = 2 μm.

kept there against chaotic influences. It is very likely that molecular motors are involved in this covert motility.

Microtubule-Based Intracellular Motility

Many cell types use the microtubular lattice in order to move material around. Protozoans are among the most complex cells: many have vast arrays of highly ordered microtubules (Sleigh, 1989). The Golgi apparatus and endoplasmic reticulum in fibroblasts provide examples of this. The Golgi apparatus is arranged as distinct stacks of membranes arranged along microtubules: if the microtubules are

disrupted by colchicine or other agents, the Golgi is seen to collapse quickly and fragment. The endoplasmic reticulum also is arranged along microtubules in a dynamic manner which can be seen in reconstitution experiments (Figure 7). A most spectacular manifestation of microtubule motility is chromatophore movement in the pigment cells of the skin (McNiven and Ward, 1988). Microtubule motors under phosphorylation control cause dispersion and aggregation of the pigment granules to and from the cells' center.

Microtubule-Based Motor Proteins

Just as myosins are able to move along microfilaments, there are motor proteins that move along microtubules. Microtubules, like microfilaments, are polar polymeric assemblies, but unlike actin-myosin interactions, microtubule-based motors exist that move along microtubules in either direction. A constant traffic of vesicles and organelles is visible in cultured cells especially using time-lapse photography. The larger part of this movement takes place on microtubules and is stimulated by phorbol ester (an activator of protein kinase C), and over-expression of N-*Ras* oncoprotein (Alexandrova et al., 1993).

The nerve cell is the most extreme example of a polarized cell morphology (Figure 6). The sciatic nerve of large animals, for example, can be in the region of feet in length, presenting problems in getting material, such as neurotransmitters from the cell body (where they are synthesized) to nerve terminus (where they are released). This outward flow of material, known as anterograde transport has a fast component (up to 4 μm/sec) and a slow component (less than 0.05 μm/sec) (Okabe and Hirokawa, 1989). Retrograde transport (returning to the cell body) is intermediate at about 1.5 μm/sec. Two microtubule-based motor proteins have been isolated from axoplasm and characterized: cytoplasmic dynein (MAP1c) and kinesin. *In vitro* studies suggest that these are at least partly responsible for retrograde and anterograde movement, respectively. Actin filaments also exist within the axoplasm and may additionally facilitate transport of material (Bearer et al., 1993).

The Meiotic and Mitotic Spindles

The processes of meiosis and mitosis involve many motile events, from the separation of the daughter chromosomes to the final act of cell separation at cytokinesis (Wadsworth, 1993). DNA replication itself may be considered as a motile event, because the polymerase complex moves along the linear DNA. One of the most obvious motile events is the separation of the chromosomes along the mitotic spindle at anaphase. Details of the structure and polarity of microtubules in the spindle apparatus in meiosis and mitosis are known through electron and light microscopy, but it is not yet clear whether the chromosomes are pushed, pulled or

both. Microtubule-based motors are known to be associated with these structures (reviewed in Sawin and Endow, 1993). Most of these are related to kinesin but a cytoplasmic dynein is also known. The kinesin-related group are both plus- and minus-end directed while cytoplasmic dynein is minus-end directed. As the large number of kinesin-related motor proteins have been characterized from many different organisms it is not clear how many different motors function in any given spindle or if this heterogeneity merely reflects differences between species.

Microfilament-Based Intracellular Motility

Some cell types that lack cytoplasmic microtubules remain able to carry out intracellular motility. *Labyrinthula* is an organism which appears very similar to the giant amoeba *Reticulamyxa* (Koonce and Schliwa, 1986), a well studied microtubule dependent motile cell, yet *Labyrinthula* seems to perform parallel functions using microfilament bundles. Even in cells with a well developed microtubule system, microfilaments are thought to be responsible for the local delivery of material, brought to the cell periphery by microtubules. In addition to being implicated in force production in cell locomotion, the nonmuscle myosins probably have a role in intracellular motility. Myosin-I is capable of moving organelles along actin cables at least *in vitro* (Adams and Pollard, 1986), and is assumed to perform this function *in vivo* (Adams and Pollard, 1989; Coudrier et al., 1992). It is possible that the microfilament and microtubule systems are responsible for moving the same cargoes at least in yeast where, unexpectedly a defective myosin-I gene was found to be rescued by over-expression of a kinesin-like motor protein (Lillie and Brown, 1992). The protozoan contractile vacuole has no obvious vertebrate homologue. These vacuoles perform a renal function in removing excess water, by slowly filling and then violently contracting, resulting in pumping water out of the cell. A particular type of myosin-I, myosin-IC is specifically localized to the contractile vacuole (Yonemura and Pollard, 1992; Baines et al., 1992). This myosin isotype has a membrane binding tail: it is envisaged that this tail is immobilized at the cytoplasmic surface of the contractile vacuole membrane allowing the motor domain to interact with actin filaments known to be concentrated around the vacuoles.

Cytokinesis

Cytokinesis is the separation of daughter cells at the completion of the microtubule directed separation of duplicate chromosomes at mitosis or meiosis. This is usually accomplished by a purse string mechanism, whereby daughter cells become separated by the gradual constriction of a structure composed of actin microfilaments and myosins-II (Fujiwara and Pollard, 1976; Satterwhite and Pollard, 1992).

A second type of cytokinesis has come to light in mutants of *Dictyostelium* lacking mysoin-II. These cells are able to crudely separate by traction mediated cytokinesis, a process by which the bi- (or multi-) nucleate cells tear themselves apart in an attempt to crawl in two opposite directions.

Motility and Parasitic Protozoans

Humans play unwilling hosts to a variety of protozoans, especially in the tropics. Three genera of amoebae *Entamoeba*, *Naegleria* and *Acanthamoeba* are pathogenic in humans. *Entamoeba* is a common and serious problem infecting some 480 million people world-wide producing debilitating dysentery and life threatening abscesses. *Naegleria* and *Acanthamoeba* cause meningoencephalitis: because of the unfamiliarity of this disease, one suspects that it is very much more of a problem than is currently appreciated. These amoebae actively invade human tissues by crawling movements similar to those of vertebrate cells (but considerably faster). Sporozoan protozoans, which include the malaria producing *Plasmodium* and *Toxoplasma*, are all endo-parasites whose infective stage is a typically elongate cell that moves by a rather mysterious gliding motility (King, 1988). This motility is important for the active invasion of host cells: the machinery of this actin-based system may provide an attractive target for chemotherapeutic drugs (King, 1988).

CHALLENGES AND FUTURE PROSPECTS

Despite over a century of accumulated knowledge in the field of cell motility, many questions remain unanswered. How molecular motors produce physical movement from nucleotide hydrolysis, and what exactly constitutes a motor? Are microtubules and microfilaments themselves motors under some circumstances? The recent solution of the three-dimensional structure of the force producing head of myosin-II (Rayment et al., 1993) provides a structural basis for investigating the so-far elusive conformational changes that are proposed during force production. As far as cell locomotion itself is concerned, the dynamic actin filament system, in concert perhaps with the unconventional myosins, is likely to provide the motive force. However, if the Brownian ratchet model is correct, actin polymerization alone may be sufficient to produce protrusive force, in which case other functions will be sought for these myosin motors. Major challenges include discovering if this is the case, and if so, to elucidate controlling mechanisms. Knowledge concerning the signaling pathways for the inhibition of cytoskeletal activity (upon contact inhibition for example), and activation (upon chemotactic stimulation) might provide suitable targets for drugs to control metastatic tumor spread, parasitic infection, and to stimulate wound healing.

REFERENCES

Abercrombie, M., & Heaysman, J.E.M. (1954). Social behavior of cells in tissue culture. II. Monolayering of fibroblasts. Exp. Cell Res. 6, 293–306.

Abercrombie, M. (1979). Contact inhibition and malignancy. *Nature* 281, 259–262.

Abercrombie, M. (1980). The crawling of metazoan cells. Proc. Roy. Soc. London 207, 129–147.

Adams, R.J., & Pollard, T.D. (1986). Propulsion of organelles isolated from *Acanthamoeba* along actin filaments by myosin-1. Nature 322, 754–756.

Adams, R.J., & Pollard, T.D. (1989). Membrane bound myosin-I provides new mechanism in cell motility. Cell Mot. Cytoskel. 14, 178–182.

Adelman, M.R., & Taylor, E.W. (1969). Further characterization of slime mold myosin and slime mold actin. Biochemistry, 8, 4976–4988.

André, E., Brink, M., Gerisch, G., Isenberg, G., Noegel, A., Schleicher, M., Segall, J.E., & Wallraff, E. (1989). A *Dictyostelium* mutant deficient in severin, an F-actin fragmenting protein, shows normal motility and chemotaxis. J. Cell Biol. 108, 985–995.

Alexandrova, A.Y., Dugina, V.B., Paterson, H., Bershadsky, A.D., & Vasiliev, J.M. (1993). Motility of intracellular particles in rat fibroblasts is greatly enhanced by phorbol ester and by overexpression of normal $p21^{N\text{-}ras}$. Cell Mot. Cytoskel. 25, 254–266.

Allan, V.J., & Vale, R.D. (1994). Movement of membrane tubules along microtubules *in vitro*: Evidence for specialized sites of motor attachment. J. Cell Sci. (in press).

Arencibia, I., Pedari, L., & Sundqvist, K.G. (1987). Induction of motility and alteration of surface membrane polypeptides in lymphocytes by contact with autologous and allogenic fibroblasts. Expt. Cell Res. 172, 124–133.

Baines, I.C., Brzeska, H., & Korn, E.D. (1992). Differential localization of *Acanthamoeba* myosin I isoforms. J. Cell Biol. 119, 1193–1203.

Bearer, E., DeGiorgis. J.A., Bodner, R.A., Kao, A.W., & Reese, T.S. (1993). Evidence for myosin motors on organelles in squid axoplasm. Proc. Natl. Acad. Sci. USA 90, 11252–11256.

Bhowmick, D.K. (1967). Electron microscopy of *Trichamoeba villosa* and ameboid movement. Exp. Cell Res. 45, 570–589.

Bray, D., & White, J.G. (1988). Cortical flow in animal cells. Science, 239, 883–888.

Bretscher, M.S. (1987). How animal cells move. Sci. Amer. 257, 44–51.

Bretscher, M.S. (1989). Particle migration on cells. Nature 341, 491–492.

Brown, A.F. (1982). Neutrophil granulocytes: Adhesion and locomotion on collagen substrata and in collagen matrices. J. Cell Sci. 58, 455–467.

Cao, L.-G., Fishkind, D.J., & Wang, Y.-L. (1993). Localization and dynamics of non-filamentous actin in cultured cells. J. Cell Biol. 123, 173–181.

Chen, W.T. (1981). Surface changes during retraction-induced spreading of fibroblasts. J. Cell Sci. 49, 1–13.

Chu, Y-W., Runyan, R.B., Oshima, R.G., & Hendrix, M.J.C. (1993). Expression of complete keratin filaments in mouse L cells augments cell migration and invasion. Proc. Natl. Acad. Sci. USA 90, 4261–4265.

Cooper, J.A. (1991). The role of actin polymerization in cell motility. Ann. Rev. Physiol. 53, 585–605.

Cooper, M.S., & Schliwa, M. (1986). Motility of cultured fish epidermal cells in the presence and absence of direct current electric fields. J. Cell Biol. 102, 1384–1399.

Cortese., J.D., Schwab III, B., Frieden, C., & Elson, E.L. (1989). Actin polymerization induces a shape change in actin-containing vesicles. Proc. Natl. Acad. Sci. USA 86, 5773–5777.

Couchman, J.R., & Rees, D.A. (1979). The behavior of fibroblasts migrating from chick heart explants: Changes in adhesion, locomotion and growth, and in the distribution of actomyosin and fibronectin. J. Cell Sci. 39, 149–165.

Coudrier, E., Durrbach, A., & Louvard, D. (1992). Do unconventional myosins exert functions in dynamics of membrane compartments? FEBS Lett. 307, 87–92.

Cox, D., Condeelis, J., Wessels, D., Soll, D., Kern, H., & Knecht, D.A. (1992). Targeted disruption of the ABP-120 gene leads to cells with altered motility. J. Cell Biol. 116, 943–955.

Cunningham, C.C., Stossel, T.P., & Kwiatowski, D.J. (1991). Enhanced motility in NIH 3T3 fibroblasts that overexpress gelsolin. Science, 251, 1233–1236.

Cunningham, C.C., Gorlin, J.B., Kwiatowski, D.J., Hartwig, J.H., Janmey, P.A., Byers, R., & Stossel, T.P. (1992). Actin-binding protein requirement for cortical stability & efficient locomotion. Science, 255, 325–327.

De Biasio, R.L., Wang, L.-L., Fisher, G.W., & Taylor, D.L. (1988). The dynamic distribution of fluorescent analogs of actin and myosin in protrusions at the leading edge of migrating Swiss 3T3 fibroblasts. J. Cell Biol. 107, 2631–2645.

de Bruyn, P.P.H. (1947). Theories of amoeboid movement. Quart. Rev. Biol. 22, 1–24.

DeLozanne, A., & Spudich, J.A. (1987). Disruption of the *Dictyostelium* myosin heavy chain gene by homologous recombination. Science 236, 1086–1091.

Dunlevy, J.R., & Couchman, J.R. (1993). Controlled induction of focal adhesion disassembly and migration in primary fibroblasts. J. Cell Sci. 105, 489–500.

Edwards, J.G., Cambell, G., Carr, M., & Edwards, C.C. (1993). Shapes of cells spreading on fibronectin: Measurement of the stellation of BHK21 cells induced by raising cyclic AMP, and of its reversal by erum and lysophosphatidic acid. J. Cell Sci. 104, 399–407.

Euteneuer, U., & Schliwa, M. (1984). Persistent, directional motility of cells and cytoplasmic fragments in the presence of absence. Nature 310, 58–61.

Euteneuer, U., & Schliwa, M. (1992). Mechanism of centrosome positioning during the wound response in BSC-1 cells. J. Cell Biol. 116, 1157–1166.

Ford-Hutchinson, A.W., Bray, M.A., Doig, M.V., Shipley, M.E., & Smith, M.J.H. (1980). Leukotreine B, a potent chemokinetic and aggregating substance released from polymorphonuclear leukoctyes. Nature, 286, 264–265.

Franck, Z., Footer, M., & Bretscher, A. (1990). Microinjection of villin in cultured cells induces rapid and long-lasting changes in cell morphology but does not inhibit cytokinesis, cell motility or membrane ruffling. J. Cell Biol. 111, 2475–2485.

Fujiwara, K., & Pollard, T.D. (1976). Fluorescent antibody localization of myosin in the cytoplasm, cleavage furrow and mitotic spindle of human cells. J. Cell Biol. 71, 848–875.

Fukui, Y.F., Lynch, T.J., Brzeska, H., & Korn, E.D. (1989). Myosin I is located at the leading edges of locomoting *Dictyostelium* amoebae. Nature, 341, 328–331.

Gingell, D., & Owens, N. (1992). How do cells sense and respond to adhesive contacts? Diffusion-trapping of laterally mobile membrane proteins at maturating adhesions may initiate signals leading to local cytoskeletal assembly response and lamella formation. J. Cell Sci. 101, 255–266.

Grimstad, I.A. (1987). Direct evidence that cell locomotion contributes importantly to invasion. Expt. Cell Res. 173, 515–523.

Guelstein, V.I., Ivanova, O.Y., Margolis, L.B., Vasiliev, J.M., & Gelfand, I.M. (1973). Contact inhibition of movement in the culture of transformed cells. Proc. Natl. Acad. Sci. USA 70, 2011–2014.

Handel, S.E., Hendry, K.A.K., & Sheterline, P. (1990). Microinjection of covalently cross-linked actin oligomers causes disruption of existing actin filament architecture in PtK2 cells. J. Cell Sci. 97, 325–333.

Haston, W.S., & Shields, J.M. (1984). Contraction waves in lymphocyte locomotion. J. Cell Sci. 68, 227–242.

Haston, W.S., Shields, J.M., & Wilkinson, P.C. (1982). Lymphocyte locomotion and attachment on two-dimensional surfaces and in three-dimensional matrices. J. Cell Biol. 92, 747–752.

Hatano, S., & Oosawa, F. (1966). Isolation and characterization of plasmodium actin. Biochim. Biophys. Acta 127, 488–498.

Heath, J.P., & Holifield, B.F. (1991). Cell locomotion: New research tests old ideas on membrane and cytoskeletal flow. Cell Mot. Cytoskel. 18, 245–257.

Heath, J.P. (1983). Behavior and structure of the leading lamella in moving fibroblasts. I. Occurrence and centipetal movement of arc-shaped microfilament bundles beneath the dorsal cell surface. J. Cell Sci. 60, 331–354.

Höner, B., Citi, S., Kendrick-Jones, J., & Jockusch, B.M. (1988). Modulation of cellular morphology and locomotory activity by antibodies against myosin. J. Cell Biol. 107, 2181–2189.

Hülsmann, N., & Haberey, M. (1973). Phenomena of amoeboid movement. Behavior of the cell surface in *Hyalodiscus simplex*. (Wohlfarth-Bottermann). Acta Protozoologica 7, 71–82.

Huxley, H.E. (1973). Muscular contraction and cell motility. Nature 243, 445–449.

Isenberg, G., Rathke, P.C., Hulsmann, N., Franke, W.W., & Wolfarth-Botterman, K.E. (1976). Cytoplasmic actomyosin fibrils in tissue culture cells. Direct proof of contractility by visualization of ATP-induced contraction in fibrils isolated by laser microbeam dissection. Cell Tiss. Res. 166, 427–444.

Jay, P.Y., Pham, P.A., & Elson, E.L. (1993). A function of myosin II in *Dictyostelium* cell motility. Mol. Biol. Cell 4, 54a.

Katzin, W.E., & Gershman, H. (1984). Control of cell mobility by cyclic AMP. J. Cell Sci. 65, 177–192.

Keller, H.U., & Cottier, H. (1981). Crawling-like movements and polarization in non-adherent leukocytes. Cell Biol. Int. Rep. 5, 3–7.

King, C.A., & Maciver, S.K. (1987). Spatial organization of microfilaments in amoeboid locomotion. Soc. Gen. Microbiol. 23, 101–118.

King, C.A. (1988). Cell motility of sporozoan protozoa. Parasitology Today 4, 315–319.

Knecht, D.A., & Loomis, W.F. (1987). Antisense RNA inactivation of myosin heavy chain gene expression in *Dictyostelium discoideum*. Science 236, 1081–1086.

Kolega, J., & Taylor, D.L. (1993). Gradients in the concentration and assembly of myosin II in living fibroblasts during locomotion and fiber transport. Mol. Biol. Cell 4, 819–836.

Komnick, H., Stockem, W., & Wohlfarth-Botterman, K.E. (1973). Cell motility: Mechanisms in protoplasmic streaming and amoeboid movement. Int. Rev. Cytol. 34, 169–249.

Koonce, M.P., & Schliwa, M. (1986). Reactivation of organelle movements along the cytoskeletal framework of a giant amoeba. J. Cell Biol. 103, 605–612.

Lackie, J.M. (1982). Aspects of the behavior of neutrophil leucocytes. In: Cell Behavior (Bellairs, R., Curtis, A., & Dunn, G., eds.), Cambridge University Press, Cambridge.

Lackie, J.M. (1986). Cell Movements and Cell Behavior. Allen & Unwin, London.

Lackie, J.M., & Wilkinson, P.C. (1984). Adhesion and locomotion of neutrophil leukocytes on 2-D substrata and in 3-D matrices. In: White Cell Mechanics: Basic Clinical Aspects (Meiselman, H.J., Lichtman, M.A., & LaCelle, P.L., eds.), pp. 237–254. Alan R. Liss, New York.

Lassing, I., & Lindberg, U. (1988). Evidence that the phosphatidyl-inositol cycle is linked to cell motility. Expt. Cell Res. 174, 1–15.

Lewis, W.H. (1931). Locomotion of lymphocytes. Bull. Johns Hopkins Hosp. 49, 29–36.

Lee, J., Ishihara, A., & Jacobson, K. (1993). How do cells move along surfaces? Trends Cell Biol. 3, 366–370.

Lillie, S.H., & Brown, S.S. (1992). Suppression of a myosin defect by a kinsein-related gene. Nature 356, 358–361.

Luis Roderíguez Fernández, J., Geiger, B., Salmon, D., & Ben-Ze'ev, A. (1993). Suppression of vinculin expression by antisense transfection confers changes in cell morphology, motility, and anchorage-dependent growth of 3T3 cells. J. Cell Biol. 122, 1285–1294.

Maciver, S.K. (1987). Cytoskeletal Organization in Amoeboid Locomotion. Ph.D thesis. University College, London.

Maciver, S.K. (1991). The actin filament severing protein actophorin promotes the formation of rigid bundles of actin filaments crosslinked with α-actinin. J. Cell Biol. 115, 1621–1628.

Maciver, S.K. (1994). Actin binding proteins. In: Encylopaedia of Molecular Biology (Kendrew, J., & Lawrence, E., eds.), pp. 7–14. Blackwell, Oxford.

Maciver, S.K. (1995). Actin organization and actin-binding proteins. In: Treatise on the Cytoskeleton (Hesketh, J.E., & Pryme, I.F., eds.), JAI Press Inc., Greenwich.

Manstein, D.J., Titus, M.A., DeLozanne, A., and Spudich, J.A. (1989). Gene replacement in *Dictyostelium*: Generation of myosin null mutants. EMBO J. 8, 923–932.

McNiven, M.A., & Ward, J.B. (1988). Calcium regulation of pigment transport *in vitro*. J. Cell Biol. 106, 111–125.

Nagasaki, T., Chapin, C.J., & Gundersen, G.G. (1992). Distribution of detyrosinated microtubules in motile NRK fibroblasts is rapidly altered upon cell-cell contact: Implications for contact inhibition of locomotion. Cell Mot. Cytoskel. 23, 45–60.

Newell, P.C., Europe-Finner, G.N., & Small, N.V. (1987). Signal transduction during amoebal chemotaxis of *Dictyostelium discoideum*. Microbiol. Sci. 4, 5–11.

Okabe, S., & Hirokawa, N. (1989). Axonal transport. Curr. Opin. Cell Biol. 1, 91–97.

O'Neill, C., Riddle, P., & Rosengurt, E. (1985). Stimulating the proliferation of quiescent 3T3 fibroblasts by peptide growth factors or by agents which elevate cellular cyclic AMP level has opposite effects on motility. Expt. Cell Res. 156, 65–78.

Pasternak, C., Spudich, J.A., & Elson, E.L. (1989). Capping of surface receptors & concomitant cortical tension are generated by conventional myosin. Nature 341, 549–551.

Peskin, C.S., Odell, G.M., & Oster, G.F. (1993). Cellular motions and thermal fluctuations: The brownian ratchet. Biophys. J. 65, 316–324.

Pollard, T.D., Doberstein, S.K., & Zot, H.G. (1991). Myosin I. Ann. Rev. Physiol. 53, 653–681.

Rayment, I., Rypniewski, W.R., Schmidt-Bäse, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G., & Holden, H.M. (1993). Three-dimensional structure of myosin subfragment-1: A molecular motor. Science. 261, 50–65.

Redmond, T., & Zigmond, S.H. (1993). Distribution of F-actin elongation sites in lysed polymorphonuclear leukocytes parallels the distribution of endogenous F-actin. Cell Mot. Cytoskel. 26, 7–28.

Ridley, A.J., & Hall, A. (1992). The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. Cell 70, 389–399.

Ridley, A.J., Paterson, H.F., Johnston, C.L., Diekermann, D., & Hall, A. (1992). The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. Cell, 70, 401–410.

Rot, A. (1993). Neutrophil attractant/activation protein-1 (interleukin-8) induces *in vitro* neutrophil migration by haptotactic mechanism. Eur. J. Immunol. 23, 303–306.

Safer, D. (1992). The interaction of actin with thymosin β4. J. Muscle Res. Cell Mot. 13, 269–271.

Satterwhite, L.L., & Pollard, T.D. (1992). Cytokinesis. Current Biology. 4, 43–52.

Sanders, M.C., & Wang, Y.L. (1990). Exogenous nucleation sites fail to induce detectable polymerization of actin in living cells. J. Cell Biol. 110, 359–365.

Sawin, K.E., & Endow, S.A. (1993). Meiosis, mitosis and microtubule motors. BioEssays 15, 399–407.

Schmalsteig, F.C, Rudloff, H.E., Hillman, G.R., & Anderson, D.C. (1986). Two-dimensional and three-dimensional movement of human polymorphonuclear leukocytes: Two fundamentally different mechanisms of locomotion. J. Leukocyte Biol. 40, 677–691.

Schliwa, M., & Höner, B. (1993). Microtubules, centrosomes and intermediate filaments in directed cell movement. Trends Cell Biol. 3, 377–380.

Sheetz, M.P., Turney, S., Qian, H., & Elson, E.L. (1989). Nanometre-level analysis demonstrates that lipid flow does not drive membrane glycoprotein movements. Nature 340, 284–288.

Sinard, J.H., & Pollard, T.D. (1989). Microinjection into *Acanthamoeba castellanii* of monoclonal antibodies to mysoin II slows but does not stop cell locomotion. Cell Mot. Cytoskel. 12, 42–52.

Skoufias, D.A., & Scholey, J.M. (1993). Cytoplasmic microtubule-based motor proteins. Curr. Opinion Cell Biol. 5, 95–104.

Sleigh, M.A. (1989). Protozoa and Other Protists. Edward Arnold, London.

Small, J.V. (1989). Microfilament-based motility in non-muscle cells. Curr. Opin. Cell Biol. 1, 75–79.

Smith, S.J. (1988). Neuronal cytomechanics: The actin-based motility of growth cones. Science 242, 708–715.

Sträuli, P., & Weiss, L. (1977). Cell locomotion and tumor penetration. Eur. J. Cancer 13, 1–12.
Stoker, M., Gherardi, E., Perryman, M., & Gray, J. (1987). Scatter factor is a fibroblast-derived modulator of epithelial cell motility. Nature, 327, 239–242.
Taylor, W.R., Greenberg, A.H., Turley, E.A., & Wright, J.A. (1993). Cell motility, invasion, and malignancy induced by overexpression of K-FGF or bFGF. Expt. Cell Res. 204, 295–301.
Theriot, J.A., & Mitchison, T.J. (1991). Actin microfilament dynamics in locomoting cells. Nature, 352, 126–131.
Theriot, J.A., & Mitchison, T.J. (1992). The nucleation-release model of actin filament dynamics in cell motility. Trends Cell Biol. 2, 219–222.
Tilney, L.G., & Inoue, S. (1982). The acrosomal reaction of *Thyone* sperm. II. The kinetics and possible mechanism of acrosomal process elongation. J. Cell Biol. 93, 820–827.
Tilney, L.G., & Tilney, M.S. (1993). The wily ways of a parasite: induction of actin assembly by *Listeria*. Trends Microbiol. 1, 25–31.
Titus, M.A., Wessels, D., Spudich, J.A., & Soll, D. (1993). The unconventional myosin encoded by the myoA gene plays a role in *Dictyostelium* motility. Mol. Biol. Cell 4, 233–246.
Van Roy, F., & Mareel, M. (1992). Tumor invasion: effects of cell adhesion and motility. Trends Cell Biol. 2, 163–169.
Vikstrom, K.L., Borisy, G.G., & Goldman, R.D. (1989). Dynamic aspects of intermediate filament networks in BHK-21 cells. Proc. Natl. Acad. Sci. USA 86, 549–553.
Wadsworth, P. (1993). Mitosis: Spindle assembly and chromosome motion. Curr. Opin. Cell Biol. 5, 123–128.
Wagner, M.C., Barylko, B., & Albanesi, J.P. (1992). Tissue distribution and subcellular localization of mammalian myosin I. J. Cell Biol. 119, 163–170.
Wallraff, E., Schleicher, M., Modersitzki, M., Reiger, D., Isenberg, G., & Gerish, G. (1986). Selection of *Dictyostelium* mutants defective in cytoskeletal protein: Use of an antibody that binds to the ends of alpha-actinin rods. EMBO J. 5, 61–67.
Wang, E. (1985). Are cross-bridging structures involved in the bundle formation of intermediate filaments and the decrease in locomotion that accompany cell aging? J. Cell Biol. 100, 1466–1473.
Wang, Y.L. (1985). Exchange of actin subunits at the leading edge of living fibroblasts: Possible role of treadmilling. J. Cell Biol. 101, 597–602.
Weeds, A.G., & Maciver, S.K. (1993). F-actin capping proteins. Curr. Opin. Cell Biol. 5, 63–69.
Wessels, D., Soll, D.R., Knecht, D., Loomis, W.F., De Lozanne, A., & Spudich, J. (1989). Myosin IB null mutants of *Dictyostelium* exhibit abnormalities in motility. Cell Mot. Cytoskel. 20, 301–315.
Wilkinson, P.C. (1986). The locomotor capacity of human lymphocytes and its enhancement by cell growth. Immunology 57, 281–289.
Wilkinson, P.C. (1987). Leukocyte locomotion: Behavioral mechanisms for accumulation. J Cell Sci. Suppl. 8, 103–119.
Yonemura, S., & Pollard, T.D. (1992). The localization of myosin-I and myosin-II in *Acanthamoeba* by fluorescence microscopy. J. Cell Sci. 102, 629–642.
Zigmond, S.H. (1993). Recent quantitative studies of actin filament turnover during cell locomotion. Cell Mot. Cytoskel. 25, 309–316.

RECOMMENDED READINGS

Bray, D. (1992). Cell Movements. Garland, New York.
King, C.A., & Preston, T.M. (1992). Cell Motility. In: Fundamentals of Cell Biology, Vol. 5B Membrane Dynamics and Signalling, pp. 145–182. JAI Press Inc, Greenwich.
Lackie, J.M. (1986). Cell Movements and Cell Behavior. Allen & Unwin, London.
Preston, T.M., King, C.A., & Hyams, J.S. (1990). The Cytoskeleton and Cell Motility, Blackie, Glasgow.
Stossel, T.P. (1993). On the crawling of animal cells. Science 260, 1086–1094.

Chapter 5

Mitochondrial Oxidations and ATP Synthesis in Muscle

D.M. TURNBULL and H.S.A. SHERRATT

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 107–154.

ISBN: 1-55938-808-0

INTRODUCTION

ATP is the immediate source of chemical energy for muscle. During vigorous short-term exercise when the oxygen supply is limiting, this is provided by the conversion of muscle glycogen to lactate by glycolysis. ATP is provided much more efficiently during sustained moderate exercise associated with the oxidation of the metabolic fuels by dioxygen (O_2) of which glucose, non-esterified fatty acids, and ketone bodies (3-hydroxybutyrate and acetoacetate) are the most important. This process is called oxidative phosphorylation and only occurs in the mitochondria and depends critically on their structure. In this chapter, we discuss mitochondrial oxidations and ATP synthesis, and some topics are treated in more detail in appendixes. There is a very large literature on mitochondria much of which is impenetrable, sometimes even to specialists! Most textbooks unfortunately contain misleading statements about the stoichiometry of ATP synthesis, so-called high energy bonds and thermodynamic efficiency. The books by Harold (1986), Nicholls and Fergusson (1992), and Tyler (1992) are strongly recommended. Information about experimental methods is given by Darley-Usmar et al. (1987), Sherratt et al. (1988), and Birch-Machin et al. (1993).

MITOCHONDRIAL STRUCTURE

The mitochondrion has an outer and an inner membrane (Figure 1). The outer membrane contains pores formed from a protein, porin, which allow exchange of molecules with molecular weights up to about 2,000 M_r between the cytosol and the intermembrane space. The inner membrane is extensively invaginated to increase its surface area. It has a different lipid composition from the outer membrane and is rich in the acidic phospholipid cardiolipin (diphosphatidyl-glycerol) which is only found in animal cells in mitochondria. Cardiolipin confers good electrical insulating properties on the inner membrane which is impermeable

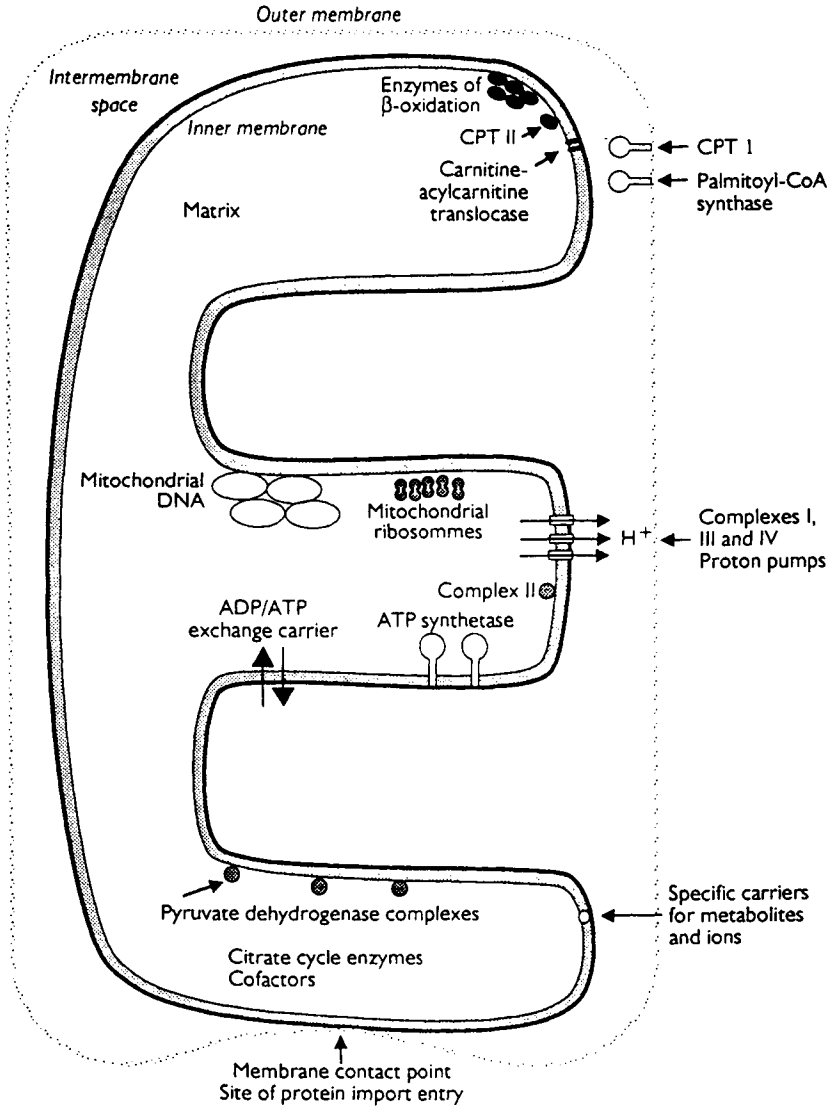


Figure 1. Structural and functional organization of the mitochondrion.

to most polar molecules except for O_2, CO_2, H_2O and NH_3. The cytochromes and the redox carriers of the respiratory chain, ATP synthase, and some other enzymes are integral components of the inner membrane or are closely associated with it. Those substrates involved in mitochondrial metabolism enter the matrix on specific carriers (Table 1). The matrix compartment contains a high concentration of enzymes and cofactors necessary for substrate oxidations.

Table 1. Some Metabolite Transporters in the Inner Membrane of Muscle Mitochondria

Transporter for	*Cytosol Compartment*	*Matrix Compartment*	*Remarks*
Adenine nucleotides	ADP^{3-}	ATP^{4}	Electrogenic antiport, ADP^{3-} uptake favored by the membrane potential: can also mediate electroneutral exchange of ADP^{3-}_{OUT} for ADP^{3-}_{IN} and ATP^{4-}_{OUT} for ATP^{4-}_{IN}
Phosphate (P_i^-)	P_i^- + H^+		Electroneutral symport, has also been formulated as P_i^-/OH^- antiport
Dicarboxylates/P_i^{2-} exchange	$Malate^{2-}$ or $succinate^{2-}$	P_i^{2-}	Electroneutral antiport
2-Oxoglutarate/malate exchange	$2\text{-Oxoglutarate}^{2-}$	$Malate^{2-}$	Electroneutral antiport
Pyruvate	Pyruvate	HO^-	Electroneutral antiport
Glutamate/aspartate exchange	$Glutamate^{2-}$ + H^+	$Aspartate^{2-}$	Electrogenic antiport, can only operate in the direction of aspartate efflux as import of H^+ is a requirement for the malate/aspartate shuttle
Acyl-carnitine exchange	Acyl-carnitine	Carnitine	Electroneutral antiport, both species zwitterions, at pH 7–8. Presumably can exchange $acyl\text{-}carnitine_{OUT}$ for $acyl\text{-}carnitine_{IN}$ or $carnitine_{OUT}$ for $carnitine_{IN}$
Calcium/sodium exchange	Ca^{2+}	$2Na^+$	Electroneutral antiport
Calcium	Ca^{2+}		Electrogenic uniport, driven by the membrane potential

Note: These are examples of important transporters involved in substrate and ADP uptake into the matrix compartment as indicated, and most are reversible. These transporters are proteins and several have been isolated and sequenced. Other specific carriers occur in mitochondria from other tissues. The inner membrane does not allow rapid exchange of NAD^+ or CoA but there are mechanisms for the slow uniport of cofactors synthesized extramitochondrially.

Mitochondria have their own DNA (mtDNA) and genetic continuity. This DNA only encodes 13 peptide subunits synthesized in the matrix that are components of complexes I, III, IV, and V of the respiratory chain. Most mitochondrial proteins are synthesized on cytoplasmic ribosomes and imported by specific mechanisms to their specific locations in the mitochondrion (see below).

Skeletal muscle contains three types of fiber: fast-twitch oxidative glycolytic (type 2A), fast-twitch glycolytic (type 2B), and slow-twitch oxidative fibers (type 1). The proportion of each fiber type varies in different muscles. Different fiber types contain different isoforms of myosin, although there is no evidence that their mitochondria differ qualitatively. It has been reported that there are differences between subsarcolemmal mitochondria and those deeper in the same fiber but this has been questioned (see Sherratt et al., 1988 for references).

PATHWAYS OF SUBSTRATE OXIDATIONS

Glucose and Glycogen Metabolism

Glucose enters muscle cells from the plasma, mediated by specific glucose transporters in the plasma membrane, and insulin increases the rate of glucose transport. Glucose is stored in polymeric form as glycogen in resting muscle. Glycogen synthesis is promoted by insulin, and glycogen degradation to glucose 6-phosphate is stimulated by epinephrine and during intense exercise. The activities of the enzymes involved in glycogen synthesis and breakdown are modified by ATP-dependent phosphorylation and dephosphorylation which are under hormonal control (Hue and Van der Werve, 1981). Glucose 6-phosphate is formed from glucose by hexokinase:

$$\text{Glucose} + \text{ATP} \rightarrow \text{Glucose 6-phosphate} + \text{ADP} + H_2O.$$

Glucose 6-phosphate is then converted by phosphoglucomutase to fructose 6-phosphate which enters the glycolytic pathway, or by phosphoglucose isomerase to glucose 1-phosphate which is a precursor for glycogen synthesis. The conversion of one molecule of glucose to two molecules of lactate yields two molecules of ATP and the conversion of each glucose unit of glycogen to two molecules of lactate yields three molecules of ATP. The complete oxidation of one molecule of glucose may yield up to about 38 molecules of ATP, although the actual number is not known (see p. 150). During glycolysis pyruvate is reduced to lactate by NADH (generated at the stage of glyceraldehyde 3-phosphate oxidation to 1,3-bisphosphoglycerate).

$$CH_3COCO_2^- + \text{NADH} + H^+ \rightleftharpoons CH_3CH(OH)CO_2^- + \text{NAD}^+$$

If the glycolytic flux is slow much of the pyruvate formed enters the mitochondria and is oxidized by the citrate cycle and reducing equivalents (2H) from NADH are oxidized indirectly (see below). When the flux is fast there is net production of

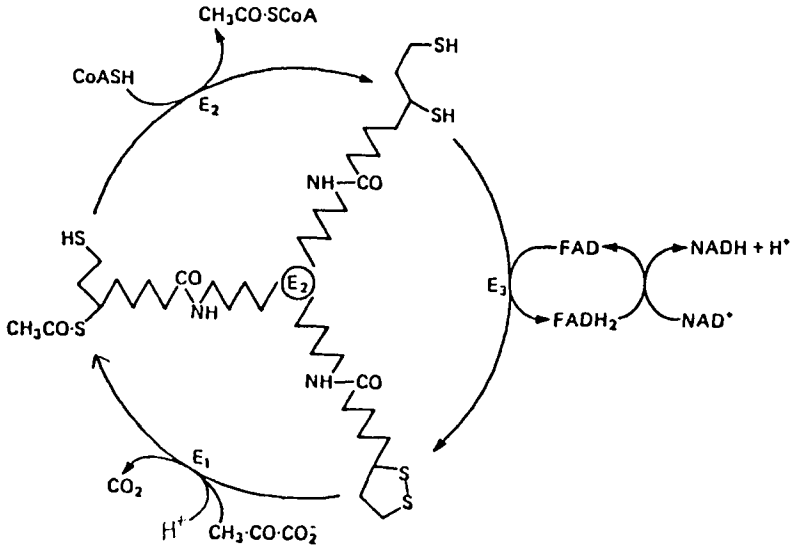


Figure 2. Mechanism of PDH. The three different subunits of the PDH complex in the mitochondrial matrix (E_1, pyruvate decarboxylase; E_2, dihydrolipoamide acyltransferase; E_3, dihydrolipoamide dehydrogenase) catalyze the oxidative decarboxylation of pyruvate to acetyl-CoA and CO_2. E_1 decarboxylates pyruvate and transfers the acetyl-group to lipoamide. Lipoamide is linked to the group of a lysine residue to E_2 to form a flexible chain which rotates between the active sites of E_1, E_2, and E_3. E_2 then transfers the acetyl-group from lipoamide to CoASH leaving the lipoamide in the reduced form. This in turn is oxidized by E_3, which is an NAD-dependent (low potential) flavoprotein, completing the catalytic cycle. PDH activity is controlled in two ways: by product inhibition by NADH and acetyl-CoA formed from pyruvate (or by β-oxidation), and by inactivation by phosphorylation of E_2 by a specific ATP-dependent protein kinase associated with the complex, or activation by dephosphorylation by a specific phosphoprotein phosphatase. The phosphatase is activated by increases in the concentration of Ca^{2+} in the matrix. The combination of insulin with its cell surface receptor activates PDH by activating the phosphatase by an unknown mechanism.

lactate by muscle. Some lactate diffuses out of muscle cells and is then transported by the circulation to the liver where it is largely reconverted to glucose by gluconeogenesis.

Pyruvate Oxidation

Many metabolic fuels are oxidized in the mitochondrial matrix. Pyruvate is oxidatively decarboxylated to acetyl-CoA by the pyruvate dehydrogenase complex (PDH)

$$CH_3COCO_2^- + CoASH + NAD^+ \rightarrow CH_3COSCoA + NADH + H^+ + CO_2$$

as described in Figure 2.

Mitochondrial Fatty Acid Oxidation

Straight-chain even numbered fatty acids are converted to acetyl-CoA by β-oxidation. β-Oxidation of acyl-CoA esters is a cyclic process involving four successive types of reactions: flavoprotein-linked 2,3-dehydrogenation; hydration; NAD^+-linked dehydrogenation; and CoA-dependent thiolysis, to give acetyl-CoA and an acyl-CoA with two carbon atoms less. Each type of reaction is catalyzed by two or more enzymes with different, but partly overlapping, chain-length specificities (Figure 3). Electrons derived from the flavoprotein dehydrogenation at the 2,3 position of the acyl-CoA esters of fatty acids and of some branched-chain acyl-CoA esters involved in the catabolism of leucine, isoleucine, and valine are transferred by the (apparently) soluble electron-transferring flavoprotein (ETF) in the matrix to ETF dehydrogenase on the inner face of the inner mitochondrial membrane and then via ubiquinone to complex III. The oxidation of polyunsaturated fatty acids (including linolenate, linoleate) requires in addition two other enzymes, the NADPH-linked 4-enoyl-CoA reductase and 2,3-enoyl-CoA isomerase (Schulz, 1985; Osmundsen and Hovick, 1988). Some long-chain fatty acids are partly oxidized in the peroxisomes, although this is not discussed here.

Transport of Fatty Acid Acyl Groups Into the Mitochondrial Matrix

Long-chain fatty acids (e.g., palmitate C_{16}) diffuse through pores in the outer mitochondrial membrane, and then form long-chain acyl-CoA esters catalyzed reversibly by palmitoyl-CoA synthase (assumed to be on the inner face of the outer membrane).

$$CH_3(CH_2)_{14}CO_2H + ATP + CoASH \rightleftharpoons$$

$$CH_3(CH_2)_{14}COSCoA + AMP + PP_i + 2H_2O$$

Long-chain acyl-CoA esters are then converted to acylcarnitine esters by readily reversible reactions with L-carnitine catalyzed by carnitine palmitoyltransferase I (CPT I).

$$CH_3(CH_2)_{14}COSCoA + \begin{array}{l} CH_2N^+(CH_3)_3 \\ | \\ HOCH \\ | \\ CH_2CO_2^- \end{array} \rightleftharpoons CH_3(CH)_{14}COOCH \begin{array}{l} CH_2N^+(CH_3)_3 \\ | \\ \\ | \\ CH_2CO_2^- \end{array} + CoASH$$

CPT I spans the outer membrane with its catalytic site on the inner face. Malonyl-CoA, which is an intermediate in fatty acid synthesis inhibits CPT I by binding to a regulatory site on the outer face (Murphy and Pande, 1987). This is thought to limit fatty acid oxidation in liver in fed animals. It is not known whether malonyl-CoA regulates CPT I activity in tissues other than liver. Long-chain acylcarnitine esters enter the matrix in exchange for carnitine on the carnitine-acylcarnitine exchange carrier in the inner membrane, and the acyl groups are then transferred to CoA in the matrix by the reverse reaction catalyzed by carnitine palmitoyltransferase II (CPT II) on the inner face of the inner membrane. Long-chain acyl CoA esters then undergo β-oxidation; the acetyl groups formed are oxidized by the citrate

---

Figure 3. Mitochondrial fatty acid oxidation. Long-chain fatty acids are converted to their CoA-esters as described in the text, and their fatty-acyl-groups transferred to CoA in the matrix by the concerted action of CPT I, the acylcarnitine/carnitine exchange carrier and CPT (**A**) as described in the text. Medium-chain and short-chain fatty acids (C_8 or less) diffuse directly into the matrix where they are converted to their acyl-CoA esters by a acyl-CoA synthase. The mechanism of β-oxidation is shown below (**B**). Each cycle of β-oxidation removes $-CH_2-CH_2-$ as an acetyl unit until the fatty acids are completely converted to acetyl-CoA. The enzymes catalyzing each stage of β-oxidation have different but overlapping specificities. In muscle mitochondria, most acetyl-CoA is oxidized to CO_2 and H_2O by the citrate cycle (Figure 4); some is converted to acylcarnitine by carnitine acetyltransferase (associated with the inner face of the inner membrane) and exported from the matrix. Some acetyl-CoA (if in excess) is hydrolyzed to acetate and CoASH by acetyl-CoA hydrolase in the matrix. Enzymes:

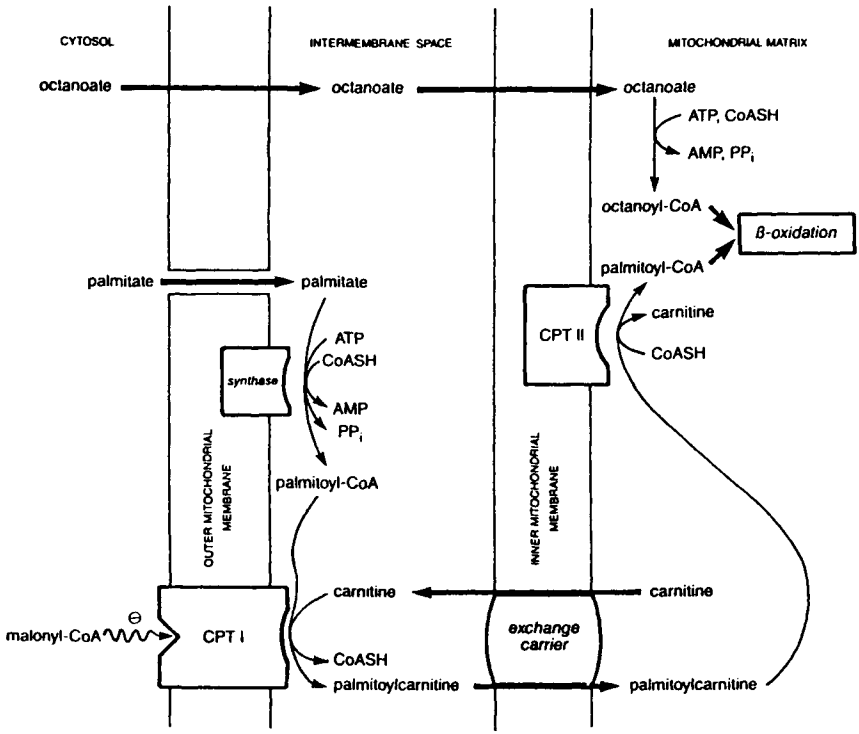
1. Very-long-chain acyl-CoA dehydrogenase
2. Long-chain acyl-CoA dehydrogenase
3. Medium-chain acyl-CoA dehydrogenase
4. Short-chain acyl-CoA dehydrogenase

(1–4) Contain FAD prosthetic groups. These dehydrogenases are limited to the respiratory chain at the level of complex III by ETF (5) dehydrogenase (6) and ubiquinone.

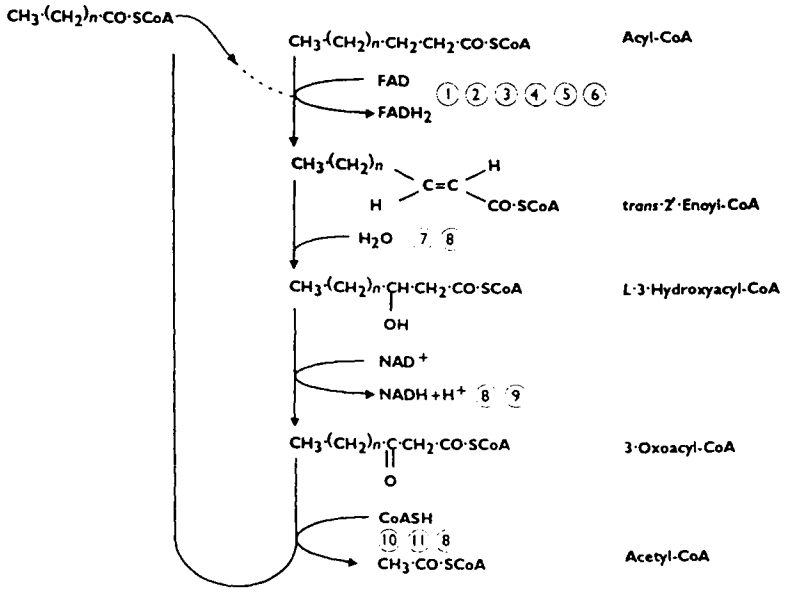
5. Electron-transfer flavoprotein (ETF)
6. ETF dehydrogenase
7. Long-chain 2-enoyl-CoA hydratase
8. Short-chain 2-enoyl-CoA hydratase
9. Long-chain 3-hydroxyacyl-CoA dehydrogenase
10. Short-chain 3-hydroxyacyl-CoA dehydrogenase
11. Acetoacetyl-CoA thiolase
12. Medium-chain 3-oxoacyl-CoA thiolase
13. Long-chain acyl-CoA thiolase

Enzymes 7, 9, and 13 form a trifunctional protein associated with the inner face of the inner mitochondrial membrane. Very-long-chain acyl-CoA dehydrogenase is also associated with other inner mitochondrial membranes while the other enzymes are in the matrix and may be loosely associated with the inner face of the inner membrane. A medium-chain 2-enoyl-CoA hydratase may also be present in the mitochondrial matrix.

A. Mitochondrial Fatty Acid Oxidation



B. Pathway of β-oxidation



cycle in most tissues, but are mainly converted to acetoacetate and 3-hydroxybutyrate in liver. Octanoate (a medium-chain fatty acid, C_8) can bypass CPT I by diffusing directly into the matrix where octanoyl-CoA formed by medium-chain acyl-CoA synthase is oxidized. However, contrary to the usual belief, octanoyl-CoA can also be formed by palmitoyl-CoA synthase and the octanoyl group transferred to CoA in the matrix as its carnitine ester (Otto, 1984).

There is controversy about the nature of CPT I and CPT II; one view is that they may be the same protein and that CPT I has an additional subunit with a binding site for malonyl-CoA. However, recent evidence strongly suggests that they are distinct proteins with CPT I consisting of a single peptide chain (about 88 kDa) containing a malonyl-CoA binding site and that CPT II is a single peptide chain (about 71 kDa). CPT I has different isoforms in skeletal muscle, heart, and liver. but CPT II appears to be the same (Esser et al., 1993).

Oxidation of Acetoacetate and 3-Hydroxybutyrate (Ketone Bodies)

Most of the acetyl-CoA formed by β-oxidation in liver is converted to acetoacetate by the 3-hydroxy-3-methylglutaryl-CoA pathway (Guzmán and Gelen, 1993). Acetoacetate is reversibly converted to D-3-hydroxybutyrate by D-3-hydroxybutyrate dehydrogenase in the mitochondrial matrix in all tissues.

$$CH_3COCH_2CO_2^- + NADH + H^+ \rightleftharpoons CH_3CH(OH)CH_2CO_2^- + NAD^+$$

In peripheral tissues acetoacetate exported by the liver reacts with succinyl-CoA formed in the citrate cycle to give acetoacetyl-CoA and succinate catalyzed by a specific CoA transferase.

$$CH_3COCH_2CO_2^- + {}^-O_2CCH_2CH_2COSCoA \rightleftharpoons CH_3COCH_2COSCoA + {}^-O_2CCH_2CH_2CO_2^-$$

The acetoacetyl-CoA then reacts with CoASH to yield two molecules of acetyl-CoA, catalyzed by acetoacetyl-CoA thiolase

$$CH_3COCH_2COSCoA + CoASH \rightleftharpoons 2CH_3COSCoA$$

which are then oxidized by the citrate cycle.

Metabolism of Amino Acids

The metabolism of amino acids is complex and is described in standard text books. These are usually converted by aminotransferases to the corresponding 2-oxoacids which are partly oxidized in the matrix of muscle mitochondria and partly exported to the liver. Glutamate and aspartate yield 2-oxoglutarate and oxaloacetate, respectively, which enter the citrate cycle directly, and other 2-

oxoacids can be metabolized to products including acetyl-CoA, acetoacetate, fumarate, and succinyl-CoA. Some glutamate is converted to 2-oxoglutarate and NH_3 by glutamate dehydrogenase.

The Citrate Cycle

The citrate cycle is the final common pathway for the oxidation of acetyl-CoA derived from the metabolism of pyruvate, fatty acids, ketone bodies, and amino acids (Krebs, 1943; Greville, 1968). This is sometimes known as the Krebs or tricarboxylic acid cycle. Acetyl-CoA combines with oxaloacetate to form citrate which then undergoes a series of reactions involving the loss of two molecules of CO_2 and four dehydrogenation steps. These reactions complete the cycle by regenerating oxaloacetate which can react with another molecule of acetyl-CoA (Figure 4).

The Mitochondrial Matrix

There are very high concentrations of proteins (estimated at 40%) and cofactors (total adenine nucleotides, NAD, NADP, CoA and its esters, and glutathione each of the order of 5 mM) (about 25 mM) in the matrix. Many of the enzymes catalyzing substrate utilization may be soluble in the matrix, including 2-oxoglutarate dehydrogenase, which is a complex of multiple copies of component enzymes resembling PDH (Yeaman, 1989), or loosely associated with the inner membrane. Very gentle sonication or osmotic shock may damage the membranes but leave some enzymes of the citrate cycle attached (Robinson and Srere, 1985). Some of the enzymes involved in β-oxidation appear to bind to the inner membrane, for example, the long-chain trifunctional enzyme, enoyl-CoA hydratase: 3-hydroxyacyl-CoA dehydrogenase; 3-oxoacyl-CoA thiolase (Uchida et al., 1992). Succinate dehydrogenase (complex II) is an integral component of the inner membrane (Figure 2). Because of the extensive invagination of the inner membrane, particularly in skeletal muscle mitochondria, all the soluble matrix proteins are within a few molecular diameters of the inner face of the membrane.

It has often been questioned whether the rates and kinetics of purified enzymes, determined in very dilute solutions with high concentrations of their substrates, but not always of their cofactors, can be extrapolated to the conditions prevailing in the matrix. Much of the mitochondrial water will be bound to protein by hydrogen bonds and electrostatically, but there is also a pool of free water which may only be a fraction of the total water (Gitomer, 1987). The molar concentrations of intermediates of the citrate cycle and of β-oxidation are very low, usually less than those of most enzymes (Srere, 1987; Watmough et al., 1989; Sumegi et al., 1991). The extent to which cofactors and intermediates bind specifically or nonspecifically to enzymes is not known. It is therefore difficult to estimate concentration of these

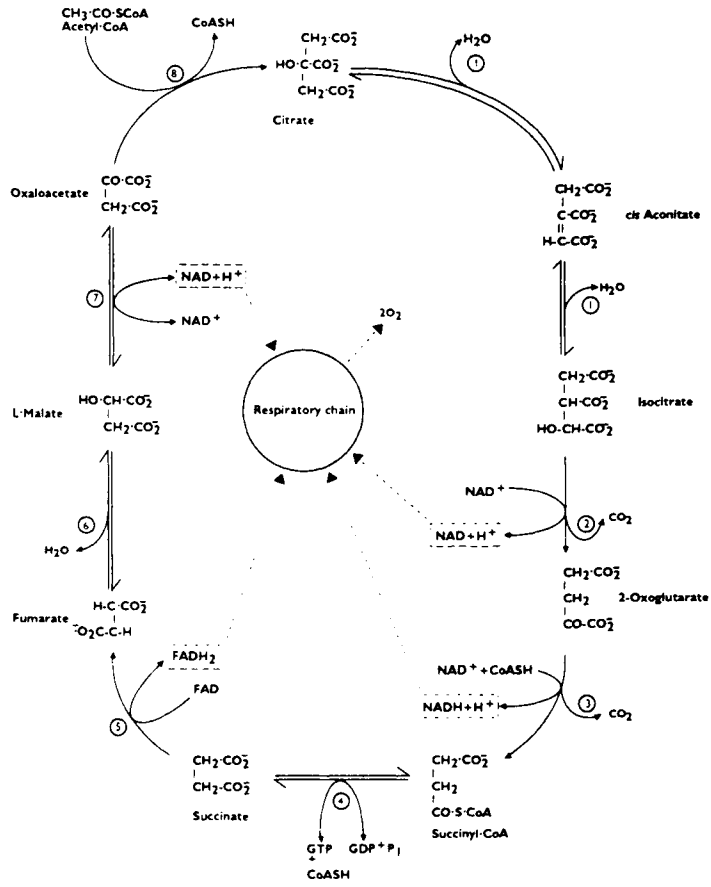


Figure 4. The citrate cycle. There is complete oxidation of one molecule of acetyl-CoA for each turn of the cycle $CH_3COSCoA + 2O_2 \rightarrow 2CO_2 + H_2O + CoASH$. The rate of the citrate cycle is determined by many factors including the ADP/ATP ratio, $NAD^+/NADH$ ratio, and substrate concentrations. During muscle contraction, Ca^{2+} is released from cellular stores (mainly the sarcoplasmic reticulum) and then taken up in part by the mitochondria (see Table 2). Ca^{2+} activates 2-oxoglutarate and isocitrate dehydrogenases (Brown, 1992). Succinate dehydrogenase may be effectively irreversible. Enzymes:

1. Aconitase
2. Isocitrate dehydrogenase
3. 2-Oxoglutarate dehydrogenase
4. Succinyl-CoA synthase
5. Succinate-dehydrogenase (complex II)
6. Fumarase
7. Malate dehydrogenase
8. Citrate synthase

Table 2. Components of Electron Transport and ATP Synthesis in Mammalian Mitochondria

Component	*Molecular Size*	*Number of Peptide Subunits*	*Prosthetic Groups*	*Topology in Inner Membrane*	*Abundance in Inner Membrane (nmol per mg Protein. Data for Cardiac Mitochondria)*	*Proton Movements (the Stoichiometry is Discussed in Appendix 3)*
Complex I	850 kDa (probably a dimer in membrane)	About 40	1 FMN covalently bound, bound 16–24 Fe-S atoms in 5 to 7 centers	Spans membrane, NADH site on matrix face, UQ site in membrane	0.06 UQ	Pumps protons out of matrix during electron transport/$2e^-$
Complex II	120 kDa	4	1 FAD covalently bound, 8 Fe-S atoms in 3 centers	In inner membrane, succinate site on matrix face. UQ site in membrane.	0.19 UQ	None
Ubiquinone 10	830 kDa	—	—	Diffuses within lipid phase of inner membrane	6–8	—
Complex III	280 kDa	11	2B type hemes (b_i and b_o) bound to same mitochondrially coded peptide 1 C heme (cytochrome c_1) 1Fe-S center Rieske factor	Spans membrane, cytochrome b_1 and b_0 in membrane, cytochrome c, and Fe-S center on outer face	0.25–0.53	Pumps protons out of matrix during electron transport/$2e^-$
Cytochrome c	12.3 kDa	1	C type heme	Inter membrane space, loosely associated with inner membrane	0.8–1.02	—
Complex IV	200 kDa (probably as a dimer)	13	2 A type hemes (a_1, a_3) 2 or 3 Cu	Spans inner membrane, cytochrome c site on outer face	0.16–1.00	Pumps protons out of matrix during electron transport/$4e^-$
Complex V	370 kDa	About 16	3 bound adenine nucleotides	Base (F_o) spans membrane, connected to F_1 on inner face	0.52–0.54	Translocates protons across membrane, is associated with ATP synthesis, or with ATP hydrolysis

Note: Most data collected by Capaldi (1982). Some of the nuclear-coded peptides of complexes I, III, and IV have isoforms which may differ from different tissues.

components in the free water, and to calculate rates from reported values of K_m and V_m. It has been suggested that the product of one enzyme may not dissociate and then diffuse through the aqueous phase to the next enzyme in the metabolic sequence, but that the enzyme-product complex may be the substrate for the next enzyme. This is known as dynamic substrate channeling (Westerhoff and Welch, 1992). Some experimental results may provide evidence that there is substrate channeling between glycolytic enzymes in solution (Ovadi, 1988) and in the stage of hepatic urea synthesis that occurs in the mitochondrial matrix (citrulline synthesis) (Watford, 1991). These ideas may appear at variance with the idea that high rather than low K_ms facilitate rapid catalysis (Fersht, 1985), so that a product may dissociate rapidly from the active site of an enzyme before the complex would have time to diffuse to and dock with a second molecule.

It has also been suggested that the enzymes of a pathway such as the citrate cycle may be organized in a single supramolecular complex or metabolon by analogy with the PDH or fatty acid synthase complexes with direct transfer of metabolites between enzymes (Srere, 1987). This concept of static channeling (Westerhoff and Welch, 1992) has many difficulties for glycolytic, citrate cycle, or β-oxidation enzymes because of the large number of different enzymes involved and their occurrence in non-stoichiometric amounts (Maretski et al., 1989). In the long-chain trifunctional β-oxidation enzyme complex, the second enzyme in the reaction sequence, 3-hydroxyacyl-CoA dehydrogenase, can be assayed directly so that its substrates, 3-hydroxyacyl-CoA and NAD^+, and the products, 3-oxoacyl-CoA and NADH, have access to and leave the active site, respectively, and therefore any channeling cannot be absolute. There is disagreement about the possible kinetic advantages of dynamic channeling in facilitating metabolic pathways. Cornish-Bowden (1991) maintains that it may not explain low concentrations of intermediates while Westerhoff and Welch (1992) and Mendes et al. (1992) maintain that it does. Clearly, there is still much to be discovered about enzyme-catalyzed reactions in their cellular environment.

THE MITOCHONDRIAL ELECTRON-TRANSPORT CHAIN

NADH and reduced substrate dehydrogenase-flavoproteins (FPH_2) must be continually reoxidized for mitochondrial oxidations to proceed. This is achieved by the electron transport chain (respiratory chain) which is a series of redox carriers of graded redox potential in the inner mitochondrial membrane (Appendix 1) that catalyzes the net reactions:

$$2(NADH + H^+) + O_2 \rightarrow 2NAD^+ + 2H_2O$$

$$2FPH_2 + O_2 \rightarrow 2FP + 2H_2O$$

Electrons from NADH, together with two protons, are transferred to ubiquinone to form ubiquinol by complex I (NADH ubiquinone oxidoreductase). Complex I

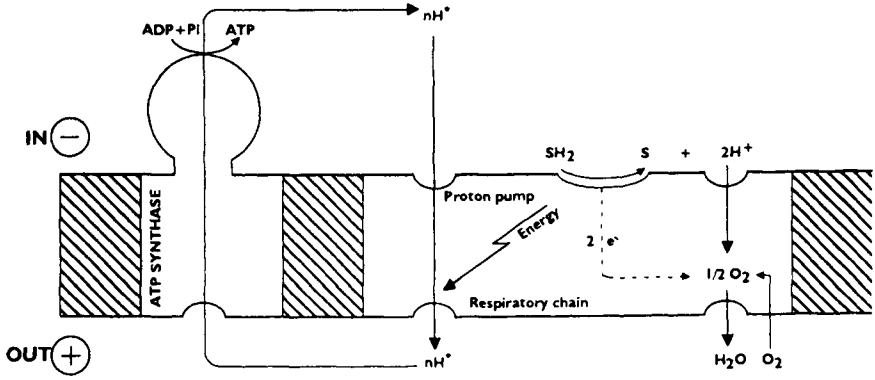


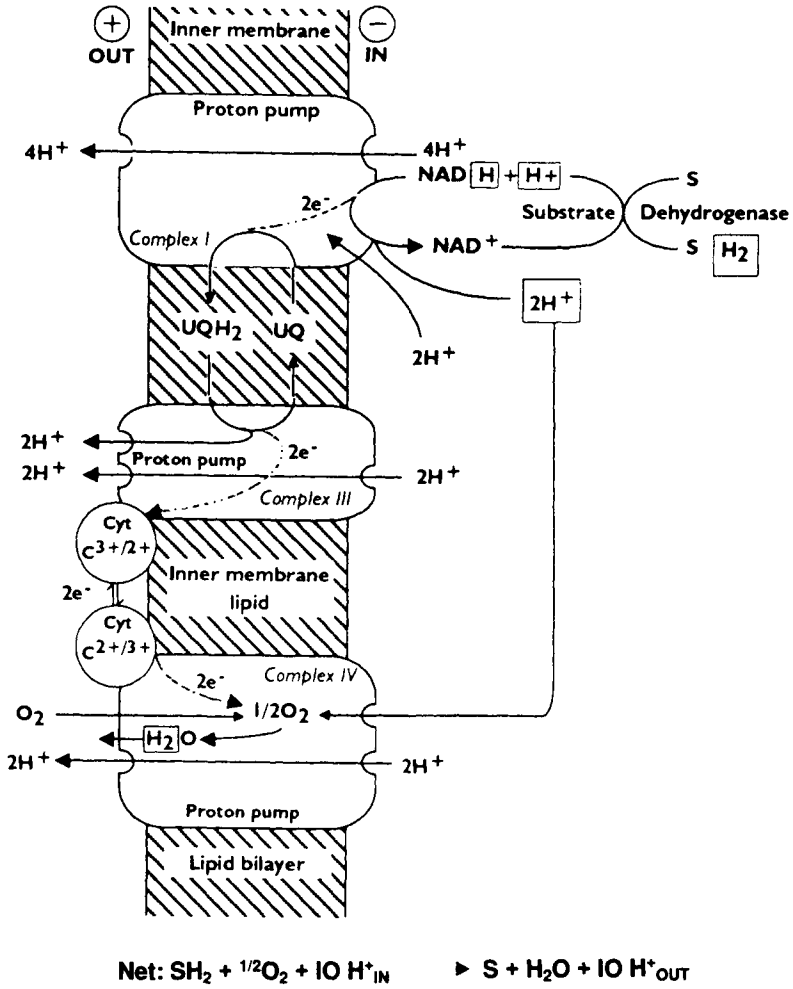
Figure 5. Coordination of ATP synthesis with electron transfer.

spans the inner-membrane and the binding site for NADH is on the inner face. Complex II has a binding site for succinate on the inner face. Both donate electrons to ubiquinone and release protons from their substrates. Electrons from the reduced flavoprotein acyl-CoA dehydrogenases are transferred to ubiquinone via the electron transfer flavoprotein (ETF and ETF dehydrogenase). The ubiquinol formed is oxidized by complex III. There are relatively small differences in the amounts of complex I, II, III, and IV in mitochondria from different tissues compared with ubiquinone and its reduced form, ubiquinol, which are in large molar excess (Table 2). Ubiquinone and ubiquinol diffuse between complexes I, II, and III in the hydrophobic interior of the inner membrane. Ubiquinol then transfers electrons via complex III and cytochrome c as an intermediate carrier to cytochrome c oxidase (complex IV) by single electron transfers. Complex III is a transmembrane protein and donates electrons to cytochrome c loosely associated with the outer face of the inner membrane. Complex IV, which also spans the inner membrane, accepts electrons from cytochrome c (which is a single electron carrier). Complex IV accumulates four electrons, which are then donated to dioxygen with the consumption of four protons to form water:

$$4e^- + O_2 + 4H^+ \rightarrow 2H_2O.$$

All the complexes consist of several subunits (Table 2); complex I has a flavin mononucleotide (FMN) prosthetic group and complex II a flavin adenine dinucleotide (FAD) prosthetic group. Complexes I, II, and III contain iron-sulphur (FeS) centers. These centers contain either two, three, or four Fe atoms linked to the sulphydryl groups of peptide cysteine residues and they also contain acid-labile sulphur atoms. Each center can accept or donate reversibly a single electron.

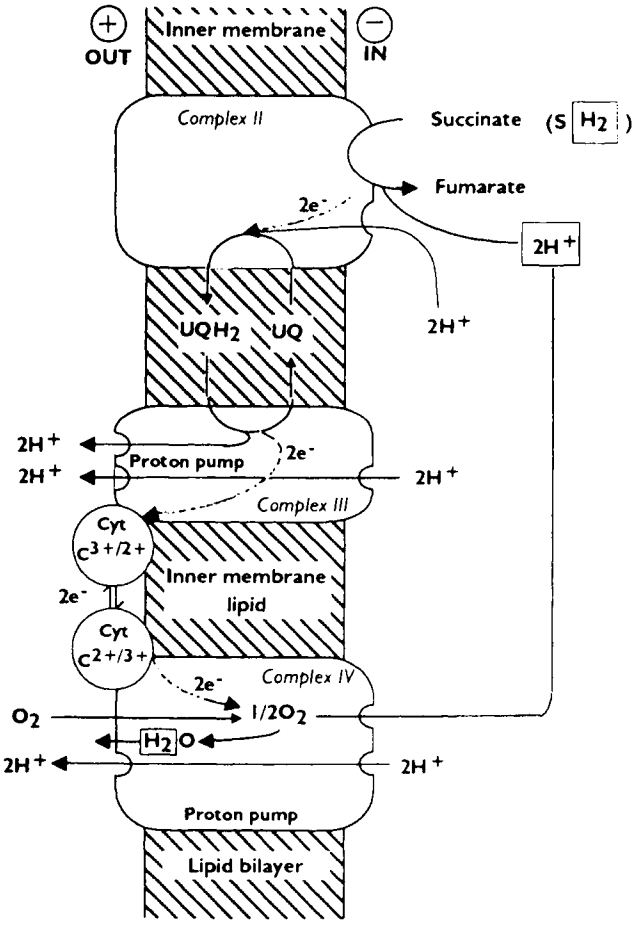
(a)



(continued)

Figure 6. Pathways of protons and electrons during mitochondrial oxidations. The diagrams show the pathways of electrons which enter the electron chain at the level of complex I (**a**) or complex II (**b**). Complexes I, III, and IV use the free energy of electron transport to pump protons out of the matrix. This diagram also distinguishes formally between protons released by dehydrogenation and those which are pumped out of the matrix, although they all enter or leave the same pool.

(b)



Net: Succinate + $^{1/2}$ O_2 + $6H^+_{IN}$ ⟶ Fumarate + H_2O + $6H^+_{OUT}$

Figure 6. (Continued) Note that these sections show the dehydrogenation of one molecule of substrate: two molecules must be dehydrogenated to give four electrons necessary to reduce dioxygen. Other versions of this scheme are possible.

Complexes III and IV have Fe-porphyrin prosthetic groups (hemes), complex IV also contains copper atoms which are involved in electron transport. Complexes I, III, and IV use the energy of electron transport to pump protons out of the matrix so as to maintain a pH gradient and an electrical potential difference across the inner membrane required for ATP synthesis (see below and Appendix 3). It is important to remember that all dehydrogenations of metabolic substrates remove two protons as well as two electrons and that a corresponding number of protons are consumed in the final reduction of dioxygen (Figures 5, 6).

ENERGY CONSERVATION BY THE ELECTRON-TRANSPORT CHAIN

The Chemiosmotic Hypothesis

The tendency of electrons to pass from one redox pair to another depends on the difference between their standard redox potentials and this is a measure of the free energy change (ΔG) of such a reaction (Appendix 2). The mechanism of ATP synthesis by oxidative phosphorylation has long been debated but the chemiosmotic hypothesis of Mitchell (1961) is now generally accepted and only this will be described here. ATP synthesis can be easily demonstrated in mitochondrial fractions prepared by the differential centrifugation of homogenates of many tissues, when incubated appropriately with substrates, oxygen, ADP, and orthophosphate (P_i), provided that care is taken to ensure that the mitochondria are structurally intact (Sherratt et al., 1988). Mitchell (1961) proposed that active pumping of protons across the inner membrane occurs when electrons flow along the electron-transport chain. This generates an electric potential and a pH gradient across the inner membrane, the free energy of these together comprising the proton motive force (PMF) (ΔP) (Appendix 3). The transported protons are derived from the electrolytic dissociation of H_2O:

$$H_2O \rightleftharpoons H^+ + OH^-$$

It was suggested that the PMF drives protons through mitochondrial ATP synthase (complex V, the proton-translocating ATPase) which spans the inner membrane and that the difference in the free energy of protons across the membrane is utilized to synthesize ATP from ADP and P_i (Figure 3). It was later shown experimentally that respiring mitochondria actively pump out protons from the mitochondrial matrix (Figure 5) (Mitchell and Moyle, 1967) (although originally it was suggested arbitrarily that hydroxyl ions were pumped out (Mitchell, 1961)).

When isolated mitochondria are incubated with substrate, P_i, and oxygen, in the absence of ADP (state 4, Chance and Williams, 1956), the proton gradient is not dissipated and electron transport is slow since the redox potential is balanced by the PMF. Addition of ADP (State 3, Chance and Williams, 1956) stimulates

electron transport when the PMF is utilized for ATP synthesis. The ADP concentration in cells is one of the factors controlling the rate of mitochondrial oxidations (see below) and, *in vivo*, the mitochondria are usually between state 3 and state 4. The equilibrium of the reaction catalyzed by ATP synthase in isolation is close to complete hydrolysis of ATP. In the absence of a PMF (due to anoxia or to the presence of respiratory chain inhibitors) in intact mitochondria ATP synthase hydrolyzes ATP and pumps protons out of the matrix unless hydrolysis is blocked by specific inhibitory proteins (see below). Mechanical or chemical disruption also destroys the ability of mitochondria to make ATP by making them permeable to protons and collapsing the membrane potential; such mitochondria are said to be uncoupled.

The mechanism of ATP synthesis discussed here assumes that protons extruded during electron transport are in the bulk phase surrounding the inner mitochondrial membrane (intermembrane and extramitochondrial spaces). An alternative view is that there are local proton circuits within or close to the respiratory chain and complex V, and that these protons may not be in free equilibrium with the bulk phase (Williams, 1978), although this has not been supported experimentally (for references see Nicholls and Ferguson, 1992). The chemiosmotic mechanism is both elegant and simple and explains all the known facts about ATP synthesis and its dependence on the structural integrity of the mitochondria, although the details may appear complex. This mechanism will now be discussed in more detail.

Proton Concentrations in the Matrix and ATP Synthesis

It may be asked if the chemiosmotic mechanism is consistent with the absolute concentrations of protons (as H_3O^+) at pH values between 7–8, of about 0.01–0.1 μM. A single mitochondrion has a volume (all compartments) of about 10^{-16} μl and might be expected to contain about 1–10 protons. Each mitochondrion contains several thousand respiratory complexes and it may appear unlikely that a scheme such as that illustrated in Figure 6 could be true. However, there is a large reserve of bound protons in protonated species in rapid equilibrium with free protons. Protonation reactions of carboxylate groups and basic nitrogen atoms are among the fastest chemical reactions known (diffusion controlled) with second order rate constants of about $5 \times 10^{10}\ s^{-1}M^{-1}$. Protonation of OH^- ions to give H_2O has a rate constant of $1.4 \times 10^{11}\ s^{-1}M^{-1}$ at 25 °C (Fersht, 1985). The deprotonation of H_2O to give H^+ (as H_3O^+) and OH^- is slower with a first-order rate constant of $1.4 \times 10^4\ s^{-1}$, but the absolute rate is high since the concentration of water is about 33 M in the matrix (pure water is 55 M). The concentration of protons described by the pH values is therefore valid statistically since the turnover of protons is much more rapid than those of the other reactions involved.

Complexes of the Mitochondrial Electron-Transport Chain

Complex I (NADH Ubiquinone Oxidoreductase)

Complex I is a large complex of about 40 peptides, 7 coded by mtDNA, which spans the inner membrane. Some subunits, including one with covalently bound flavin mononucleotide (FMN) which reacts with NADH, protrude into the matrix space. Complex I also contains about seven iron-sulphur centers each of which can accept or donate a single electron. Ubiquinone, in solution in the lipid phase of the inner membrane, is the acceptor for electrons from complex I. Fe-sulphur centers form part of the electron-conducting pathway between FMN and the ubiquinone binding site (assumed to be buried in the inner membrane). In intact mitochondria the oxidation of one molecule of NADH is coupled to the pumping of protons from the matrix to the intermembrane space by an unknown mechanism. The present consensus view is that four protons are extruded by complex I, although this is not rigorously established and formidable technical difficulties have prevented final resolution of this apparently simple problem.

Formally, when electrons are transported from a metabolic intermediate to dioxygen, protons are released (into the matrix) which eventually are consumed when dioxygen is reduced to water by the reaction catalyzed by complex IV (cytochrome c oxidase) (see below). NADH reacts with FMN to give $FMNH_2$, presumably deriving the extra proton from the matrix. Oxidation of $FMNH_2$ then releases two protons into the matrix as two electrons are transported through the complex to the Q-binding (ubiquinone-binding) site, where they are accepted by ubiquinone, probably with ubisemiquinone (UQH) as intermediate. This suggested mechanism also implies that two protons from the mitochondrial matrix are also taken up to form ubiquinol at the Q-binding site (Figure 6). This scheme shown in Figure 6 also accounts for all the movements of protons and the stoichiometry of their uptake and release by various reactions during electron transport from the intermediates.

Complex II (Succinate Dehydrogenase; Succinate Ubiquinone Oxidoreductase)

Complex II contains four peptides, the two largest form succinate dehydrogenase, the largest has covalently bound flavin adenine dinucleotide (FAD) which reacts with succinate, and the other has three iron-sulphur centers. Smaller subunits anchor the two larger subunits to the membrane and form the UQ binding site. Ubiquinone is the electron acceptor but complex II does not pump protons (see below).

Complex III (Cytochrome bc_1*; Ubiquionol-Cytochrome c Oxidoreductase)*

Complex III contains about 11 peptides and spans the inner membrane. One binds two B-type hemes with different redox potentials b_{566} and b_{562} (cytochromes b_{566} and b_{562}; b_L and b_H; b_{OUT} and b_{IN}) (see below), and this peptide is coded by mtDNA. The other peptides are coded by nuclear DNA. It is generally accepted that the flow of two electrons through the complex causes pumping of four protons out of the matrix, which is achieved by an elaborate mechanism termed the Q-cycle (Mitchell, 1976). A mechanism is shown in Figure 7, although other formulations are possible. The interconversion of ubiquinone and ubiquinol involves the intermediate ubisemiquinone which is stabilized by binding to specific sites within the complex. Cytochrome c is a small soluble protein (about 12 kDa) in the intermembrane space and is readily lost from mitochondria when the outer membrane is disrupted. Both complex III and complex IV have cytochrome c binding sites.

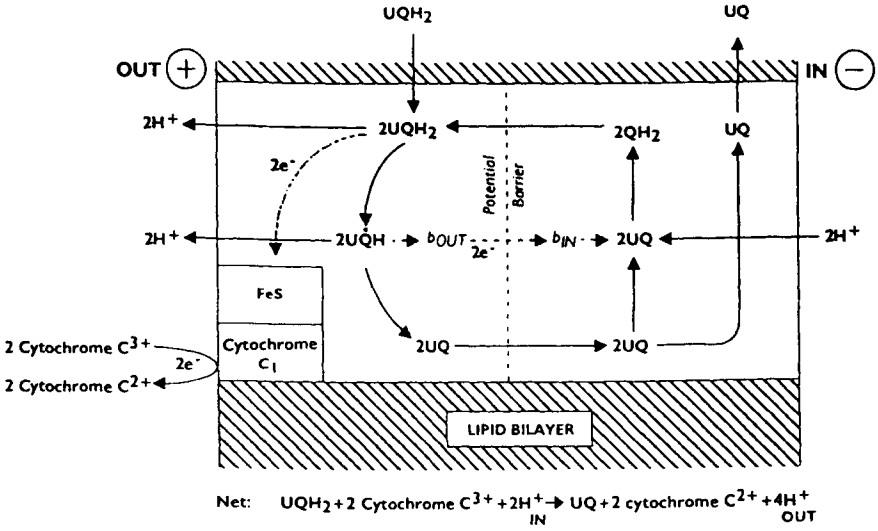


Figure 7. Mechanism of the proton-translocating ubiquinol: cytochrome c reductase (complex III) Q cycle. There is a potential difference of up to 150 mV across the hydrophobic core of this complex (potential barrier represented by the vertical broken line). Cytochromes b_{OUT} and b_{IN} are heme groups on the same peptide subunits of complex III which can transfer electrons across the hydrophobic core. The movement of two electrons provides the driving force to transfer two protons from the matrix to the cytosol. Diffusion of UQ and UQH_2, which are uncharged, in the hydrophobic core, and lipid bilayer of the inner membrane is not influenced by the membrane potential (see Nicholls and Ferguson, 1992).

Complex IV (Cytochrome c Oxidase)

Complex IV consists of 13 peptides, two heme A groups (cytochrome a and a_3) and two or three Cu atoms (Table 2). It spans the inner membrane and protrudes into the intermembrane space. Complex IV catalyzes the reduction of dioxygen by oxidized cytochrome c, and four protons derived from the matrix are consumed in the reaction.

$$4 \text{ cytochrome } c^{2+} + O_2 + 4H^+ \rightarrow 4 \text{ cytochrome } c^{3+} + 2H_2O$$

The three largest subunits are coded by mtDNA. There is a high-affinity binding site for cytochrome c on the intermembrane face of complex IV. Four single-electron redox groups (heme a, heme a_3, Cu_A, and Cu_B) are bound to two of the larger subunits, and these appear to catalyze the redox reactions. It is possible that the hemes a and a_3 and Cu_B are associated with the largest subunits and Cu_B with another large subunit. The functions of the other mtDNA coded subunit and the 10 nuclear DNA coded subunits in the mammalian enzyme are unknown; some may have a role in proton pumping. Some bacterial cytochrome c oxidases have only three subunits corresponding to the mitochondrial encoded subunits of the human complex IV.

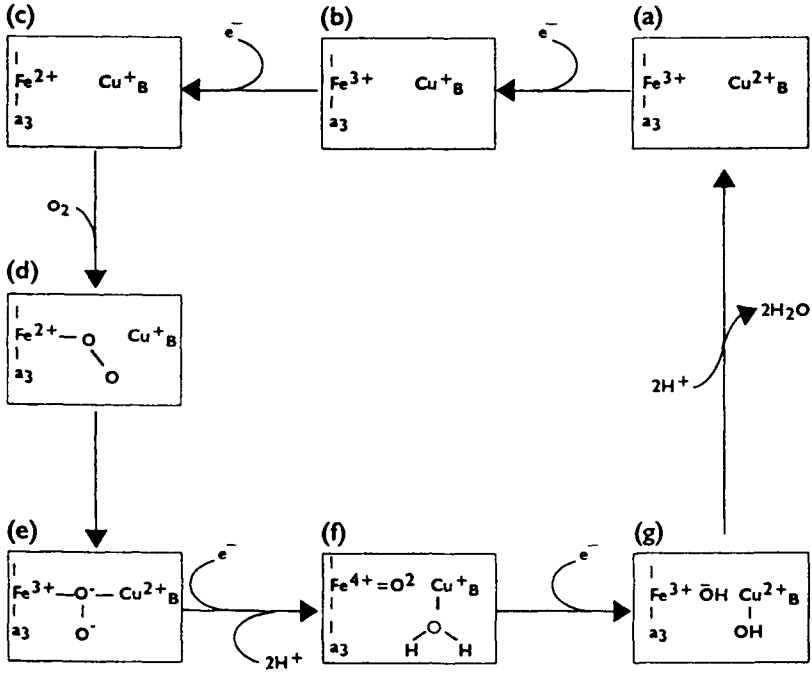


Figure 8. Mechanism of cytochrome c oxidase. Explanation given in text.

Four single electrons donated successively by cytochrome c^{2+} and four protons from the matrix are required for the reduction of a molecule of dioxygen. A tentative scheme for this has been given by Babcock and Wikström (1992) (Figure 8). It is assumed that Cu_A or heme a is the initial electron acceptor and that heme a_3 and Cu_B are associated in a binuclear center (a). The Cu_B^{2+} of this center accepts an electron from an initial center (Cu_A^{2+} (cupric) or heme a·Fe^{2+} (ferrous)) reducing it to Cu_B^{2+} (cuprous) (b), and the heme·a_3·Fe^{3+} (ferric) accepts a second electron reducing it to heme a_3·Fe^{2+} (ferrous) (c). This enables binding of dioxygen as a bridge between Cu_B and a_3 (d). Two protons from the matrix and uptake of a third electron results in the fission of the O-O bond with the formation of heme a_3·Fe^{4+} (ferryl) (e). A fourth electron is then taken up (f). Finally, the reaction cycle is completed by the uptake of two protons with the formation of two molecules of water and regeneration of (a). The cytochrome c oxidase reaction is the only reaction (with the possible exception of succinate dehydrogenase) involved in oxidative phosphorylation that is irreversible.

The mechanism of proton pumping coupled to the oxidation of cytochrome c^{2+} is unknown, but it probably involves reversible conformational changes in some of the peptide subunits driven by changes in the redox state of complex IV. Cytochrome c oxidase has a high affinity for dioxygen which enables its efficient use (apparent K_m less than 1 μM). It is generally believed that eight protons are pumped out of the matrix for the reduction of one molecule of dioxygen; however, it is conventional to discuss stoichiometry in terms of one atom of oxygen (four protons pumped) since reduction of dioxygen requires oxidation of two molecules of substrate (SH_2).

The Stoichiometry of Proton Pumping by the Mitochondrial Electron-Transport Chain

The diagrammatic representations of the mammalian respiratory chain (Figure 6) show the generally accepted view that the oxidation of one molecule of NADH is associated with the pumping of 10 protons from the matrix (Hinkle et al., 1991; Brown et al., 1993), although complete certainty cannot be claimed (Nicholls and Ferguson, 1992). The electron-transport chain is organized in different ways in plant mitochondria and bacteria as outlined by Nicholls and Ferguson (1992).

Complex V (ATP Synthase, Mitochondrial Proton-Translocating ATPase)

Complex V catalyzes the synthesis of ATP from ADP and P_i utilizing the energy of the proton motive force across the inner membrane (Senior, 1988, 1990).

$$nH^+_{OUT} + ADP_{IN} + P_{i_{IN}} \rightarrow nH^+_{IN} + ATP_{IN} + H_2O$$

The present consensus is that translocation of three protons (n = 3) drives the synthesis of one molecule of ATP, although it has sometimes been supposed that only two are necessary. Complex V consists of two parts, F_o (160 kDa) which spans the inner membrane and F_1 (370 kDa) which protrudes into the matrix and is connected to F_o by a short stalk. Complex V contains up to eight types of peptide, possibly with the stoichiometry abc_{10-12} $\alpha\beta\delta\varepsilon$. The peptides comprising F_o are hydrophobic and appear to form a specific protein channel or well through the inner membrane. F_1 has three copies of the larger α-subunits and β-subunits, and there is a catalytic site on each β-subunit. The β, δ, and ε subunits may form the stalk connecting F_o and F_1. Some ADP or ATP is always bound to F_1. Protons are thought to interact with F_o at its interface with F_1 and then reach the matrix at a lower electrical potential by an undefined route. ADP and P_i approach complex V and ATP leaves from the matrix side. There is no covalent phosphorylated intermediate, and complex V only conducts protons from the membrane space into the matrix side during ATP synthesis. Protons are conducted out of the matrix during hydrolysis of ATP when the PMF is not large enough for ATP synthesis.

$$nH^+_{IN} + ATP_{IN} + H_2O \rightarrow nH^+_{OUT} + ADP_{IN} + Pi_{IN}$$

The details of ATP synthesis are not well understood but a plausible mechanism has been reviewed by Cross (1981) (Figure 9). This is that each of the three catalytic sites (on the β-subunits) exists successively during catalysis in a conformational state with low affinity for ATP, a state with intermediate affinity, and a state with a very high affinity (10^{-12} M) (Cross, 1981; Senior, 1988) (Figure 9). ADP and P_i bind tightly at one catalytic site on a β-subunit where they reversibly form ATP which is also tightly bound (sub-micromolar range) (the equilibrium constant for this reaction between bound reactants being near unity, as distinct from that for reactants in free solution, which is about 10^{-5} M). Energy to release bound ATP comes from the proton gradient, presumably as a consequence of the translocation of three protons into the matrix through F_o. Binding of ADP and P_i to a second catalytic site is also necessary for release of ATP. Energy from the proton gradient causes concerted changes of binding affinity at each site to convert state A to state B (Figure 9): site (1) becomes the lowest affinity site and ATP released is replaced by ADP and P_i from the matrix, but which bind weakly (millimolar range) because of the conformational changes. Site (2) now becomes the high affinity site and reversible ATP synthesis occurs. Site (3) becomes an intermediate affinity site (2). Translocation of another three protons will release a second ATP and convert stage B to state C. During steady state ATP synthesis each β-subunit cycles through repeated conformation changes.

With this model, the energy-requiring step is not the formation of ATP but the conformational change that allows release of tightly bound ATP. The role of the α-subunits may be to maintain the functional conformation of the β-subunits. Another subunit is sometimes associated with F_1; this may regulate ATP synthase

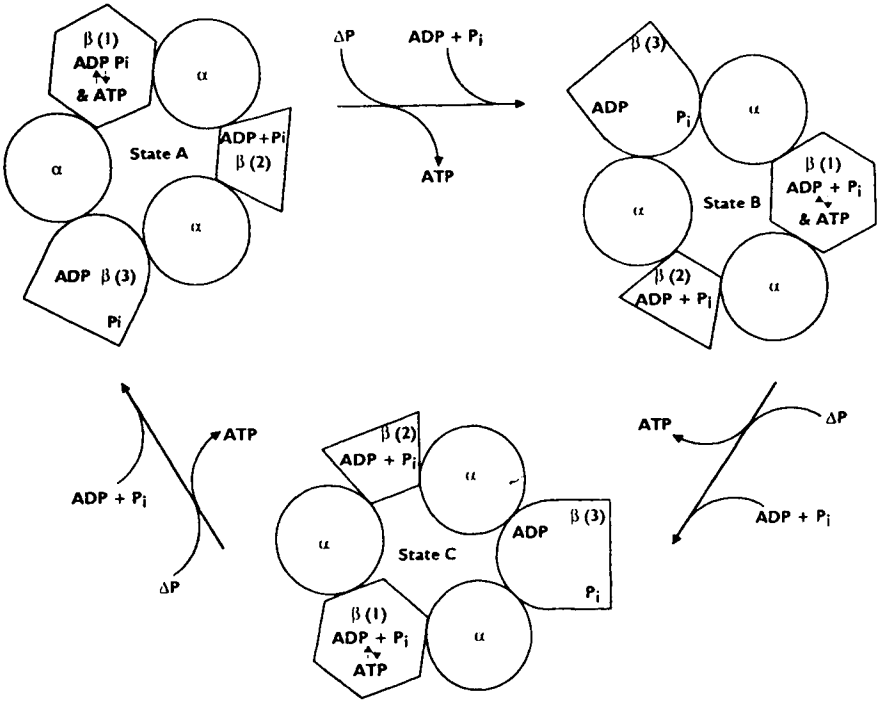


Figure 9. Proposed cyclic mechanism for ATP synthesis by complex V involving all three catalytic sites of F_1. In this scheme only the α and β subunits of F_1 are shown; these are connected by a short stalk to F_o in the inner membrane. Proton translocation through F_o driven by the proton motive force (ΔP) causes sequential conformational changes in each of the β-subunits and ATP synthesis as described in the text: hexagons, high-affinity sites; semicircles, low affinity sites; parallelepipeds, intermediate-affinity sites (with no movement of F_1).

activity and limit ATP hydrolysis in the absence of a membrane potential (see Harris and Das, 1991; Pederson and Amzel, 1993).

Mitochondrial Transport of Adenine Nucleotides

ATP and ADP, but not AMP, cross the inner membrane by means of a specific carrier, the adenine nucleotide translocator. This transmembrane protein, which is probably the most abundant protein in the inner membrane, consists of two identical 30 kDa subunits. It mediates a compulsory one-to-one exchange between either one molecule of ADP_{OUT} and one molecule of ADP_{IN}, ATP_{OUT} and ADP_{IN}, ATP_{IN} and ATP_{OUT}, or ADP_{IN} and ADP_{OUT}. Exchanges between ADP and ATP are electrogenic because at physiological pH values ADP has three negative charges and ATP has four (Klingenberg, 1980). Most ATP made by oxidative phosphorylation is exported to the cytosol in exchange for ADP and because this exchange is electro-

genic it is favored. ATP synthesis also requires the electroneutral import of P_i^- with a proton by the phosphate carrier. The net stoichiometry for the synthesis of one molecule of cytosolic ATP from one molecule of cytosolic ADP is four (strictly, n = 3+1).

$$ADP^{3-}_{OUT} + P^-_{i_{OUT}} + 4H^+_{OUT} \rightarrow ATP^{4-}_{OUT} + H_2O + 4H^+_{IN}$$

It may be noted that when ATP is used within the matrix the ADP and P_i formed can reform ATP without the need to cross the inner membrane and the stoichiometry is:

$$ADP^{3-}_{IN} + P^-_{i_{IN}} + 3H^+_{OUT} \rightarrow ATP^{4-}_{IN} + 3H^+_{IN}.$$

Oxidation of Cytosolic NADH

During intense muscle activity, particularly in type I fibers, more pyruvate is formed by glycolysis than can be oxidized. NADH generated at the glyceraldehyde 3-phosphate dehydrogenase step is used to reduce pyruvate to lactate reversibly. Oxidation of lactate during the recovery phase (not all is removed by circulation) requires oxidation of cytosolic NADH. The inner mitochondrial membrane is impermeable to NADH and oxidation by the respiratory chain is achieved indirectly by two or three pathways. It may appear paradoxical that the NAD^+/NADH couple in the matrix is more reduced (by 30 mV) than in the cytosol; however, this difference is maintained by the membrane potential (Davis et al., 1980). For this reason energy is expended in transferring reducing equivalents from cytosolic NADH to matrix NAD^+. This is achieved (in part) by the malate-aspartate shuttle, which involves the malate/2-oxoglutarate and glutamate plus H^+/aspartate ex-

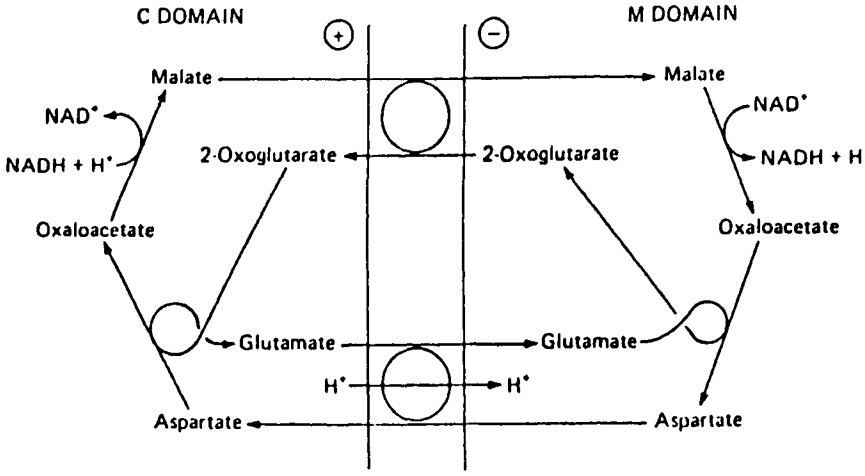


Figure 10. The glutamate/aspartate shuttle.

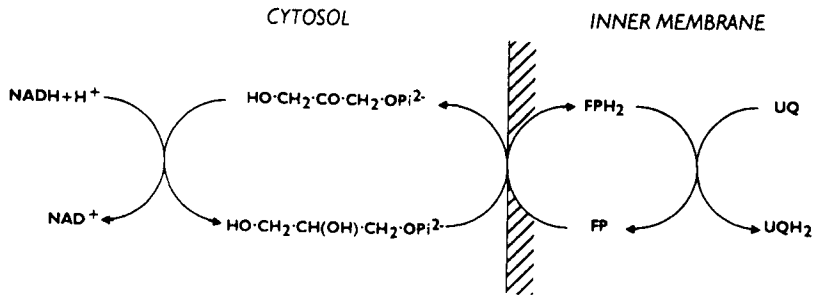


Figure 11. The L-glycerol 3-phosphate shuttle.

change carriers (Table 2), and the glutamate-oxaloacetate aminotransferase and malate dehydrogenases in both the cytosolic and matrix compartments (Figure 10). A key feature of this shuttle is that ejection of aspartate is coupled to the entry of a proton. This reaction is therefore a ratchet so that the shuttle only works unidirectionally to import reducing equivalents (LaNoue and Schoolwerth, 1979).

Another pathway is the L-glycerol 3-phosphate shuttle (Figure 11). Cytosolic dihydroxyacetone phosphate is reduced by NADH to s,n-glycerol 3-phosphate, catalyzed by s,n-glycerol 3-phosphate dehydrogenase, and this is then oxidized by s,n-glycerol 3-phosphate: ubiquinone oxidoreductase to dihydroxyacetone phosphate, which is a flavoprotein on the outer surface of the inner membrane. By this route electrons enter the respiratory chain from cytosolic NADH at the level of complex III. Less well defined is the possibility that cytosolic NADH is oxidized by cytochrome b_5 reductase in the outer mitochondrial membrane and that electrons are transferred via cytochrome b_5 in the endoplasmic reticulum to the respiratory chain at the level of cytochrome c (Fischer et al., 1985).

Stoichiometry and Efficiency of ATP Synthesis

It is conventional to discuss the stoichiometry of proton extrusion driven by electron transport as $H^+/2e^-$ ratios, although there are two-, one-, and four-electron reductions at different stages of the respiratory chain. Most textbooks assert that the flow of two electrons from NADH to oxygen is associated with the synthesis of three molecules of ATP, and from succinate to oxygen with the synthesis of two molecules, any lower values obtained experimentally being due to ion transport or partial uncoupling competing for the PMF, or to errors. It is usually said, for example, that the complete oxidation of one molecule of glucose yields 36 (or sometimes 38) mols of ATP.

$$C_6H_{12}O_6 + 6O_2 + 36ADP + 36P_i \rightarrow 6CO_2 + 42H_2O + 36ATP$$

with the conservation of 38% of the Gibbs (free) energy of oxidation of glucose as the Gibbs energy of ATP. This statement is very misleading (see Appendix 3).

We may calculate the ATP synthesized/$2e^-$ ratio assuming various stoichiometries for proton extrusion by the respiratory chain and for ATP synthesis by complex III. Such calculations are unreal limiting cases where the proton gradient is not dissipated in any other way (see below). If we accept the current consensus values of 10 protons extruded for the oxidation of NADH by dioxygen, six for oxidation of succinate, three used for ATP synthesis, and one to exchange cytosolic ADP^{3-} for ATP^{4-} (Figure 6), then for NADH oxidation the yield of ATP is 10/4 = 2.5, and for succinate oxidation the yield of ATP is 6/4 = 1.5. These values were supported experimentally by Hinkle et al. (1991) using polarographic measurements of oxygen uptake. If the ATP is used within the matrix a proton is not used for ADP/ATP exchange and the yield of ATP for NADH oxidation is 10/3 = 3.33, although this point is not often mentioned. This calculation can be repeated for any other assumed stoichiometries. A lower yield would be expected for ATP synthesis when electrons are derived from cytosolic NADH.

The membrane potential is also used to transport some ions into the matrix, particularly Ca^{2+}, and to import most mitochondrial proteins which are synthesized on extramitochondrial ribosomes (see below). These processes therefore compete with ATP synthesis. Further, some NADH generated in the matrix or cytosol is not oxidized by the electron-transport chain and is used for other biochemical reactions (see Sherratt, 1981). In state 4 (absence of ADP) isolated mitochondrial preparations (provided with substrates) consume oxygen at about 10–20% of the maximum ADP-stimulated rate. The chemiosmotic mechanism predicts that a high membrane potential will oppose electron flow to oxygen. It has been estimated from measurements with liver mitochondria of the distribution of positive and negative ions between the matrix and the suspending medium that the membrane potential in state 4 is about 220 mV (Nicholls, 1974). At this value there is some dielectric breakdown and proton leakage back into the matrix is significant, causing the slow rate of electron transport mentioned above and dissipation of some of the energy of the proton gradient as heat. During rapid electron transport and ATP synthesis (state 3) the membrane potential will fall to 180–200 mV which appears to limit dielectric breakdown so that the potential is used more effectively for biochemical work. The partitioning of the free energy of the PMF between ATP synthesis, ion transport, and proton leakage depends on specific physiological conditions. For all these reasons, in addition to uncertainties about the stoichiometries of proton pumping, the conventional text book calculations of the yield of ATP during the complete oxidation of a molecule of any substrate are not valid (Appendix 3).

The actual Gibbs energy ($\Delta G'$) required for ATP synthesis as distinct from the standard Gibbs energy ($\Delta G^{o\prime}$) depends on the prevailing ATP/ADP ratio and, with some conditions, there are only small energy losses. It is meaningless to compute an efficiency of energy conservation by dividing the standard values of Gibbs energy, $\Delta G^{o\prime}$, of ATP synthesis by that of NADH oxidation as is done in many text books (Atkinson, 1977; Cornish-Bowden, 1983) (see Appendix 3).

Side Reactions of Electron Transport

Reduced flavoproteins and complex III may undergo side reactions directly with dioxygen, to a limited extent, forming highly reactive and toxic products, hydrogen peroxide, and superoxide.

$$FPH_2 + O_2 \rightarrow FP + H_2O_2$$

$$FPH_2 + 2O_2 \rightarrow FP + 2H^+ + 2O_2^{\bullet -}$$

$$b_o^{2+} + O_2 \rightarrow b_o^{3+} + O_2^{\bullet -}$$

Normally, these reactive species are destroyed by protective enzymes, such as superoxide dismutase in mitochondria and cytosol and catalase in peroxisomes, but if a tissue has been anoxic the respiratory chain is very reduced and reoxygenation allows dangerous amounts to be formed. Muscle also contains significant quantities of the dipeptide, carnosine (β-alanylhistidine) (10–25 mM). The functions of carnosine are obscure although it has been suggested to be an effective antioxidant (Pavlov et al., 1993).

Inhibitors and Uncouplers of Mitochondrial Reactions

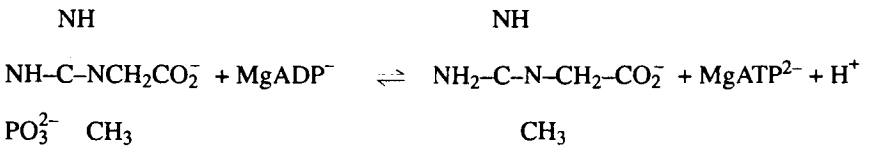
Many inhibitors of substrate oxidations, substrate transport, electron transport, and ATP synthesis are known including many well-known toxins (see Sherratt, 1981; Harold, 1986; Nicholls and Ferguson, 1992). These are not discussed here except to mention specific uncouplers of oxidative phosphorylation. Classic uncouplers such as 2,4-dinitrophenol have protonated and unprotonated forms, both of which are lipid soluble and cross the inner mitochondrial membrane discharging the proton gradient. This prevents ATP synthesis and stimulates respiration.

INTEGRATED MITOCHONDRIAL METABOLISM

The rate of mitochondrial oxidations and ATP synthesis is continually adjusted to the needs of the cell (see reviews by Brand and Murphy 1987; Brown, 1992). Physical activity and the nutritional and endocrine states determine which substrates are oxidized by skeletal muscle. Insulin increases the utilization of glucose by promoting its uptake by muscle and by decreasing the availability of free long-chain fatty acids, and of acetoacetate and 3-hydroxybutyrate formed by fatty acid oxidation in the liver, secondary to decreased lipolysis in adipose tissue. Product inhibition of pyruvate dehydrogenase by NADH and acetyl-CoA formed by fatty acid oxidation decreases glucose oxidation in muscle.

ADP/ATP Concentrations and the Membrane Potential

In resting muscle the high concentration of ADP does not decrease the proton gradient effectively and the high membrane potential slows electron transport. ADP, formed when ATP is hydrolyzed by myosin ATPase during contraction, may stimulate electron transport. However, the concentration of ATP (largely as its Mg salt) is buffered by its readily reversible formation from creatine phosphate catalyzed in the intermembrane space, and in other cell compartments, by the various isoenzymes of creatine kinase (reviewed by Walliman et al., 1992).



In resting skeletal muscle the total concentration of ATP is in the range 2–5 mM, which could only sustain maximum concentration during exercises for a few seconds, while creatine phosphate is about 20–35 mM and creatine 5–10 mM, depending on the muscle and fiber type. The maximal activity of creatine kinase is several times that of the maximal rates of ATP synthesis and so that the ATP/ADP ratio (normally about 100) does not fall too far. This, together with binding of ADP to F-actin, keeps the ADP concentrations in the range where increases can stimulate mitochondrial respiration (20–30 μM ADP may cause half-maximal rate of ATP synthesis (Chance et al., 1986)). The mitochondrial isoenzyme of creatine kinase located in the intermembrane space tends to form octomers binding to both membranes at the contact sites between them. It has been suggested that mitochondrial creatine kinase has privileged access to ATP exported by the adenine nucleotide carrier in the inner membrane (Walliman et al., 1992). Creatine phosphate is reformed from creatine and ATP during recovery from exercise. Creatine phosphate and creatine are smaller molecules than ATP and ADP and may therefore diffuse more rapidly to sites of ATP utilization (Walliman et al., 1992).

In rat liver mitochondria, in state 4, the ΔP was estimated to be about 220 mV, with the membrane potential representing about 90% of this (Nicholls, 1974; Appendix 3). Similar values have been reported for human and rat skeletal muscle mitochondria in state 4 (Stumpf et al., 1982). The control of the rate of electron transport is not only determined by the availability of ADP, but also of P_i oxidizable substrates, and oxygen. There is evidence for futile cycling of protons in intact normal rat hepatocytes (Brand et al., 1993). Recently, Porter and Brand (1993) found a correlation between the proton permeability of the inner membrane of liver mitochondria and body size in animals from the mouse (20 g) to horses (150 kg) with a decrease in permeability with increasing weight of several-fold at a constant

membrane potential. If this also holds for mitochondria from other tissues, there would also be a correlation with the mass specific metabolic rate so that more heat production in mitochondria at or near state 4 would help to maintain body temperature in small animals. An extreme example of this in humans is Luft's disease where hyperthermia is associated with muscle mitochondria whose respiratory rate appears maximal and is not stimulated by ADP (Luft et al., 1962). Presumably, in this disease the inner membrane is abnormally permeable to protons but ATP synthesis can compete for the PMF with little effect on the total rate of oxygen uptake.

In both intermediate and maximum rates of respiration, control is distributed between several different steps, including the activity of the adenine nucleotide translocator (Groen et al., 1983). It is now recognized that the idea of a simple rate-limiting step for a metabolic pathway is simplistic and that control is shared by all steps although to different extents (Kacser and Burns, 1978; Fell, 1992). Each step in a pathway has a flux control coefficient (FCC) defined as:

$$\frac{\Delta J/J}{\Delta E/E} = \frac{\partial J \cdot E}{\partial E \cdot J} \quad \text{or } \partial \ln J / \partial \ln E$$

where J is the flux and E the amount of enzyme (in appropriate units). The FCC describes the effect on the flux of changing the amount of an enzyme if, for example, an enzyme is in apparent excess any small change in amount would have a very small effect. Similarly, if an enzyme is functioning near its maximum rate a small change in amount would have a much larger effect on flux. Thus, if an enzyme in a particular metabolic state has a FCC of 0.5 a 1% increase in amount causes an 0.5% increase in flux. The sum of FCCs in a linear pathway can be shown to equal unity. Steps with a high control coefficient (for example >0.8) have often been described misleadingly as rate-limiting. Control may be shared between several stages of oxidative phosphorylation including the proton leak, substrate supply, activity of complexes of the respiratory chain, ATP synthase, and both intra- and extramitochondrial ATP utilization. Some of the steps may exert very small control. In isolated liver mitochondria in state 4, the FCC for proton leak is about 0.9 and substrate transport about 0.15 (addition of FCCs to more than unity is assumed due to experimental error). In state 3, where the rate of electron transport is about 5 times greater than in state 4, the FCCs of overall phosphorylation and of substrate transport are about 0.5, while that of the proton leak is only about 0.03%. At intermediate rates (about state 3.5) the FCC of phosphorylation is about 0.7, substrate transport about 0.1, and proton leak about 0.2 (Hafner et al., 1990). Such analyses of control are necessarily limited in that they apply to specifically designed experimental conditions with other factors which can influence rates absent or constant. However, a very important conclusion from such studies is that the FCC of any step varies widely with different metabolic conditions.

Mitochondrial Enzyme Activities

In experiments with isolated mitochondria the rate of oxygen uptake and ATP synthesis can be increased several-fold by adding ADP if substrate, ADP, P_i, and O_2 are not limiting. However, the ADP concentration is not always the most important factor *in vivo* and the rate of oxidations may be influenced by other factors including the activities of some enzymes (Brown, 1992). Many hormones influence Ca^{2+} concentrations in the cytosol which in turn influence Ca^{2+} concentrations in the matrix. Free Ca^{2+} in the matrix is thought to be in the range 0.1–1.0 μM. During muscle contraction the concentration of Ca^{2+} is increased and this stimulates the activities of pyruvate, isocitrate, and 2-oxoglutarate dehydrogenases, which is assumed to increase the flux through the citrate cycle and to increase the NADH/NAD^+ ratio (Moreno-Sanchez et al., 1991). Increased uptake of Ca^{2+} may also increase the matrix volume as suggested by Davidson and Halestrap (1989) and stimulate the oxidation of many substrates, particularly fatty acids.

With some conditions, for example in heart muscle, the rates of ATP synthesis can vary without significant changes in ADP, ATP, P_i, or ΔP, suggesting that the activity of ATP synthase may be regulated directly (see Harris and Das, 1991). This may be achieved by two regulatory proteins in the matrix, IF_1 (intramitochondrial protein, potential-dependent inhibitor of F_1) and CaBI (intramitochondrial protein, Ca^{2+}-dependent inhibitor) which can react directly with the F_1 component of ATP synthase. IF_1 (a 12.5 kDa protein) partly inhibits ATP synthesis and ATP hydrolysis when the ΔP is low (for example during anoxia) to conserve ATP that is still available. CaBI (a 12.5 kDa dimer) also partly inhibits ATP synthesis. Ca^{2+} imported into the matrix when muscle activity is increased causes inhibition of ATP synthase by combining with CaBI and dissociating it into monomers which are not inhibitory (Harris and Das, 1991).

BIOSYNTHESIS OF MITOCHONDRIAL PROTEINS

Nuclear-Coded Mitochondrial Proteins

In humans the mitochondrial genome only encodes 13 proteins, all subunits of the respiratory chain. All other proteins are coded by nuclear genes and this includes the majority of the subunits of the respiratory chain. There is a gradual progress in understanding of the mechanism of the targeting, import, and assembly of these nuclear-encoded proteins and these advances have been the subject of several excellent reviews (Pfanner et al., 1991; Glover and Lindsay, 1992; Pfanner et al., 1992). Much of the pioneering work has been done in lower organisms (e.g., yeasts) rather than mammalian tissues because of the ease of genetic manipulation. However, it is likely that similar mechanisms occur in higher animals. It is also important to recognize that this is a very active area of research and that there is not necessarily

agreement between researchers on the specific mechanisms involved. One area in which there is considerable debate is whether mitochondrial proteins are transferred into the mitochondria by a posttranslational or cotranslational mechanism (Verner, 1992; Fujiki and Verner, 1993).

Leader Sequences

Most newly synthesized mitochondrial proteins have N-terminal presequences which are 15–70 amino acids in length. There is no obvious homology for these presequences, although they are all rich in positively charged and hydroxylated amino acids. In addition, they all seem capable of forming an amphiphilic α-helix when in contact with a lipid bilayer. This presequence appears to be important in the targeting of the precursor protein to the mitochondria, perhaps also for recognition at specific receptors on the outer mitochondrial membrane and for the passage of protein through the membranes. Not all mitochondrial proteins have a cleavable presequence to target them to mitochondria, for example cytochrome c. For these proteins it is assumed that the targeting sequence is contained within the sequence of the mature protein. Similarly, those proteins destined for the outer mitochondrial membrane do not have a presequence.

Cytosolic Factors Involved in Translocation and Import of Precursors

The efficient uptake of precursor proteins depends on their presentation in a translocation competent state. This is maintained *in vivo* by the specific interaction with a highly conserved group of proteins, the heat-shock or stress related proteins (hps70s). These act as molecular chaperones and interact with the proteins to maintain them in a correctly folded state, a process which is ATP dependent.

Import Receptors and Membrane Translocation of Precursor Peptides

While it has been known for many years that the N-terminal presequence is sufficient to promote mitochondrial targeting and assembly, the subsequent interaction of the precursor molecule with the outer mitochondrial membrane and the uptake of the protein is still an area of active research. There seems little doubt, however, that there are proteins on the outer mitochondrial membrane which are required for the import process. The function of these proteins is uncertain, but they may act as receptors with the subsequent transfer through the membrane at proteinous pores located at contact sites between the inner and outer membranes. Several proteins have been identified which seem to play an important role as either receptor proteins or part of the import channel (Pfanner et al., 1991). Again, not all proteins seem to depend on this mechanism. Cytochrome c, which is loosely associated with the outer aspect of the inner mitochondrial membrane, can cross

the outer mitochondrial membrane even when it has been treated with proteases. The import of this protein depends on the heme lyase located in the intermembrane space. The probable mechanism is that the attachment of the heme forms a nucleus around which the protein is folded, thus pulling it across the outer mitochondrial membrane. A membrane potential is essential for driving the import process. However, protein translocation still occurs in diseased mitochondria and therefore the effective potential can be small (20–40 mV). The membrane potential is probably involved in the translocation of the positively charged leader sequence across the mitochondrial inner membrane.

Mitochondrial Factors Involved in Proteolytic Cleavage and Assembly of Proteins

The maturation of the precursor protein involves their proteolytic cleavage. There are two proteins important in this cleavage, so-called processing protease and protease enhancing peptide. These are now believed to be nonidentical subunits of the same enzyme. The structural requirements for recognition of the cleavage site are not fully understood and except for a positively charged residue at position (-2) there is no consensus sequence around this site.

A mitochondrial matrix hsp70-like protein has been identified which seems to be important in the import process. This appears to interact as soon as polypeptides undergoing translocation enter the mitochondrial matrix compartment. Release of the translocation intermediate from hsp70 is an ATP-mediated reaction which may be associated with the final transport of the imported polypeptide in the matrix. A mitochondrial matrix hsp60 may have an important role in the folding and assembly of the multimeric protein complexes. It is not yet known if the hsp70 transfers proteins directly to the hsp60; however, one possibility is that the hsp70 maintains the proteins in a loose-folded state and then transfers them to hsp60 for their final assembly into functional complexes (Manning-Krieg et al., 1991).

Translocation of Proteins to Their Correct Intramitochondrial Location

The nuclear-encoded proteins are inserted into both inner and outer mitochondrial membranes, the intermembrane space, and the matrix and there are several different mechanisms involved. As mentioned above there is no apparent requirement for a presequence on proteins which insert specifically into the mitochondrial outer membrane. For proteins destined for the inner mitochondrial membrane, a stop-transfer mechanism is proposed. Thus some information in the peptide must stop the complete transfer of the protein into the mitochondrial matrix, enabling the protein to remain in the inner mitochondrial membrane. For some proteins in the intermembrane space (for example the Rieske iron-sulphur protein associated with the outer face of complex III), a particularly complicated import pathway

exists (Hartl and Neupert, 1990). These proteins have N-terminal presequences and are transferred across both membranes into the mitochondrial matrix. Part of the presequence is cleaved and then the protein is reexported across the inner membrane where it is cleaved to its final form by a separate membrane-bound protease. It is possible that some inner membrane proteins are also processed in the mitochondrial matrix and then transferred into the inner membrane. Those proteins that are destined to remain in the mitochondrial matrix are transferred across the two membranes and usually contain a presequence.

Proteins Encoded by Mitochondrial DNA

Mitochondria are unique organelles in man and higher animals in that they contain their own genome. Mitochondrial DNA (mtDNA) in humans is a small (16.5 kb), circular genome that encodes only 13 proteins, 22 transfer RNA (tRNA), and 2 ribosomal RNA (rRNA) molecules. mtDNA is inherited only from the mother and is present in multiple copies within one mitochondrion.

Mitochondrial DNA is transcribed as a polycistronic RNA which is subsequently cleaved to generate the various mature mRNA, tRNA, and rRNA (Clayton, 1984). The 13 proteins encoded by mtDNA are all components of the respiratory chain and are seven subunits of complex I, one subunit of complex III, three subunits of complex IV, and two subunits of complex V.

SUMMARY

The synthesis of ATP by mitochondrial oxidative phosphorylation is described. The outer membrane encloses an intermembrane space and is relatively permeable to small molecules. The inner membrane encloses the matrix which contains the enzymes necessary for the oxidation of pyruvate by the citrate cycle and of the β-oxidation of fatty acids. It has good electrical insulating properties and contains specific transporters for polar compounds including substrates, ADP/ATP, and phosphate. Electrons from substrate oxidations are transferred to dioxygen by the electron-transport chain in the inner membrane. The electron-transport chain contains four multimeric complexes containing redox centers (flavins, cytochromes, iron-sulphur centers, and Cu) which span the membrane and the mobile electron carriers, cytochrome c, and ubiquinone. Complexes I, III, and IV use the free energy of electron transport to pump protons out of the matrix to generate a membrane potential and pH gradient across the inner membrane (the proton motive force or ΔP). ΔP drives protons (probably three) through complex V (ATP synthase) in the inner membrane, and this uses the energy of ΔP to convert ADP + P_i to ATP on its matrix face. The ADP/ATP exchange across the inner membrane also requires the uptake of a proton.

The control of the rate of ATP synthesis is shared by several processes including the demand for ATP, the rate of ADP/ATP exchange, the activities of the enzymes

of substrate oxidation, substrate supply, and the activity of complex V, when their relative contributions depend on the rate. When the rate of ATP synthesis is slow ΔP is higher and more protons leak back into the matrix by-passing complex V. The exact stoichiometry of ATP synthesis and extent of the proton leak varies with the metabolic state.

APPENDIX 1: OXIDATION AND REDUCTION

Two very important classes of chemical reactions are oxidation-reduction (redox) reactions and acid-base reactions, which are defined by molecules or ions accepting and donating electrons or protons, respectively.

Redox: Electron donor $\rightleftharpoons e^-$ + electron acceptor

Acid/base: Proton donor $\rightleftharpoons H^+$ + proton acceptor

Formally, in redox reactions there is transfer of electrons from a donor (the reductant) to the acceptor (the oxidant), forming a redox couple or pair. Oxidations in biological systems are often reactions in which hydrogen is removed from a compound or in which oxygen is added to a compound. An example is the oxidation of ethanol to acetaldehyde and then to acetic acid where the oxidant is NAD^+, catalyzed by alcohol dehydrogenase and acetaldehyde dehydrogenase, respectively.

$$CH_3CH_2OH + NAD^+ \rightleftharpoons CH_3CHO + NADH + H^+$$

$$CH_3CHO + NAD^+ + H_2O \rightleftharpoons CH_3COOH + NADH + H^+$$

All dehydrogenations involve the transfer of two electrons (where 2H is equivalent to $2H^+$ plus $2e^-$). NAD^+ or $NADP^+$ are the cofactors in most substrate dehydrogenations; these accept two electrons while a proton is released.

$$SH_2 + NAD(P)^+ \rightleftharpoons S + NAD(P)H + H^+$$

Dehydrogenation may involve an intermediate step with the formation of a partly reduced free radical, for example, ubisemiquinone (Figure 7).

$$UQ + e^- + H^+ \rightleftharpoons UQH^\bullet$$

$$UQH^\bullet + e^- + H^+ \rightleftharpoons UQH_2$$

Similarly, the reduction of riboflavin prosthetic groups (FP) of some flavoprotein dehydrogenases may also involve a free radical intermediate.

$$FP + e^- + H^+ \rightleftharpoons FPH^\bullet$$

$$FPH^\circ + e^- + H^+ \rightleftharpoons FPH_2$$

Such free radicals may be stabilized by binding to proteins. Redox reactions may also occur between ionic species, for example the oxidation of reduced cytochrome c by hexacyanoferrate (ferricyanide) ions.

$$\text{cytochrome } c^{2+} + Fe(CN)_6^{3-} \rightleftharpoons \text{cytochrome } c^{3+} + Fe(CN)_6^{4-}$$

Electrons are transferred from a state of higher potential energy (lower or more negative electrical potential) to one of lower potential (more positive electrical potential) and energy is released (exergonic reaction). Some reactions occur spontaneously and others require a catalyst. The tendency of a redox couple AH_2/A or M^{2+}/M^{3+} to lose or gain electrons is given by its standard oxidation potential (E_0') which is the electromotive force in millivolts (mV) in a half cell in which the reductant and oxidant are both present (with a catalyst if needed) at unit activity at 25 °C and pH 7.0, in equilibrium with an inert electrode capable of reversibly accepting electrons. This potential is measured relative to the standard hydrogen electrode.

$$2H^+ + 2e^- \rightleftharpoons H_2$$

with H_2 gas at 1 atmosphere pressure, pH = 7 ($a_{H^+} = 10^{-7}$ M) at 25 °C which has a value of –420 mV.

The observed potential of a redox couple, E_h', depends on the concentration of the oxidized and reduced forms and is given by:

$$E_h' = E_0' + \frac{2.303RT}{nF} \log \frac{A[\text{electron acceptor}]}{D[\text{electron donor}]}$$

where R is the gas constant (8.31 $Jdeg^{-1}M^{-1}$), T is the absolute temperature, F is the Faraday constant (96,406 JV^{-1}), n is the number of electrons transferred, and A and D are the thermodynamic activity coefficients of the electron acceptor and donor, respectively. At 37°C 2.303 RT/nF = 61 when n = 1 and 30.5 when n = 2. When the concentrations (strictly activities) of the acceptor and donor are equal, the equation reduces to $E_h' = E_o'$. When mixtures of two redox couples are made, electrons are transferred from the couple with the lowest E_h' to the other until their redox potentials (E_h') are equal. Spectrophotometric techniques are also used to determine redox potentials (E_h') of components of the electron-transport chain by measuring the extent of their reduction, since the reduced and oxidized forms often have different absorption spectra when incubated in media of known redox potential (see Nicholls and Ferguson, 1992). When two redox couples are mixed the equilibrium concentration of each oxidized and reduced form can be calculated from a knowledge of the standard potentials and the total concentrations of each couple (reduced plus oxidized). The oxygen water couple,

$$O_2 + 4e^- + 4H^+ \rightleftharpoons 2H_2O, \text{ has } E_o' = 816 \text{ mV}$$

Table 3. A Standard Oxidation Potential for Several Redox Couples

Couple	*Number of Electrons Transferred per Reaction*	E'_0(mV)
$NAD^+/NADH + H^+$	2	−320
Fumarate/succinate	2	+30
Ubiquinone/ubiquinol	2	+40
Cytochrome c^{3+}/cytochrome c^{2+}	4	+220
$O_2/2H_2O$ (in practice irreversible)	4	+816

and thus oxygen is a good oxidizing agent. In the electron-transport chain (see text) each carrier forms a redox couple and there is successive transfer of electrons from carriers of lower potential to carriers of higher potential. The nature and organization of the chain in the inner mitochondrial membrane only allows transfer of electrons from a carrier to that with the next highest potential. The midpoint potential of components of the respiratory chain have been determined indirectly and some values of E'_0 are given in Table 3. The differences in midpoint potentials ($\Delta E'_h$) (state 4) across segments of the respiratory chain are given in Figure 12.

Reference: Nicholls, D.G. & Ferguson S.J. (1992). Bioenergetics 2, Academic Press, London.

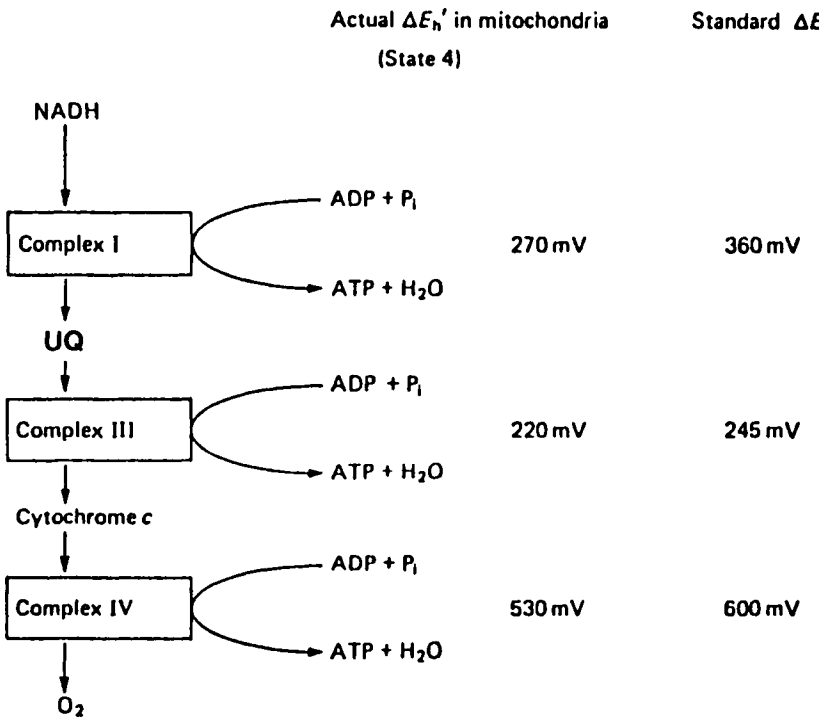


Figure 12. The $\Delta E'_h$ of the respiratory chain.

APPENDIX 2: FREE ENERGY ($\Delta G'$) OF ATP SYNTHESIS AND UTILIZATION

Energy expended by living cells for maintenance is expressed quantitatively in appropriate units, for example kJ Kg^{-1} s^{-1}, and in animals it is largely provided as ATP. In this chapter, we outline how this is achieved, although our thermodynamic treatment lacks formal rigor. Further information on classical thermodynamics is given in textbooks of physical chemistry.

Work can only be done by chemical reactions (exergonic or energy releasing reactions) when the reactants are displaced from equilibrium. Those processes which require an input of energy (endergonic reactions) to be pushed away from equilibrium must be coupled to exergonic reactions (Figure 13). For a simple reaction, $A \rightleftharpoons B$, the equilibrium constant (K) is defined by $K = [B]/[A]$, where [B] and [A] denote concentrations of the reactants at equilibrium. At non-equilibrium concentrations, the mass action ratio, Γ, is defined by the prevailing values of [B]/[A], and the reaction will proceed to equilibrium either spontaneously or in the presence of an appropriate catalyst.

Only some of the energy, released by an exergonic reaction, can do useful work. To treat this quantitatively we define a system. This is a quantity of matter that is distinct from its environment: it may be completely isolated, but of more interest are closed and open systems. Closed systems exchange energy with their surroundings and open systems exchange both energy and matter. At equilibrium the entropy of a system is maximum, and all systems tend to equilibrium unless external work is done on them. Entropy (S) is a measure of heat unable to do work and increases during any exergonic reaction. The Gibbs energy (free energy) (G) is a measure of the maximum possible energy able to do work. The Gibbs energy is defined as:

$$G = H - TS$$

where H is the heat content or enthalpy in biology and T the absolute temperature. However, we are more interested in changes in the Gibbs energy (ΔG) during biochemical reactions, which is a measure of the driving force. The changes in entropy (ΔS) are related to changes in enthalpy (ΔH) and to the temperature, so that $\Delta G = \Delta H - T\Delta S$. Most biological reactions occur isothermally. Heat released during metabolism is lost from the system so that the temperature remains approximately constant. The standard Gibbs energy change, $\Delta G°$, of a reaction is related to the equilibrium constant (K), and is defined under standard conditions as $\Delta G° = -RT \ln K$ or $-2.303\ RT \log K$, where R is the gas constant (8.3 $Jmol^{-1}$) with all the reactants being present initially at 1M activity (often approximated to 1M concentration). In biological systems one of the reactants may be a proton and the standard Gibbs energy ($\Delta G°'$) is defined for $\{H^+\} = 0.1\ \mu M$ (pH 7) (where the brackets represent the activity of the hydrogen ion).

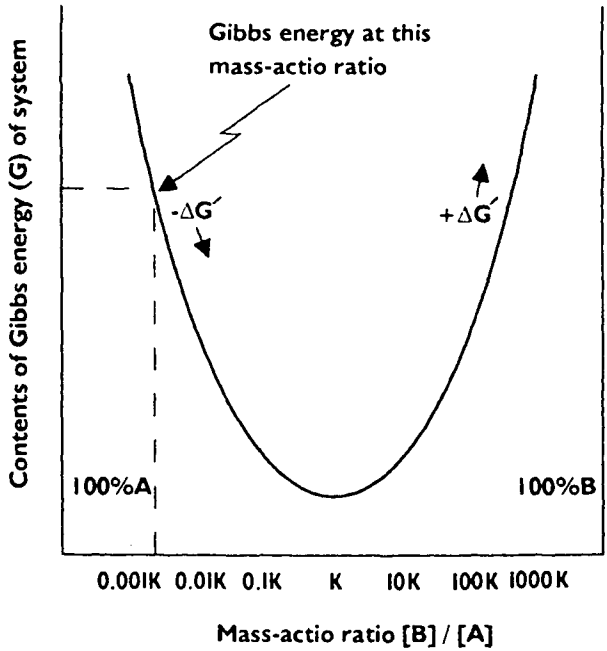


Figure 13. The Gibbs energy available from a reaction, A $\rightleftharpoons$ B, depends on its displacement from equilibrium when [B]/[A] = K. The ΔG value is plotted against the mass-action ratio, and this is the value when [B]/[A] is maintained constant in the steady state if the rate of substrate supply and substrate removal is constant.

The activity of water is taken as one (rather than 55M in a dilute solution) and is therefore not included in the equations. It can be shown that at any actual concentrations of the reactants, ΔG′ is given by

$$\Delta G' = \Delta G^{o\prime} + 2.303\ RT \log \{[B]/[A]\} = \Delta G^{o\prime} + 2.303\ RT \log \Gamma$$

This relation also holds for reactions with several reactants and products but K must be defined so as to be dimensionless. We can now illustrate the relation between ΔG′ and the displacement of the reaction from equilibrium (see Figure 13).

When the reaction is set at equilibrium, K = Γ and G is minimum and S is maximum. Any displacement from equilibrium would require an input of Gibbs energy and cannot occur spontaneously. The numerical value of ΔG′ depends on the extent of the displacement from equilibrium; at 25 °C a reaction that is one order of magnitude from equilibrium (10B > A) has a value of the 2.303 RT log Γ term of 5.71 kJ mol^{-1}, and one that is two orders of magnitude from equilibrium has a value of 11.5 kJ mol^{-1}. The value of the Gibbs energy change is negative (exergonic) when the mass-action ratio is less than the equilibrium constant and positive

(endergonic) when it is greater, since $\Delta G' = \Delta G^{o\prime} + -2.303\ RT \log K/\Gamma$. It must be stressed that these values of $\Delta G'$ only represent those at the stated concentrations of reactants and products. As will be seen from Figure 13, the value of $\Delta G'$ decreases when the reaction is allowed to proceed to equilibrium. However, in cells, reactions usually proceed in a steady state with continual supply of the reactant and removal of product so that $\Delta G'$ remains constant.

ATP is involved in many endergonic reactions necessary for cell metabolism, for example, the first stage of glucose utilization catalyzed by hexokinase:

$$\text{Glucose} + \text{ATP} \rightarrow \text{Glucose 6-phosphate} + \text{ADP} + H_2O$$

where $K = 360$ and $\Delta G^{o\prime} = -21\ \text{kJ mol}^{-1}$. Formally, this reaction can be treated as if it is the sum of the reactions

$$\text{Glucose} + P_i = \text{Glucose 6-phosphate} + H_2O,\ \Delta G^{o\prime} = 12.5\ \text{kJ mol}^{-1}$$

$$\text{ATP} + H_2O = \text{ADP} + P_i,\ \Delta G^{o\prime} = -33.5\ \text{kJ mol}^{-1}$$

because the free energy difference depends only on the initial and final states, without assuming any reaction mechanism. ATP is often, inaccurately, regarded as having a high-energy phosphate bond. Its chemical properties are due to the whole molecule and the value of ATP as an energy currency depends on maintenance of non-equilibrium concentrations in cells by oxidative phosphorylation that can drive many biochemical reactions when catalyzed by appropriate enzymes. The apparent equilibrium constant for the hydrolysis of ATP by complex V

$$K_{app} = [\text{ADP}][P_i]/[\text{ATP}]$$

depends on the concentrations of Mg^{2+} (since Mg^{2+} binds to phosphate groups), ATP, ADP, P_i, and on the pH (which influences ionization of phosphate groups), where the terms in brackets are total concentrations of all species. K_{app} may have a value of about 10^5 for conditions in the matrix.

Reference: Nicholls, D.G. & Fergusson, S.J. (1992). Bioenergetics 2, pp. 39–54, Academic Press, London.

APPENDIX 3: THERMODYNAMICS AND STOICHIOMETRY OF MITOCHONDRIAL ATP SYNTHESIS

Thermodynamics

We have outlined in the text the generally accepted mechanism for ATP synthesis. Electron transport pumps protons out of the matrix maintaining an electrical potential (negative inside) and a pH gradient (alkaline inside) so that both gradients form the proton motive force (PMF, ΔP). The PMF drives the synthesis of ATP by complex V. This mechanism necessarily implies that the inner membrane is a closed structure surrounding the matrix. The energy changes involved can be treated quantitatively and this involves five basic ideas:

1. The $\Delta G'$ of the redox span of segments of the electron-transport chain concerned with proton-pumping (Appendix 2) which is given by $\Delta G^\circ = -nF\Delta E^\circ$ in the standard state, or by

$$\Delta G' = -nF\Delta E'_h$$

for specific conditions, where n is the number of electrons transferred, and F is Faraday's constant (96,548 $kJmV^{-1}M^{-1}$, defined as the electrical charge of 1 mol of electrons (6.023×10^{23} electrons)).
2. The $\Delta G'$ required for ATP synthesis, which depends on the conditions (Appendix 2).
3. The difference in the electrical potential of protons across the inner membrane ($\Delta\psi$) in mV.
4. The $\Delta G'$ of the concentration (strictly the activity) difference of protons across the inner membrane.
5. The PMF (ΔP), which drives ATP synthesis and which has an electrical component ($\Delta\psi$) and a concentration (osmotic) component (ΔpH).

It is important to note that the PMF is usually expressed in millivolts, and $\Delta G'$ as kJ mol^{-1}. The redox potential difference $\Delta E'_n$ (where $\Delta E'_h = E_h'_A - E'_{hB}$ between two redox couples A_{red}/A_{ox} and B_{red}/B_{ox} is often expressed in electrical units, but it can also be expressed as kJ mol^{-1}

$$\Delta G' = -nF\Delta E^{\circ\prime}$$

where n is the number of electrons transferred.

During ATP synthesis, protons move down these gradients from outside $[H^+_{OUT}]$ into the mitochondrial matrix $[H^+_{IN}]$, each proton doing both electrical and osmotic work (due to the concentration difference) so that the Gibbs energy change is

$$-\Delta G'_{total} = -\Delta'_{electrical} - \Delta G'_{osmotic}$$

The electrical component is given by

$$\Delta G'_{electrical} = -F\,\Delta\psi$$

where $\Delta\psi$ is the membrane potential ($\psi_{OUT} - \psi_{IN}$) defining ΔP as positive, $\Delta P = -\Delta G'/F$ (note that ΔP is sometimes defined as negative, for example by Harold (1986)). The osmotic component is the work done in transferring a proton from the outside to the mitochondrial matrix.

$$\Delta G'_{osmotic} = 2.303\ RT \log \frac{[H^+_{in}]}{[H^+_{out}]}$$

$$= 2.303\ RT\ (pH_{in} - pH_{out}) = 2.303\ RT\Delta\ pH$$

(pH is defined as $-\log[H^+]/[H^+]^\circ$, where $[H^+]^\circ$ is a standard state equal to unity to make the definition dimensionally correct, but $[H^+]^\circ$ is usually omitted for simplicity). Hence,

$$\Delta G'_{total} = -F\Delta\psi + 2.303\ RT\Delta pH$$

converting to electrical units and multiplying by –1

$$\Delta G'/F = \Delta P = \Delta\psi - 2.303\ RT/F\ \Delta pH$$

At 37 °C, 2.303 RT/F = 61.5mV. Therefore, $\Delta P = \Delta\psi - 61.5\ \Delta pH$

This equation shows that the contribution of the two components of the PMF differ with different conditions. In state 4, the electrical potential gradient across the inner membrane can be as high as 300,000 Vcm^{-1} and the Δ pH difference one unit. ATP synthesis only occurs when the PMF is sufficiently large. The phosphorylation potential ($\Delta G'_{ATP}$) is lower for ATP synthesis in the matrix ($\Delta G'_{ATP\ IN} = 3\Delta P$) for ATP exported to the cytosol ($\Delta G'_{ATP\ OUT} = 4\Delta P$) because an extra proton is consumed in importing ADP into the matrix (see text).

Most textbooks give a calculation for the efficiency (E) of the utilization of Gibbs energy by a coupled biochemical reaction, for example the ATP-dependent phosphorylation of glucose (Appendix 2) by dividing the standard Gibbs energy, $\Delta G^{o\prime}$, of glucose 6-phosphate (which is formed), by that of ATP, $\Delta G^{o\prime}_2$ (which is used); E = 12.5/33.5 = 37%.

But, $E = \Delta G^{o\prime}_1/\Delta G^{o\prime}_2 = 2.303\ RT\log K_1/2.303RT\log K_2 = \log K_1/\log K_2$ where K_1 and K_2 are the corresponding equilibrium constants. Taking antilogarithms $K_2^E = K_1$. This defines efficiency as the power to which one equilibrium constant must be raised to give another, which is nonsense (Atkinson, 1977; Cornish-Bowden, 1983). Similar dubious calculations are often made for this efficiency of complete metabolic pathways; for example, the oxidation of 1 mol of glucose to CO_2 and H_2O ($-\Delta G^{o\prime}$ = 2868 kJ mol^{-1}) assumed to couple to the traditional value of 36 mol of ATP synthesized ($\Delta G^{o\prime}_{ATP}$ = 33.5 kJ mol^{-1}, (for 36 mol; $\Delta G^{o\prime}_{ATP}$ = 1206 kJ mol^{-1}) so that E = 1206/2868 = 42%. However, ATP synthesis usually occurs in a steady state where the $\Delta G'$ value for synthesis depends on how far the mass action ratio $[ADP][P_i]/[ADP]$ is displaced from equilibrium, depending on physiological conditions. Further, mitochondrial preparations are able to maintain the mass action ratio as low as 10^{-5} M, which is 10 orders of magnitude from equilibrium (Slater et al., 1973). The $\Delta G'$ for electron transport from NADH to oxygen is 220 kJ mol^{-1} and $\Delta G'$ for ATP synthesis may vary between –43 kJ mol^{-1} and –65 kJ mol^{-1}. Therefore, the values for the synthesis of 2.5mols of ATP would vary between 108 kJ mol^{-1} and 163 kJ mol^{-1} (Slater et al., 1973). The true efficiency, according to irreversible thermodynamics, is given by power output over power input (Stucki, 1983; 1991).

Stoichiometry

It is conventional to discuss the stoichiometry for proton extrusion as $H^+/2e^-$ ratios, although there are two-, one-, and four-electron reductions at different stages in the respiratory chain. Most textbooks still assert that the flow of two electrons

from NADH to oxygen is associated with the synthesis of 3 mol of ATP, and from succinate to oxygen with the synthesis of 2 mol. However, there has been considerable debate during the last 15 years about these values.

There now appears to be a consensus that the oxidation of one molecule of NADH is linked to the extrusion of 10 protons, that three protons are required to synthesize one ATP and that one proton is required to exchange ADP^{3-} in the cytosol for ATP^{4-} (Hinkle et al., 1991). This would give a maximum ATP/$2e^-$ ratio of 2.5 (Brand et al., 1993). Many recent measurements strongly indicate that ATP/$2e^-$ ratios of slightly less than 2.5, are in agreement with this prediction. Many older determinations apparently giving higher values may be subject to error (Senior and Sherratt, 1968). However, very careful determinations using the now unfashionable manometric techniques have given values of 2.7 to 2.9 (Aldridge, 1957; Senior and Sherratt, 1968). The reasons for these discrepancies are unknown, but the higher values, if valid, would require revision of current ideas of proton stoichiometries. Thus if two protons were necessary to make one ATP, the ATP/$2e^-$ ratio (including ADP^{3-}/ATP^{4-} exchange) would be 10/3 = 3.33. Alternatively, as has sometimes been claimed, more protons than the consensus value of 10 would have to be pumped by the electron transport chain. Unfortunately, technical problems have so far prevented an unequivocal resolution of these problems.

The textbook statements about the number of mols of ATP synthesized when a mol of glucose or of a fatty acid is oxidized completely to CO_2 and water are therefore invalid. Most state that the complete oxidation of glucose yields 36 mols of ATP (or 38 depending on the pathway by which NADH generated by glycolysis is oxidized); this assumes that all the ATP is delivered to the cytosol, that there is no uncoupling and that ΔP is not used for the transport of ions or proteins into the matrix. Making these assumptions, but using revised values for proton stoichiometries, Brand et al. (1993) quote a value of 29.5 mols ATP for the complete oxidation of glucose. Some ATP is used in the matrix (more in hepatocyte mitochondria than in muscle mitochondria), when no protons are required for ADP/ATP exchange when ADP is converted to ATP. So, assuming consensus values, the oxidation of one mol of NADH would give 10/3 = 3.33 mol ATP.

Sometimes it is stated that the extrusion of two protons from the matrix is associated with the oxidation of one molecule of ubiquinol by complex III and four with the oxidation of two molecules of reduced cytochrome c by complex IV (Hinkle et al., 1991). For the oxidation of ubiquinol by complex III in isolation (ubiquinol:cytochrome c reductase) the reaction is thought to be

$$UQH_2 + 2 \text{ cytochrome } c^{3+} + 2H_{IN} \rightarrow UQ + 2 \text{ cytochrome } c^{2+} + 4H_{OUT}$$

with a net translocation of only two protons from the matrix, since the other two protons are derived from added UQH_2. If the oxidation of reduced cytochrome c is also considered in isolation the net reaction is

$$2 \text{ cytochrome } c^{2+}_{out} + 1/2O_2 + 4H^+_{IN} \rightarrow 2\text{cytochrome } c^{3+}_{out} + H_2O + 2H^+_{out}$$

Although only two protons are pumped out of the matrix, two others from the matrix are consumed in the formation of H_2O. There is therefore a net translocation of four positive charges out of the matrix which is equivalent to the extrusion of four protons. If four protons are required by the chemiosmotic mechanism to convert cytosolic ADP + P_i to ATP, then 0.5 mol ATP is made for the oxidation of one mol of ubiquinol and one mol ATP for the oxidation of 2 mols of reduced cytochrome c. These stoichiometries were obtained experimentally when ubiquinol was oxidized when complexes I, II, and IV were inhibited by rotenone, malonate, and cyanide, respectively, and when reduced cytochrome c was oxidized with complex III inhibited by antimycin (Hinkle et al., 1991). (In these experiments, of course, no protons were liberated in the matrix by substrate oxidation.) However, in the scheme illustrated in Figure 6, with the flow of two electrons through the complete electron transport chain from substrate to oxygen, it also appears valid to say that four protons are extruded by complex I, four by complex III, and two by complex I.

Reference: Nichols, D.G. & Ferguson, S.J. (1992). Bioenergetics 2, Academic Press, London.

ACKNOWLEDGMENTS

We thank Dr. I.C. West for his valuable comments and Dr. A.E. Senior for information.

REFERENCES

Aldridge, W.N. (1957). Liver and brain mitochondria. Biochem. J. 67, 423–431.

Atkinson, D.E. (1977). In: Cellular Energy Metabolism and its Regulation. Academic Press, New York.

Babcock, G.T. & Wikström, M. (1992). Oxygen activation and the conservation of energy in cell respiration. Nature 356, 301–309.

Birch-Machin, M.A., Jackson, S., Singh Kler, R.S., & Turnbull, D.M. (1993). Study of skeletal muscle mitochondrial dysfunction. Methods Toxicol. 2, 51–69.

Brand, M.D., Harper, M.E., & Taylor, H.C. (1993). Control of the effective P/O ratio of oxidative phosphorylation in liver mitochondria and hepatocytes. Biochem. J. 291, 739–748.

Brand, M.D. & Murphy, M.P. (1987). Control of electron flux through the respiratory chain in mitochondria and cells. Biol. Rev. 62, 141–193.

Brown, G.C. (1992). Control of respiration and ATP synthesis in mammalian mitochondria and cells. Biochem. J. 284, 1–13.

Capaldi, R.A. (1982). Arrangement of proteins in the inner mitochondrial membrane. Biochim. Biophys. Acta 694. 291–306.

Chance, B. & Williams. G.R. (1956). The respiratory chain and oxidative phosphorylation. Adv. Enzymol. 17. 65–134.

Chance, B., Leigh, J.S., Kent, J., McCully, K., Nioka, S., Clark, B.J., Marris, J.M., & Graham, T. (1986). Multiple controls of oxidative metabolism in living tissues as studied by phosphorous magnetic resonance. Proc. Natl. Acad. Sci. USA 83, 9458–9462.

Clayton, D.A. (1984). Transcription of the mammalian mitochondrial genome. Ann. Rev. Biochem. 53, 573–594.

Cornish-Bowden, A. (1983). Metabolic efficiency: Is it a useful concept? Biochem. Soc. Trans. 11, 44–45.

Cornish-Bowden, A. (1991). Failure of channelling to maintain low concentrations of metabolic intermediates. Eur. J. Biochem. 195, 103–108.

Cross, R.L. (1981). The mechanism and regulation of ATP synthesis by F_1-ATPases. Ann. Rev. Biochem. 50, 689–714.

Darley-Usmar, V.M., Rickwood, D., & Wilson, M.T. (eds.) (1987). Mitochondria: A Practical Approach. IRL Press, Oxford, Washington.

Davidson, A., & Halestrap, A.P. (1989). Inhibition of mitochondrial-matrix inorganic pyrophosphatase by physiological [Ca^{2+}], and its role in the hormonal regulation of mitochondrial matrix volume. Biochem. J. 258, 817–821.

Davis, E.J., Bremer, J., & Akerman, K.E. (1980). Thermodynamic aspects of translocation of reducing equivalents by mitochondria. J. Biol. Chem. 225, 2277–2283.

Esser, V., Britton, C.H., Weiss, B.C., Foster, D.W. & McGarry, J.D. (1993). Cloning, sequencing and expression of a cDNA encoding rat liver carnitine palmitoyltransferase I. Direct evidence that a single polypeptide is involved in inhibitor interaction and catalytic function. J. Biol. Chem. 268, 5817–5822.

Fell, D.A. (1992). Metabolic control analysis: A survey of its theoretical and experimental development. Biochem. J. 286, 313–330.

Fersht, A. (1985). Enzyme Structure and Mechanism. 2nd edn., W.H. Freeman & Co., New York.

Fujiki, M. & Verner, K. (1993). Coupling of cytosolic protein synthesis and mitochondrial protein import in yeast. J. Biol. Chem. 268, 1914–1920.

Glover, L.A. & Lindsay, J.G. (1992). Targeting proteins to mitochondria: A current overview. Biochem. J. 284, 609–620.

Gitomer, W.L. (1987). The activity of mitochondrial water. Biochem. J. 248, 997–998.

Greville, G.D. (1968). In: Carbohydrate Metabolism and its Disorders (Dickens, F., Randle, P.J., & Whelan, W.J., eds.), pp. 297–334, Vol. 1, Academic Press, London.

Gröen, A.K., Wanders, R.J.A., Westerhoff, H.V., van der Meer, R., & Tager, J.M. (1983). Quantification of the contribution of various steps to the control of mitochondrial respiration. J. Biol. Chem. 257, 2754–2757.

Guzmán, M. & Geelen, M.J.H. (1993). Review. Regulation of fatty acid oxidation in mammalian liver. Biochim. Biophys. Acta 1167, 227–241.

Hafner, R.P., Brown, G.C., & Brand, M.D. (1990). Analysis of the control of respiration rate, phosphorylation rate, proton leak rate and proton motive force in isolated mitochondria using the "top-down" approach of metabolic control theory. Eur. J. Biochem. 188, 313–319.

Harris, D.A. & Das, A.M. (1991). Control of mitochondrial ATP synthesis in heart. Biochem. J. 280, 561–573.

Harold, F.M. (1986). The Vital Force: A study of Bioenergetics, W.H. Freeman and Co., New York.

Hartl, F.-U. & Neupert, W. (1990). Protein sorting to mitochondria: Evolutionary conservations of folding and assembly. Science 247, 930–937.

Hinkle, P.C., Kumar, M.A., Rasetar, A., & Harris, D.L. (1991). Mechanistic stoichiometry of mitochondrial oxidative phosphorylation. Biochemistry 30, 3576–3582.

Hue, L. & Van de Werve, G. (eds.) (1981). In: Short-Term Regulation of Liver Metabolism. Elsevier/North Holland Biochemical Press, Amsterdam.

Kacser, H. & Burns, J.A. (1978). Molecular democracy: Who shares the control? Biochem. Soc. Trans. 7, 1149–1160.

Klingenberg, M. (1980). The ADP-ATP translocation in mitochondria, a membrane potential controlled transport. J. Memb. Biol. 56, 97–105.

Krebs, H.A. (1943). The intermediary stages in the biological oxidation of carbohydrate. Adv. Enzymol. 3, 191–252.

LaNoue, K.E. & Schoolwerth, A.C. (1979). Metabolite transport in mitochondria. Ann. Rev. Biochem. 48, 871–922.

Luft, R., Ikkos, D., Palmieri, G., Ernster, L., & Afzelius, B. (1962). A case of severe hypermetabolism of non-thyroid origin with a defect in the maintenance of mitochondrial respiratory control: A correlated clinical, biochemical and morphological study. J. Clin. Invest. 41, 1776–1804.

Manning-Krieg, U.C., Scherer, P.E., & Schatz, G. (1991). Sequential action of mitochondrial chaperonins in protein import into the matrix. EMBO J. 10, 3273–3280.

Maretski, D., Reimann, B., & Rapoport, S.M. (1989). A reappraisal of the binding of cytosolic enzymes to erythrocyte membranes. Trends Biochem. Sci. 14, 93–96.

Mendes, P., Kell, D.B., & Westerhoff, H.V. (1992). Channelling can decrease pool size. Eur. J. Biochem. 204, 257–260.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. Nature 191, 144–148.

Mitchell, P. (1976). Possible molecular mechanisms of the proton motive function of cytochrome systems. J. Theoret. Biol. 62, 327–367.

Mitchell, P. & Moyle. J. (1967). Respiration-driven proton translocation in rat liver mitochondria. Biochem. J. 105, 1147–1162.

Moreno-Sáchez, R., Devars, S., López-Gómez, F., Uribe, A., & Corona, N. (1991). Distribution of control of oxidative phosphorylation in mitochondria oxidizing NAD-linked substrates. Biochim. Biophys. Acta 1060, 284–292.

Murthy, M.S.R. & Pande, S.V. (1987). Malonyl-CoA binding site and the overt carnitine palmitoyltransferase activity reside on the opposite sides of the outer mitochondrial membrane. Proc. Nat. Acad. Sci. USA 84, 378–382.

Nicholls, D.G. (1974). The influence of respiration and ATP hydrolysis on the proton electrochemical gradient across the inner membrane of rat liver mitochondria as determined by ion distribution. Eur. J. Biochem. 50, 305–315.

Nicholls, D.G. & Ferguson, S.J. (1992). In: Bioenergetics 2, Academic Press, London.

Osmundsen, H. & Hovik, R. (1988). β-Oxidation of polyunsaturated fatty acids. Biochem. Soc. Trans. 16, 420–422.

Otto, D.A. (1984). Relationship of the ATP/ADP ratio to the site of octanoate activation. J. Biol. Chem. 259, 5490–5194.

Ovádi, J. (1988). Old pathway—new concept: Control of glycolysis by metabolite-modulated dynamic enzyme associations. Trends Biochem. Sci. 13, 486–490.

Pavlov, A.R., Revina, A.A., Dupin, A.M., Baldyrev, A.A., & Yaropolow, A.I. (1993). The mechanism of interaction of carnosine with superoxide radicals in water solutions. Biochem. Biophys. Acta 1157, 304–314.

Pedersen, P.L. & Amzel, L.M. (1993). ATP syntheses. Structure reaction center, mechanism and regulation of one of nature's most unique machines. J. Biol. Chem. 268, 9937–9940.

Pfanner, N., Söllner, T., & Neupert, W. (1991). Mitochondrial import receptors for precursor proteins. Trends Biochem. Sci. 16, 63–67.

Pfanner, N., Rossow, J., van der Klei, I.J., & Neupert W. (1992). A dynamic model of the mitochondrial protein import machinery. Cell 68, 999–1002.

Porter, K. & Brand, M.D. (1993). Body mass dependence of H^+ leak in mitochondria and its relevance to metabolic rate. Nature 362, 628–630.

Robinson, J.B. & Srere, P.A. (1985). Organization of Krebs tricarboxylic acid cycle enzymes in mitochondria. J. Biol. Chem. 260, 10800–10805.

Schulz, H. (1985). Oxidation of fatty acids. In: Biochemistry of Lipids and Membranes (Vance, D.E. & Vance, J.E., eds.), pp. 116–142, Benjamin/Cummings Publishing Co., Menlo Park, California.

Senior, A.E. (1988). ATP synthesis by oxidative phosphorylation. Physiological Rev. 68, 177–230.

Senior, A.E. (1990). The proton-translocating ATPase of *Esherichia coli*. Ann. Rev. Biophys. Chem. 19, 7–41.

Senior, A.E. & Sherratt, H.S.A. (1968). Biochemical effects of the hypoglycaemic compound pent-4-enoic acid and related non-hypoglycemic fatty acids. Oxidative phosphorylation and mitochondrial oxidation of pyruvate, 3-hydroxybutyrate and tricarboxylic acid-cycle intermediates. Biochem. J. 110, 499–509.

Sherratt, H.S.A. (1981). Inhibition of gluconeogenesis by non-hormonal hypoglycaemic compounds. In: Short-Term Regulation of Liver Metabolism (Hue, L. & Van de Werve, G., eds.), pp. 119–227, Elsevier/North Holland Biochemical Press, Amsterdam.

Sherratt, H.S.A., Watmough, N.J., Johnson, M., & Turnbull, D.M. (1988). Methods for study of normal and abnormal skeletal muscle mitochondria. Met. Biochem. Analysis 33, 243–335.

Slater, E.C., Rosing, J., & Mol, A. (1973). The phosphorylation potential generated by respiring mitochondria. Biochim. Biophys. Acta 292, 534–553.

Srere, P.A. (1987). Complexes of sequential metabolic enzymes. Ann. Rev. Biochem. 56, 89–124.

Stucki, J.G. (1991). Non-equilibrium thermodynamic sensitivity of oxidative phosphorylation. Proc. R. Soc. B. 244. 197–202.

Stumpf, D.A., Haas, R., Eguren, L.A., Parks, J.K., & Eilert, R.E. (1982). Proton motive force in muscle mitochondria. Muscle and Nerve 5, 14–19.

Sumegi, B., Porpaczy, Z., & Alkonyi, I. (1991). Kinetic advantage of the interaction between the fatty acid β-oxidation enzymes and the complexes of the respiratory chain. Biochim. Biophys. Acta 1081, 121–128.

Tyler, D.D. (1992). The Mitochondria in Health and Disease. VCH, Cambridge.

Uchicda, Y., Izai, K., Orii, T., & Hashimoto, T. (1992). Novel fatty acid β-oxidation enzymes in rat liver mitochondria. II. Purification and properties of enoyl-coenzyme A (CoA) hydratase/3-hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase trifunctional protein. J. Biol. Chem. 267, 1034–1041.

Verner, K. (1992). Early events in yeast mitochondrial protein targeting. Mol. Microbiol. 6, 1723–1728.

Wallimann, T., Wyss, M., Brdicza, D., Nicolay, K., & Eppenberger, H.M. (1992). Intracellular compartmentation, structure and function of creatine kinase isoenzymes in tissues with high and fluctuating energy demands: The "phosphocreatine circuit" for cellular energy homeostasis. Biochem. J. 281, 21–40.

Watford, M. (1991). The urea cycle: A two system compartment system. Essays Biochem. 26, 49–48.

Watmough, N.J., Turnbull, D.M., Sherratt, H.S.A, & Bartlett, K. (1989). Measurement of the acyl-CoA intermediates of β-oxidation by hplc with on-line radiochemical and photodiode-array detection. Application to the study of [U-^{14}C]hexadecanoate by intact rat liver mitochondria. Biochem. J. 262, 261–269.

Westerhoff, H.F. & Welch, G.R. (1992). Enzyme organization and the direction of metabolic flow: Physiochemical considerations. Curr. Topics Cell. Reg. 33, 361–390.

Williams, R.J.P (1978). The multifarious couplings of energy transduction. Biochim. Biophys. Acta 505, 1–44.

Yeaman, S.J. (1989). The 2-oxo acid dehydrogenase complexes: Recent advances. Biochem. J. 257, 625–632.

RECOMMENDED READINGS

Abrahams, J.P., Leslie, A.G.W., Lutter, R., & Walker, J.E. (1994). Structure at 2.8Å resolution of F_1-ATPase from bovine heart mitochondria. Nature 370, 621–628.

Agius, L., & Sherratt, H.S.A. (eds.) (1996). Channelling and intermediary metabolism. Portland Press, London.

Friedrich, T., Steinmuller, K., & Weiss, H. (1995). The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. (Review). FEBS Letters 367, 107–111.

Chapter 6

Regulation and Activity of Smooth Muscle

LLOYD BARR

INTRODUCTION

Smooth muscle cells constitute the most diversified class of muscle cells. They are the parenchymal cells of many organs, including the uterus and other reproductive organs, gall bladder, urinary bladder, respiratory passages, etc. In all these cases,

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 155–200.

ISBN: 1-55938-808-0

as the parenchymal cell type, smooth muscle cell actions characterize the action of the organ. However, there are complexities within a particular organ because there are often very significant differences between smooth muscle cells and their innervation from region to region. Finally, most muscle cells in animals are smooth muscle cells, despite the exceptional mass of skeletal muscle fibers whose responsibilities are specialized for gross animal movement and cardiac myocytes, which are also highly specialized.

Like cardiac muscle, smooth muscle does not work against a bony skeletal lever system. Instead, it is usually part of the wall of a hollow organ and, as such, it is often involved in regulation of flow of the contents of a tube either as a sphincter or as a peristaltic pump. This requires coordination in time between different regions. Smooth muscle cells are specialized according to the quickness and duration of their contractions. Various smooth muscles are often classified as either tonic or phasic, according to quickness of onset of contraction and the duration of the contractile event. Muscles which contract slowly and maintain their contraction over long periods are called tonic muscles, e.g., the vascular tree, while those which respond quickly are called phasic muscles, e.g., parts of the gastrointestinal tract. Tonic muscles are usually associated with sphincters, while the phasic muscles often provide the mechanical energy for the pumps.

The integration of smooth muscle cells into organs requires their interaction with (a) each other, (b) nerve cell processes, and (c) connective tissue cells, fibroblasts, etc. In general, smooth muscle cells are interconnected by gap junctions (nexuses) which provide direct passage ways between the cell interior for ions and small molecules. Thus, in those organs where smooth muscle cells are organized into sheets or bands, there is always the possibility that there is functional continuity between the smooth muscle cells of one region and those of another region separated by a connective tissue barrier, e.g., between layers in the gastrointestinal tract. Only in those situations where the smooth muscle cells are either closely associated or widely dispersed is the issue of functional continuity perfectly understood. Although electrical activity does spread between smooth muscle cells directly, the regulation and coordination of the smooth muscle of various organs is regionally controlled by neural activity. In many cases, e.g., the gastrointestinal tract, the neurons involved are part of extensive and relatively independent peripheral plexuses.

Smooth muscle cell activity is in general under neural control. Thus, the many transmitters of the autonomic nervous system are paired with receptors on the smooth muscle cell membrane. One of the current questions about smooth muscle function is: What intracellular processes are the different transmitters modulating in the smooth muscle cells, in addition to their effects on the contractile state?

Investigation of how smooth muscles work is most frequently carried out using experimentally reduced systems of some sort. The *in vivo* relevance of processes observed in such preparations can be divided according to the degree of complexity

of the preparations used. While there have been observations made on enzymatic preparations for decades, and whole animal experiments for even longer, in the recent past the number of levels of experimental reduction which are available has greatly increased. For mechanical studies these include bundles of smooth muscle cells permeabilized in various ways, single cells in culture, and protein filament systems. In a similar fashion, electrical studies can be made on various cultured systems, including single cells, pharmacologically altered bundles, and membrane patches, including those which contain only a single channel.

DIVERSITY AND STRUCTURE OF SMOOTH MUSCLE CELLS

Although it is difficult to make absolute categories for different kinds of smooth muscles, it is relatively easy to identify attributes which can be used to differentiate between large groups of smooth muscles. For example, in addition to being tonic or phasic, smooth muscle cells usually are physiologically either activated by cholinergic or adrenergic stimuli. So different smooth muscle cells express different signal-effector pathways, including the receptors, the complex of connecting components, and the contractile machinery in a particular smooth muscle cell. In addition, there are nonmechanical effector systems modulated by the signal transduction network. Very significantly for vascular smooth muscle, cellular metabolism may shift into a proliferation mode. These constitutive differences between smooth muscle cells are also manifest at the cytological and histological levels.

Smooth muscle cells are either spindle shaped (fusiform) or less frequently stellate, as they are in some blood vessels. Their nuclei are centrally located and ovoid. At the ends of the nuclei, the Golgi systems and the mitochondria together form caps which taper following the contour of the cytoplasmic filaments. Mitochondria are also found in the cell cortex, subjacent to the plasma membrane. As smooth muscle cells diverge in their molecular and microscopic morphology, this in turn leads to differences in their function. As an example of smooth muscle differentiation, the degree to which smooth muscle cells in a tissue exhibit synchronous contractions has served as a basis to classify a particular tissue as being either unitary, operating as a single unit or multiunit, with smaller groups of cells operating independently. In the case of unitary smooth muscles, they are usually triggered to synchronized activity by the passage of action potentials from cell to cell via connexons at gap junctions (nexuses) (Figure 1).

Electrical continuity between cell interiors is provided by the tight association of heximeric doughnut shaped protein units (hemi-connexons) residing in neighboring smooth muscle cell membranes. The six monomers (connexins) are arranged to give six-fold symmetry, like a regular hexagon. These associated, monomeric connexins leave a hole in the center of a hemi-connexon, akin to a doughnut hole. Where two hemi-connexons stick to each other so that the doughnut holes are continuous, an actual aqueous pathway between cell interiors is constructed. The openings are not static and are regulated by a variety of processes (vidé infra).

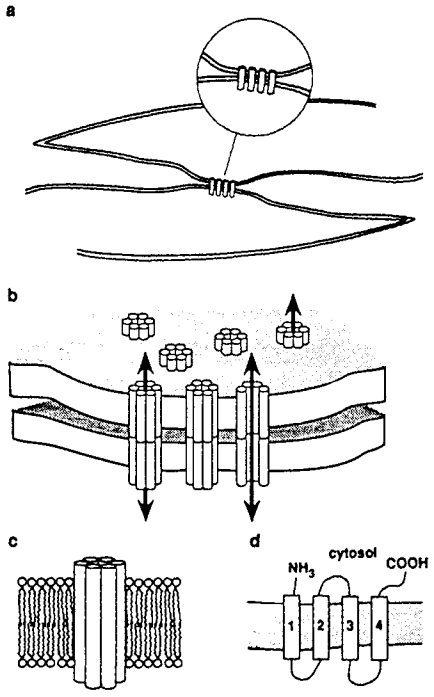


Figure 1. A gap junction (or nexus) between smooth muscle cells is an array of connexons which provide pathways for the movement of ions and small molecules between cell interiors. In **(a)** the structure of a nexal region as seen in electron micrographs is shown schematically. The adjacent cell membranes adhere to each other by associations between intrinsic proteins in both membranes. In the regions surrounding the connecting proteins, the cell membranes are separated by aqueous channels continuous with the extracellular space, the so-called gaps. In **(b)** connexons, the connecting protein units, are shown to be formed by heximeric, hemi-connexons in each of the cell membranes which bind to each other. The narrow, 1.5 nm, aqueous, cytosol-to-cytosol channel is formed by the arrangement of the monomers, connexins, in a slightly expanded circle. The cartoon in **(c)** shows the currently understood relationship between the connexins of a single hemi-connexon and the lipid bilayer structure of a cell membrane. Finally, **(d)** is a cartoon of the presently held substructure of a single connexin molecule. There are four membrane spanning alpha helices and both the carboxyl and amino termini are on the cytoplasmic side. The two extracellular loops, connecting helix 1 to helix 2 and helix 3 to helix 4, are thought to be responsible for the formation of connexin from hemi-connexins.

Different cells express different connexin isoforms, which make for great variations in binding affinities between different hemi-connexons. Connexons do form between cells of different types but not all heterogeneous hemi-connexon combinations support the formation of connexons between different cells. This may be

important for the transition areas between different kinds of smooth muscle cells in various organs, for example, at sphincters.

The extracellular space between smooth muscle cells is not amorphous and inert. The components of the extracellular space are specific to each tissue. First, there are a number of different kinds of collagen and elastin molecules which make up the major constituents of the filaments of the extracellular space. In the case of collagen, the number of molecular forms is clearly more than a dozen. Collagen filaments between smooth muscle cells are largely made of the Type III and Type V molecules and this results in less rigidity than in some other cases. Type III collagen of arteries, the intestine, and the uterus is synthesized by smooth muscle cells. There is convincing evidence that Type I and Type V collagen are also synthesized by smooth muscle cells, but to what extent is not yet clear. The collagen filaments tend to run parallel to the axes of the smooth muscle cells in a given tissue, but are seen in electron micrographs to follow precisely the contours of basal lamina where the latter are folded. Basal lamina comprise the collection of thinner filaments and ground substance intimately associated with the cell membrane. The extracellular Type IV collagen is linked by multiadhesive proteins, (e.g., the glycoprotein, laminin), to intrinsic special receptor proteins in the cell membrane. Multiadhesive proteins also link various collagen and elastin filaments which together provide the mechanical stability of the extracellular matrix to the basal lamina. Certain smooth muscles, especially in the cardiovascular system, have a high proportion of elastin filaments surrounding them. They are prevalent in regions subject to frequent, large changes of stress, such as the arteries, the pulmonary tree, etc. Proteoglycans, together with the glycoproteins, another class of extracellular protein with an even larger relative content of carbohydrate, make up the so-called ground substance of the extracellular matrix. Proteoglycans, in addition to whatever mechanical function they have, bind significantly to a large variety of first messenger substances. These functionalities probably derive from their high esterified sulfate content.

Inside the typical smooth muscle cell, the cytoplasmic filaments course around the nuclei filling most of the cytoplasm between the nuclei and the plasma membrane. There are two filamentous systems in the smooth muscle cell which run lengthwise through the cell. The first is the more intensively studied actin-myosin sliding filament system. This is the system to which a consensus of investigators attribute most of the active mechanical properties of smooth muscle. It will be discussed in detail below. The second system is the intermediate filament system which to an unknown degree runs in parallel to the actin-myosin system and whose functional role has not yet been completely agreed upon. The intermediate filaments are so named because their diameters are intermediate between those of myosin and actin. These very stable filaments are functionally associated with various protein cytoarchitectural structures, microtubular systems, and desmosomes. Various proteins may participate in the formation of intermediate filaments, e.g., vimentin,

desmin, etc. These subunit proteins have similar enough structural motifs that they will copolymerize. Desmin is characteristic of the intermediate filaments of smooth muscle but vimentin is also present in large amounts. The intermediate filaments segregate into exclusive bundles and do not mingle with the filaments of the actin-myosin system. These two kinds of bundles of different kinds of filaments run in parallel through the myoplasm of smooth muscle cells. The mechanical significance of intermediate filaments, although unknown, has often been the object of speculations.

The primary contractile proteins of smooth muscle, actin and myosin, are very similar to those of other muscles. Although actin molecules organize into filaments that are very similar to their counterparts in other muscles, the differences in the myosin molecules in many cases may lead to organization into filaments which may be quite different. In skeletal muscle, the golf club shaped myosin molecules, are stacked together into the myosin filaments, and are each slightly turned and inch along the filament axis. Thus, a helix is made of the handles of the golf clubs and the myosin heads trace out an equivalent helix on the surface of the filament. One half of the myosin filament is symmetrical to the other half, because all the handles point to the midpoint of the filament. There is mechanical polarity for each end of a myosin filament because the helically arranged crossbridges, (golf club heads), symmetrically angle away from the midpoint.

The filament structure in smooth muscle may be based on an entirely different principle from that of skeletal muscle. Indeed, current evidence indicates that smooth muscle myosin filaments are side-polar. That is, the myosin molecules are not arranged helically around a myosin filament, but instead the molecules all along one side would point in one direction and all the molecules on the opposite side would point in the other direction. Moreover, as a consequence of the details of the molecular stacking, they have crossbridges coming out on opposite sides of the filament only. Thus, the crossbridges on each side point toward opposite ends of adjacent actin filament. Consequent to this arrangement of myosin heads, the geometry of how the myosin filaments interact with actin filaments should be quite different. Still, we should remember that length-tension and force-velocity curves of all muscles are very similar. Thus, at the actual transduction level, it may be that the set of interactions between myosin crossbridges and actin binding sites are quite similar.

The structure of the contractile apparatus of smooth muscle at the next higher level is also characteristically different from other muscles. The concentrations of actin and myosin in smooth muscle are about three times higher for actin and four times lower for myosin than in skeletal muscle. Correspondingly, in smooth muscle the ratio of the numbers of moles of actin to moles of myosin, and the ratio of the number of actin filaments to those of myosin filaments, are about 12 times larger than for other muscles. Thus, the arrangements of the two sets of filaments are bound to be quite different just on the basis of numbers of actin and myosin

filaments. The different arrangements of these filaments is what gives different muscles their characteristic histological appearance.

First, in the striated muscles, the cross-sectional organization of filaments is highly ordered in a hexagonal pattern commensurate with the ratio of actin to myosin filaments and the distribution of active myosin heads, S-1 segments, helically every 60 degrees around the myosin filament. In smooth muscle, with perhaps 13 actin filaments per myosin filament, many actin filaments appear to be ranked in layers around myosin filaments. It is not known how the more distant actin filaments participate in contraction.

Second, striated muscles are striated because the myosin filaments and the actin filaments are held in register across the cell diameter by myofibrillar protein structures, the M-line and the Z-line, respectively. On the other hand, in smooth muscle, there are no equivalent bulkheadlike structures to align filaments of the same kind to yield the cross striations characteristic of other muscles. Myosin filaments of smooth muscle apparently have no midpoint to bind M-proteins. It is not clear how they are organized within an actin-myosin bundle. The situation for actin filaments is clearer since they are organized around dense bodies. Dense bodies are spindle shaped, α-actinin containing structures into which bundles of actin filaments penetrate from both sides along the spindle axes. The actin filaments are thereby polarized and myosin S-1 fragments bound to them always point away from the dense bodies. α-Actinin is one of a number of actin cross linking proteins occurring in various muscle types. It also binds to vinculin, a protein thought to bind, in turn, to integral membrane proteins in desmosomal regions. In addition to these other proteins, it appears that together with the myosin filaments, these structures provide the mechanical continuity necessary for the actin-myosin system to be able to bear a load.

FUNCTIONAL PATTERNS

There are smooth muscles in the simplest of animals, where there are no striated muscles. Myoepithelial cell activity is often integrated with smooth muscle activity. All the mechanical functions carried out by animals, excluding only blood pumping and skeleton moving, are executed by smooth muscles. Thus, the functional diversity of smooth muscle action is very broad. While many smooth muscles generate action potentials autorhythmically, contraction in other smooth muscle cells seem to be controlled, entirely independently of membrane potential changes, by G-protein cascades. In those smooth muscles which generate action potentials, they tend to do so in bursts or continuous trains. Variations and combinations of ionic currents and the associated channel proteins similar to those found in other tissues account for some of these activity patterns and for the wide variety of action potential shape (time-courses). However, the different electrical pacemaking activities are based on processes which at the present time are only dimly understood and may be peculiar to smooth muscle.

Mechanical Properties

Again, while the principles are the same, individual muscles vary greatly. The simplest mechanical property, compliance, describes the force required to stretch a muscle to a certain length. The relationship between increments of length and stretching force, stress, is not a constant but in fact varies with length. The compliance of a muscle, as a function of length, determines how much longer it becomes when subjected to an incremental elongating stress. Stretched muscle is less compliant. Skeletal muscles exert little resistance to stretch until their *in vivo* length is exceeded. On the other hand, most inactive smooth muscles are less compliant than inactive skeletal muscle over their whole range of lengths, but most noticeably below the typical *in vivo* lengths. Perhaps this is due to the relatively larger extracellular spaces, hence a stronger network of connective tissue filaments opposing stretch, and also a contribution from the intermediate filament system. When muscles are stimulated they become much stiffer and comparisons between them are confounded by the issues related to the details of the action of the contractile apparatus. Once activated, the contractile machinery is the primary contributor to the resistance to stretch at all but extreme lengths where the in parallel, extracellular filaments bear most of the stress.

However, it is clear that the resting compliance is relatively more important to the springlike behavior of many active smooth muscles. Each smooth muscle must be analyzed individually since the extracellular compartment of each smooth muscle as a histological entity is different.

To the extent that a muscle has viscosity, it exerts a retarding force proportional to the rate at which it is stretched. A completely inactive smooth muscle is mostly like a spring; it has very little viscosity. The viscouslike properties of active smooth muscle are dominated by the influence of rate of change of length on the behavior of the contractile apparatus. So in summary, the active chemomechanical transducing processes are very important in understanding both mechanical properties of active smooth muscle, but the resting component (parallel elastic component) may contribute to the springlike properties of many smooth muscles.

Functional Roles of Smooth Muscle

Smooth muscles are associated with epithelial sheets. They provide the mechanism to propel the contents of hollow organs to move and the means, via sphincters, to gate the flow through tubular organs. Although in many systems the motility function is dominant, in the cardiovascular system the regulation of flow by the sphincters of the microvasculature is clearly the primary function of vascular smooth muscle. Generally, smooth muscle cells are organized into sheets or bands which run in parallel with epithelial layers. Where smooth muscle structures are organized to act as sphincters, the axes of the smooth muscle cells tend to run circularly around the lumen of the organ, i.e., normal to the direction of flow and the axis of the organ. Elsewhere, as in the intestines, muscles which run in parallel

to the axis and also around the circumference of the lumen are concerned with mixing and moving of material down the tube. The functions of moving material down different portions of the tract are largely regulated by the autonomic nervous system or its extensions. In fact, all the influences of the nervous system on the activity of smooth muscle occur by way of the autonomic nervous system. In the peripheral vasculature, the neural control is more central and involves the regulation of patterns of arteriolar contractions and venous pooling. In the gastrointestinal tract, the neural control has a more peripheral component as it involves the sensing of the lumenal content and the organization of propulsive contractions of the smooth muscle layers in the intestinal wall.

Analytic Equivalents

The passive mechanical behavior of a muscle can be described quite well using analogies with idealized physical components, force generators, springs, and dashpots. Since the contractile apparatus is intracellular and some of the springlike behavior is clearly due to extracellular filaments, it is common to speak of a parallel elastic component and series elastic component as if they were extracellular. However, these analogies must be used with caution; for example, the series elastic component resides partially not only in series with the bridges that pull but also in the myosin crossbridges themselves, since the rest of the crossbridge is in series with the attachment sites. Moreover, intracellular intermediate filaments, etc., may contribute to the parallel elastic component. In the analysis of muscle as a machine which does mechanical work, it is important to understand the relationship between the length of the muscle and the force required to prevent the muscle from shortening, i.e., the length-tension curve.

If a muscle did not change its mechanical properties on activation, particularly the force it exerts at a given length, no external work would be possible. Therefore, activation has actually been defined as the change in the force needed to stretch a muscle when activation occurs, i.e., after stimulation. A response of a muscle held at a constant length is called an isometric contraction. Thus, during an isometric contraction, it is the increase in external extending force required to prevent an activated muscle from shortening that is recorded. If the force opposing an activated muscle is set constant experimentally, but is smaller than the peak isometric force, then after a transient period the muscle shortens, isotonically.

Consider a muscle in general physical terms. In order to stretch a body which has some constant internal structural restraints, a progressively larger stretching force must be applied to elongate the body more and more. When this kind of experiment is performed on an inactive muscle preparation, the relationship between length and force differs considerably from the behavior of a simple spring. The length of a simple spring increases in proportion to the elongating force. However, muscle gets stiffer at longer lengths so that relatively more force is required per unit length increment at longer lengths (Figure 2). This nonlinear springlike behavior seems to be due to the stretching of the network of extracellular

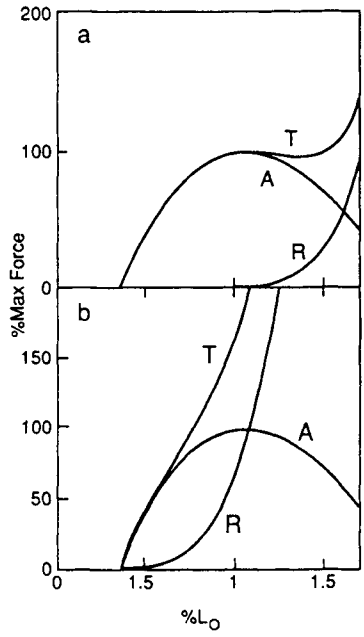


Figure 2. The relationship between a muscle's length and the force it exerts is usually plotted in three different ways. First, the force required to stretch (or maintain) an unstimulated muscle at an arbitrary length is plotted to show the resting force-length relationship. Second, the force required to maintain a stimulated muscle at a certain length is also plotted. Third, their difference, the active force-length relationship, is usually plotted as well. Some smooth muscles have force-length relationships quite similar to those of striated muscle, while in others the relationship is quite different. In **(a)** is shown the relationships for an idealized smooth muscle whose behavior, like many others, is similar to that of striated muscle. Curve R, resting, shows there is little resistance to stretch until the muscle is stretched past the length at which stimulation results in the maximum increment to force, L_o. Curve T, total force, is essentially the same as the active force, curve A, below L_o. The active force falls to zero at long lengths in a fashion resembling the behavior of striated muscle. At short lengths the active force remains higher, and is still significant at lengths where in striated muscle the myosin filaments begin to meet the Z-line. In **(b)** the relationship for an idealized smooth muscle whose behavior exemplifies many other smooth muscles. The stiffness of the muscle at rest, (curve R), is significant at lengths much below L_o, and at L_o, the total force exerted by the muscle when stimulated (curve T) is only twice the force exerted by the muscle at rest. Significantly for the contractile mechanism, the active force, curve A in both **(a)** and **(b)**, is similar in all smooth muscles.

filaments. In skeletal muscle, the network becomes much more inextensible above the length of the muscle in the body. Although it is more difficult to obtain a completely relaxed smooth muscle preparation, the resting length-tension curves of all muscles seem to be very similar.

Resting Tension and the Parallel Elastic Component

The analytic validity of an abstract parallel elastic component rests on an assumption. On the basis of its presumed separate physical basis, it is ordinarily taken that the resistance to stretch present at rest is still there during activation. In short, it is in parallel with the filaments which generate active force. This assumption is especially attractive since the actin-myosin system has no demonstrable resistance to stretch in skeletal muscle. However, one should keep in mind, for example, that in smooth muscle cells there is an intracellular filament system which runs in parallel with the actin-myosin system, the intermediate filament system composed of an entirely different set of proteins, (vimentin, desmin, etc.), whose mechanical properties are essentially unknown. Moreover, as already mentioned, different smooth muscles have different extracellular volumes and different kinds of filaments between the cells.

Irrespective of the detailed physical basis, it is useful and consistent to propose that mechanically, a smooth muscle is analogous to a simple mechanical device consisting of three components: a contractile apparatus, a parallel elastic component, and a series elastic component. There are two possible arrangements of the three components, given the series component is by definition in series with the contractile apparatus. Although there is no experimental method using external mechanical measurements to distinguish between the two possible arrangements of the three components, in smooth muscles which have no tendons the Maxwell model (Figure 3) has the clear conceptual advantages of (a) identifying the parallel component with the noncontractile filaments inside and outside of the cells, and (b) the total force is the sum of the parallel elastic force and the length-dependent active force. Moreover, the resting length-tension curve completely defines the Maxwell Parallel Elastic Component (PEC). The Voigt components are not so easily handled but would be a better approximation for the case that extracellular structures between the membrane attachments of the actin-myosin system are significantly stretched during contraction.

In order to determine the elasticity of the PEC (using the Maxwell Model), one must establish a resting length-force curve (Figure 2). Several problems emerge. First, comparison of different muscles from equivalent sources requires the use of some kind of normalized length, L_o. Since sarcomeric length cannot be used for smooth muscles, the less precise length at the peak of the active tension curve is usually used. This brings forward the second problem. There are no standard protocols or even criteria to insure that the contractile apparatus of a smooth muscle is not partially turned on. Therefore, various investigators have used different means to inactivate the contractile machinery. While some of these techniques seem to be effective, irreversibility tends to creep in with the efficacy of the inhibitory method. The current situation is, therefore, vague and each circumstance must be

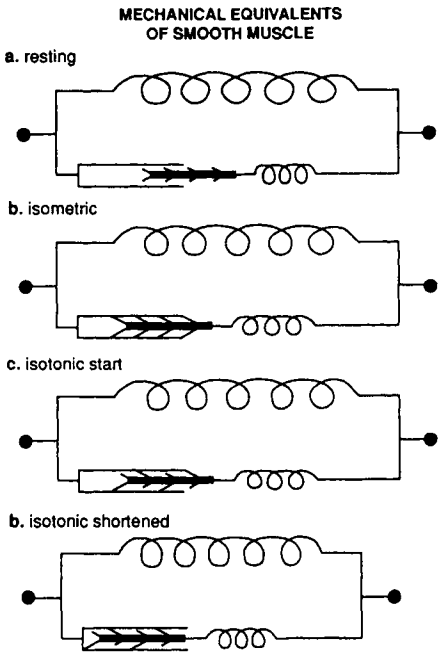


Figure 3. The abstract, Maxwell mechanical equivalent of a smooth muscle is shown in cartoons of four different physiological states. In each case it has two parallel limbs. One limb is a spring, the Parallel Elastic Component (PEC). The other limb is made of the Contractile Apparatus (CA), in series with the Series Elastic Component (SEC). At rest **(a)**, all of the force exerted by the muscle is due to the PEC. The crossbridges are not attached and the SEC is at its zero force length. When, as in **(b)**, a muscle is stimulated but not allowed to shorten, the PEC stays in the same configuration and exerts the same force. However, the CA shortens until the force exerted by the SEC prevents it from any further shortening. The total force in this model is the sum of forces exerted by the two limbs. If, as shown in **(c)**, at the peak of an isometric contraction a muscle is released to a smaller force, the muscle shortens slightly as the springs readjust, and then the crossbridges slide past each other with only as many attached as are necessary to bring the total force up to the load. And **(d)**, as the muscle shortens further, the contractile apparatus eventually reaches a length at which the active force of the contractile apparatus is equal to the correct extending force and shortening stops.

evaluated on its own merits as to whether or not a clear experimental definition of the PEC really exists.

Active Tension

After a stimulus, the force a muscle exerts increases; that is, the force required to prevent a muscle from shortening increases. Thus, the force measured by a

transducer holding the muscle at constant length increases. The time-course measured by such a transducer is the time-course of the force of an isometric contraction following that particular stimulus. Note that if the force opposing shortening is too small, then the muscle will shorten and the force developed by the muscle is compromised by the fact of shortening itself. Defining the force a shortening muscle can maintain is a separate and more complicated issue than finding the determinants of the isometric force. When the active tension is referred to it usually means the tension observed during an isometric contraction above the resting tension.

The so-called great range of shortening so often mentioned in discussions of smooth muscle mechanics is the result of (a) the ability of smooth muscle to develop force (and thus shorten further) over a larger range of lengths below the length of maximum active force development, and (b) using too short a standard length in the absence of the calibrating sarcomere.

The increment to force during activation. Presumably, the structures which maintain the resting force are not those which are activated or even changed following a stimulus. Therefore, activation involves the parallel addition of a nascent force resulting from the emergent interaction of the actin and myosin filaments. Active force is therefore an incremental force added to the resting force. This is a fundamental assumption of almost all analyses of contraction from A.V. Hill onward.

Just as important as the maximum active force a muscle can exert at various lengths, is the rate at which the muscle shortens as a function of the force load, i.e., the force-velocity curve. Both the length-tension curve and the force-velocity curve vary according to the degree of activation of a muscle. The rate at which crossbridges cycle is an inverse function of the load force (Figure 4).

Hill found a rectangular hyperbola fitted this mechanical relationship quite well:

$$(F + a)(v + b) = (F_{max} + a)b = a(v_{max} + b)$$

where F = force, v = shortening velocity and a and b are constants characteristic of a given muscle. In earlier times this relationship was endowed with great theoretical significance because the rate of energy utilization by a muscle is certainly related to the rate of external work being done, ($F \times V$). However, more recent studies have made it necessary to consider the detailed biochemical event, as well as overall energy conservation.

The compliance in series with the active force. Force exerted by the activated elements must be transmitted or borne by whatever structural elements are in series with them. In skeletal muscle there is clearly a tendon in series but not so with smooth muscle. In smooth muscle, the total length of contractile apparatus is broken up into individual cells with intercalating extracellular connective structures. In addition, the portions of the crossbridges in series with the pulling site must also be stretched before force can rise to isometric levels. Taken together, the

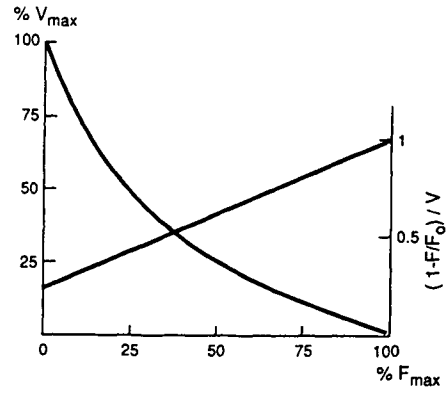


Figure 4. When a muscle contracts isotonically or a constant resisting force is imposed on it during a contraction, the velocity at which it shortens quickly comes to a constant. The force-velocity curve shows the relationship between the force applied to a muscle and the steady-state velocity of shortening. As in all other muscles, the force-velocity curve of smooth muscle is a rectangular hyperbola for all positive shortening velocities. In order to compare the behavior of muscles of different lengths and diameters, it is common to normalize force and velocity by dividing each by its maximum value and expressing the result as a percentage, $\%F_{max}$ and $\%V_{max}$, respectively. That data from a particular muscle conform to the Hill equation is often illustrated by plotting the normalizing function $(1 - F/F_o)/V$ against $\%F_{max}$ to obtain a confirming straight line.

slack or compliance which must be taken up is greater for smooth muscle than for skeletal muscle. A quick release from different isometric tensions to progressively shorter lengths is the most usual observation used in the estimation of the series elastic compliance. Good experimental evidence of an inverse correlation between the strength of a contraction and compliance in this way has been interpreted to mean that both are proportional to the number of crossbridges. This has in turn resulted in a consensus of opinion that the series elastic component resides largely in the crossbridges themselves.

It is implicit in the idea of a series elastic component that there should be a lag between the activation of the contractile apparatus and the rise of force measured between the ends of the muscle. The observed time lag is also commensurate with the idea that the series component is largely due to the crossbridges themselves.

Contractile Apparatus

In order to exert a force from one end of a muscle to the other, a structure must be continuous from one end to the other. In smooth muscle tissue, this structure is a system of alternating myosin and actin filaments within a cell, firm attachments first to the cell membrane, and beyond that extracellular attachments to the

extracellular matrix of collagen and elastin fibers. This is the macromolecular substrate for the abstract force generator and the abstract series elastic component of smooth muscle.

Molecular Structure

The contractile apparatus may be thought of as the sum of those intracellular components which constitute the machinery of chemomechanical transduction. It is the set of proteins which convert the chemical energy of the terminal phosphate ester bond of ATP into mechanical work. The structure of the contractile apparatus is determined by the connections between the various protein molecules via specific binding sites or, in a minority of cases, via labile covalent linkages. The kinetics of the contractile machinery are determined by the regulation of changes in these connections.

Component proteins. The major proteins of the contractile apparatus are the major proteins of the actin and myosin filaments, plus the proteins which anchor the filaments to the membrane. Certain proteins prominent in striated muscle, troponin from the actin filaments and M-proteins from the myosin filaments, are absent from smooth muscle. Myosin, the enzyme which catalyzes the hydrolysis of ATP, is the main part of the minimal contractile machine in all muscle. It is the conveyor of the bond energy of the terminal phosphate ester bond of ATP to the mechanically active system. As mentioned above, myosin of smooth muscle polymerizes into a side polar filament instead of a helical filament organized about a midpoint. Smooth muscle myosin is, also, a hexamer composed of three pairs of subunits. Two identical 200 kDa subunits, the heavy chains, entwine to form a coiled-coil α-helix about 140 nm long. This constitutes the rigid rod-like handle of the golf club. At the head end, there are two 20 kDa regulatory light chains and two 17 kDa essential light chains organized into two equivalent heads. The light chains are more globular than the heavy chains and form the binding sites for ATP and actin. The heads protrude sideways from the shaft of the myosin filaments. The myosin heads are prolate spheroids, 15–20 nm long and 6 nm wide. The heads plus some as yet undetermined length of the heavy chains constitute the crossbridges.

Actin is a 42 kDa bent dumbbell-shaped globular monomer which is found in most eukaryotic cells. It is the primary protein of the thin (or actin) filaments. Also, by mass or molarity, actin is the largest constituent of the contractile apparatus, actually reaching millimolar concentrations. Actins from different sources seem to be more similar than myosins from the same sources. Actin binds ATP which is hydrolyzed to ADP, if the monomeric actin polymerizes. The backbone structure of the actin filament is a helix formed by two linear strands of polymerized actins like two strings of actin beads entwined.

Tropomyosin is a long thin rodlike protein (42×2 nm) of about 30 kDa, composed of two nonidentical subunits. It associates with actin in a ratio of about

7:1. Tropomyosin is thought to lie in the groove formed between the associated actin strands. The sites at which the myosin crossbridges attach are affected by the relationship between tropomyosin and the actin strands. The role of tropomyosin in smooth muscle is completely undefined while in striated muscle it is clearly involved in the activation of contraction. The difference is made clear by the absence from smooth muscle of the protein, troponin, which in striated muscle provides the binding site for the triggering calcium.

Another actin binding protein, the large 100 kDa α-actinin, crosslinks actin filaments together at the dense bodies and near the points of actin filament attachment to the cell membranes. α-actinin is also associated with still another actin binding protein, vinculin, which may stabilize both the Z-line like dense bodies and the membrane attachments.

Finally, a 150 kDa calmodulin binding protein, caldesmon is found bound to actin filaments at low calcium concentrations. The ratio of caldesmon to actin molecules seems to vary from 1:20 in the aorta to as much as 1:200 in chicken gizzard. It has been reported and disputed that caldesmon crosslinks actin filaments into large bundles.

Organization into macromolecular structures. There are no apparent templates necessary for the assembly of muscle filaments. The association of the component proteins *in vitro* is spontaneous, stable, and relatively quick. Filaments will form *in vitro* from the myosins or actins from all three kinds of muscle. Yet *in vitro* smooth muscle myosin filaments are found to be stable only in solutions somewhat different from *in vivo* conditions. The organizing principles which govern the assembly of myosin filaments in smooth muscle are not well understood. It is clear, however, a filament is a sturdy structure and that individual myosin molecules go in and out of filaments whose structure remains in a functional steady-state. As described above, the crossbridges sticking out of one side of a smooth muscle myosin filament are all oriented and presumably all pull on the actin filament in one direction along the filament axis, while on the other side the crossbridges all point and pull in the opposite direction. The complement of minor proteins involved in the structure of the smooth muscle myosin filament is unknown, albeit not the same as that of skeletal muscle since C-protein and M-protein are absent.

The superstructure of smooth muscle actin filaments is differentiated from those of striated muscle by the absence of the troponins and the lateral organization by association of the filaments with dense bodies instead of with the Z-line. How these differences are encoded is again not at all clear. However, the myofibrillar structure and the alignment of the alternating actin and myosin filaments is apparently due primarily to dense bodies and the actin-actinin macrostructures. As the bent dumbbell shaped actins assemble into filaments they are all oriented in the same direction. The S-1 fragments of myosin will bind to actin filaments *in vitro* and in

doing so reflect the polarity of the actin filaments by all pointing toward the same end of the actin filament, like feathers at the end of an arrow.

Biochemical Activities

Although the fundamental chemomechanical transduction processes seem to be the same in all types of vertebrate muscle, contraction in smooth muscle is characterized by much greater involvement of enzymatically catalyzed control reactions. In smooth muscle the control reactions themselves involve the use of phosphorylation-dephosphorylation cycles. Moreover, they are futile in the sense they cause the expenditure of bond energy without a tangible work resultant, i.e., compounds synthesized or external work done.

Enzymatic activities. The hydrolysis of ATP by actin-activated myosin is the characteristic enzymatic activity of muscle, smooth muscle included. All forms of smooth muscle myosin are slower than those of other muscles. The binding site for ATP and a reduced enzymatic activity are still present in monomeric myosin. The enzymatic activity of monomeric myosin is altered by a conformational change, (the 10S–6S transition) and the species of cations present in the reaction mixture. These differences relate to the possible mechanisms of regulation.

Of the several kinase activities which are important in smooth muscle, myosin light chain kinase, MLCK, is the one responsible for activation of the actin-myosin system to *in vivo* levels. MLCK is present in the other nonmuscle cell types which have the actin-myosin contractile system and all of these are probably activated in a manner similar to smooth muscle rather than by way of the Ca^{2+}-troponin mechanism of striated muscle. MLCK from smooth muscle is about 130 kDa and is rather variable in shape. It is present in smooth muscle in 1–4 μM concentrations and binds with an equally high affinity to both myosin and actin. Thus, most MLCK molecules are bound to actin. Myosin light chain serine-19 is the primary target of smooth muscle myosin light chain kinase.

MLCK itself is phosphorylated by cyclic-AMP activated protein kinase, (protein kinase A) and cyclic-GMP activated protein kinase, (protein kinase G). Protein kinase A will phosphorylate MLCK at two sites and protein kinase G at one in some cases and two in others. These differences seem to be important in how the individual smooth muscle cells are regulated.

Finally, if the phosphorylation of myosin is the activation mechanism, then dephosphorylation is likely to be the deactivation mechanism, and so it seems. However, there are several myosin phosphatases in smooth muscle cells and they have some range of substrate specificities. Thus, there are several possible candidates for a regulatory role.

Regulatory mechanisms. There is a large body of evidence which indicates that the contractile apparatus of smooth muscle is turned on as a result of the phospho-

rylation of serine-19 of the regulatory myosin light chain. Smooth muscle myosin is peculiar in that when it is phosphorylated the heavy chains fold and the angle of the heads relative to the heavy chains is increased greatly. This is referred to as the 10s to 6s transition, since the sedimentation constant in a centrifugal force field increases greatly. The fact that the regulatory switches of smooth muscle and striated muscle are on the myosin filament and actin filament, respectively, correlates with the absence of Ca^{2+}-binding troponin from the actin filament of smooth muscle. Similarly, the phosphorylation of myosin from striated muscle by MLCK neither activates it nor uses it to change shape. It is not clear how the bending of myosin following phosphorylation is related to the increased binding of myosin crossbridges to the actin filament.

Regulation of Crossbridge Activity

The smooth muscle cell does not respond in an all-or-none manner, but instead its contractile state is a variable compromise between diverse regulatory influences. While a vertebrate skeletal muscle fiber is at complete rest unless activated by a motor nerve, regulation of the contractile activity of a smooth muscle cell is more complex. First, the smooth muscle cell typically receives input from many different kinds of nerve fibers. The various cell membrane receptors in turn activate different intracellular signal-transduction pathways which may affect (a) membrane channels, and hence, electrical activity; (b) calcium storage or release; or (c) the proteins of the contractile machinery. While each have their own biochemically specific ways, the actual mechanisms are for the most part known only in outline.

From compliance studies, myosin crossbridges at rest are thought to be unattached to actin filaments. The first step then of the activation process is a shift in the likelihood that a crossbridge is attached. However, the total transduction process involves a cycling of crossbridges through a series of states and any of the transitions between states might be regulated.

Phosphorylation of Myosin

Although in *in vivo* circumstances an intracellular free calcium increase apparently acts as the primary modulator of contraction, it can be bypassed in highly permeabilized smooth muscle preparations where the active subunit of MLCK can be introduced to phosphorylate myosin and induce contraction. The MLCK catalyzed phosphorylation of serine-19 is seen as the necessary event in the activation of smooth muscle myosin to form crossbridges. Thus, the rising phase of force during an isometric smooth muscle contraction follows an increase in the degree of phosphorylation of myosin, and that in turn follows the transient rise of (a) cytosolic free Ca^{2+}, (b) Ca-calmodulin complexes, and (c) the active form of MLCK. The regulation of the intracellular calcium is discussed below. The dynam-

ics of the whole myosin phosphorylation-dephosphorylation cycle which occurs during a contraction can be summarized by the following equations. The ATP used in this cycle is thought not to contribute energy to chemomechanical transduction. Therefore, the regulatory myosin phosphorylation cycle is usually thought of as being "futile."

1. $Ca_{SR\ or\ Ex} \Leftrightarrow Ca_{cyto}$
2. Ca_{cyto} + calmodulin $\Leftrightarrow Ca_1$-calmodulin
3. $Ca_{cyto} + Ca_1$-calmodulin $\Leftrightarrow Ca_2$-calmodulin
4. $Ca_{cyto} + Ca_2$-calmodulin $\Leftrightarrow Ca_3$-calmodulin
5. $Ca_{cyto} + Ca_3$-calmodulin $\Leftrightarrow Ca_4$-calmodulin
6. Ca_4-calmodulin + apoMLCK $\Leftrightarrow$ MLCK
7. myosin + ATP $\overset{MLCK}{\Leftrightarrow}$ Pmyosin + ADP
8. Pmyosin $\overset{MLCP}{\Leftrightarrow}$ myosin + P_i

Calmodulin provides the calcium sensitivity to the system. When it binds calcium ion, it becomes an activator of MLCK. It is the regulatory subunit of MLCK holoenzyme. Calmodulin is present in all eukaryotic cells, and is a member of a strongly homologous family of calcium-binding proteins which play regulatory roles in a wide variety of intracellular stimulus response sequences. Calmodulin is the regulatory subunit of a large number of holoenzymes, including certain cyclic nucleotide diesterases, adenylate cyclase, and phosphorylase kinase. Calmodulin plays the central regulatory role in actin-myosin contractile systems everywhere in vertebrates outside striated muscle. Indeed, the presently accepted scheme outlined here was first proposed to explain the behavior of platelet myosin. Calmodulin is a highly conserved, 17 kDa, monomeric protein. It has many similarities including sequence homologies with troponin C. Both of these proteins share the EF hands characteristic of the calcium-binding sites of the family. Calmodulin has four Ca-binding sites, the first two of which have a higher affinity for Ca^{2+} and cause the larger part of the total conformational change. It is not clear how many of these sites must be occupied either to induce the critical shape change, to drastically increase the binding affinity to apoMLCK or to activate MLCK. However, various kinds of evidence indicate that perhaps species with either three or four sites occupied are active. In turn this would imply that there are more than one species of MLCK to be considered. As an overall indicator, the dependence of force on Ca^{2+} concentration can be used to calculate the effective "n" of Equation 3. Calmodulin binds to the apoMLCK 1:1, and is necessary for the activity of the holoenzyme. The fraction of calmodulin which is bound and unavailable for activation of MLCK is controversial. However, the great excess of calmodulin over MLCK, about 40 μM to 4 μM, insures the possibility of saturation of MLCK. A

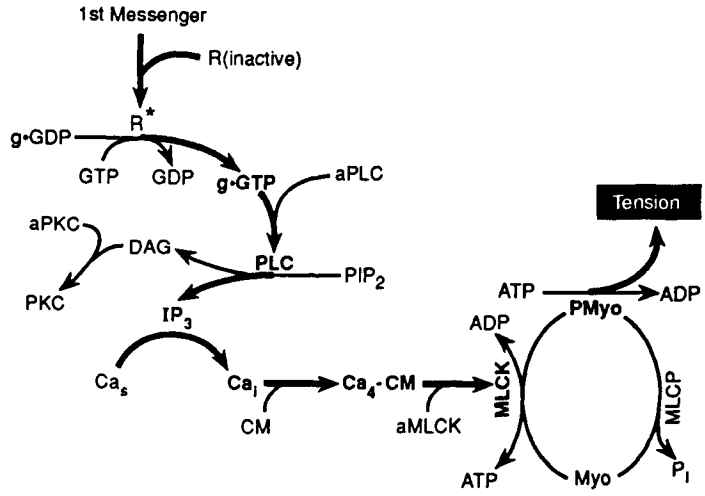


Figure 5. This provides the current view of how smooth muscle is activated by first messengers (transmitters, etc.) via an intracellular, signal-transduction pathway that results in the phosphorylation of the regulatory light chains of myosin. The thickened arrows indicate the hypothetical pathway along which stimulus induced transients of concentration propagate. The first step in this pathway is the activation of receptors by the binding of the appropriate ligand to a site on the receptor. Following the occupation of this site, the receptor changes its affinity for a complementary G-protein. In a cyclic process a series of G-proteins are themselves activated which involves the exchange of GTP for GDP. The α-subunit of the G-protein then acts as an activating subunit for phospholipase C (PLC). PLC catalyzes the hydrolysis of phosphatidyl inositol (1,4) bisphosphate into inositol (1,4,5) trisphosphate (IP_3). Calcium channels in the SR membrane open when IP_3 binds to them. Calcium ions which pour out into the cytoplasm combine with cytoplasmic protein calmodulin thereby forming a complex which is the regulatory subunit of myosin light chain kinase. The phosphorylation of myosin is then accelerated, the chemomechanical processes of contraction are set in motion and the muscle converts ATP bond energy into mechanical work. The degree of contractile activation according to this model is determined by the amount of phosphorylated myosin relative to the total myosin.

large number of compounds complex with calmodulin and interfere with its activation of various enzymes. These include various antipsychotic phenothiazines, e.g., trifluoperazine and sulfonamide derivatives, such as W-7. Trifluoperazine acts allosterically, binding to the helical part of calmodulin. The multiple actions of these inhibitors call attention to the fact that however useful a compound is, there may be sites on other proteins which can bind it. It follows that physiological specificity is relatively rare and use of inhibitors ought to be carefully controlled.

Correlated with the wide distribution of the actin-myosin system found in smooth muscle, MLCK is also found in neural, epithelial, connective, and blood

cells. There may be a number of MLCK isozymes since the reported molecular weight seems to vary widely around 125 kDa according to the source. MLCK has three functionalities: (a) the catalytic site, (b) calmodulin binding site, and (c) cAMP activated protein kinase phosphorylation site. Segments containing these individual sites can be separated following partial trypsin digestion.

There is controversy about the location of MLCK in the cell. The present body of evidence points to the conclusion that most of it is bound to actin filaments. If this is true, then given that there are perhaps 15 myosins for each MLCK molecule in a smooth muscle cell, there is a problem visualizing how more than a small fraction of myosins can be phosphorylated by a tethered MLCK. Obviously, if a significant fraction of MLCK is ultimately found to be soluble, this problem disappears.

Stress

All muscles, including smooth muscles, exhibit the Fenn Effect. That is, when activated muscles are opposed by forces progressively smaller than the isometric force, they hydrolyze larger amounts of ATP and transduce more chemical energy. Surprisingly, as a result they do more external work. This must mean that critical rate constants in the cyclic sequence of events in the crossbridge generation of force are sensitive to the force each crossbridge experiences. As a muscle shortens, sites on the actin filament must be pulled along by crossbridges. If the shortening is greater than the length between crossbridge attachment sites, an attached crossbridge will be pulled backward as a site is passed and will experience a negative force. This may be the mechanism by which the rate of ATP hydrolysis is accelerated. Events at this level are just beginning to be worked out.

Other Mechanisms of Regulation

If MLCK activates contraction by increasing myosin phosphorylation, then an increase in the activity of myosin light chain phosphatase, MLCP, by decreasing the fraction of myosin which is phosphorylated, should lead to relaxation from the active (contractile) state. Cyclic adenosine monophosphate (AMP) is a strong inhibitor of smooth muscle contraction and it has been suggested that activation of MLCP could result from its phosphorylation via cAMP activated protein kinase (see Figure 5).

When MLCK is phosphorylated by cAMP activated protein kinase, it itself is harder to activate. Molecule for molecule, being phosphorylated does not diminish the effectiveness of MLCK in catalyzing the phosphorylation of myosin. However, phosphorylated MLCK has a much smaller affinity for the Ca-calmodulin complex, which activates it, than the uninhibited, nonphosphorylated form. Thus, phosphorylation of MLCK by protein kinase decreases the number of activated MLCK

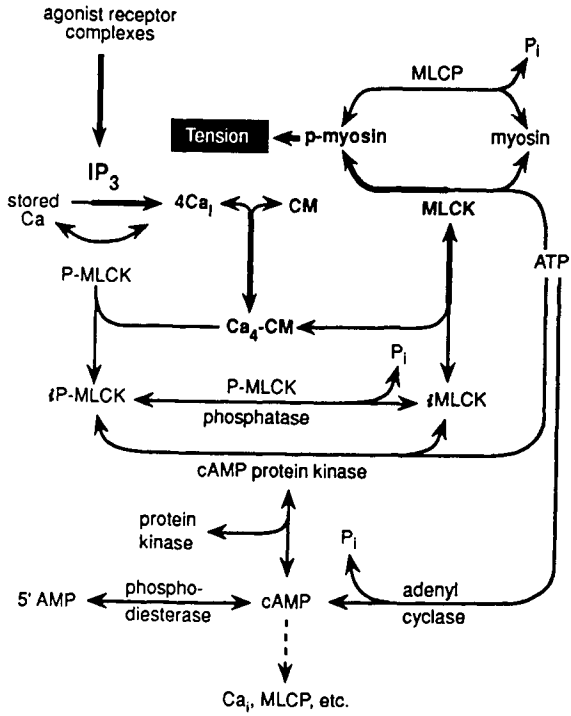


Figure 6. A hypothetical scheme for the control of the number of active crossbridges in smooth muscle. Following the activation of a smooth muscle by an agonist, the concentrations of intermediates along the main route begins to build up transiently. This is shown by the thickened arrows. Also, cAMP is generated which is universally an inhibitor in smooth muscle. Cyclic AMP in turn combines with protein kinase A, which accounts for most of its action. The downstream mechanisms, however, are not well worked out and at least three possibilities are likely in different circumstances. First, protein kinase A is known to catalyze the phosphorylation of MLCK, once phosphorylated MLCK has a relatively lower affinity for Ca-calmodulin so that for a given concentration of Ca-calmodulin, the activation downstream is reduced. The law of mass action predicts that this inhibition should be reversed at high calcium concentrations. Other cAMP inhibitory mechanisms for which there is evidence include interference with the SR Ca^{2+} storage system, and activation of a MLC phosphatase.

molecules, and thereby causes a reduction in the rate of myosin phosphorylation and an attendant decrease in contraction or the contractile response to other excitatory stimuli (Figure 6).

In *in vitro* experiments, caldesmon can inhibit the activation of myosin by actin and this inhibition can be reversed by calmodulin. Thus, there is a potentiality for

the regulation of the myosin-actin interaction by caldesmon. The suggestion has been made on the basis of the *in vitro* studies that as the Ca-calmodulin concentration rises following a stimulus, the caldesmon-Ca-calmodulin complexes which are formed tend to dissociate caldesmons from their actin binding sites, resulting in a disinhibition of the crossbridge-actin system. However, it appears that the calcium levels needed to do this are not often reached *in vivo*. Moreover, there are other problems with this control mechanism, as will be discussed below.

Particularly, because under a variety of circumstances in various smooth muscles, myosin phosphorylation, isometric force, and the rate shortening at zero load do not vary together, a variety of second regulatory mechanisms have been considered by a large number of investigators. In particular, the case often occurs that force remains high after intracellular calcium and myosin phosphorylation have dropped toward resting levels. Unfortunately, although alternative or additional mechanisms of regulation of crossbridge behavior have been frequently suggested, there is presently no clear demonstration of how they act physiologically. One of the most frequently suggested mechanisms is the phosphorylation of myosin by protein kinase C. Phorbol esters elicit contractions, notably persistent ones in various smooth muscles. Phorbol esters activate protein kinase C. Protein kinase C has very broad activity and it has been suggested alternatively that it might (a) activate myosin via the same site as MLCK, (b) initiate a protein kinase cascade which in as yet unknown fashion induces a contraction, or (c) activate the intermediate filament system, which runs in parallel with the actin-myosin system, to bear a portion of the external load. This last possibility seems unlikely since any participant in the ordinary contractile process must be capable of doing external work, even if that only involves stretching the series elastic element. However, there are no known energy releasing reactions which could be activated via the intermediate filament system. Without such a basis, any involvement of the intermediate filament system is not credible.

The possibility of a second site of action of Ca^{2+} has not been excluded to the satisfaction of all. However, there has never been a convincing demonstration of the validity of a second mechanism either. Some of the recurrent suggestions include regulation via binding of Ca^{2+} ion directly to myosin and regulation via proteins that in turn bind to the actin filament. Evidence that *in vitro* the affinity of myosin for Ca^{2+} ion is too low and skinned muscle preparations lack sensitivity to Ca^{2+} ion independent of phosphorylation, has mitigated against these alternatives. Since there is tropomyosin but not troponin in smooth muscle, it has often been suggested that some other actin filament based control mechanism might exist which utilizes other proteins that bind to actin. Two proteins have been found which bind to actin and inhibit actin-activated P-myosin ATPase activity. The more popular candidate for various mechanisms, caldesmon, is a 87 kDa protein which can be phosphorylated. Thus, there is the possibility that phosphorylated caldesmon has a smaller affinity for actin, and hence, its inhibition of actin-activated myosin

may be relieved by it being phosphorylated. Phorbol esters can induce the phosphorylation of caldesmon.

It has also been proposed that, since Ca-calmodulin combines with caldesmon and the complex has little affinity for actin, this may also contribute to the activation process. The notion here is that, in parallel with the activation of MLCK disinhibition, regulation via caldesmon may occur. Unfortunately, such a mechanism would work only at cytosolic Ca^{2+} levels too high in respect of typical contraction.

One should note overall, that while some of these suggested mechanisms may in the future prove to have a role in the control of smooth muscle contraction, in chemically skinned preparations maximum force development follows activation by the MLCK active subunit in extremely low Ca^{2+} ion concentrations. The conclusion can hardly be avoided that phosphorylation alone is sufficient for activation, and if another mechanism is involved, it is not necessary for the initial genesis of force. If such mechanisms are operative, then they might be expected to run in parallel or consequent to myosin phosphorylation. A possible example of this category of effect is that a GTP-dependent process (G-protein) shifts the force vs. Ca^{2+} ion concentration relationship to lower Ca^{2+} ion concentrations. This kind of mechanism calls attention to the divergence of signals along the intracellular control pathways.

Kinetic Models of Crossbridge Regulation

For the purpose of discussion, crossbridge regulation can be split into three overlapping sets of reactions: (a) the Ca-calmodulin cascade (MLCK activation), (b) the phosphorylation-dephosphorylation cycle (the Four State Model), and (c) actin-myosin cycle (chemomechanical transduction).

Kinetics of activation. The chemical stoichiometric equations of the Ca-calmodulin cascade (Equations 1–8, p. 173) correspond to an equal number of differential equations, as well as reduce to algebraic equations at steady states including equilibrium. Thus, transient as well as steady-state behavior is implicitly predicted by such sets of chemical equations. Even if an arbitrary set of equations can not be solved analytically, the behavior of the system and the solutions of the differential equations can be approximated using computer techniques. This procedure is often called modeling and involves cyclical evaluations and adjustments quite similar to those of experimental studies. Modeling requires more precise meanings to hypotheses and provides more precise predictions of experimental outcomes than proposing cartoons or hypothetical pathways. Its power comes from the quantitative testing it allows of theoretical statements.

In a smooth muscle cell under steady state conditions, the reactions described by the set of Equations 2–6 are probably very close to equilibrium in the sense that they are not continuously held far away from equilibrium by ongoing metabolic reactions, e.g., ATP hydrolysis. The resting concentrations of the reactants of these

reactions will be determined by their equilibrium constants and the concentrations of cytosolic Ca^{2+} ion, ATP, ADP, and P_i. The cell membrane Ca pump-leak system, Equation 1, determines the steady-state cytosolic Ca^{2+} concentration. In the steady-state of the futile cycle comprising the reactions of Equations 7 and 8, the ratio of Pmyosin:myosin is proportional to the ratio of the enzymatic activities MLCK:MLCP. In many smooth muscle cells, in what passes for the resting state, the level of cytosolic Ca^{2+} ion is high enough for a significant amount of MLCK to be active, and an important fraction of crossbridges to be activated. Consonant with that, inhibitors of MLCP (Equation 8) cause many smooth muscles to contract at physiologically meaningful rates.

There is nothing in Equations 1–8 which is an all-or-none situation. There are no positive feedback loops which might cause some kind of flip-flop of states of operation of the system. There are some possibilities for saturation phenomena but all relationships are graded. Overall, transient or steady-state, the changes of concentration of P-myosin are continuous, monotonic functions of the intracellular Ca^{2+} ion concentration. On this basis it is more appropriate to say that smooth muscle contraction is modulated rather than triggered by Ca^{2+} ion.

The binding reactions in the Ca-calmodulin cascade leading to the activation of MLCK, Equations 2–6, are coupled by shared intermediates. However, the P-myosin and myosin concentrations have no backward influence along the cascade. They are decoupled by the fact that the linkage between the futile phosphorylation-dephosphorylation cycle and the Ca-calmodulin cascade involves no more than an enzyme, MLCK. This conclusion from the model depends on the validity of the hidden assumption that the fraction of the enzyme in the substrate complex is small. Experimentally, this simplification seems to be justified and the cascade appears to act as a valve to regulate the rate of flow (phosphorylation) from the myosin into the P-myosin pool. If the MLCP concentration is unchanged, the system will begin to evolve toward a new and higher ratio of P-myosin:myosin. If in an experimental situation the reactions of the cascade are fast enough to approach equilibrium before those of the P-myosin:myosin cycle proceed very far, then the P-myosin levels (isometric force) will tend to follow the Ca^{2+} ion concentration time-course with a single exponential time course. This is actually seen in many smooth muscles and is particularly striking for relaxation curves which are often seen to be superimposable. In the light-sensitive sphincter pupillae where the stimulus can be turned on and off very quickly, Equations 1–8 are sufficient to predict the time-courses of force during isometric responses over a wide range of stimulus strength and duration.

Phosphorylation-dephosphorylation. The site on myosin which is phosphorylated is not the same as the site by which it attaches to actin. Therefore, there are two geometrically separate reactions in regulation and from the Law of Reversibility there must be at least some myosin molecules in at least four different states:

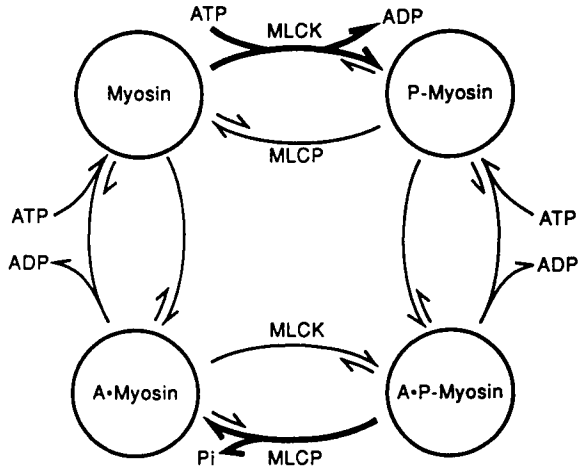


Figure 7. This figure depicts the eight overall reactions that are involved in a Four State Model of phosphorylation regulation of smooth muscle contraction. In a simple scenario, most of the myosin in a myocyte completely at rest is neither phosphorylated nor attached. As MLCK is activated via the Ca·calmodulin cascade, myosin is phosphorylated and the P-myosin state would begin to fill. Phosphorylated crossbridges would begin to attach and cycling around the chemomechanical cycle, between P-myosin and A·P-myosin, would occur. As the number of phosphorylated crossbridges increases, the rate of formation of unphosphorylated crossbridges also increases. The resistance to stretch of a muscle would seem to be dependent on the number of attached crossbridges, averaging the different degrees of stretch of the crossbridges attached at an instant.

A·M, A·P-M, M, and P-M (Figure 7). The following kinetic scheme describes the possible transitions.

Since traversing the cycle in a clockwise direction results in the hydrolysis of ATP, there always is a clockwise flow of myosin molecules driven by the energy of the high cytosolic ATP concentration. What changes during activation is the fraction of myosin molecules (crossbridges) in each of the four states. There is no evidence that the energy of hydrolysis of ATP in this cycle is transduced into mechanical energy, so this is also a futile cycle. Although the event regulating the total activity of myosin molecules is a change of the fraction of myosin molecules phosphorylated, none of the four states in the scheme should ever be thought to be completely empty. This is especially true for those smooth muscles which have a significant active component of tension at rest.

The existence of two attached states, A·M and A·P-M, has provided the basis for a family of attractive hypotheses to explain the separation of the maintenance of maximum force and maximum shortening velocity. As in other systems, continuous occupation of the receptor sites produces only a transient elevation of the

intracellular calcium in certain vascular smooth muscle cells. Experimentally, as Murphy's group first found, if concentrations of first messengers are maintained at elevated levels for periods in the range of minutes, the intracellular concentration of calcium ion falls. The level of phosphorylation also falls, following the predictions of Equations 1–8. However, surprisingly, the force observed isometrically does not fall proportionate to phosphorylation, while the force-velocity curves determined by release experiments shift toward lower velocities. In particular, the maximum (unloaded) shortening velocity apparently correlated with the degree of phosphorylation of myosin. Moreover, in a variety of situations the maximum force is generated if only a small fraction of the myosin is phosphorylated. If the force a muscle bears is proportional to the number of bridges attached, it would then appear that the fraction of myosin molecules in the attached states is not a simple function of myosin phosphorylation. The first hypothesis to incorporate the four state schema as an explanation of this was the "Latch Bridge Hypothesis" developed by Murphy and his collaborators. The essence of this hypothesis is that the unphosphorylated myosin crossbridge attachment rate is zero and that the detachment rate is very low (Figure 8). The consequence of this is seen to be an accumulation of crossbridges in the "Latch State" which dissipates slowly even if MLCK activation has decayed. However, there are problems with the concept of a crossbridge which does not go through the chemomechanical cycle. These include the experimentally unfulfilled prediction of an internal mechanical load caused by latch bridges which would have to be broken for shortening to occur. Also, the force exerted by nonphosphorylated crossbridges which can not reattach should only decay, and not contribute to the redevelopment of tension after a release, as they seem to do.

Turnover experiments, in particular, have led to an alternative, the "Cooperative Four State Hypothesis." The notion is that the rate constants in the chemomechanical cycle are the same for both kinds of crossbridge but values of the rate constants increase with the fraction of the myosin phosphorylated. Part of the basis for this view is the observation that 100% of the exchangeable ADP probe prebound to myosin in skinned fibers is released by less than 20% thiophosphorylation of the myosin. The conclusion drawn here is that a small fraction of crossbridges phosphorylated can cause all crossbridges to turn over in a few seconds. Thus, it appears that all the crossbridges are activated to go through the attachment process, and presumably could bear a load even though only a small fraction of the crossbridges are phosphorylated. However, the experimentally measured rate of ATP hydrolysis is proportional to the fraction of myosin thiophosphorylated. Thus, the rate at which the crossbridges turn over is increased by phosphorylation of more crossbridges. The rate of crossbridge turnover and ATPase hydrolysis would be expected to correlate with the maximum velocity of shortening.

The "Cooperative Four State Hypothesis" makes the separation of the time-courses of isometric force and maximum shortening velocity easier to understand. Moreover, it avoids the faulty predictions of the Latch Hypothesis described above.

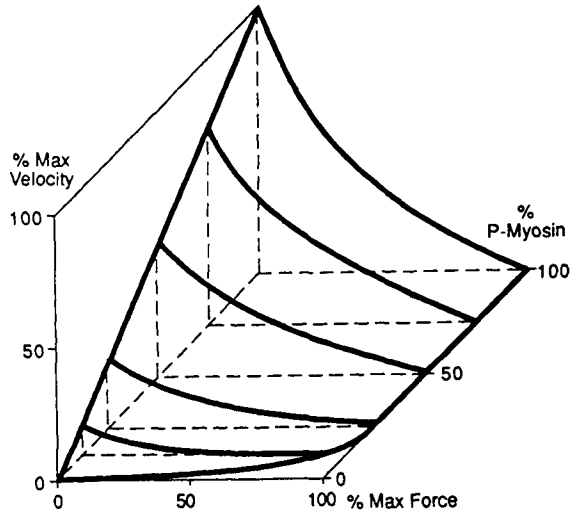


Figure 8. For any set of conditions, the greatest velocity that a muscle can shorten is attained when the total force opposing shortening is zero. Empirically, the maximum velocity of shortening increases with the degree of phosphorylation of myosin. This is seen as the straight line in the velocity-phosphorylation plane. The maximum force that a smooth muscle can develop is not increased by phosphorylation beyond about 25% phosphorylation. It seems therefore that past a point, phosphorylation regulates the rate at which work is being done rather than the force that can be developed. The force a muscle can develop if 25% myosin is phosphorylated is maximal and saturated; however, the rate of doing work is not saturated and continues to increase with further phosphorylation.

For these reasons the Cooperative Hypothesis seems to be a plausible alternative to the Latch Bridge Hypothesis. Given that computer calculations of the behavior of various hypothetical schemes are now possible, an independent, noninvasive measure of distribution of myosin among the states would be of great use for further understanding of crossbridge kinetics.

The genesis of force and the movement of myosin heads along the actin filament. The molecules involved in the transduction of ATP bond energy into mechanical work in smooth muscle are very similar to their equivalents in striated muscle. Actins from striated and smooth muscles can be used interchangeably in a variety of hybrid reduced systems, from actomyosin ATPases to chimeric filament systems. Myosins from both smooth and striated muscle have the same number of subunits of about the same molecular weight. Even though the organization into filaments is different, it seems that the structure of the crossbridges themselves, the catalytic sites, the actin binding sites, etc., may also be quite similar. X-ray

diffraction studies indicate that the crossbridges are the same distance apart on the myosin filaments in both muscle types. The states that myosin go through and the intermediates which are formed during the transduction cycle are taken to be the same.

The mechanical behavior of the contractile apparatus of smooth muscle is also very similar to that of striated muscle. So that to the extent that the force-velocity curves reflect the interaction of mechanical force and the rate of enzymatic catalysis, the steps of the chemomechanical transduction cycles in the two muscles are apparently modulated in similar ways. Also relationships between the active isometric force and muscle length are very similar (except as noted above for shorter lengths).

The fundamental event for work to be done by muscle must be an increase in the force exerted by an attached crossbridge, concomitant with the advent of sliding of the filaments past each other. Taking the individual steps in the process to be the same in both smooth and striated muscle, by analogy the sequence can be summarized by Figure 9. Combining this scheme from skeletal muscle with the Phosphorylation Theory, the resting smooth muscle may be expected to be characterized by having most of its myosin molecules in the M-ADP·P_i state. During activation, MLCK catalyzes the phosphorylation of myosin molecules which are thereby converted into a state, $_pM$-ADP·P_i, that attaches at a significant rate. Once attached, the crossbridge, A·$_pM$·ADP·P_i, is most likely to follow the pathway of descending free energy along which the conformation of myosin in the next state, A·$_pM$·ADP, is changed (45°→90°), and then force is developed. As force is developed, a crossbridge remains attached. Indeed, the binding is said to go from weak to strong. As the filaments move relative to each other, work is done, and the increased strain is relieved. Then, ADP is released. The empty nucleotide site on the crossbridge presumably increases the likelihood of ATP binding. Binding of ATP, in turn, results in a 90°→45° conformation change, and consequently, detachment of the crossbridge from the actin filament. The actual catalytic step, the hydrolysis of the ATP to ADP and P_i, is considered to occur at this point, where the myosin is not attached to actin. However, the hydrolytic products, ADP and P_i, are only very slowly released and most of the myosin molecules in the resting muscle pile up in this state. The activation event which also allows crossbridges to bind to actin accelerates the rate of debinding of first P_i, and then ADP. These characteristics which are common to most present day analyses of the chemomechanical reaction sequence are based on rapid reaction rate measurements of isolated segments of the sequence.

Regulation of Intracellular Free Calcium Concentration (an n-Compartment System)

As discussed above, for all smooth muscle cells, the terminal sequence of the signal transduction pathways which regulates contraction seems to be the same. As

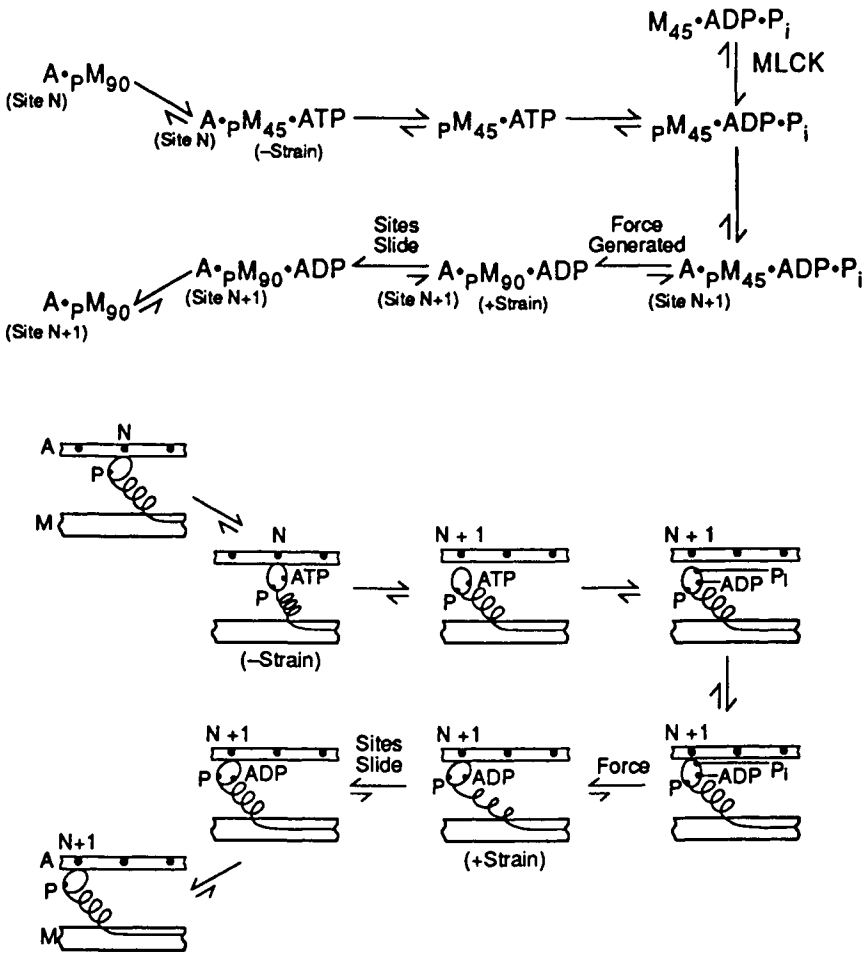


Figure 9. The chemomechanical cycle of the actin-myosin systems of muscle seem to have in common a set of steps which result in the conversion of chemical energy into mechanical energy. The cycle is both a chemical cycle and a mechanical one. As an arbitrary the first step, the effect of binding ATP is shown to place a negative strain on the low energy attached cross bridge. Dissociation of actin and myosin is the result and the essential hydrolysis of ATP follows. However, the hydrolytic products do not separate immediately but instead there is a reassociation of actin and myosin first. After the crossbridge reattaches the inorganic phosphate product is released with yet another attendant conformational change. At this point, a force develops between the two filaments. Following that, as the filaments slide past each other, any strain will be relieved. Any crossbridge will then experience a negative strain if its ATP binding site is occupied or if sliding goes far enough.

activating perturbations pass along various smooth muscle cells' control pathways, the first event in the common terminal sequence of events is a rise in intracellular free calcium. This is the reason that a rise in intracellular free calcium is often referred to as the trigger of contraction. Physiologically, a rise in intracellular free calcium is a necessary event to elicit contraction. There are three possible sources of trigger calcium following a stimulus. These are: (a) extracellular space which has a very high calcium concentration, (b) the so-called sarcoplasmic reticulum which has a very high apparent calcium concentration and a large buffered reserve, and (c) the mitochondria which can have a very large amount of calcium associated with them. Any inflow of Ca^{2+} ion into the cytoplasm from these sources would be expected to trigger contraction. However, of these three, only the first two seem to play significant physiological roles. Despite much research effort there are very few data which indicate that mitochondrial calcium is released in normal physiological states. The reason is simply that the concentration range at which mitochondria load and unload is much higher than that which triggers contraction.

Transport Across the Plasma Membrane

The steady-state balance of the Ca^{2+} pump and plasma membrane leaks of Ca^{2+} determines the resting intracellular free Ca^{2+} concentration. Kinetically, all the other membrane bound compartments and their transport processes are analogous to buffer systems with various rates of binding and release. The essential point is that all the other pools must come to steady-state with the intracellular free Ca^{2+} concentration. Thus, the plasma membrane Ca^{2+}-pump for the Ca^{2+} economy of the cell has primacy.

The plasma membrane Ca^{2+}-pump. During activity of the plasmalemmal Ca^{2+}-pump, the energy of hydrolysis of the terminal phosphate of ATP is coupled to Ca^{2+} extrusion via a protein phosphorylation step. This classifies it as a P-type cation pump. The pump seems to consist of a single large, 130 kDa protein with 10 transmembrane helices. It is activated by Ca-calmodulin and perhaps regulated by acidic phospholipids as well. Ca^{2+} is moved out of the cytosol against a concentration of perhaps 10,000 times by the Ca^{2+}-pump. The calcium ion equilibrium potential is therefore very large and positive, given the usual conventions.

$$E_{Ca} = RT/2F \ln(Ca_o/Ca_i) \approx 120mV$$

The movement of Ca^{2+} against its energy gradient to establish the concentration gradient is thermodynamically paid for by coupling to the bond energy of the terminal phosphate of ATP, as is true in so many other cases. The stoichiometry of the plasmalemmal pump reaction is not agreed upon. Therefore, we write simply:

$$2Ca_i + ATP \Leftrightarrow 2Ca_o + ADP + P_i$$

to indicate that at least the ratio of Ca transported to ATP hydrolyzed is known, even if any cotransport is not completely ruled out.

In any case, one might expect that the rate of the Ca^{2+}-pump is rather slow relative to the processes unleashed by muscle activation. This is certainly true for the Na^{+}-K^{+}-pump, and if it is true for the Ca^{2+}-pump as well, then the significance of the amount of charge which may or may not be transported concomitantly with Ca^{2+} becomes rather unimportant. In short, the rate at which the pump works is more important for contraction than the detailed energetics of the transport reaction. How much of the variability of smooth muscle responsivity is in fact determined by transients in the Ca^{2+} distribution, e.g., in smooth muscle myocytes, remains to be discovered.

Plasma membrane Ca^{2+} channels. The most common mechanism for the movement of Ca^{2+} into smooth muscle cells from the extracellular space is the electrodiffusion of Ca^{2+} ions through highly selective channels. This movement can be significant in two quite different ways. First, Ca^{2+} ions carry two positive charges and, in fact, most of the inward charge movement across the plasma membrane of smooth muscle myocytes is carried by Ca^{2+}. Most smooth muscle action potentials are known to be Ca^{2+} action potentials. And second, the concentration of intracellular free calcium, the second messenger, is increased by inward calcium movement.

Calcium channels are members of the large family of proteins, including Na^{+} and K^{+} channels, which become incorporated into plasma membranes, and which form intermittent aqueous pathways through which ions can move. The channels open and close. As is the case generally for membrane spanning proteins, a Ca^{2+} channel is formed by a set of helical units, in this case seven, which associate to form the channel.

Voltage gating. In smooth muscle only two types of transmembrane voltage regulated Ca^{2+} channels have been shown to play a significant role in the variation of either the transmembrane voltage or the intracellular Ca^{2+} concentration. As is now the case for all other voltage sensitive channels, voltage gating of Ca^{2+} channels is studied primarily using patch voltage clamp techniques. The essence of the technique is to seal to the membrane a small pipette over a single channel, and while holding the transmembrane voltage constant, measure the current flows to indicate the conductance of the open channels.

The most common type of Ca^{2+} channel in smooth muscle is the L-channel. It belongs, as do all known Ca^{2+} channels, to the same family of channels as the Na channel of nerve and skeletal muscle, and has a conductance of about 20pS. A cardinal characteristic of the L-channel is its affinity for and sensitivity to a number of dihydropyridine compounds. These drugs can affect the probability of L-channels being open. The inhibitory members (e.g., nifedipine) of this group of drugs decrease the L-channel open-time probability, while the excitatory members, (e.g., BAY-K 8644), increase the open-time probability. The implication here is that the

DHP drugs bind allosterically. The open L-channel is somehat more permeable to the Ba^{2+} ion than to the Ca^{2+} ion but is very much less permeable to the Na^{+} ion. Nonetheless, because Na^{+} ion concentrations are so much higher than Ca^{2+} ion concentrations, the actual fraction of charge carried by the two ions is not always so clear. There are a number of states that the L-channel can be in, aside from simply being open or closed. It is the distribution of L-channel molecules among the various states that is influenced by transmembrane voltage. From another view the rate constants between the states are functions of the transmembrane voltage.

In the steady-state, the L-channel is closed at membrane potentials close to the resting potential. It does not open significantly unless the transmembrane voltage is displaced quite far towards zero. Thus, it requires a large voltage change to be activated, i.e., the L-channel is a high threshold channel. On the other hand, after being opened the L-channel is slow to close. The rate of closure is greatly increased by the intracellular free Ca^{2+} ion concentration. Thus, we may say that, in the ordinary course of events, the L-channel is opened by voltage and closed by a rise in the intracellular free Ca^{2+} ion concentration.

Many smooth muscles will develop force proportionate to the frequency of firing of action potentials. It is thought by many investigators that in those smooth muscle cells that are physiologically triggered to contract by action potentials enough Ca^{2+} enters via the L-channels to elicit contraction. On the other hand, there are a number of investigators who think that Ca^{2+} ions which enter through channels, or some other aspect of the action potential, cause the release of Ca^{2+} ions from the sarcoplasmic reticulum—the so-called Ca^{2+} induced Ca^{2+} release (CICR). The observation that a rise in Ca^{2+} ion concentration itself can cause Ca^{2+} release from the sarcoplaslmic reticulum (SR) has been made in other cell types in addition to smooth muscle. In fact, there is very little direct evidence which bears on the question of whether or not action potentials trigger the release of Ca^{2+} from the SR of smooth muscles, by any mechanism. That is to say, while it is very clear that in many smooth muscles contractile activity increases with the frequency of action potentials in the smooth muscle cells themselves, the source of the triggering Ca^{2+} ion is not so clear.

The generation of an action potential itself may require a less Ca ion movement into the cytosol than does the triggering of contraction; to initiate an action potential certainly requires less. It is thus not difficult to suppose that the generation of an action potential is sufficiently triggered by quick and low threshold channels which carry some small threshold inward current. Since the L-channel is slow and turned on only by big potential changes, it seems that in smooth muscle there is another kind of Ca^{2+} ion channel, the T-channel, which may act as the fast, sensitive, transient factor needed to initiate an action potential. Knowing that T-channels turn on closer to the resting potential, and hence early in the time course of an action potential, is not sufficient to implicate them as the triggers. The product of the individual channel conductances multiplied by the density of the channels in the

membrane determines the conductance of a particular set of channels. For the T-channels the product is small and it is still problematic how much of a role the T-channel plays in any particular cell. In the general case, the T-channel may provide a necessary instability to the smooth muscle membrane potential for the explanation of the spontaneous and elicited generation of action potentials. That is very much a speculation. However, at the present time, addition of these kinds of channels to typical ion channel models has not been shown to be enough to provide an explanation of the frequency of firing of action potentials in smooth muscle cells.

Ligand regulation. There are no clear cases for smooth muscle where a first or a second messenger molecule binding to a Ca^{2+} channel of any type causes an activation (opening) of the channel or a shift of the voltage sensitivity. However, these remain as viable possible modes of regulation.

Covalent regulation. Following occupation and activation of the M_2 acetyl choline receptors, phospholipase C (PLC), is activated and both inositol (1,4,5)-trisphosphate (IP_3), and diacylglycerol (DAG), are formed by hydrolysis of phosphatidylinositol (4,5)-bisphosphate (PIP_2).

$$PIP_2 \overset{PLC}{\Leftrightarrow} IP_3 + DAG$$

DAG is a potent activator of protein kinase C (PKC). It has been long known that ACh causes an increase in the number of effective L-channels in many smooth muscle cells. This effect is mimicked by phorbol esters (known to activate PKC) and DAG itself. Therefore, it has been suggested that one of the actions of ACh involves the activation of L-channels via the evoked increase of DAG.

The Na^+: Ca^{2+} exchanger. There are in the plasma membrane of many cells, including smooth muscle cells, proteins which facilitate the coupled movement of three Na^+ ions in one direction across the cell membrane in exchange for one Ca^{2+} ion in the opposite direction. A similar exchange process occurs in the retinal rod except that there the ions exchanged are four sodium ions for one calcium ion and one potassium ion. While the ratio may be different in different tissues, the essential point is that the movement of these ions is fixed by a coupling relationship. Since the energy of this process does not involve coupling to ATP hydrolysis, and is otherwise entirely dependent on the gradients of Ca^{2+}, Na^+, and electrical force, many investigators classify the Na^+:Ca^{2+} exchanger as an entirely passive mechanism. In the sense that for a particular Na^+ ion gradient, Ca^+ ion gradient, and membrane potential, the direction of operation of the exchange is fixed, the system is passive (dissipative); no bond energy is involved. In another sense, the exchange runs by virtue of the Na^+ ion gradient established by the plasmalemmal Na^+:K^+ pump and can act as a Ca^{2+} pump or vice versa. Indeed, the evidence seems convincing that the Na^+:Ca^{2+} exchanger when operating in the reverse mode can move enough Ca^{2+} ions inward to cause a significant increase in force generation.

Depolarization of the cell membrane, for example, favors such events. It should also be noted that any mechanism which affects the action of the Na^+:K^+ pump and the Na transmembrane gradient will also affect the Na^+:Ca^{2+} exchanger flux.

Transport Across the Sarcoplasmic Reticulum Membrane

In the sarcoplasm of smooth muscle cells there is a membrane bound compartment usually referred to as the SR by analogy with skeletal muscle. However, it is not at all clear that the interior of these membrane-bound regions are continuous as they are in skeletal muscle. The primary properties of this system seem to be quite similar to those of the endoplasmic reticulum of many other cell types. In general, calcium is concentrated into the membrane-bound reticulum and then released to initiate the characteristic action of the cell.

The SR membrane Ca^{2+}-pump. The SR Ca^{2+}-pump is another member of the family of P-type cation pumps. It is, therefore, an intrinsic membrane protein which has the properties of both a channel and an enzyme. The combination of these properties, as in other pumps, results in the coupling of ion transport against an electrochemical gradient at the expense of ATP bond energy. The detailed energetics of this overall reaction are not yet known. Indeed, even the net obligatory movement of ions is not clear. However, it does seem clear that the following stoichiometric relationship for ATP and Ca^{2+} is good for a first approximation.

$$2Ca_{cyto} + ATP \Leftrightarrow 2Ca_{SR\text{-}Lumen} + ADP + P_i$$

The absence of any indication of a counter ion movement calls attention to the likelihood that there is no SR transmembrane voltage difference.

Sarcoplasmic reticulum Ca^{2+}-channels. In many smooth muscle cells the rise of intracellular calcium which triggers contraction comes from the flow of calcium from the SR through Ca^{2+} channels. In others, the SR contributes some unknown fraction of the triggering calcium relative to the amount which comes from the extracellular space through the plasma membrane Ca^{2+}-channels. There are at least two kinds of Ca^{2+}-channels in the SR.

Voltage-gated SR Ca^{2+}-channels. Although action potentials trigger contractions in smooth muscle, it is not clear that they elicit a significant flux of Ca^{2+} from the SR via a plasma membrane voltage mechanism or, in fact, by any other means. It may be that in those smooth muscle cells that contract in proportion to the frequency of action potentials, all the triggering calcium may come from the extracellular space via plasma membrane calcium channels. In short, the case for or against the sufficiency of the plasma membrane fluxes to explain such contractions is unresolved. The scanty evidence available at present does not bear significantly on the question of whether there are SR channels which are activated by voltage.

Ligand-gated SR Ca-channels. On the other hand, there are a number of small molecules which are clearly ligands to sites on the SR membrane capable of influencing the release of calcium from the SR lumen. Ryanodine and caffeine are two pharmacological agents which are important in terms defining molecular substrate for release. Inositol triphosphate, cyclic adenosine diphosphate ribose, and calcium itself are candidates for actual physiological roles in the activation of smooth muscle cells.

Pharmacological intervention. Caffeine is a plant alkaloid, a member of the methylxanthine family and long known as a central nervous system stimulant and a vascular smooth muscle relaxant. In skinned smooth muscle preparations caffeine elicits a contraction which can be shown to result from release of calcium from the SR. It is characteristic of caffeine that a supramaximal dose seemingly exhausts the SR calcium so that subsequent applications of caffeine are ineffective. Caffeine apparently opens SR Ca^{2+} channels by increasing the opening likelihood. SR calcium is thereby released. Caffeine stimulates on first application but inhibits in the long term. This action is apparently independent of its inhibition of phosphodiesterases, which of course would increase the myoplasmic concentrations of the inhibitory cyclic nucleotides, cAMP and cGMP.

Ryanodine, another plant alkaloid, can also open SR channels and thereby induce contraction. Ryanodine differs from caffeine in that it seems to have a larger effect on the closure rate constants than on opening rate constants. Thus it is often said in the literature that caffeine can open channels but ryanodine merely keeps them open. The actions of caffeine and ryanodine are mutually competitive (occlusive). However, they are experimentally completely independent of the effects of IP_3.

Physiological release of Ca^{2+} into the cytosol. Neurotransmitters induce contractions via signal-transduction cascades, independent of any change in the membrane potential. Indeed, they can cause smooth muscles to contract, even those bathed in strongly depolarizing solutions, e.g., isotonic K_2SO_4. The reason this occurs is that the occupied receptor proteins are also activators of G-proteins which in turn activate phospholipase C, which catalyzes the hydrolysis of phosphatidylinositol (4,5)-bisphosphate (PIP_2) to yield IP_3 and DAG, both of which in turn act as second messengers. In a large number of cell types, including smooth muscle, IP_3 binds to a site on a protein channel on the SR, the IP_3-gated channel, and causes it to greatly increase its open time. These actions have been most easily studied in skinned fiber preparations where the *in vitro* bathing solutions are coextensive with the cytoplasmic compartment.

However, the total regulatory system is not so simple and linear. In skinned muscle preparations especially, it can be shown that there are calcium stores which cannot be released by IP_3 but which are released by elevated levels of calcium itself. That is, by the mechanism of calcium induced calcium release (CICR). The CICR

channel is specifically affected by ryanodine and caffeine. Moreover it is now known that cyclic adenosine diphosphate ribose, an NAD metabolite, can apparently release calcium via the same channel. Since each of the two release mechanisms can only operate on a fraction of the total SR calcium, independently, it is thought by most investigators that there are two membrane-bound regions, one whose membrane has IP_3 receptor activated channels, and the other whose membrane contains CICR (Ca, ryanodine, caffeine, cADP ribose activated channels). It is not entirely clear whether or not some fraction of the smooth muscle SR has both kinds of channels in its membrane, thereby making the SR membrane system tripartite. At the present time, Ca^{2+} ion itself is the best candidate to be the ligand which physiologically activates the CIRC mechanism. In addition, perhaps, cADPR or other modulators, set the sensitivity of the system to Ca^{2+} ion. If that is the case, then the CICR mechanism is essentially piggybacking on either plasma membrane channels or the IP_3 cascade mechanism, and can be seen as a kind of variable gain device for those activation pathways.

Second Messengers: Their Production and Their Action

The sensitivity of cells to perturbations of their environments depends on the transduction of external variations in concentration or physical variables, such as light or voltage, into changes of intracellular concentration of modulator substances, viz. second messengers. In smooth muscle, the prominent second messengers include Ca^{2+} ion, cyclic nucleotides, phospholipid hydrolysis products, and activated kinases. A family of membrane-bound proteins, the G-proteins, play important roles in transduction processes, especially those which involve small molecular weight second messengers. The G-proteins are trimeric and possess larger, more specific α-units, and smaller less specific β and γ subunits. These subunits act as regulatory subunits of various channels and enzymes. Similarly, the G-proteins themselves are activated by complexing with activated membrane receptor proteins (Figure 10).

In general, the receptor·G-proteins complexes exchange bound GDP for GTP. In turn, the two, smaller subunits of the G-protein components of these complexes are released and the receptor protein dissociates. The remaining G-protein·GTP complex then complexes with and activates a specific enzyme. It is very significant to note that G-proteins therefore have at least three specific binding sites: (a) for nucleotides, (b) for a receptor protein, and (c) an effector protein.

Inositol trisphosphate; Receptor/G-protein cascades. As discussed above, IP_3 is one of the products of the hydrolysis of PIP_2. To say that it acts as a second messenger means that a rise in its concentration occurs as a result of some meaningful event and that the rise causes some other significant event. In terms of information flow, the signal immediately preceding the rise in IP_3 is a rise in the concentration of active PLC. This rise is due to the binding of a subset of G-proteins

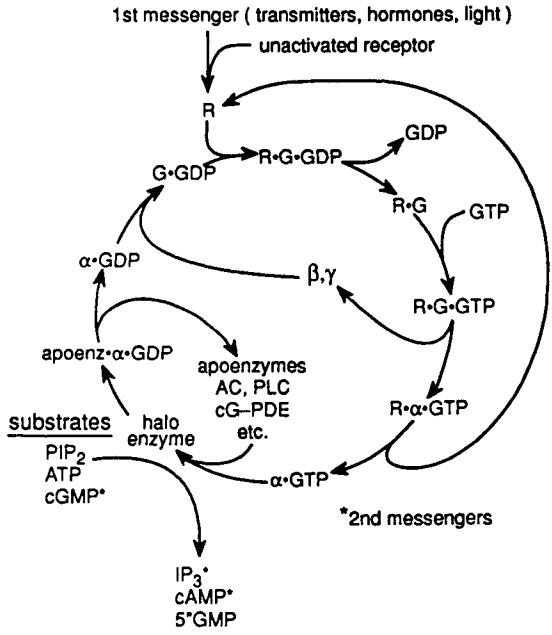


Figure 10. The G-protein cascades in smooth muscle catalyze the exchange GDP for GTP on G-protein. Following the binding of GTP, the trimeric G-protein splits into an α-GTP part and a β-γ part. The α-GTP part ordinarily then combines with its specific apoenzyme to constitute the active enzyme. For the activation of the contractile activation path, the enzyme is phospholipase C and the second messenger products are IP_3 and DAG. The IP_3 in the myoplasm binds to Ca^{2+} channels in the SR membrane, opening them. Other second messengers include the inhibitors of contractile activity, cGMP and cAMP.

which regulate the apoenzyme. In turn, they are activated by the plasma membrane receptors occupied by various transmitters of the autonomic nervous system. The receptor types include: (a) acetyl choline receptors, (b) catecholaminergic receptors, (c) histamine receptors, and (d) ATP receptors. The experimental evidence now available that IP_3 is a physiological trigger is overwhelming. Contractions and Ca^{2+} ion release can be induced by microinjection of IP_3 and in skinned preparations by concentration changes. The concentrations required for activation are commensurate with the dissociation constants measured for isolated IP_3 receptor protein and for microsome preparations.

The molecular weight of isolated IP_3 receptor protein is about 225 kDa but the actual receptor is apparently a tetramer. The IP_3 receptor is slightly smaller in size than the ryanodine receptor which is also a tetramer. Also, like the ryanodine receptor, Ca^{2+} ion binding has a synergistic effect on the IP_3 receptor. As one might expect, these two receptors have a large sequence homology. There is not much

evidence which bears on the issue of voltage sensitivity. However, increases in K^+ concentration do not release SR Ca^{2+} ion. The IP_3 receptor flickers between an open and a closed state. The fraction of the time it spends in the open state is increased by IP_3 binding.

Other metabolites of inositol phospholipids, e.g., inositol (1,3,4,5)-tetrakisphosphate (IP_4) may have additional signal transduction roles, particularly in accelerating the uptake of extracellular Ca^{2+} ion into the SR following a contraction.

The Coordination of Mechanical Activity: Plexuses

Whether by releasing transmitters which modulate the generation of action potentials, or by the release of non-electrogenic transmitters, the motor nerve endings of the autonomic nervous system determine the geometry and the extent of contraction in smooth muscle. Which autonomic motor nerves fire is determined by a combination of interactions between the central nervous system and the peripheral plexuses. The outflow of electrical activity from the brain stem nuclei controls the pattern of peripheral smooth muscle mechanical activity. For the smooth muscles of the vessels of the cardiovascular system there is relatively little peripheral integration, while in the gastrointestinal tract the peripheral plexuses have been called an entirely separate nervous system. In short, the peripheral expression and coordination of central autonomic control of smooth muscle varies greatly.

Action potentials, self-propagating. Action potentials of smooth muscle differ from the typical nerve action potential in at least three ways. First, the depolarization phases of nearly all smooth muscle action potentials are due to an increase in calcium rather than sodium conductance. Consequently, the rates of rise of smooth action potentials are slow, and the durations are long relative to most neural action potentials. Second, smooth muscle action potentials arise from membrane that is autonomously active and tonically modulated by autonomic neurotransmitters. Therefore, conduction velocities and action potential shapes are labile. Finally, smooth muscle action potentials spread along bundles of myocytes which are interconnected in three dimensions. Therefore the actual spatial patterns of spreading of the action potential vary.

On the other hand, it is important to emphasize that the most fundamental features of electrical activity are the same in all excitable systems, smooth muscle included. For example, in active regions of membrane, the membrane batteries are directed positive side inward, while in those regions which are at rest the membrane batteries are directed positive side outward. Current flow between the active and adjacent resting regions is associated with the depolarization of resting regions by adjacent active regions. This results in the propagation of action potentials. The local circuit current loops are strikingly similar to those of other excitable tissues.

The ionic basis of action potentials. Pacemaking. Action potentials in pacemaker regions of smooth muscle arise autonomously from the fluctuating balance of membrane ion channels as they jump between various states. Certain combinations of channels constitute stable electrical oscillators. Action potentials in smooth muscles are not elicited as they are in most neurons or skeletal muscles by a postsynaptic excitatory potential. The action potentials of smooth muscles are not driven or entrained by nerve activity but instead their autonomous frequencies of firing are modulated by neurotransmitters released broadly into the extracellular space. Action potentials from different smooth muscles vary a great deal. Some are as short as 15 milliseconds, while others are more than a 100 milliseconds in duration. Once initiated, action potentials may propagate through the smooth muscle tissue by local circuit current which flows between active and resting regions. If the pacemaker oscillations of membrane voltage do not reach threshold, action potentials do not occur and the pacemaker potentials, often called slow waves, can be viewed as separable processes. Thus, in the normally functioning smooth muscle, action potentials may either arise from autonomous pacemaker potentials or are elicited by electrotonic spread from active regions.

The time course of an action potential reflects net current flow and thus the balance of open ion channels. The rate of change of membrane voltage is proportional to transmembrane current flow, according to the equation:

$$dV_m/dt = (1/C_m)I_m$$

where C_m is the membrane capacitance and I_m is the membrane current. In smooth muscle the depolarizing phase of action potentials is caused by an excess inward flux of calcium ions through L-channels. This inward movement of positive charge discharges the negative charge excess on the in side of the membrane capacitance and piles up a net positive charge inside. Thus inward current tends to drive the membrane potential toward the EMF of the L-channels. Near the peak of the action potential in smooth muscle cells, the transmembrane voltage reverses its sign as it does in other excitable cells. Empirically, the EMF of L-channels is about +40 mV and the midpoint of activation lies in the neighborhood of –40 mV. They are inactivated slowly and were named L-channels to signify that they are long lasting. Current evidence indicates they may be inactivated not by depolarizing voltages but by slowly increasing intracellular calcium concentrations. However, it is worth noting that L-channels of different smooth muscle cells are not identical. In addition to L-channels there are apparently other kinds of channels in various smooth muscle membranes. Perhaps the most significant other channel is the "T" or transient channel. The T-channel turns on at more negative voltages and therefore is seen by some as a candidate for the initiation of slow waves. Except, notably, for uterine myometrial cells there are no few instances where voltage gated sodium channels play a significant role in action potential generation in smooth muscle.

The repolarization phase of smooth muscle action potentials is driven by currents through potassium channels. There are many different kinds of potassium channels but the three most frequently identified in smooth muscle are (a) the calcium dependent K^+-channel, (b) the delayed rectifying K^+-channel, and (c) the anomalous rectifying K^+-channel. The tendency of the calcium dependent K^+-channel to open is increased by both positive membrane voltages and increased Ca^{2+} ion concentrations. Smooth muscle delayed rectifying K^+-channels are similar to those found in other cell types. Both of these potassium channels seem to function primarily as conduits for the repolarizing currents which terminate action potentials. The anomalous rectifying K^+-channel, on the other hand, tends to introduce destabilizing behavior to the membrane voltage, and it may play a role in action potential generation by turning off stabilizing potassium current if the membrane potential drifts toward positive values. This kind of event would tend to occur as calcium is pumped out between action potentials.

Local circuits. The extracellular space is electrically a low resistance medium. The intracellular space is similarly a low resistance medium, running in parallel with and separated from the extracellular space by the cell membrane. Electrical continuity of the intracellular space between the cell interiors of individual cells is provided by gap junctions or nexuses consisting of connexons rafted together. When all the cell membranes in a tissue are at the same membrane potential, there can be no net driving force for current flow anywhere. If, however, two regions of membrane are in different states and there is difference of membrane EMF between them, then a circuit exists such that current will flow between the regions. If one of the regions had been resting and the other supporting an action potential, then the resting region will be depolarized and driven toward threshold. A loop of current in such a circumstance is said to flow around a local circuit and there is a spread of activity concomitantly. The fundamental issues are the same as along nerve fibers.

In a nerve process or skeletal muscle fiber, the spread of activity is essentially only in one dimension, along the fiber. However, in smooth muscle the situation is rather more complex geometrically, and all three dimensions are involved. Action potentials conduct electrotonically just as they do in nerve fibers but they do so in three dimensions. *In situ*, regions supporting action potentials are not pointlike but tend to be large and the spread from them is more like a surface, approximating a plane.

NEURAL CONTROL

Although many smooth muscles are still rhythmically active when separated from extrinsic innervation, most are quiescent if completely denervated. Net excitatory neural modulation of smooth muscle is the rule. To make things more complex, the influences of innervation to smooth muscle are carried by a rather large number of different transmitters, whose effects are still being investigated. This presumably

accounts for a large fraction of the complexity of the electrical and mechanical behavior of smooth muscle, and presumably also will account for the switching between other physiological activities of smooth muscle, e.g., proliferation or protein secretion.

The autonomic nervous system (ANS) innervates the viscera. It is the motor system which controls smooth muscle organs and the vascular smooth muscle tissue of all organs of other sorts. The spinal neurons of the autonomic system do not directly innervate effector cells but instead synapse in the periphery with another, the final efferent neuron. At each level, the cell bodies of these neurons are in the lateral horn of the spinal cord. They exit the spinal cord as part of the mixed spinal nerve. As a first approximation, there is a separation of function of the nerve fibers leaving the central nervous system in the cranial or sacral regions relative to those leaving in the thoraco-lumbar region. The cranial-sacral divisions of the ANS constitute the parasympathetic nervous system and the thoraco-lumbar divisions form the sympathetic nervous system. The set of transmitters (acetyl choline, etc.) of the parasympathetic nervous system tend to have antagonistic action relative to the transmitters of the sympathetic nervous system (norepinephrine, etc.). However, this generalization must be applied only as a guide because the influence of a particular transmitter is a function of what its receptor does, and this varies. The polypeptide neurotransmitters do not segregate simply between the two subdivision of the ANS.

The ANS seems to have evolved into its present configuration in the elasmobranchs. Thus, in animals at the level of the teleosts and above, the role of the autonimic system in the innervation of an organism is relatively predictable. However, a particular myocyte has receptors for the many different transmitters released by autonomic nerve endings and an integration of inputs unique to the smooth muscle gives it its peculiar response complex. This further points to the circumstance that each myocyte response, which may be defined by the cell's multiple outputs, is determined by the cellular state, a function of genome expression and the transfer functions between each receptor and each output. Thus, the responses of myocytes to transmitters released by stimulation of various nerves may be said to depend on their physiological state, which is a function of the history of prior stimulations. In short, the multiple innervation of the ANS provides a conveyance for different influences which the signal transduction system of the myocyte then resolves into the actions of the effector systems of the myocyte. Thus, there is no common path prior to the myocyte itself and the integrative functions performed by the smooth muscle myocyte are reminiscent of those done by neurons.

Parasympathetic Nervous System

The parasympathetic nervous system is capable of maintaining gastrointestinal motility, etc., even when connections to the central nervous system have been

severed. The basis for this is its extensive set of peripheral ganglia. The plexuses in the periphery are miniature nervous systems. Afferent fibers which influence the activity of the parasympathetic pathway come back to the central nervous system either by way of parasympathetic nerves, (e.g., the vagus) or by way of typical spinal nerves. Many make connections in the peripheral ganglia first, and thereby initiate local reflexes or response loops. This is especially true for the gastrointestinal tract.

Acetyl choline is the characteristic transmitter of the parasympathetic nervous system. Thus, the action potentials of the postganglionic, parasympathetic efferent neurons characteristically cause the release of acetyl choline. In addition they release a wide variety of polypeptide neurotransmitters. On the average, a particular one of these polypeptide transmitters may be more significantly associated with acetyl choline release than with norepinephrine release. However, the same polypeptide transmitter may be co-released with an adrenergic transmitter at another site. Thus, the understanding of a particular neural control system first requires an identification of the various transmitters released as a function of the stimulus. In order to predict the response to the stimulation of a certain distribution of nerve endings one must both know which transmitters are released and the receptor distribution on the smooth muscle. Then, only following that, the integrated response may be anticipated.

As distinct from the acetyl choline receptor of the neuromuscular junction, the acetyl receptors of the viscera are not blocked by nicotine but are blocked by muscarine. Moreover, based on differences in the binding of the muscarinic antagonist, pirenzapine, the muscarinic acetyl choline receptors (mAChRs), are separated into two classes, viz. high affinity m_1 receptors, and low affinity m_2 receptors. The latter predominates in the heart, cerebellum, and smooth muscle broadly. These different receptors mediate quite different actions.

Dominance of the activity of the parasympathetic nervous system is usually indicative of vegetative activity of the organism. For example, when the parasympathetic influence on the gastrointestinal tract is dominant, the organism is readied for digestion. Cholinergic nerve impulses promote slowing of the heart, dilation of the blood vessels, and increased gastrointestinal motility. It has been shown in a number of smooth muscle types that occupation of m_2 receptors by agonists activates PLC, and leads to increased intracellular IP_3 levels.

Sympathetic Nervous System

The small molecular transmitter characteristic of the sympathetic nervous system is norepinephrine. However, epinephrine from the adrenal medulla is an important partner in systemic sympathetic actions. The actions of the sympathetic system are unique because in different smooth muscles responses are diametrically opposed: they may be either by contraction or inhibition of contraction and

relaxation. The basis for this dichotomy is, on one hand, the preponderance in the cell membranes of systemic vascular smooth muscle of α-receptors, which have a greater affinity for epinephrine than for norepinephrine. On the other hand, in the membranes of most other types of smooth muscle cells there is a preponderance of β-receptors, which have a much greater affinity for norepinephrine. While α-receptors are thought to activate a G-protein-IP_3-Ca^{2+} release pathway β-receptors are thought to activate a G-protein-adenylate cyclase-cAMP release pathway. Cyclic AMP is an inhibitor of contraction. Correspondingly, activation of α-receptors causes contraction and activation of β-receptors causes relaxation. Because there are both kinds of receptors expressed in most smooth muscle myocytes, either of these effects can be seen in most smooth muscles following pharmacological blockade of the opposing receptor type.

Because of its motor, i.e., activating effect on vascular smooth muscle and its inhibitory effect on intestinal smooth muscle, the sympathetic nervous system has been cast into the role of the component of the nervous system that executes control of visceral function in times of physical emergency for the organism. The phrase "fight or flight" has been often used to describe the circumstances in which the adrenergic transmitters of the sympathetic system are dominant over the cholinergic parasympathetic system. This concept is perhaps oversimplified but it has the utility of a first approximation of how the two components of the ANS interact in the periphery. Sensory inputs which lead to increased blood pressure, for example, activate the sympathetic pathways.

OTHER PHYSIOLOGICAL STATES

It is usually presumed that smooth muscle cells have only one kind of activity, contraction, and that the only alternative to contractile activity is a kind of estivating resting state (Figure 11). The actual situation is of course more complicated. For example, smooth muscles synthesize extracellular filament protein. They also proliferate, particularly in the cardiovascular system. Both of these processes require a considerable amount of control of the cellular economy.

The Proliferative State

Very little is known about the mechanisms of control of smooth muscle cell proliferation. In most organs, smooth muscle cells are not normally mitotically active. However, vascular tissues are under continuous control and a number of pathologic states involve proliferation of smooth muscle cells. Several examples of possible control mechanisms can be given which indicate the direction of current thought. First, many growth factors elicit contractions by turning on the IP_3 cascade, and cAMP synthesis is also often stimulated. So there are hints of interaction between contraction and growth control mechanisms but not much beyond speculation. Second, angiotensin II synthesis inhibitors are effective blockers of the

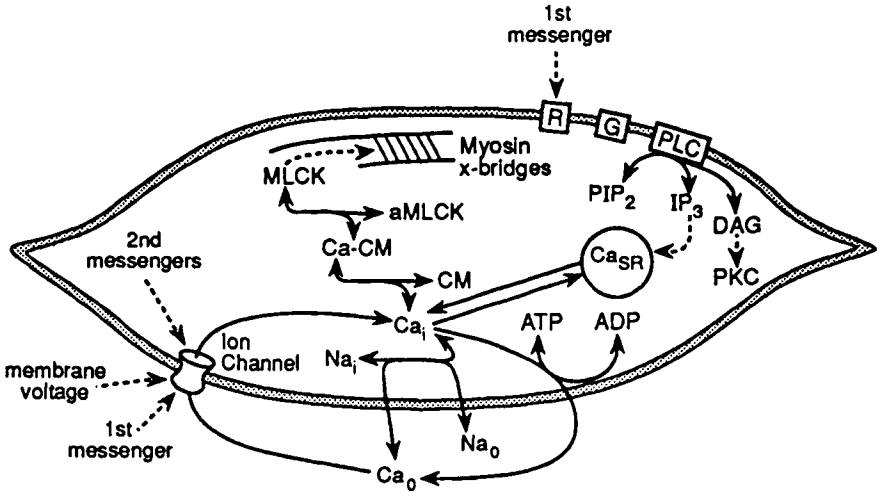


Figure 11. A summary cartoon illustrating the relationships between the first messengers and the release of Ca^{2+} ions from the SR, the various pathways that influence the intracellular Ca^{2+} ion concentration, and the activation of MLCK, which leads to contraction.

cardiac hypertrophy associated with hypertension. Third, insulin-like growth factor (IGF), platelet-derived growth factor (PDGF), macrophage-derived growth factor (MDGF), endothelial-derived growth factor (ECGF), and transforming growth factor β all have been suggested to influence proliferation of vascular smooth muscle cells. Fourth, lipoprotein a is a stimulant of cell proliferation, apparently by way of its constituent, the glycoprotein apo a. Lipoprotein a acts by competitive inhibition of surface-activated plasminogen and the disinhibition of TGF-β.

Secretion

The secretion of extracellular matrix proteins is also a function of smooth muscle cells but, since it occurs concurrently with other activities, it does not seem to constitute a physiological state. However, the fraction of the cellular resources which are devoted to it must be regulated; these regulatory mechanisms are virtually unknown. In addition, it should be anticipated that autocrine activity occurs as well, involving peptides, prostaglandins, cytokines, and nitric oxide.

SUMMARY

Contraction in smooth muscle is triggered by an increase in the intracellular free Ca^{2+} concentration. Trigger Ca^{2+} may come from intracellular compartments, from

influx through membrane voltage-dependent channels, or both. Release from intracellular stores is modulated by plasma membrane receptors via a G-protein cascade. Plasma membrane receptors are activated by the binding of specific first messengers. Primarily, first messengers are neural transmitters, which may modulate membrane channels as well.

Once the intracellular Ca^{2+} concentration begins to rise, calmodulin-calcium binding also rises and MLCK, which is dependent on calmodulin activation, rises in turn. The next step in this cascade is the phosphorylation of myosin. Finally, the phosphorylation of myosin results in the activation of the crossbridges and the accompanying transduction of ATP energy into mechanical work. Despite its differences in regulation, smooth muscle behaves mechanically much like other muscles.

The behavior of smooth muscle crossbridges is modulated by more than a simple phoshorylation switch. Crossbridges show more behavioral states than rest-off and active-on. What the other regulatory mechanisms are (involving cyclic nucleotides, protein kinases, phosphatases, transferases, other second messengers, etc.) remains to be clarified. The ability to maintain tension at a low metabolic cost is peculiar to crossbridge regulation in smooth muscle and probably involves extra regulation.

Smooth muscles do more than contract and the regulation of the mitotic activity of vascular smooth muscle is important in the maintenance of the oxygen supply to tissues and the genesis of hypertension of non-renal origin. Secretion activities of smooth muscle are just beginning to be appreciated.

REFERENCES

Barr, L. (1989). The photochemical responses of the vertebrate sphincter pupillae. Crit. Rev. Neurobiol. 4, 325–366.

Hai, C.-M., & Murphy, R.A. (1989). Ca^{2+}, crossbridge phosphorylation, and contraction. Ann. Rev. Physiol. 51, 285–298.

Hartshorne, D.J. (1987). Biochemistry of the contractile process in smooth muscle. In: Physiology of the Gastrointestinal Tract. 2nd Edition (Johnson, L.R. ed.), pp. 423–482, Raven Press, New York.

Hille, B. (1992). Ionic Channels of Excitable Membranes. 2nd edn., Sinauer Associates, Sunderland, MA.

Meiss, R.A. (1989). Mechanics of gastrointestinal smooth muscle. In: Handbook of Physiology, Section VI. The Gastrointestinal System, Volume 1, Part 1 (Wood, J.D. ed.), pp. 273–329, American, Physiological Society, Bethesda, MD.

Ruegg, J.C. (1988). Calcium in Muscle Contraction. 2nd edn., Springer-Verlag, Heidelberg.

Sanders, K.M., & Publicover, N.G. (1989). Mechanics of gastrointestinal smooth muscle. In: Handbook of Physiology, Section VI. The Gastrointestinal System, Volume 1, Part 1 (Wood, J.D. ed.), pp. 273–329, American Physiological Society Bethesda, MD.

Siegman, M.J., Butler, T.M., Vyas, T., Mooers, S.U., & Narayan, S. (1991). Cooperative mechanisms in the regulation of smooth muscle contraction. In: Regulation of Smooth Muscle Contraction (Moreland, R.S., ed.), pp. 77–85. Plenum, New York.

Somlyo, A., & Somlyo, A.V. (1989). Flash photolysis studies of excitation-contraction coupling. Ann. Rev. Physiol. 52, 857–875.

Chapter 7

The Cellular and Molecular Basis of Skeletal and Cardiac Muscle Contraction

MICHELLE PECKHAM

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 201–237.

ISBN: 1-55938-808-0

INTRODUCTION

Muscle contraction has been studied since the early 1800s (see review by A.F. Huxley, 1980). We now know that actin and myosin are the two proteins essential for force production in skeletal muscle. Each of these proteins is organized into a separate filament, myosin (molecular weight about 500,000) in the thick filaments and actin (molecular weight about 42,000) in the thin filaments. Each thick filament (1.6 μm long) contains about 300 myosin molecules and each thin filament (1.0 μm long) about 380 actin molecules. These filaments are organized into a repeating structure called the muscle sarcomere that is 2.2 μm long in frog muscle. A frog sartorius muscle is typically about 30 mm long and contains about a thousand myofibrils that will each contain about 13,500 sarcomeres arranged along their length. *In vitro*, a single myosin crossbridge will interact with an actin filament to produce 3–4 pN of force and 11 nm of movement (Finer et al., 1994). *In vivo*, a whole muscle can produce forces of up to 200 N mm^{-2} (force per cross-sectional area) and/or movement of the muscle.

Muscle is able to generate such forces and/or movement by the cyclical interaction of myosin crossbridges in the thick filaments with actin in the thin filaments. The fuel for this process is ATP (ΔG is 60 kJ mol^{-1}). Although each crossbridge can only generate about 3–4 pN of force, or 11 nm of movement, the large amounts of force or movement that a whole muscle can produce is generated by the repetitive action or cycling of the crossbridges which contribute a small amount of force or motion each time they attach. They subsequently detach and repeat this cycle after reattaching further along the thin filament. This is considered to be the mechanism of contraction in both cardiac and skeletal muscle. To understand the molecular basis of muscle contraction, one must therefore investigate the molecular interaction of the myosin crossbridge with actin, and how this process is driven by ATP hydrolysis.

BASIC PROPERTIES OF MUSCLE

Development

Skeletal muscle is made up of many muscle fibers (Figure 1) each of which is a multinucleated cell that was formed during development by the fusion of many cells (myoblasts). Skeletal muscle is formed from precursor myoblasts which arise

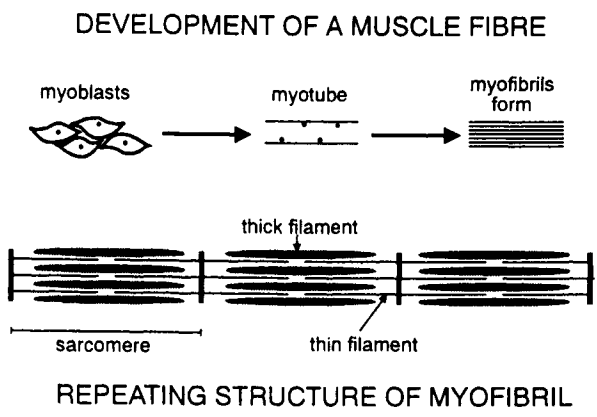


Figure 1. Muscle development. A skeletal muscle fiber is formed by the fusion of many single cells (myoblasts) into a multinucleated myotube. Myotubes then develop into the muscle fiber (see text). Sarcomeres form in longitudinal structures called myofibrils. The repeating structure of the sarcomere contains interdigitating thick and thin filaments.

in the mesodermally derived somites of the developing embryo. Myoblasts migrate to the developing limb and ventral body wall where after a proliferative stage they cease dividing and fuse to form myotubes in which the myofibril develops. Muscle specific proteins are activated at the time of myoblast fusion. In developing myotubes these proteins assemble to form the sarcomeres and myofibrils. Primary myotube formation is followed later by a secondary round of myotube formation alongside the primary myotubes. These then develop into the muscle fibers and muscle blocks. Each fiber is packed full of myofibrils which run from one end of the muscle to the other, and contain the contractile proteins of the muscle organized into repeating units called muscle sarcomeres. Each fiber is innervated (connected to and stimulated by) a nerve through neuromuscular junctions. The contractile properties of individual fibers in a single muscle are often different, but fall into two main classes: fast and slow twitch.

Contractile Properties

The contractile properties of muscle can be investigated using single muscles or even single muscle fibers *in vitro*. In a typical experiment a single frog muscle, such as the sartorius, is dissected out and mounted between two hooks, one connected to a tension transducer, which measures force, and the other to a motor, which can change the length of the muscle. The muscle is bathed in an isotonic Ringer's solution and is kept at 0 °C. When it is stimulated by a single electrical shock, which lasts less than 1 ms, the muscle contracts and relaxes (Figure 2), and this single

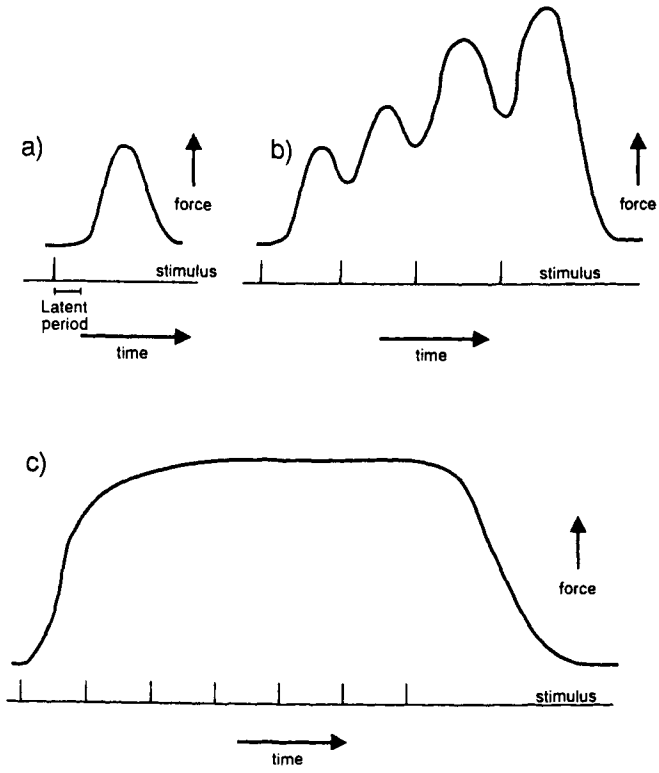


Figure 2. Muscle stimulation. **a**) a single nerve impulse (stimulus) causes a single contraction (a twitch). There is a small delay following the stimulus before force rises called the latent period. **b**) A train of stimuli at a low frequency causes an unfused tetanus. Force increases after each progressive stimulus towards a maximum, as calcium levels in the myofibrillar space increase. But there is enough time between each stimulus for calcium to be partially taken back up into the sarcoplasmic reticulum allowing partial relaxation before the next stimulus occurs. **c**) A train of stimuli at a higher frequency causes a fused tetanus, and force is maximum. There is not enough time for force to relax between stimuli. In the contractions shown here, the ends of the muscle are held fixed; the contractions are isometric.

contraction is called a twitch. The electrical shock causes an action potential which depolarizes the muscle fibers. There is a small delay, called the latent period, between the action potential and the contractile response. The latent period is the time required for enough Ca^{2+} to be released into the myofibrillar space to allow contraction to occur.

Calcium levels in the myofibrillar space are usually low, to prevent contraction. The calcium ions are stored in an internal membrane system called the sarcoplasmic

reticulum (SR). The time-course of the rise and fall of force observed during the twitch arises from the release of Ca^{2+} from the SR into the myofibrillar space, followed by binding of Ca^{2+} to the regulatory proteins that prevent actin and myosin from interacting in the absence of Ca^{2+}, and finally by re-uptake of calcium into the SR. It is also determined by the kinetics of the interaction between myosin crossbridges in the thick filaments and actin in the thin filaments (see below). The regulatory proteins prevent actin and myosin interacting in the absence of Ca^{2+} is by a combination of steric blocking and an inhibition of the activation of myosin ATPase by actin. The molecular mechanism of these changes is unclear. As might be expected, fast twitch muscle fibers have a fast tension response to a nerve stimulus and the length of the tension response is short. Slow twitch fibers have a slower and longer tension response.

If the muscle is stimulated repeatedly, each stimulus causes more Ca^{2+} to be released, causing more force to be produced (Figure 2). If the time between the stimuli is decreased, eventually there is not enough time for the calcium to be taken back up into the SR, Ca^{2+} reaches a maximum level, the muscle does not relax between stimuli, and a single unfused tetanus is produced (Figure 2). Tension is maintained at a high level until stimulation ceases. In this type of contraction, the ends of the muscle are held fixed, and the muscle length does not change: this is called an isometric contraction.

In terms of muscle function, muscle is very adaptable. Depending on the type of stimulation, muscle can either twitch or contract tetanically for a variable length of time. If the ends are held fixed, then it contracts isometrically and the force produced is maximal. If one or both ends of the muscle are not held fixed then the muscle is able to shorten. The muscle can shorten at a fixed load (isotonic contraction) where the velocity of shortening is also constant. Power output (force $\times$ velocity) is maximum where the velocity of shortening is about one third of the maximal rate. Finally, the muscle can shorten at maximum velocity (unloaded shortening). However, the molecular basis of the interaction of myosin with actin to produce force, or shortening, is the same in each case.

Very little is known about how cardiac muscle cells form during development. A region of the embryo becomes committed early in development to form the heart, but the exact basis of this is still unclear. Cardiac muscle cells do not migrate but form early in embryogenesis in the region in which the heart will develop and subsequently mature to form the early atria and ventricles. Cardiac cells do not fuse during development but remain as single cells with a single nucleus that are closely connected by intercalated discs. Unlike skeletal muscle, cardiac cells can dedifferentiate and grow during development, or in response to stress on the heart. Each cardiac cell is not connected to a nerve, but electrical stimuli can pass from cell to cell via the intercalated discs (there are also specialized contraction pathways in the heart). *In vivo*, cardiac muscle twitches but cannot contract tetanically, and does not undergo such a large range of force/shortening conditions as does skeletal

muscle. The sarcoplasmic reticulum is much less abundant in cardiac muscle than in skeletal muscle (Ruegg, 1988, for review), and the amount of calcium that it can take up and release is only enough to induce one twitch contraction. Despite these differences, the molecular basis of contraction is thought to be very similar.

MUSCLE PROTEINS AND SARCOMERE STRUCTURE

Muscle contains two types of filament; the actin-containing thin filament and the myosin-containing thick filament, that are organized into a regular array in muscle sarcomeres (Figure 3). Longitudinally, a muscle fiber has a series of dark and light bands when viewed under polarized light in a light microscope. The dark bands have been termed anisotropic or A-bands, and the light bands isotropic or I-bands. They alternate; in the center of each light band is a dark line termed the Z-line. The term muscle sarcomere describes the area between two Z-lines. The structure observed in this region is repeated all along the muscle length. A myofibril is a series of repeating muscle sarcomeres and runs along the length of the muscle fiber. A single muscle fiber is composed of many myofibrils which can be branched in cardiac muscle.

The A-bands contain both the myosin-containing thick filaments and the actin-containing thin filaments. In the A-bands, each thick filament is surrounded by six thin filaments (Figure 3) such that the two types of filament overlap, although the

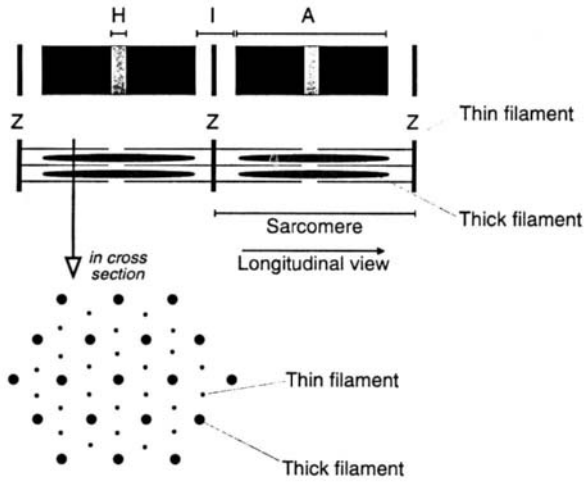


Figure 3. Structure of a muscle sarcomere. In a polarizing microscope muscle appears to have dark (A) and light (I) bands. The I-band region only contains thin filaments. The A-band region contains both thick and thin filaments. One sarcomere is the distance between two Z-lines. In cross section, the hexagonal packing of the thick and thin filaments can be seen.

overall ratio of thick to thin filaments is 2:1 in cross section. The I-bands contain only the actin-containing thin filaments which insert into a complex structure in the Z-line (Figure 3).

Myosin can be extracted from muscle by high salt (0.3 M KCl, 0.15 M KP_i) and recovered by precipitating the protein at low salt. It has two globular heads flexibly joined to a tail of about 156 nm in length as shown by electron microscopy (Elliott and Offer, 1978). The heads appear to move freely with respect to the tail. When myosin is digested with various enzymes, different proteolytic fragments can be recovered. The entire myosin molecule is able to self-associate to form filaments; on the globular heads it has two sites that can bind to actin, and two sites that can hydrolyze ATP. If it is digested with trypsin, for example (Figure 4a), two proteolytic fragments are produced; light meromyosin (LMM) and heavy meromyosin (HMM). LMM is still able to self-associate, but cannot bind to actin or hydrolyze ATP. HMM cannot self-associate but can bind to actin (two sites) and hydrolyze ATP (two sites). If myosin is digested with papain, for example, two fragments are produced; the rod and subfragment-1 (S-1). The rod can self-associate, but does not have any actin or ATP sites. The S-1 cannot self-associate and has one actin binding site and one ATP hydrolysis site. HMM can be digested into S-1 and subfragment-2 (S-2). S-2 does not self-associate. The myosin molecules aggregate in a bipolar fashion to form the thick filament (Figure 4c) which results in a bare zone at the center of the thick filament where there are no myosin molecules, which is about 0.2 μm long. The exact structure of the thick filament is still not clearly known.

Each myosin molecule is made up of two alpha-helical polypeptide chains, termed heavy chains, which wind around each other in a left handed coil (Figure 4b). The myosin heavy chains wrap around each other in the rod to give a coiled-coil structure. At the end of the rod region, the heavy chains diverge and, together with the myosin light chains, fold separately to form S-1. Each myosin molecule contains four light chains, two associated with each globular head (S-1). There are two kinds of light chains: regulatory and essential. The regulatory light chain (also called DTNB or LC2), contains a single Ca^{2+}/Mg^{2+} ion binding site and can be removed reversibly from myosin by DTNB [5,5′-dithio-bis-(2-nitrobenzoic acid)]. The essential light chain (also called alkali or LC1 and LC3) can be removed by alkali. LC1 and LC3 are derived from the same gene, but different exons (coding regions of the gene) are spliced out to produce the two isoforms. LC1 and LC3 are essentially the same, but LC3 lacks 41 amino acids at the amino-terminus. LC1 has 191 amino acids and LC3 has 149 amino acids. A myosin S-1 molecule contains one regulatory and one essential light chain, either LC1 or LC3. The light chains may have a subtle role in regulating the amount of force produced in skeletal muscle either by phosphorylation or, in some nonmammalian skeletal muscles, such as those from scallop, by regulating the ATPase activity.

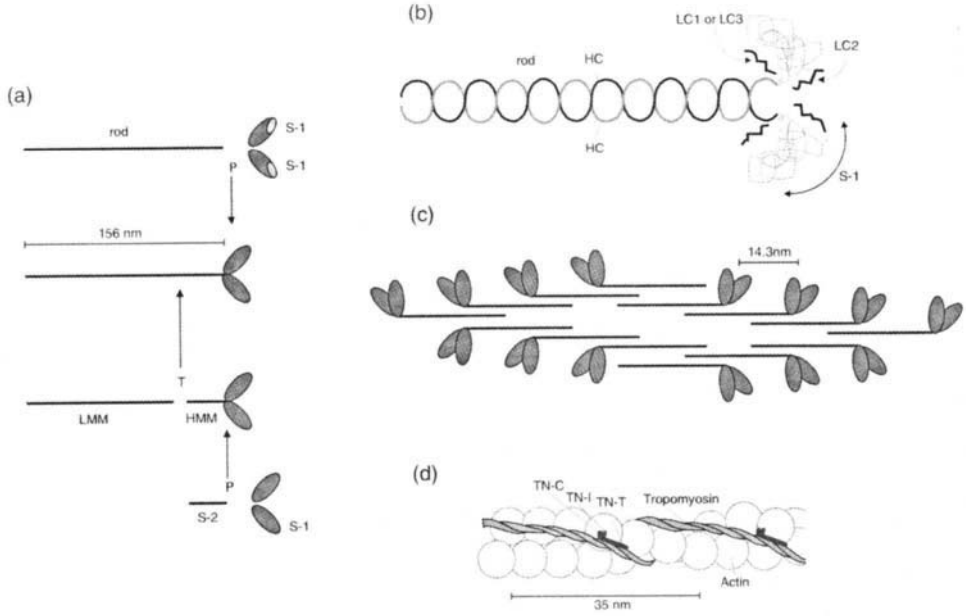


Figure 4. Structure of myosin and its organization into the thick filament, actin and its organization into the thin filament. **a**) The myosin molecule consists of a rod and two globular heads (S-1). Proteolytic digestion with papain (P) splits the molecule into the rod and the two heads (S-1). Proteolytic digestion with trypsin (T) splits the molecule into two different fragments known as LMM and HMM. Further digestion of HMM with papain splits HMM into S-2 and the two heads (S-1). **b**) In the rod region of the myosin molecule, two alpha-helical polypeptide heavy chains (HC) wrap around each other in a left-handed coil to give a coiled-coil structure (see text). At the end of the rod region the heavy chains diverge, and together with the myosin light chains fold separately to form S-1. Two light chains are associated with each globular head (S-1), one essential (LC2) and one regulatory (LC1 or LC3). **c**) The myosin molecules aggregate in a bipolar fashion to form the thick filament. **d**) Actin monomers aggregate to form thin filaments which can be described as two chains of monomers twisted around each other to form a twin threaded coil with a complete twist every 13–14 monomers. Troponin and tropomyosin bind to actin in the thin filament. One molecule of tropomyosin binds alongside seven actin monomers. One troponin complex is found for every tropomyosin molecule.

Both the thick and thin filaments contain other proteins. For example, the thick filament contains titin (molecular weight about 3,000,000) and the thin filament contains nebulin (although not in cardiac muscle), and the regulatory proteins troponin (molecular weight about 33,000) and tropomyosin (molecular weight about 70,000). Nebulin and titin are thought to be ruler proteins, that is, they determine the overall length of the thin and the thick filament, respectively. The

length of the thin filament in heart muscle is more variable possibly because it does not contain nebulin. M-protein, myomesin, creatine kinase, and C-protein are also found in the thick filaments. M-protein and C-protein possibly have a role in maintaining the structure of the sarcomere. The way in which these accessory myosin binding proteins such as titin and C-protein assemble into the thick filament is unknown. In the thin filaments, α-actinin and β-actinin can also be found which link actin filaments to the Z-line and cap (bind to the end and stop polymerization of actin filaments, respectively.

The regulatory proteins tropomyosin and the troponin-complex bind to actin in the thin filament (Figure 4d). Tropomyosin forms a coiled-coil structure similar to the structure found for the myosin rod. One tropomyosin molecule binds alongside seven actin monomers. The amino-terminus of this molecule is thought to be involved in regulation and the carboxy-terminus in binding to troponin-T of the troponin complex. The troponin complex consists of three subunits: troponin-T, which binds tropomyosin; Tn-I, which binds to Tn-C and actin and inhibits the ATPase activity of actomyosin; and Tn-C, which is globular and binds Ca^{2+} ions. TN-T is 20 nm long, half as long as tropomyosin and is 40% α-helix.

DEVELOPMENT OF THE SLIDING FILAMENT THEORY OF MUSCLE CONTRACTION

Early ideas about how muscle contracts were mainly based on the mechanical properties of muscle, since little was known about these protein components, although myosin had been named in the mid 1800s. Muscle was thought to be like a spring which could store energy and then release it on contraction. But these kinds of ideas suggested that only a fixed amount of energy was available to the muscle, which was shown not to be true by Fenn in the 1920s. He observed that there was an increase in energy output when muscle shortened above that for an isometric tetanus, and that the amount of energy available to a muscle was therefore not fixed, but variable (Fenn, 1924). These ideas were amended to better describe the observed properties of muscle (i.e., Hill, 1938) but were still based on mechanics.

In the 1950s several observations gave rise to a new sliding filament or independent force generator theory which still provides the basis for modern ideas about how muscle works. First, electron microscopy revealed the interdigitating thick and thin filaments and that there were projections from the thick filaments. Second, electron and light microscopy showed that when a muscle shortens the lengths of these two filaments does not change (A.F. Huxley and Niedergerke, 1954; H.E. Huxley and Hanson, 1954), suggesting that shortening must occur by the filaments sliding past each other (Huxley and Hanson, 1954; Huxley, 1957). Third, HMM had been shown to bind to actin and to have an ATPase activity (Szent-Györgyi, 1953). A.F. Huxley used these observations to develop a mathematical model which could explain many of the experimental observations of the mechanical and energetic properties of muscle (A.F. Huxley, 1957).

A.F. Huxley's 1957 Theory

In his theory, A.F. Huxley suggested that a myosin crossbridge attaches to actin and produces an amount of force which is proportional to the amount of strain on the crossbridge, that is to the amount by which it is distorted (Figure 5). He also suggested that crossbridges were only able to attach over a range of preferred distortions. From these basic ideas he derived mathematical expressions for force

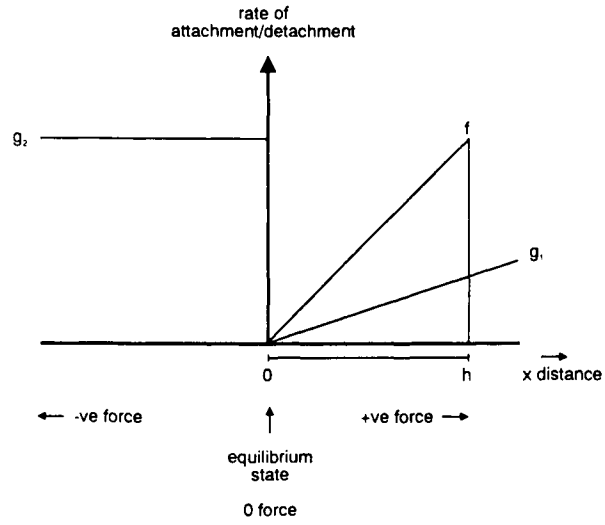


Figure 5. Huxley's theory used mathematics to predict how many crossbridges were attached under various conditions. Very briefly, in this theory, the crossbridge can be in either of two states, attached or detached. As soon as a crossbridge attaches there is simultaneous production of force. The rate at which crossbridges attach is f and the rate at which they detach is g. The detached crossbridge moves randomly (Brownian motion), longitudinally with respect to the actin filament, and has a limited range over which it can attach to actin from x = 0 to x = h. The position x = 0 is the equilibrium state of the crossbridge; attachment to actin at x = 0 does not produce any net force. The position x = h is the maximum displacement from the equilibrium position at which the crossbridge can attach. Attachment between x = 0 and x = h produces net positive force. In this region, Huxley made the attachment and detachment rate constants, f and g, depend linearly on x, although there is no experimental basis for this. The crossbridge cannot attach at negative values of x, as negative force would result from the attachment. The attachment rate constant f is zero in this region, and the detachment rate constant (g_2) is large and has a constant value. Huxley gave values of about 10 nm for h, $200s^{-1}$ for g_2, and about $50s^{-1}$ for f. He was able to use this theory to calculate the dependence of the fraction of crossbridges attached on x during isometric contraction and during different speeds of shortening.

production, shortening velocity, and energy liberation describing the observations that had been made, for example those of A.V. Hill on the contractile properties of shortening muscle (Hill, 1938). In these experiments, the muscle length is not held constant but allowed to shorten at a constant velocity (isotonic contraction) after maximum force (Po) is reached during tetanic stimulation. The muscle length is allowed to decrease such that the force is reduced to a new steady level (P), and the length decreases with a constant velocity (V). There is a hyperbolic relationship between the amount of force maintained for different velocities of shortening which Hill described in 1938 (Figure 6). The relationship shows that heavier loads are lifted more slowly than light loads.

The experimentally derived force-velocity relationship can be explained on the basis of Huxley's crossbridge theory in the following way. Steady state shortening must involve many crossbridge cycles as each crossbridge can remain attached to actin only over a limited range and so it has to detach and reattach several times to allow the muscle to shorten over large distances. As the velocity increases, fewer crossbridges will be able to attach to actin as the filaments slide past each other more quickly, and so force will be lower. Furthermore, each crossbridge will be slackened during sliding so it will produce less force than in an isometric contraction.

This theory was also able to explain the energetic properties of muscle. Hill had found in 1938 that the heat produced by a muscle was proportional to the shortening distance and Huxley was able to derive this relationship from his mathematical expressions. However, Hill found later (Hill, 1964), that the rate of energy output did not increase at a constant rate as the velocity increased, as he had originally found, but declined at high velocities. This could not be explained by Huxley's 1957 theory.

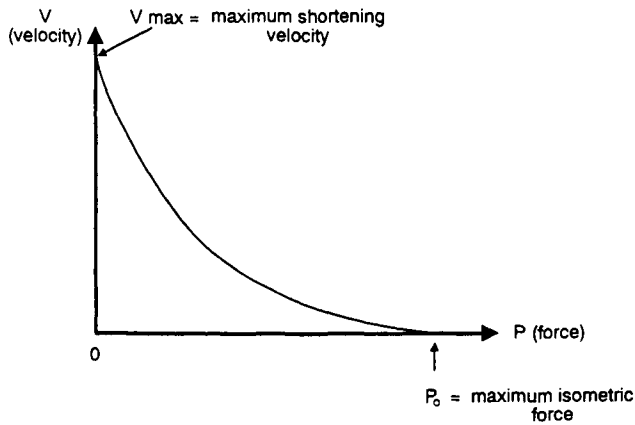


Figure 6. Force velocity relationship. A schematic diagram of the experimental relationship between load (force) and velocity for muscle during contraction.

It had already been observed (Ramsey and Street, 1940) that the amount of active tetanic force a muscle produces depends on its length in experiments in which the muscle contracts isometrically. This dependence of force on muscle length is one of the most important pieces of evidence for the sliding filament theory. It was subsequently confirmed more precisely by Gordon, Huxley, and Julian in 1966, who measured force at different muscle lengths in which the sarcomere length was known. If force production is solely due to the interaction of myosin crossbridges from the thick filament with actin in the thin filament, then by reducing the overlap between thick and thin filaments, the numbers of crossbridges which could interact with actin should also be reduced, and the force should decrease. It was found that this was indeed the case (Figure 7). Force was maximal between the two sarcomere lengths at which filament overlap was maximal (2.0 μm to 2.2 μm). (Due to the

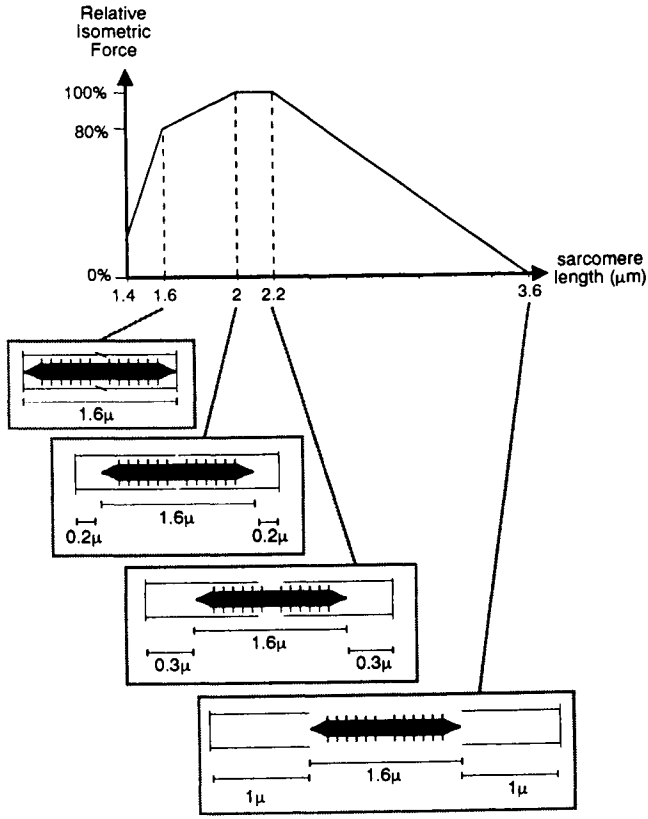


Figure 7. Length tension relationship. A schematic diagram showing how force varies with sarcomere length, and how this is explained by the relative amount of overlap between the thick and the thin filaments, and hence the numbers of myosin cross-bridges in the thick filaments that can interact with actin in the thin filaments.

bare zone, about 0.2 μm long, in the middle of the thick filament which does not contain crossbridges, a change in length of 0.2 μm when overlap between thick and thin filaments is maximal does not change the number of interacting crossbridges.) As sarcomere length is increased further, force declines linearly with the increase in sarcomere length reaching zero force at a sarcomere length of 3.6 μm (no filament overlap). At sarcomere lengths below 2.0 μm, force declines because thin filaments begin to overlap each other, reducing the potential interactions between thick and thin filaments. Finally at shorter sarcomere lengths (below 1.6 μm), the thick filament begins to hit the Z-line and starts to buckle.

These experiments have been repeated often, and most workers find this relationship between sarcomere length and force. But the measurements have to be made very carefully, particularly when sarcomere length is increased beyond 3.0 μm, because sarcomere length can vary substantially along the length of a muscle fiber, and while it may appear that the sarcomere length of the region measured is 3.0 μm, other regions may contain sarcomeres which are shorter in length, have a greater degree of filament overlap, and are thus able to produce more force, which can cause an overestimate of the overall force. When errors due to these problems are reduced, the predicted length tension curves are obtained.

Finally, Gordon, Huxley, and Julian also observed that the maximum velocity of shortening is independent of sarcomere length above 2.0 μm, where overlap between thick and thin filaments is maximum. Therefore, this velocity must be independent of the numbers of crossbridges that attach and it must be related to the rate of movement of each crossbridge. That is, each crossbridge independently will have the same force velocity curve as all the crossbridges acting together.

Further Structural Approaches

Two structural techniques, electron microscopy and X-ray diffraction, have given rise to more ideas about the detailed structure of muscle, the crossbridge, and how contraction might occur. Electron microscopy of isolated myosin filaments showed that the crossbridge was likely to be heavy meromyosin and not light meromyosin because aggregates of LMM did not show projections, whereas aggregates of HMM did. These projections were of similar size to those seen in intact muscle (H.E. Huxley, 1963). Electron microscopy also showed the polarity of the thick filaments was such that crossbridges always pointed away from the middle of the thick filament (bare zone) indicating that all crossbridges in one half of the A-band have the same polarity which is opposite to that in the other half. This polarity is the one that is appropriate to pull the thin filaments towards the center of the A-band during muscle shortening.

X-ray diffraction of live muscle (H.E. Huxley and Brown, 1967) showed the structure of the thick and thin filaments and how they changed when the muscle contracted, or was put into rigor (in rigor muscle, ATP is absent from the muscle,

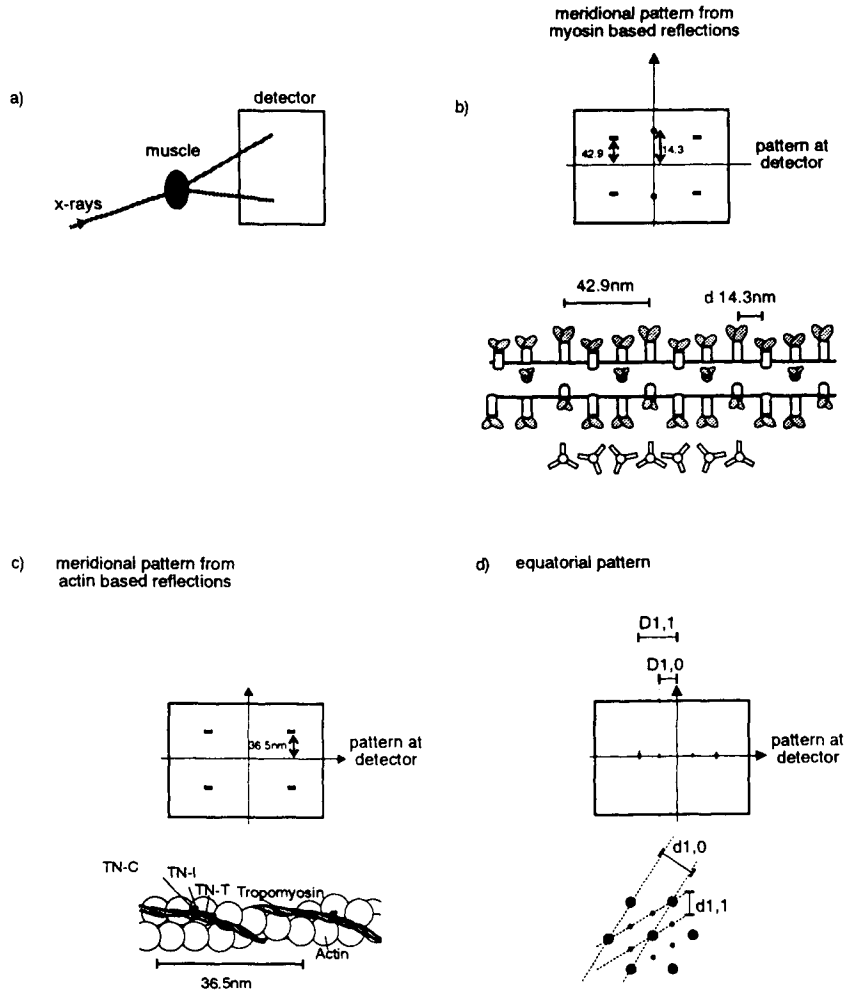


Figure 8. A schematic representation of the elements of the X-ray diffraction pattern from relaxed muscle. These reflections are interpreted to arise from various repeating structures in the muscle. Bragg's law, which states that

$$d = \frac{n\lambda}{2\sin(1/2\tan^{-1}(D/L))}$$

where d is the lattice spacing in the crystal/muscle, D the spacing of the diffraction spots at the detector, L the distance between the muscle and the detector, and n is the order, shows that the spacing of the spots at the detector is inversely proportional to the lattice spacing in the muscle. This figure shows the sources of some of the major reflections separately, but all these reflections would be seen together on a single two dimensional detector. **a**) X-ray diffraction of muscle. **b**) Meridional reflections from myosin. The most intense reflection seen along the meridian is the 14.3 nm spot.

and almost all the crossbridges bind very tightly to the actin filament). The X-ray pattern is split into two regions; the equator and the meridional. The equator runs horizontally with the page, and the meridional runs vertically (Figure 8a) for X-rays shone at a muscle positioned with the fibers (and hence the muscle sarcomeres) running vertically. The equatorial part of the pattern comes from repeating structures that are transverse to the muscle (a muscle cross-section), and the meridional pattern comes from repeating structures that run parallel to the muscle fiber axis.

The helical structure of the thick filament is such that at a given level there are three pairs of crossbridges projecting out from the thick filament. The next level is 14.3 nm distant and rotated by 120° with respect to the first level. This arrangement is repeated such that the first structure is repeated 3 × 14.5 nm later. These repeats can be seen as a 14.3 nm spot on the meridian and layer lines at 42.9 nm off the meridian (Figure 8b). The intensity of these layer lines in resting muscle showed that the crossbridges were not attached to actin but only stuck out partway. The thin filament is made up of G-actin monomers into a nonintegral helix. The subunit repeats of the helix are 5.46 nm long along either side of two chains which wind around each other and are staggered to each other by half a unit (2.73 nm). The crossover repeat is 36.5 nm giving a helix pitch of 73 nm.

Hugh Huxley's 1969 Theory

When a relaxed muscle (ATP present and crossbridges are not bound to actin) was put into rigor (ATP absent, all crossbridges are bound to actin), the main change in the meridional reflections that was observed by Huxley and Brown was that the myosin layer lines (42.9 nm) disappeared but the 14.3 nm meridional spot remained, and its spacing was unchanged. The actin based layer lines closest to the equator increased in intensity. From these observations, Huxley suggested that crossbridges could move away from the backbone of the thick filament and match up with the actin filament without losing their 14.3 nm spacing by swinging out and changing

Figure 8. (Continued). As described above, the packing of myosin molecules into the thick filament is such that a layer of heads is seen every 14.3 nm, and this reflection is thought to derive from this packing. Off the meridian the 42.9 nm myosin based layer line is shown. This arises from the helical pitch of the thick filament, due to the way in which the myosin molecules pack into the filament. The helical pitch is 42.9 nm. **c**) Meridional reflections from actin. Actin based layer lines can be seen at 35.5 nm, 5.9 nm and 5.1 nm (1st, 6th, and 7th layer lines) and they all arise from the various helical repeats along the thin filament. Only the 35.5 nm layer line is shown here.The 5.9 nm and 5.1 nm layer lines arise from the monomeric repeat. The 35.5 nm layer line arises from the long pitch helical repeat and is roughly equivalent to seven actin monomers. A meridional spot at 2.8 nm can also be seen. **d**) The equatorial reflections, 1,0 and 1,1 which arise from the spacings between crystal planes seen in cross section of muscle.

their radial and aximuthal positions to attach to actin without changing their axial positions (H.E. Huxley, 1969). Similar, but smaller, changes were observed when a relaxed muscle was activated. The myosin based layer lines decreased in intensity but the decrease was less than when the muscle was put into rigor. The 14.3 nm meridional spot decreased in intensity by a smaller amount than the myosin based layer lines. Finally, the actin based layer lines increased in intensity but the increase was smaller than when the muscle was put into rigor.

Huxley suggested that crossbridges can move out in this way and bind to actin because S-2 of HMM acted as a flexible link between LMM in the thick filament backbone and S-1. This was based on the observation that heavy meromyosin could be digested by chymotrypsin into two further subfragments (Lowey et al., 1966), S-1 and S-2, as described above, and that S-1 contained the ATPase and actin binding sites, whereas S-2 did not; moreover, S-2 did not self-aggregate, as did the rod or LMM portion of myosin.

The ability of S-2 to act as a flexible link also explained another problem in muscle contraction. When muscle contracts its volume remains constant. As a muscle shortens, and the filaments slide past each other, the spacing between the filaments increases as part of this constant volume behavior. Therefore, the cross-bridges have to be able to interact with actin over a wide range of filament spacings. The presence of the flexible link in S-2 would allow this to occur.

Finally, this mass transfer between the thick and thin filaments, and the change in radial distribution of crossbridges between relaxed and rigor muscle, was confirmed by looking at the intensities of reflections in the equatorial pattern (Hanson, 1968). The most intense reflections observed for the X-ray diffraction pattern of muscle are seen along the equator; they arise from the ordered arrangement of the myosin and actin containing filaments into a crystalline lattice in the muscle fiber (Figure 8b). The two main X-ray reflections or spots are the 1,0 and the 1,1 reflections which arise from the 1,0 and 1,1 crystal planes. From simplifying Bragg's law, which is used to derive the spacing of the crystal planes, it can be shown that the lattice spacing in the crystal, (d), is inversely proportional to the spacing of the diffraction spots at the detector (D). That is, if the lattice expands, the spacing of the spots will decrease.

Changes in the intensities of the 1,0 and 1,1 reflections have been interpreted to arise from a large transfer of mass of about 30% from the thick to the thin filaments when relaxed muscles are put into rigor. In relaxed muscle, the intensity of the 1,0 reflections is greater than that of the 1,1 reflections. When relaxed muscle is put into rigor, the intensity of the 1,0 reflection decreases and the intensity of the 1,1 reflection increases. Similar but smaller changes are observed when the muscle contracts. These intensity changes are reversed when the muscle relaxes. They were interpreted to show that crossbridges do indeed move out radially and attach to actin when relaxed muscles are put into rigor (or activated).

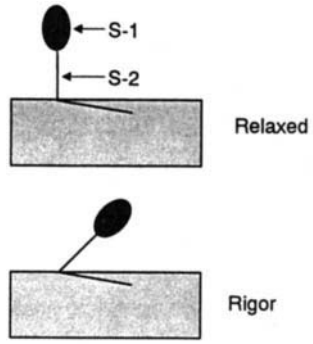


Figure 9. A schematic representation of crossbridge orientation assumed from electron micrographs of insect flight muscle in relaxed and rigor states by Reedy et al. (1965). The crossbridge is thought to have an orientation of 90° to the thick filament axis in the relaxed state and an orientation of 45° to the thick filament axis in rigor.

Finally, from observations of crossbridges in electron micrographs, the idea arose that while S-2 was acting as a flexible link, S-1 was able to rotate, or change its angle relative to actin, and that it was this rotation that caused S-1 to move the actin filament, or produce force (Reedy et al., 1965). The best electron micrographs of muscle come from the insect flight muscle of a giant water bug (*Lethocerus*), because the structure of these muscles is more nearly crystalline than that of any vertebrate muscles. Electron micrographs of these muscles in relaxed and rigor states showed that the crossbridges appeared to stick out at 90° to the thick filament axis in relaxed muscle and were more angled at 45° to the thick filament axis in rigor muscle (Figure 9). The conclusion was that the crossbridges attached to actin at 90°, generating force or shortening when they changed their orientation to 45°, finally detaching from the thin filaments. The flexible link of S-2 could allow the kinds of strain produced by this orientation change to be transmitted.

In summary, by the end of the 1960s, a combination of mechanical, biochemical, and structural approaches had shown that muscle contraction arises from the cyclical interaction of myosin crossbridges with actin. The myosin crossbridges are formed by the S-1 part of myosin which contains both the ATP binding site and the actin binding site, linked to the thick filament backbone by S-2. Force generation was suggested to occur by a rotation of the crossbridge while it is attached to actin.

TRANSIENT MECHANICAL PROPERTIES

Velocity Transients

The next step was to increase the time resolution of the velocity measurements to look at early changes in the velocity record (Civan and Podolsky, 1966). When

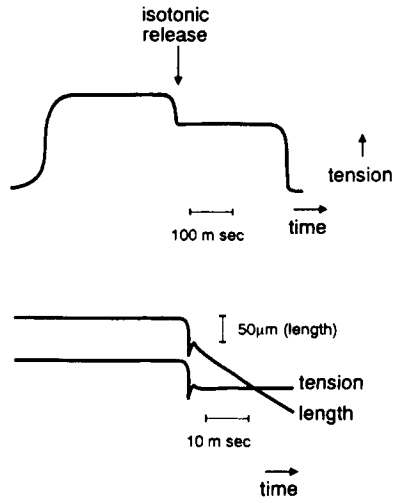


Figure 10. A schematic representation of the complex time-course for changes in velocity (change in length/time) which occur when the length of an isometrically contracting muscle is suddenly released and the muscle is then allowed to shorten against a constant new tension level. Following the release, the velocity oscillates before approaching a constant value. The upper diagram shows the force only on a slow timescale. The lower diagram shows the force and length changes on a rapid timescale.

the tension is released quickly, the initial changes in the velocity of shortening which occur after the tension has reached its new level follow a complex time course (Figure 10). This velocity transient shows that the crossbridges take a finite, measurable, amount of time to change from being in a steady-state before the release, when the muscle was contracting isometrically and the velocity of shortening was zero, to entering a new steady-state following the release, when the velocity of shortening has reached a constant rate. However, velocity transients are more difficult to interpret in terms of crossbridge behaviors than the tension transients described below.

Tension Transients

In this case, the length of an isometrically contracting muscle is changed rapidly by a small amount (+/−0.3 to 1.5%) and the resulting change in tension is measured (Figure 11a). The change in tension which follows a rapid length change has four phases. In phase 1, there is a rapid change in tension which occurs at the same time as the change in length. This is due to an elastic element in the crossbridge. During phases 2–4, tension recovers towards its original level before the length change and

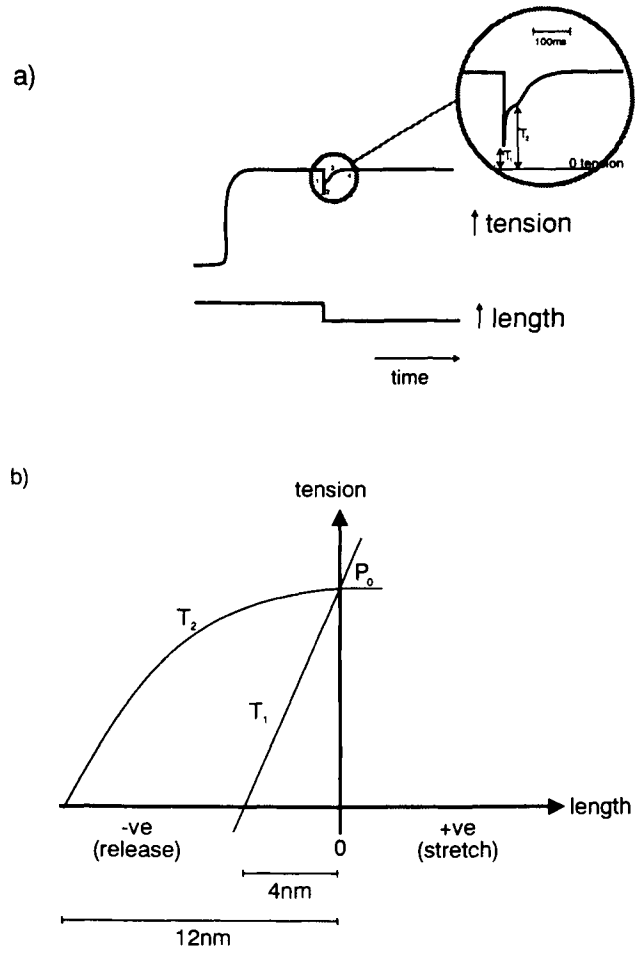


Figure 11. **a**) A schematic representation of experiments which show the tension change in response to a rapidly applied length change in isometrically contracting fibers. The tension change is shown in more detail in the inset. The tension response can be described by four phases. In the first phase the tension changes instantaneously with the change in length. T1 is the tension difference measured between the isometric tension (P_o) and the maximum tension change (t1) during the length step: ($T1=P_o-t1$) In the second phase, immediately following the length step, tension recovers to a new level (T2). **b**) A diagram which shows the relationship between T1 and T2 measured for different sizes of length step, mainly in this case for releases.

the time courses and amplitudes of these phases can be related to steps between crossbridge states. If a similar length change is given to rigor muscle, only phase 1, the elastic component, is seen. Because phases 2–4 do not occur, this implies that they arise from properties of actively contracting crossbridges.

Phases 1 and 2 were measured by Huxley and Simmons for different sizes of stretches and releases. The amplitude of the force change produced by the sudden change in length during phase 1 of the tension transient is linearly related to the amplitude of the length change, and the slope of this plot (the T1 curve) measures the stiffness of the crossbridges (Figure 11b). The elastic property (or stiffness) of the crossbridge can be described by Hooke's law which states that the force (F) required to extend a spring a certain distance (x), is directly proportional to the stiffness of the spring (k) and the amount of extension (x); $F = kx$. The change in force is therefore linearly related to the extension of the spring, and the slope of a plot of F against x is the stiffness k. When this curve is measured for muscles at different sarcomere lengths, in which the overlap between thick and thin filaments is different, it is found that stiffness is proportional to the amount of overlap, and therefore to the numbers of interacting crossbridges. This is the only evidence that this elasticity must arise from the crossbridges. Unlike the T1 curve, the T2 curve, the relationship between the amplitude of force recovery during phase 2 to the size of the step, is highly nonlinear. For releases, as the size of a release increases, the amplitude of the tension recovery is reduced until the size of the release reaches about 12–14 nm per half sarcomere. The rate of this tension response also varies with the size of the step. This tension recovery step is so fast that it is unlikely to be caused by a change in the number of attached crossbridges. The form of the T2 curve suggests that it might be due to active recovery of force by an element of finite extent. This could be something like a rotation in the crossbridge.

A.F. Huxley and Simmons's 1971 Theory

Huxley and Simmons were able to model the T2 curve by assuming that the crossbridge produces isometric work in a small number of steps and that the elasticity of the crossbridge is such that the crossbridge can go through these steps without displacing the thick and thin filaments relative to each other. The steps could be seen as a progressive rotation of the crossbridge which fitted nicely with Hugh Huxley's model of the rotating crossbridge. They modeled the transients using three steps of 4 nm each to give 12 nm movement in total. A model of the complete crossbridge cycle would also include the attachment and detachment processes of the Huxley 1957 model.

Muscle Stiffness

Is the stiffness measured by the T1 curve proportional to the number of crossbridges attached to thin filaments? In a rigor muscle, it might be expected that all the crossbridges are attached because the binding constant of myosin S-1 for actin is very high in the absence of ATP. The stiffness of a rigor muscle is indeed very high, much more than that of a relaxed muscle. The stiffness of an actively

contracting muscle held isometric is about 50% to 70% of that measured for a rigor muscle. This would be expected if the crossbridges are cycling as not all the crossbridges would necessarily be attached to actin at the same time. When a muscle shortens, the stiffness measured is proportional to the velocity of shortening and is highest for an isometrically contracting muscle, but falls to about 40% of this value when the muscle is shortening at its maximum velocity. Again, it might be expected that as the muscle shortens more and more quickly, crossbridges will remain attached to actin for shorter times. If this did not happen, the attached crossbridges would resist the shortening of the muscle. Finally, stiffness measured during the rise of tension at the start of a tetanus increases before the increase in tension. From the time courses of the increase in stiffness and tension related to the start of stimulation it can be estimated that detached crossbridges attach to actin at a rate of $20s^{-1}$ but initially attach in a state that does not produce force. The attached crossbridges then move into a state in which they produce force at a rate of about $100\ s^{-1}$.

APPROACHES TO INVESTIGATE THE ENERGY SOURCE OF MUSCLE CONTRACTION

How is the generation of force by this cyclical interaction of myosin crossbridges with actin linked to the ATPase cycle? When actin, myosin and ATP are mixed together, the ATP is hydrolyzed to ADP and P_i. The energy produced by this hydrolysis drives contraction in muscle. The mechanism of ATP hydrolysis may give clues as to how muscle uses energy to do work. For example, it might be predicted that the step during the ATPase cycle during which most of the free energy change occurs must happen just before or during the working stroke of the crossbridge. Since the mixture of actin and S-1 of myosin (acto-S1-1) is soluble at physiological ionic strength, ATP hydrolysis of acto-S-1 can be studied in solution. (Ionic strength is the sum of 1/2[concentration of ions × the charge on each ion]). Solution biochemistry, which is used to study the biochemical kinetics of the actomyosin ATPase, often uses the standard conditions: rabbit actomyosin, 15 °C, pH 7 and an ionic strength of 0.02 M. (Physiological ionic strength is about 0.17 M.)

Myosin as an ATPase

Myosin alone can act as an ATPase. The change in free energy when ATP is hydrolyzed to ADP and P_i is about 50 kJ per mole, and the overall steady-state rate is very slow, about $0.03s^{-1}$. The time course of ATP hydrolysis by myosin was first measured by a technique known as quenched-flow in the late 1960s (Eisenberg and Moos, 1968; Eisenberg and Moos, 1970; Lymn and Taylor, 1971). Myosin S-1 and ATP are mixed rapidly, the reaction is stopped by acid quenching at different times after the mixing, and the mixture is analyzed for ATP and P_i concentrations. These

experiments showed that when myosin S-1 and ATP are first mixed, P_i concentration increases quickly so that about 0.7 moles P_i per mole of myosin S-1 is produced before measurements can be taken. This rapid increase in P_i has been called the phosphate burst. The phosphate burst is followed by a slow steady-state increase in P_i of about 0.03 $S\text{-}1^{-1}$ s^{-1}. Thus initial hydrolysis of ATP is rapid, but there is then a slow step following hydrolysis which causes the slow steady-state rate. The equilibrium constant for ATP is about three in this experiment, determined by using radioactively labeled ATP and measuring the ratio of ^{32}P to ^{32}ATP.

Because the equilibrium constant is close to one, this also means that the free energy does not change much when ATP is hydrolyzed. Most of the fall in free energy is associated with ATP binding to myosin as the equilibrium constant for this step is about 10^{10} M^{-1}. The binding energy is used to dissociate myosin from actin.

As ATP binding to myosin, and ATP hydrolysis, are both faster than the overall observed ATPase rate, the slow step that follows the rapid phosphate burst and that must limit the overall observed ATPase rate must be the release of phosphate, or the release of ADP. The rate at which ADP is released was measured by a displacement technique (Trentham et al., 1972), in which the rate at which ADP bound to S-1 is displaced by ATP was measured. This experiment showed that the rate of ADP release (2 s^{-1}) is greater than the overall rate of hydrolysis (0.03 s^{-1}). Thus the release of P_i, and not that of ADP, is rate limiting.

Activation of Myosin ATPase by Actin

Actin can activate the myosin ATPase (0.03 s^{-1}) by more than 100 times to about 5 s^{-1} at high actin concentrations (Eisenberg and Moos, 1968, 1970; Lymn and Taylor, 1971). This value is similar to the ATPase measured in contracting muscle. As the concentration of actin in the mixture is increased, the activation of the myosin ATPase increases up to a maximum. The effective concentration of actin in a muscle fiber is probably about 500 μM, but actin concentrations this high cannot be used in solution as actin filaments are very viscous, and rapid mixing experiments would not work very well. Lower concentrations of actin (about 50 μM) have to be used in solution experiments, but to obtain the same high levels of activation of the myosin ATPase as would be obtained in a muscle fiber (physiological ionic strength of about 0.17 M), a low ionic strength (0.02 M) has to be used. At this low ionic strength binding between actin and myosin is tighter and the ATPase is maximally activated. Under these conditions, there is a hyperbolic relationship between the ATPase and the actin concentration, which follows Michaelis Menten kinetics such that: $V = V_{max}\,[Actin]/(K_m + [Actin])$, where V is the measured ATPase rate for the actin concentration used [Actin], V_{max} is the maximum rate and K_m is the concentration of actin for which the ATPase rate is half maximal. In this experiment V_{max} is 5 s^{-1} and K_m is 20 μM.

How does actin activate the ATPase? Actin must bind to myosin during some part of the ATPase cycle and in doing so accelerate the release of P_i, which limits the myosin ATPase rate. When ATP is absent, myosin and actin in solution bind very tightly with a binding constant of about $10^8\ M^{-1}$. When ATP is added, myosin and actin bind much more weakly. The activation of the ATPase by actin was investigated using quenched-flow experiments as above in which actin, myosin S-1, and ATP are rapidly mixed (Lymn and Taylor, 1971). However, in these experiments, measurements of light scattering and fluorescence were made continuously to monitor the degree of association of actin and myosin and the progress of ATP hydrolysis. Light scattering is higher when actin and myosin are bound than when they are dissociated. Fluorescence of A.M.ADP.P_i and M.ADP.P_i are higher than that of A.M.ATP and M.ATP. Therefore, by making both these measurements, it could be shown whether actin and myosin dissociate before ATP is cleaved (K_3, in the scheme below) or after (K_4).

$$\begin{array}{ccccc} & \overset{K_1}{} & & \overset{K_2}{} & \\ \text{A.M.} + \text{ATP} & \Leftrightarrow & \text{A.M.ATP} & \Leftrightarrow & \text{A.M.ADP}\cdot P_i \\ & & K_3 \Uparrow\Downarrow & & \Uparrow\Downarrow K_4 \\ \text{M} + \text{ATP} & \underset{K_5}{\Leftrightarrow} & \text{M.ATP} & \underset{K_6}{\Leftrightarrow} & \text{M.ADP}\cdot P_i \end{array}$$

These experiments showed that when actomyosin-S1 and ATP are mixed, a phosphate burst is still observed, which is similar to that seen for myosin-S-1 alone. Therefore, as for myosin, hydrolysis of ATP occurs before the rate limiting step for actomyosin ATPase. Light scattering and fluorescence measurements showed that actomyosin-S-1 dissociates much faster (about $700 s^{-1}$ at high ATP concentrations) than the hydrolysis rate (about $20\ s^{-1}$, at high ATP concentrations) and the size of the change in light scattering suggested that actin and myosin become completely dissociated.

The simplest mechanism to explain the much faster rate of dissociation of actomyosin-S-1 by ATP than that of ATP cleavage is that actin activates the myosin ATPase by accelerating the rate at which ADP and P_i are released. That is: when ATP is added to actomyosin-S-1, ATP rapidly binds and dissociates actomyosin, myosin ATPase then hydrolyzes ATP to form myosin-ADP.P_i, this state then reattaches to actin and phosphate is released much faster from actomyosin .ADP.P_i than it is from myosin.ADP.P_i, as shown in the scheme below:

$$\begin{array}{ccccccc} & \overset{10^6\ M^{-1}s^{-1}}{} & & \overset{10\text{–}20\ s^{-1}}{} & & & \\ \text{A.M.} + \text{ATP} & \Rightarrow & \text{A.M.ATP} & \text{A.M.ADP}\cdot P_i & \Rightarrow & \text{A.M.ADP} & \Rightarrow \text{A.M.} \\ & >700\ s^{-1} \Downarrow & & \Uparrow\ 3\times10^5\ M^{-1}s^{-1} & & & \\ & \text{M} + \text{ATP} & \underset{20\ s^{-1}}{\Rightarrow} & \text{M.ADP}\cdot P_i & \underset{0.03\ s^{-1}}{\Rightarrow} & \text{M.ADP} & \Rightarrow \text{M.} \end{array}$$

Lymn and Taylor Model 1971

Lymn and Taylor suggested this scheme from their experimental evidence in 1971, showing that actin activates the myosin ATPase by accelerating the rate at which ADP and P_i are released. They then derived a model to explain how force production occurs, which linked the biochemical measurements with the ideas put forward from structural measurements by Hugh Huxley mentioned above. In this model (Figure 12), S-1 binds to actin at an angle of 45° to the long thin filament axis at the end of ATP hydrolysis when both ADP and P_i have been released (as in the rigor muscle, where ATP is absent). When ATP binds to this complex, S-1 detaches very quickly. Myosin-S-1 hydrolyzes ATP and myosin-S-1.ADP.P_i then re-attaches to actin at an angle of 90° (the angle of S-1 thought to be seen in relaxed muscle). Force production occurs during the power stroke which is correlated with the release of phosphate (the rate limiting step in the ATPase cycle) and the change in angle of the myosin crossbridge from 90° to 45°. A consideration of the energetics of the ATPase shows that there is only a small energy change associated with ATP hydrolysis itself, but a large decrease in energy associated with phosphate release, and thus with the power stroke of the crossbridge.

Although this model is very attractive because of its simplicity, further experiments have shown that there are two problems with it (Eisenberg and Hill, 1985). First, myosin.ATP binds as well to actin as does myosin.ADP.P_i such that there is a rapid equilibrium between actomyosin.ATP and myosin.ATP, and between

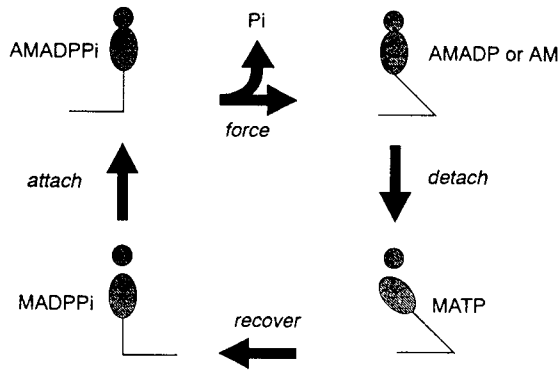
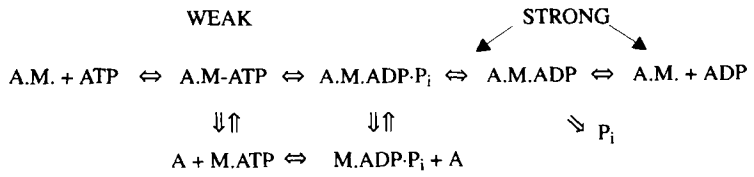


Figure 12. A schematic diagram of the Lymn-Taylor model which links structural changes in the myosin crossbridge (see Figure 9) with the ATP cycle. From the bottom right, detached crossbridges (M.ATP) hydrolyze their ATP (M.ADP.P_i). Following hydrolysis, the crossbridges can attach to actin in a perpendicular orientation (AM.ADP.P_i). Force generation occurs as phosphate is released, and the crossbridge makes the transition from a perpendicular orientation to an orientation of 45° (AM.ADP or AM.). Finally, ATP binds to this state and causes the crossbridge to detach (M.ATP).

actomyosin.ADP.P_i and myosin.ADP.P_i. This was shown from experiments which used higher concentrations of actin (80 μM) where it was found that dissociation of actomyosin by ATP was still rapid (> 700 s^{-1}) but incomplete, suggesting that the dissociation was reversible. Second, ATP could be hydrolyzed while actin was still associated with myosin; myosin did not have to dissociate from actin for ATP to be hydrolyzed. Furthermore, by varying actin concentration the association constant of A.M.ATP (A.M.ATP ⇔ M.ATP) was measured to be about 10^4 M. This association constant is much less than that for A + M ⇔ A.M. This then gave rise to the idea that myosin could either bind weakly to actin (if it had ATP bound to it), or strongly (if it did not have ATP bound to it). The Lymn-Taylor scheme was thus revised by Eisenberg and collaborators to give the scheme described below.

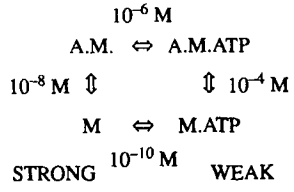
Eisenberg and Hill Model 1985



The Lymn-Taylor model linked the ATPase scheme with structural changes in the myosin crossbridge. Does the present scheme have an easy link with structural changes? The release of phosphate by actomyosin (A.M.ADP.P_i ⇔ A.M.ADP) could still be associated with a structural change in the crossbridge and force production such that A.M.ADP.P_i is the 90° state and A.M.ADP is the 45° state. There is then the potential problem that in this revised scheme, the structural change could be reversed when ATP binds to actomyosin and hydrolysis occurs without the detachment of myosin. This would mean the crossbridge would be pushed the wrong way and negative force produced. However, because the binding of actin to myosin in the presence of ATP is weak, crossbridges which try to push the wrong way (and are highly strained), would detach quickly and rebind to a different actin in which the actomyosin attachment has a lower force and strain.

How do the Predicted Free Energy Changes Fit With the Revised Model?

The binding of ATP to actomyosin weakens the binding of actin to myosin such that the dissociation constant decreases from 10^{-8} M to 10^{-4} M. There is thus a



competitive binding reaction between actin and ATP to myosin. In terms of free energy changes there is a relatively small ΔG for ATP binding to actomyosin, and there is very little free energy change associated with the hydrolysis of ATP to ADP and P_i; this step is close to equilibrium. Actomyosin release of ADP and P_i is energetically more favorable than release of ATP as the free energy of ATP in solution is much higher than that of ADP and P_i. Most of the change in free energy is associated with release of ADP and P_i. This fits with the crossbridge model in which it is expected that product release should be tightly coupled to mechanical work, and mechanical work is derived from the free energy change.

All these experiments were carried out with actin and myosin in solution, either using moderate ionic strength and low actin concentrations (Lymn and Taylor, 1971) or using low ionic strength and high actin concentrations (Stein et al., 1979). However, in muscle both the actin concentration and the ionic strength are high. To confirm these models, these same experiments need to be carried out in muscle fibers.

BIOCHEMICAL EXPERIMENTS WITH FIBERS

Skinned muscle fibers can be used for biochemical experiments. The muscle membrane is mechanically or chemically removed, which allows diffusion of chemicals into the solution surrounding the myofibrils, and so the biochemical environment of the myofibrils can be controlled. Skinned muscle fibers are activated by direct addition of Ca^{2+} in the presence of ATP, rather than by electrical stimulation which causes internal Ca^{2+} to rise in intact fibers. Indeed, the fiber can no longer be electrically stimulated once the membrane has been removed. It has been found that isometric tension and the force-velocity curve depend on MgATP concentration using skinned fibers (Ferenczi et al., 1984; Cooke and Pate, 1985). In particular, the maximum velocity of shortening (V_{max}) is dependent on MgATP concentration which shows that V_{max} is dependent on the rate at which crossbridges detach from actin. It had already been observed that the maximum velocity of shortening of different muscles was correlated with the ATPase rate measured for isolated actomyosin in solution in different muscles and therefore related to a step in the actomyosin ATPase cycle. The measurements are all steady-state measurements, as it takes some time for ATP to diffuse into the fiber when the solution is changed. The rate of diffusion limits the rate that measurements can be made. For example, the equivalent of a phosphate burst cannot be measured in fibers in this way as ATP does not diffuse into the fiber fast enough.

"Caged-ATP"

However, by carrying out experiments with skinned fibers, the composition of the solution surrounding the myofibrils can be controlled and the mechanical properties of the muscle fiber can be related more easily to the biochemistry of force

production (the ATPase of the fiber). To overcome the problem of diffusion, a compound called "caged-ATP" (P^3-1-(2-nitro)phenylethyladenosine 5-triphosphate) has been synthesized (Kaplan et al., 1978). This is a biologically inert molecule which can diffuse into fibers and equilibrate. Subsequent illumination by a high energy pulse of UV light cleaves caged-ATP to release ATP very rapidly ($118 s^{-1}$ at 20 °C). It is then possible to measure the rapid mechanical and biochemical responses to a sudden increase in ATP concentration.

Caged-ATP is usually added to a skinned muscle fiber in rigor (ATP absent). Following photolysis by UV light, ATP is produced and can be hydrolyzed. The fiber can either relax (if Ca^{2+} is absent) or contract (if Ca^{2+} is present). This allows rapid kinetic processes to be measured which can relate initial changes in force to the initial hydrolysis of ATP. When caged-ATP is released in the absence of Ca^{2+}, the tension and stiffness fall to the relaxed level (Goldman et al., 1982). The appearance of the tension transient depends on the initial force level in rigor before caged-ATP photolysis. If the force level is low, tension may transiently rise before it finally falls to the relaxed level. If the force level is high, tension falls exponentially to the relaxed level. Therefore, the fall in tension is related to the initial strain on the crossbridges in rigor before ATP release. In all cases there is a delay before tension decreases (the hump) immediately after caged-ATP photolysis. This is thought to be due to the ability of some of the rigor crossbridges that have bound ATP and detached from the thin filament to reattach and generate active tension. The crossbridges can do this because the thin filament is kept switched on by the remaining attached rigor crossbridges. Eventually enough rigor crossbridges bind ATP and detach to switch off the thin filament such that tension falls to the relaxed level. When caged-ATP is released in the presence of Ca^{2+}, the tension and stiffness initially drop (Goldman et al., 1984). The tension then rises to the higher active level.

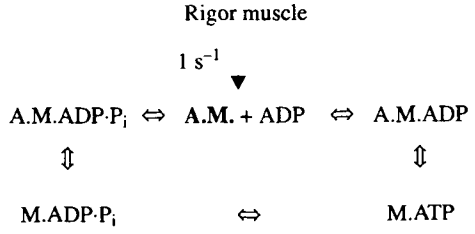
Effect of Phosphate on Force Generation

The solution biochemical studies described above suggest that force generation is probably associated with the phosphate release step. Experiments with caged-ATP in muscle fibers also suggest that the rate of force generation is either determined by the rate of ATP hydrolysis or a subsequent step such as phosphate release; viz. the transition from A.M.ADP.P_i to A.M.ADP + P_i. In rigor most, if not all, of the crossbridges are expected to be in the AM state (actomyosin with no ATP bound) and when ATP is released, the crossbridges must first bind ATP to form AM.ATP (see below), then hydrolyze ATP (on the Eisenberg and Hill model described above, detachment of crossbridges from actin before hydrolysis occurs is not necessary) before they can produce active force. However, the rise in force following ATP release into a rigor fiber in the presence of Ca^{2+} by photolysis of caged-ATP is much faster (about $100 s^{-1}$ at 20 °C for rabbit fibers) than the overall

cycling time for ATP hydrolysis of 1 s^{-1}. Therefore, the crossbridges can actively generate force before the ATPase cycle is complete.

The first of these steps, the rate of binding of ATP to rigor crossbridges, is about 1,000 s^{-1} (from solution biochemistry, see above) which is also much faster than the overall cycling time of 1 s^{-1}. This rate is derived from the rate constant for the binding of ATP to A.M. (A.M. + ATP $\Leftrightarrow$ A.M.ATP) which is $10^6 M^{-1}s^{-1}$ and the amount of ATP released in these experiments of about 1 mM (physiological concentration of ATP is about 5 mM). This rapid binding of ATP to rigor crossbridges means that in the presence of ATP, the A.M. (rigor) state is unlikely to remain occupied for a substantial length of time and is unlikely to be the force generating state in actively contracting muscle.

The second of these steps, the rate of hydrolysis at 12 °C, measured by analysis of the ADP content in fibers rapidly frozen at different times after ATP release from caged-ATP, is 40–60 s^{-1} (Ferenczi, 1986) which is similar to the rate of active force increase at 20°C once the difference in temperature has been accounted for. This rate is similar to that measured in solution however, and is not rate limiting in solution. Therefore, force generation in the caged-ATP experiments could be limited by hydrolysis, or more likely, by a step following hydrolysis such as P_i release. The idea that release of phosphate is linked to force production in muscle



can be tested by increasing the concentration of phosphate in activating solution. When the concentration of phosphate was increased up to 60 mM P_i, the amount of tetanic force a skinned fiber produced decreased linearly with the logarithm of the phosphate concentration (Cooke and Pate, 1985). When phosphate concentration is increased in this way, the equilibrium:

$$A.M.ADP.P_i \Leftrightarrow A.M.ADP. + P_i$$

is shifted towards the lefthand side, that is towards the A.M.ADP.P_i state. If force generation is associated with phosphate release, then A.M.ADP would have a higher force than A.M.ADP.P_i and shifting the equilibrium towards A.M.ADP.P_i would therefore cause the force to be decreased as the numbers of lower force A.M.ADP.P_i states are increased. This is exactly what is observed when the phosphate concentration in actively contracting skinned fibers is raised. Furthermore, there is a large free energy change associated with the release of phosphate, and so this step is a good candidate for the step in which force is produced.

Since the introduction of caged-ATP, many other caged-compounds have been synthesized, including caged-P_i. When P_i is photolytically released from caged-P_i in an isometrically contracting muscle fiber, a rapid reduction in force is observed that is again dependent on P_i concentration (Dantzig et al., 1992). The rate of this reduction is also dependent on P_i concentration. These experiments suggest that there are two steps involved in force generation and P_i release. First a nonforce generating AM.ADP.P_i crossbridge forms which then isomerizes to form a force exerting crossbridge state AM'.ADP.P_i. Phosphate is then released to form the second force generating state AM'.ADP. All three states would be significantly populated during active contraction.

In summary, therefore, solution and fiber biochemistry have provided some idea about how ATP is used by actomyosin to generate force. Currently, it seems most likely that phosphate release, and also an isomerization between two AM.ADP.P_i states, are closely linked to force generation in muscle. ATP binds rapidly to actomyosin (A.M.) and is subsequently rapidly hydrolyzed by myosin/actomyosin. There is also a rapid equilibrium between M.ADP.P_i and A.M.ADP.P_i (this can also be seen in fibers from mechanical measurements at low ionic strength). The rate limiting step in the ATPase cycle is therefore likely to be release of P_i from A.M.ADP.P_i, in fibers as well as in solution, and this supports the idea that phosphate release is associated with force generation in muscle.

"Weak" and "Strong" Crossbridges

Is there any structural evidence for weak and strong crossbridges? The original Lymn and Taylor scheme (Figure 12) suggested that the 90° crossbridges were prepower stroke (weak) crossbridges and 45° crossbridges were postpower stroke (strong) crossbridges with force generation occurring with the 90° to 45° transition, but this suggestion was derived from electron micrographs of insect flight muscle and in vertebrate muscle relaxed crossbridges are at an angle of 30° to the thick filament axis, not 90°. A different structure is reported for crossbridges in muscle fibers in low ionic strength relaxing solution (Brenner et al., 1984), which may mimic weak crossbridges, and could look more like 90° crossbridges, but there is no evidence that these crossbridges are present at physiological ionic strength. There is no reason why structures seen in relaxed or rigor muscle should be present in an actively contracting muscle. What is really needed is to define the structure of the crossbridge in active muscle, but this is proving very difficult, probably partly because the crossbridges work asynchronously, and a number of different states are populated at any one time. Attempts to synchronize the crossbridges have been made by using caged-ATP techniques (although synchrony is lost very quickly after ATP release), or by mechanical methods.

Both mechanics and biochemical studies of muscle fibers suggest there is more than one force generating state. Crossbridges are thought to attach initially as

weakly bound nonforce generating crossbridges, go through some kind of isomerization step to produce force ($AM.ADP.P_i \Leftrightarrow AM'.ADP.P_i$), release phosphate and enter a new force generating state, then finally release ADP before rebinding ATP and entering a nonforce generating state again.

MOTILITY ASSAYS

An even more recent development to try to link the biochemical properties of isolated actin and myosin in solution with the abilities of these two proteins to interact to produce force and movement is that of the motility assay. Initially, the plant *Nitella* was used (Sheetz and Spudich, 1983). This can be opened up to reveal long actin cables running from end to end along the plant cells which are normally used for organelle transport. Myosin can be bound onto beads about 1 μm in diameter and then added to the cables, with ATP as an energy source. The movement of the beads, driven by the interaction of myosin on the beads with the actin cables, could be seen in a microscope. The motility assay was then made more sophisticated. Commonly, myosin is now stuck down onto a glass coverslip coated with nitrocellulose or siliconized and the movement of fluorescently labeled actin filaments over the myosin on this surface is observed in the light microscope (Figure 13) (Yanagida et al., 1984; Kron and Spudich, 1986). The actin filaments are generally labeled with rhodamine-phalloidin which binds to actin. Rhodamine is a fluorescent compound so the actin filaments can be visualized with fluorescence microscopy. Using this type of motility assay it has been shown that myosin S-1 alone can support motion of the actin filaments, although the velocity of movement of the filaments is lower than that observed for heavy meromyosin. This is again confirmation that all the contractile properties of myosin reside in the head, or S-1, of myosin.

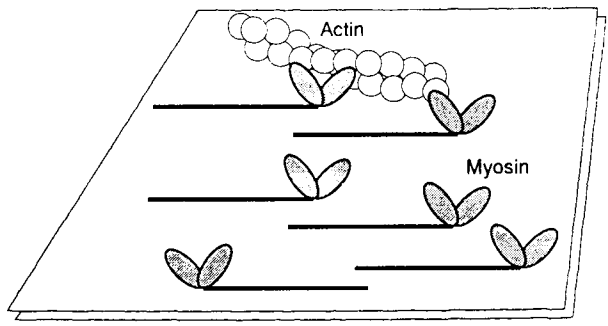


Figure 13. A schematic diagram of the motility assay. Myosin molecules (HMM or S-1 are also used) stick to glass coverslips coated with nitrocellulose. Actin, in solution, is then added to the glass coverslip and it binds to the myosin molecules. When ATP is added, actin can move over the surface, propelled by the myosin molecules.

Using this type of assay, further estimates of the working stroke of the crossbridge have been made. However, in this case the working stroke really means how far a single crossbridge can move an actin filament per ATP molecule hydrolyzed, rather than how far a single crossbridge can remain attached and generate force as discussed above. Estimates of the working stroke using the motility assay range from 10 nm (Uyeda et al., 1991) to 100 nm (Harada et al., 1990), or more. It is hard to see how a crossbridge, which is only 15–20 nm long and 4 nm wide could remain attached to actin (5 nm per monomer) over a distance of 100 nm, particularly as the mechanical transient measurements described above indicate that it can only remain attached and generate active force for about 12 nm (see Tension Transients, above). Instead, the crossbridge would have to attach many times over 100 nm and yet still only hydrolyze one ATP molecule. The major problem in determining whether this is possible or not lies partly in making estimates for this calculation, such as how many crossbridges can actually interact with the actin filament in the motility assay, and what the ATPase rate is for these interacting crossbridges, and lies partly in the conditions in which the motility assay are carried out. The ionic strength of the conditions is low, and the crossbridges are moving the actin filaments at maximum velocity as the ends of the actin filaments are not held. Mechanical transient measurements are made on isometric muscle at the higher physiological ionic strength and, in these conditions, the crossbridges are not moving actin filaments (shortening velocity is zero) but only exerting force on them. Most recently, optical traps have been used in conjunction with the motility assay to measure force and displacements arising from a single myosin molecule at 3–4 pN and 11 nm, respectively (Finer et al., 1994).

CRYSTAL STRUCTURES OF ACTIN AND MYOSIN SUBFRAGMENT-1

The crystal structures of myosin S-1 and of monomeric G-actin using co-crystals of actin-DNAse-1, were both solved (Kabsch et al., 1990; Rayment et al., 1993a) (Figure 14). Using both these crystal structures, a model of the actomyosin complex has also been proposed. Monomeric actin has four domains (Kabsch et al., 1990). Domains 3 and 1 are very similar. The four domains are held together by a nucleotide (ATP or ADP) and salt bridges. The myosin binding sites (somewhere in the region of amino acid residues 1–40 from various studies at the N-terminus of actin), are located on the outside of domain 1. How monomeric actin fits together to form a thin filament has also been modeled using this data to fit X-ray data from oriented gels of actin filaments (f-actin; actin filaments cannot be crystallized). The interaction between monomers in the filament appears to be through a loop in domain 1 (Holmes et al., 1990). The actin monomers form a double-stranded helical structure in the thin filament.

The crystal structure of myosin S-1 shows how the three subdomains (20K, 50K and 27K) of the myosin heavy chain (produced by further enzymatic digestion of

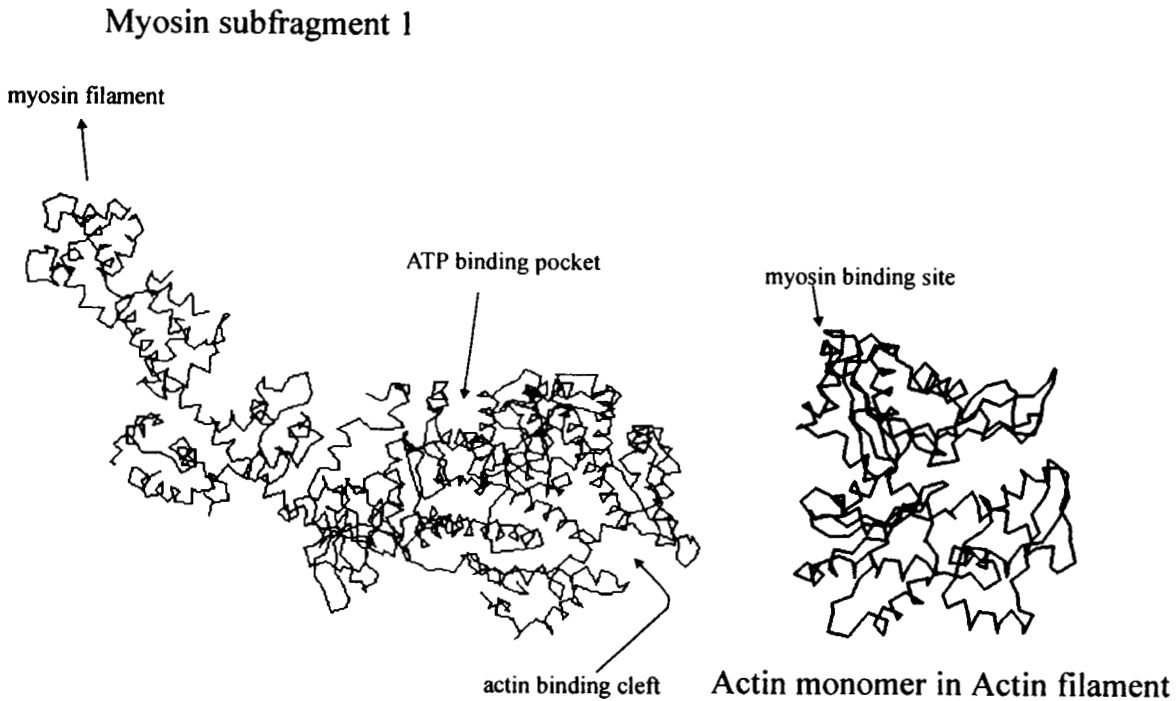


Figure 14. Crystal structures of actin and myosin subfragment-1, shown to the same scale.

S-1), fold into the head, the position of the cleft which binds ATP, and the position of the two different light chains (Rayment et al., 1993a). The molecule is 16.5 nm long, 6.5 nm wide and 4 nm thick. The 20K domain which joins S-1 to the S-2 portion of myosin runs the whole length of the head and contains a long α-helix that is 8.5 nm long, half the length of the molecule. This helix stretches from the end of S-1 where it joins S-2 (the tail region), up into the thicker part of the head. The two different light chains bind around the 20K domain in the tail region and stabilize the helix in this region. (Removal of the light chains does not inhibit ATPase activity but does eliminate motility.) The actin binding site is between the 50K and the 20K domains. The ATP binding site is contributed to by all three domains. The substrate pocket and the actin binding regions are on opposite sides of myosin separated by 3–4 nm.

There is also a cleft which separates the two halves of the 50 kDa domain which looks as though it must close when actin binds to myosin to form the actomyosin complex. In the interaction of myosin with actin, there appear to be three major contact sites (Rayment et al., 1993b; Schroeder et al., 1993). First, there is an ionic, or weak, interaction between basic residues on myosin and acidic residues on actin at the N-terminus. Second, there is a stronger hydrophobic interaction between a helix-loop-helix on myosin (50K domain) and a helix-loop-helix on actin. Finally, there is an interaction between myosin (residues 403–416) and two prolines at the C-terminus of actin. There is also a second interaction site on a second actin. Thus, there is a weak ionic binding site at the N-terminal region of actin, and a second stronger hydrophobic binding site. This fits with the idea above that binding of myosin to actin is a two step process or isomerization.

Rayment's 1993 Model

Can the crystal structures tell us anything about how myosin might change its orientation during force production? As already mentioned, this has been difficult to demonstrate in contracting muscle as the crossbridges work asynchronously, and an average distribution of crossbridge orientations can only be observed. With the caveat that the S-1 crystal structure was solved for the single conformation of S-1 in the absence of ATP, which means that if there are any conformational changes in S-1 during contraction, they can only be guessed at, then the following model can be proposed. First, there is communication between the actin binding cleft and the substrate pocket. When ATP binds to actomyosin in the substrate pocket, the substrate pocket could close and in turn cause the actin binding pocket to open, disrupting the binding of myosin to actin (Figure 15). Myosin.ATP and Myosin ADP.P_i are thought to bind weakly to actin (see above, "Caged-ATP"). Following ATP hydrolysis, when P_i is released from the substrate pocket while myosin is attached to actin, the substrate pocket could open and in turn cause the actin binding pocket to close, allowing myosin to bind tightly to actin. This release of phosphate

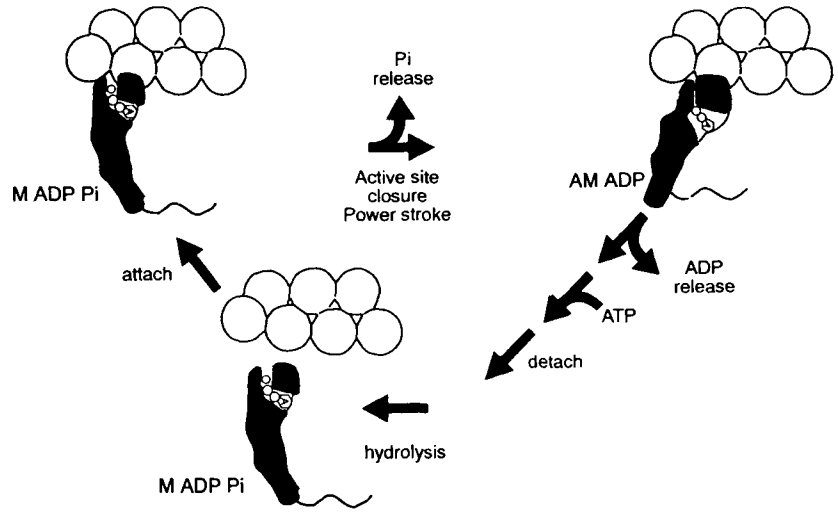


Figure 15. A schematic diagram adapted from the Rayment model which links structural changes in the myosin crossbridge with the actomyosin ATPase. In this particular case, there are two pockets or clefts, one which binds ATP and one which binds actin. As ATP is hydrolyzed and the hydrolysis products are released, the shape of the substrate cleft changes and in turn alters the shape of the actin binding cleft. The change in shape of the actin binding cleft alters the ability of myosin to bind to actin from a weak binding (AM.ADP.P_i) to a strong binding (AM.ADP) state, that in turn causes the long α-helix that forms the tail of S-1 (in the 20 kDa domain) to bend. This scheme shows how the actomyosin ATPase might be structurally linked to changes in the crossbridge structure that generate force.

is also associated with the power stroke. The closure of the actin binding pocket and the opening of the substrate pocket could be transmitted down S-1 via a long α-helix which runs through half its length. This is because the 20K domain, of which the α-helix is a part, also contributes to the actin binding site and the ATP/substrate binding site, so changes in these two sites could be communicated by the 20K domain. Opening and closing of the substrate cleft could cause the tip of the myosin molecule to move by about 5 nm, which is roughly the size of the working stroke of the crossbridge. However, all of this is highly speculative and will need to be tested further.

SUMMARY

Muscle contracts by the cyclical interaction between two proteins, actin and myosin. The energy source for this interaction is ATP. Many approaches are used to understand the molecular basis of how this interaction occurs and how is it driven

energetically by ATP, and each approach gives unique information. In this chapter a brief outline of each of the approaches used to understand muscle structure, mechanics biochemistry, and energetics was given, and how they contribute to an overall picture of how muscle works was discussed.

The earliest ideas about muscle used the analogy of a spring; these were gradually revised to give rise to a sophisticated model in which steps in the ATPase cycle alter the binding affinity of the crossbridge to actin. A change in binding strength from weak to strong may correspond with some overall change in angle in the crossbridge as it generates force or motion. The newly presented crystal structures of myosin and actin show at the molecular level what these conformational changes might be and should lead to a definition of these changes in detail.

ACKNOWLEDGMENTS

I would like to thank Professor Bob Simmons and Dr. Maxine Clarke for reading and commenting on earlier versions of this manuscript. I would also like to thank Dr. Malcolm Irving for making his lecture notes available to me.

REFERENCES

Brenner, B.M., Yu, L.C., & Podolsky, R.J. (1984). X-ray evidence for crossbridge formation in relaxed muscle fibers at various ionic strengths. Biophys. J. 46, 299–306.

Civan, M.M. & Podolsky, R.J. (1966). Contraction kinetics of striated muscle fibers following quick changes in load. J. Physiol. 184, 511–534.

Cooke, R. & Pate, E. (1985). The effects of ADP and phosphate on the contraction of muscle fibers. Biophys. J. 48, 789–798.

Dantzig, J.A., Goldman, Y.E., Millar, N.C., Lacktis, J., & Homsher, E. (1992). Reversal of the cross-bridge force-generating transition by photogeneration of phosphate in rabbit psoas muscle fibers. J. Physiol. 451, 247–278.

Eisenberg, E. & Moos, C. (1968). Adenosine triphosphate activity of acto-heavy meromyosin: A kinetic analysis of actin activation. Biochemistry 7, 1486–1489.

Eisenberg, E. & Moos, C. (1970). Actin activation of heavy-meromyosin adenosine triphosphatase. J. Biol. Chem. 245, 2451–2456.

Eisenberg, E. & Hill, T.L. (1985). Muscle contraction and free energy transduction in biological systems. Science 227, 999–1006.

Elliott, A. & Offer, G. (1978). Shape and flexibility of the myosin molecule. J. Mol. Biol. 123, 505–519.

Fenn, W.O. (1924). A quantitative comparison between the energy liberated and the work performed by the isolated sartorius muscle of the frog. J. Physiol. 58, 175–203.

Ferenczi, M.A. (1986). Phosphate burst in permeable muscle fibers of the rabbit. Biophys. J. 50, 471–477.

Ferenczi, M.A., Goldman, Y.E., & Simmons, R.M. (1984). The dependence of force and shortening velocity on substrate concentration in skinned muscle fibers from *Rana temporaria*. J. Physiol. 350, 519–543.

Finer, J.T., Simmons, R.M., & Spudich, J.A. (1994). Single myosin molecule mechanics: Piconewton forces and nanometer steps. Nature 368, 113–118.

Goldman, Y.E., Hibberd, M.G., McCray, J.A., & Trentham, D.R. (1982). Relaxation of muscle fibers by photolysis of caged-ATP. Nature 300, 701–705.

Goldman, Y.E., Hibberd, M.G., & Trentham, D.R. (1984). Initiation of active contraction by photogeneration of adenosine-5'-triphosphate in rabbit psoas muscle fibers. J. Physiol. 354, 605–624.

Gordon, A.M., Huxley, A.F., & Julian, F.J. (1966). The variation in isometric tension with sarcomere length in vertebrate muscle fibers. J. Physiol. 184, 170–192.

Hanson, J. (1968). Recent X-ray diffraction studies of muscle. Quart. Rev. Biophys. 1, 177–216.

Harada, Y., Sakurada, K., Aoki, T., Thomas, D.D., & Yanagida, T. (1990). Mechanochemical coupling in actomyosin energy transduction studied by *in vitro* movement assay. J. Mol. Biol. 216, 49–68.

Hill, A.V. (1938). The heat of shortening and the dynamic constants of muscle. Proc. Roy. Soc. B126, 136–195.

Hill, A.V. (1964). The effect of load on the heat of shortening of muscle. Proc. Roy. Soc. B159, 297–318.

Holmes, K.C., Popp, D., Gebhard, W., & Kabsch, W. (1990). Atomic model of the actin filament. Nature 347, 37–43.

Huxley, A.F. (1957). Muscle structure and theories of contraction. Prog. Biophys. Biophys. Chem. 7, 255–318.

Huxley, A.F. (1980). Reflections on Muscle. Liverpool University Press, Liverpool.

Huxley, A.F. & Niedergerke, R. (1954). Interference microscopy of living muscle fibers. Nature 173, 971–973.

Huxley, A.F. & Simmons, R.M. (1971). Proposed mechanism of force generation in striated muscle. Nature 233, 533–538.

Huxley, H.E. (1963). Electron microscope studies on the structure of natural and synthetic protein filaments from striated muscle. J. Mol. Biol. 7, 281–308.

Huxley, H.E. (1969). The mechanism of muscular contraction. Science 164, 1356–1366.

Huxley, H.E. & Hanson, J. (1954). Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. Nature 173, 973–976.

Huxley, H.E. & Brown, W. (1967). The low-angle X-ray diagram of vertebrate striated muscle and its behavior during contraction and rigor. J. Mol. Biol. 30, 383–434.

Kabsch, C., Mannherz, H.G., Suck, D., Pai, E.F., & Holmes, K.C. (1990). Atomic structure of the actin:DNase-1 complex. Nature 347, 44–49.

Kaplan, J.H., Forbush III, B., & Hoffman, J.F. (1978). Rapid photolytic release of adenosine 5'-triphosphate from a protected analogue: Utilization by the Na:K pump of human red blood cell ghosts. Biochemistry 10, 4617–4624.

Kron, S.J. & Spudich, J.A. (1986). Fluorescent actin filaments move on myosin fixed to a glass surface. Proc. Natl. Acad. Sci. USA 83, 6272–6276.

Lowey, S., Goldstein, L., Cohen, C., & Luck, S.M. (1966). Proteolytic degradation of myosin and the meromyosins by a water-insoluble polyanionic derivative of trypsin. J. Mol. Biol. 23, 287–304.

Lymn, R.W. & Taylor, E.W. (1971). Mechanism of adenosine triphosphate hydrolysis by actomyosin. Biochemistry 10, 417–424.

Ramsey, R.W. & Street, S.F. (1940). The isometric length-tension diagram of isolated skeletal muscle fibers of the frog. J. Cell. Comp. Physiol. 15, 11–34.

Rayment, I., Rypniewshi, W.R., Schmidt-Base, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkleman, D.A., Wesenberg, G., & Holden, H.M. (1993a). Three-dimensional structure of myosin subfragment-1: A molecular motor. Science 261, 50–58.

Rayment, I., Holden, H.M., Whittaker, M., Yohn, C.B., Lorenz, M., Holmes, K.C., & Milligan, R.A. (1993b). Structure of the actin-myosin complex and its implications for muscle contraction. Science 261, 58–65.

Reedy, M.K., Holmes, K.C., & Tregear, R.T. (1965). Induced changes in orientation of the cross-bridges of glycerinated insect flight muscle. Nature 207, 1276–1278.

Schroder, R.R., Manstein, D.J., Jahn, W., Holden, H., Rayment, I., Holmes, K.C., & Spudich, J.A. (1993). Three-dimensional atomic model of F-actin decorated with *Dictyostelium* myosin S1. Nature 364, 171–174.

Sheetz, M.P. & Spudich, J.A. (1983). Movement of myosin-coated fluorescent beads on actin cables *in vitro*. Nature 303, 31–35.

Stein, L.A., Schwarz, R., Chock, P.B., & Eisenberg, E. (1979). Mechanism of actomyosin adenosine triphosphatase. Evidence that adenosine 5′-triphosphate hydrolysis can occur without dissociation of the actomyosin complex. Biochemistry 18, 3895–3909.

Szent-Györgyi, A.G. (1953). Meromyosin, the subunits of myosin. Arch. Biochem. Biophys. 42, 305–320.

Trentham, D.R., Bardsley, R.G., Eccleston, J.F., & Weeds, G. (1972). Elementary processes of the magnesium ion-dependent adenosine triphosphatase activity of heavy meromyosin. Biochem. J. 126, 635–644.

Uyeda, T.Q.P., Warrick, H.M., Kron, S.J., & Spudich, J.A. (1991). Quantized velocities at low myosin densities in an *in vitro* motility assay. Nature 352, 307–311.

Yanagida, T., Nakase, M., Nishiyama, K., & Oosawa, F. (1984). Direct observation of motion of single F-actin filaments in the presence of myosin. Nature 307, 58–60.

This Page Intentionally Left Blank

Chapter 8

Muscle Fatigue

ERIC HULTMAN and LAWRENCE L. SPRIET

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 239–280.

ISBN: 1-55938-808-0

INTRODUCTION

"Fatigue, generally defined as a transient loss of work capacity resulting from preceding work, is one of the most fundamental biological problems both for research and practical application." This sentence introduced a textbook on the *Physiology of Work Capacity and Fatigue* by Ernst Simonson (1971). It describes the weakness of the muscle system as a force producing motor, i.e., the deterioration of performance with prolonged activity.

In human exercise physiology, fatigue is often defined as "an inability of a muscle or a group of muscles to sustain the required or expected force" (Edwards, 1981). This definition is very applicable for the study of factors limiting dynamic or isometric exercise in intact organisms at varying intensities.

However, before the fatigue point has been reached during submaximal intensity exercise, the maximal force generation of the muscles involved has already decreased, i.e., physiological factors underlying fatigue are already operating, possibly initiated at the onset of activity (Bigland-Ritchie and Woods, 1984). Thus, in studies with intact or isolated muscle groups and isolated fiber preparations, fatigue is defined as any reduction in force generating capacity, regardless of the force required in a given situation (Bigland-Ritchie and Woods, 1984).

POSSIBLE FATIGUE FACTORS

Localization of Fatigue Factors

Voluntary muscle contraction is initiated in the brain-eliciting action potentials which are transmitted via motor nerves to the neuromuscular junction where acetylcholine is released causing a depolarization of the muscle cell membrane. An action potential is formed which is spread over the surface membrane and into the transverse (T) tubular system. The action potential in the T-tubular system triggers Ca^{2+} release from the sarcoplasmic reticulum (SR) into the myoplasm where Ca^{2+} binds to troponin C and activates actin. This results in crossbridge formation between actin and myosin and muscle contraction.

Four different localizations of fatigue can be identified: (a) decreased central command; (b) decreased activation of the muscle membrane and the T-tubular system; (c) decreased Ca^{2+} release from the SR; and (d) decreased response to the Ca^{2+} release by the contractile proteins. The first two are partly extra-muscular while c and d are intramuscular responses to the excitation of the muscle membrane and often defined as excitation-contraction coupling.

Central Command and Motor Drive

The term central fatigue refers to conditions in which the decline in force can be related to reduced motor drive failing to maintain muscle activation. It had been suggested that the central nervous system (CNS) was not capable of recruiting all motor units maximally during voluntary effort. However, studies by Merton and by Bigland and Lippold in 1954 showed that the maximum voluntary contraction (MVC) of the adductor pollicis muscle gave the same force generation as elicited by supramaximal tetanic stimulation of the ulnar nerve. Merton (1954) also showed that the declining force, when contractions were continued, was similar during short pulses of electrical stimulation of the nerve. This maximum force could only be sustained voluntarily by highly motivated subjects. Similar results have been obtained with the quadriceps muscle of man when comparing voluntary force production and force production during stimulation of the femoral nerve or muscle surface percutaneously (Bigland-Ritchie et al., 1978; Bergström and Hultman, 1990). These results clearly show that central fatigue is not the main factor for the loss of force during short duration maximal contractions or during intermittent submaximal contractions with longer duration (Bigland-Ritchie et al., 1983a; Bergström and Hultman, 1990).

Impaired peripheral neuromuscular transmission could also cause fatigue. The mass action potential elicited by the nervous impulse can be measured with electromyography. The measured M-wave amplitude and area evoked by single maximal shocks to the nerve showed no decline in the pollicis muscle during a 3 min isometric MVC, in spite of a nearly complete loss of force. Merton et al. (1981) also found that massive direct electrical stimulation of the pollicis muscle fibers could not restore the force loss during a sustained MVC. It was concluded by Bigland-Ritchie and Woods (1984) that, in fatigue of sustained voluntary contractions executed by well-motivated subjects, the reduction in force generating capacity was not due to a decline in CNS, motor drive, or neuromuscular transmission. Alternatively, it could be attributed solely to contractile failure of the muscle involved. It should be noted that these studies were done in a laboratory with well-motivated subjects and therefore may not be relevant to exercise or other conditions with additional stress factors (e.g., muscle pain from lactate buildup).

Neuromuscular Transmission and Sarcolemma Excitability

It was shown by Bigland-Ritchie (1981) that continuous maximal electrical stimulation of the motor nerves (50–80 Hz frequency) led to a progressive decline of the M-wave and force. The M-wave and force could be largely restored by reducing the stimulus frequency to 20 Hz. The rapid fall in force during continuous high frequency stimulation is apparently due to impulse propagation failure, and has been termed high frequency fatigue. During maximal voluntary contractions there is a decrease in the neural firing rate which correlates with the slowing of the contractile process. Apparently the propagation failure seen in high frequency fatigue does not occur in voluntary muscle contractions.

However, repeated excitations do produce a Na^+ and K^+ shift over the muscle membrane, with extrusion of K^+ and uptake of Na^+. This will change the intracellular ion composition, in spite of the counteracting effect of the Na^+-K^+ pump. K_i^+ decreases of 10–20% have been observed together with 1.5-fold increases in Na_i^+ in fatigued muscle (Sahlin et al., 1978; Sjögaard et al., 1985; Lindinger and Heigenhauser, 1988). The extracellular ion composition also changes during continued stimulation. These changes have an impact both on the resting membrane potential and the action potential. The largest effect of the Na^+ and K^+ changes across membranes is likely to occur in the T-tubules due to the small volume of this compartment. The effects of these ion changes on impulse propagation in the T-tubules and its role for Ca^{2+} release will be discussed later.

Excitation-Contraction Coupling

The contractile force of the muscle is determined by the action potential transmitted from the neuromuscular junction to the interior of the fiber via the T-tubular system which is in close proximity with the SR. The SR contains a network of sacculi with a high concentration of Ca^{2+}. When the impulse reaches the sacculi, Ca^{2+} is released to the myoplasm, thereby increasing the concentration from ~50 nM in resting muscle to 1,000–5,000 nM. The result is saturation of the Ca^{2+} binding sites on troponin and hence, crossbridge formation and cycling, i.e., muscle contraction. The myoplasm free [Ca^{2+}] is dependent on an ATP-requiring Ca^{2+} pump, transporting Ca^{2+} back to the SR. When T-tubular activation ceases [Ca^{2+}] falls, Ca^{2+} is removed from troponin C, crossbridge cycling stops, and the muscle relaxes.

From this brief summary of excitation-contraction coupling it is obvious that Ca^{2+} is an important link between the activated membrane and the contractile proteins, and thus a regulator of tension development. Westerblad et al. (1991) defined three factors which explain the force decrease in fatigued muscle: reduced Ca^{2+} release from the SR, reduced Ca^{2+} sensitivity of the myofilaments, and reduced maximum Ca^{2+}-activated tension.

Energy Metabolism During Contraction

In 1807 the Swedish scientist J.J. Berzelius suggested that the amount of lactic acid in skeletal muscle was proportional to the extent it had previously exercised (citation by Lehman, 1850). This observation made ~190 years ago is still valid today for high intensity exercise. We also now know that when Ca^{2+} content increases in the myoplasm, triggered by the motor nerve impulse, a series of adenosine triphosphate (ATP) degrading processes (ATPases) are induced or augmented, i.e., crossbridge cycling, Na^{+}-K^{+} pumping and Ca^{2+} transfer back to the SR. As the ATP store in muscle is small and sufficient to sustain maximal contraction for less than 2 sec, rephosphorylation of formed adenosine diphosphate (ADP) is a prerequisite for continued contraction. This is achieved initially by the creatine kinase reaction:

$$ADP + PCr + H^{+} \leftrightarrow ATP + Cr$$

The net result of ATP degradation and resynthesis will be the formation of creatine (Cr) and inorganic phosphate (P_i) and a decrease in phosphocreatine (PCr).

The Ca^{2+} increase also activates the enzyme glycogen phosphorylase, triggering the formation of hexose phosphates which are further degraded to pyruvate. This process provides 3 ATP for each mole of glucose from muscle glycogen that is degraded during intense exercise (2 ATP if the glucose originates from the blood). The pyruvate can either be reduced to lactate with simultaneous dehydrogenation of NADH to NAD or oxidized in the mitochondrion completely to CO_2 and H_2O. The relation between these two processes is dependant on the rate of pyruvate formation, which in turn is determined by the ATP turnover rate, and on the capacity of the mitochondrion to oxidize pyruvate.

Almost all of the carbohydrate utilized during intense short duration exercise originates from muscle glycogen. During prolonged submaximal exercise, significant amounts of glucose can also enter the muscle from the blood. The ATP yield for each mole of glucose oxidized is 38 moles. Free fatty acids can also be oxidized in the mitochondrion. The result of lactate formation is a pH decrease, while oxidation produces only CO_2 which can leave the muscle. The pH decrease in muscle fatigued by intense contractions is ~0.5–0.6 units, while pH is commonly unchanged in submaximal exercise when oxidative metabolism dominates.

The three processes for ATP resynthesis—PCr degradation, anaerobic glycolysis and oxidative phosphorylation—have different maximum rates of ADP phosphorylation. PCr utilization has the highest maximal rate; approximately twice that of anaerobic glycolysis which in turn can produce ATP at a rate twice that of oxidative phosphorylation (Hultman and Sjöholm, 1986). This means that during intense contraction the rate of ATP resynthesis from oxidative pyruvate or fat utilization is insufficient to cover the demand for ATP and the dominating processes are those that do not require oxygen; that is, PCr degradation and anaerobic

glycolysis. However, the size or capacity of the energy processes are also different. The PCr store in the muscle can provide ~24 mmol ATP/kg wet muscle. Anaerobic glycolysis can provide ~2.5–3 times as much ATP as PCr degradation before muscle contractions and glycolysis are inhibited by the pH decrease. Oxidative phosphorylation of the entire glycogen store can provide 100–200 times the amount of ATP available via PCr degradation. Oxidation of the fat stored in muscle can provide ~1.5 times the ATP provided through oxidation of the glycogen store. These calculations were made for the intact human quadriceps femoris muscle (Hultman and Sjöholm, 1986).

There is also a difference between the two main fiber types in muscle, the slow twitch fatigue resistant type I fiber and the fast twitch fast fatiguable type II fiber. Type I has a higher oxidative capacity and a lower capacity for anaerobic energy release compared to type II. The maximum rate of energy production (i.e., myosin ATP-ase activity) is also higher in type II which will result in a faster loss of PCr and formation of lactate. In a recent study of single muscle fibers from *Xenopus laevis* it was shown (Van Der Laarse et al., 1991) that fatigue resistance was highly correlated to the activity ratio of the mitochondrial enzyme succinate dehydrogenase over myofibrillar ATPase. A high ratio for type I fibers and a low ratio for type II is typical.

During intense muscle contraction PCr degradation starts immediately and muscle lactate formation is also observed within the first 2 sec of tetanic electrical stimulation (Hultman and Sjöholm, 1983a). These two processes provide a sufficient rate of ATP resynthesis to keep the ATP concentration unchanged, until most of the PCr store is utilized. Then ATP starts to decrease with increasing accumulation of ADP, adenosine monophosphate (AMP), and inosine monophosphate (IMP). However, this is late in the contraction period, and the fall in ATP is no more than 30–40% of the resting value (Hultman et al., 1967). At this point the $[P_i]$ is very high due to PCr degradation, the pH is decreased, and the force generation has declined.

FATIGUE IN ISOLATED SKELETAL MUSCLE FIBERS

Metabolic Factors and Ca^{2+}-Induced Tension Development

Measurements of metabolite concentrations in muscle fibers before and after fatiguing stimulation have shown that ATP decreases from 6 to 4.6 mM and PCr decreases from 35 to 2.4 mM with a calculated increase in P_i from 3 to ~38 mM (Dawson et al., 1978; Nassar-Gentina et al., 1978). The free ADP concentration was calculated to increase from 30 to 200 μM. At the same time pH decreased from 7.0 to 6.5 (Dawson et al., 1978; Juel, 1988; Westerblad and Lannergren, 1988). The effect of these metabolic changes has been studied in skinned muscle fibers, i.e., fibers in which the cell membrane has been removed. The skinning of the fibers

results in an intact filament in which the intracellular environment can be changed by varying the composition of the bathing solution.

Fiber tension development is measured at different Ca^{2+} concentrations. The maximum Ca^{2+} activated tension is denoted P_o and the Ca^{2+} concentration needed to produce 50% of P_o is estimated and denoted Ca_{50}. These two measurements can be used to study Ca^{2+} related fatigue mechanisms. A pH decrease from 7.0 to 6.5 produced a 30% decrease in P_o and a decrease in Ca^{2+} sensitivity (2-fold increase in Ca_{50}) (Fabiato and Fabiato, 1978; Godt and Nosek, 1989). Similar changes in P_o and Ca_{50} are observed when P_i is increased from 0.9 to 17 mM.

A series of bathing solutions mimicking the intracellular environment in fatigued muscle have been studied by Godt and Nosek (1989). Variations in ATP, ADP, PCr, and Cr contents had only marginal effects on the tension development compared to decreases in pH and increases in P_i.

The effect of H^+ on the Ca^{2+} sensitivity is thought to arise from competition between H^+ and Ca^{2+}, acting either directly on troponin (Robertson and Kerrick, 1979), or indirectly by altering the net charge of the thin filament (Godt, 1981). The P_i effect on Ca^{2+} activated force is on the contractile apparatus without affecting Ca^{2+} binding. Both maximum force and Ca^{2+} sensitivity are decreased. However, Godt and Nosek (1989) concluded that some impairment of excitation-contraction coupling is likely to occur in addition to direct depression on the contractile apparatus because the maximum force of skinned muscle did not decline to zero under ionic conditions mimicking fatigue/hypoxia. They referred to a study by Allen et al. (1988) showing that Ca^{2+} concentration in the myoplasm decreased during fatigue. Additional studies by this group are summarized in a recent review by Westerblad et al. (1991).

Ca^{2+} Concentration During Stimulation

A typical pattern of tension generation and free Ca^{2+} concentration in the myoplasm is seen in Figure 1 (Westerblad et al., 1991). A 10–20% force decrease was already observed after 10–20 tetani, followed by a period with only minor force decreases and a final period with rapid force loss to ~30% of initial (Figure 1).

The first decrease in force is not due to reduced Ca^{2+} release, but probably to a reduced maximum Ca^{2+} activated tension caused by metabolic changes. There is in fact an increased Ca^{2+} concentration in this period which is probably due to Ca^{2+} binding to troponin and parvalbumin during the initial tetani. This would reduce the Ca^{2+} binding sites in the myoplasm and a larger part of the Ca^{2+} released from the SR would remain free. Some troponin C binding sites could also be occupied by H^+, further reducing Ca^{2+} binding by the myoplasm. During the last part of the stimulation period there is a rapid decline in force simultaneous with a decrease in the free Ca^{2+} concentration. This is not explained by increased buffering of myoplasmic Ca^{2+} as discussed above. Instead, there is a further decrease in Ca^{2+}

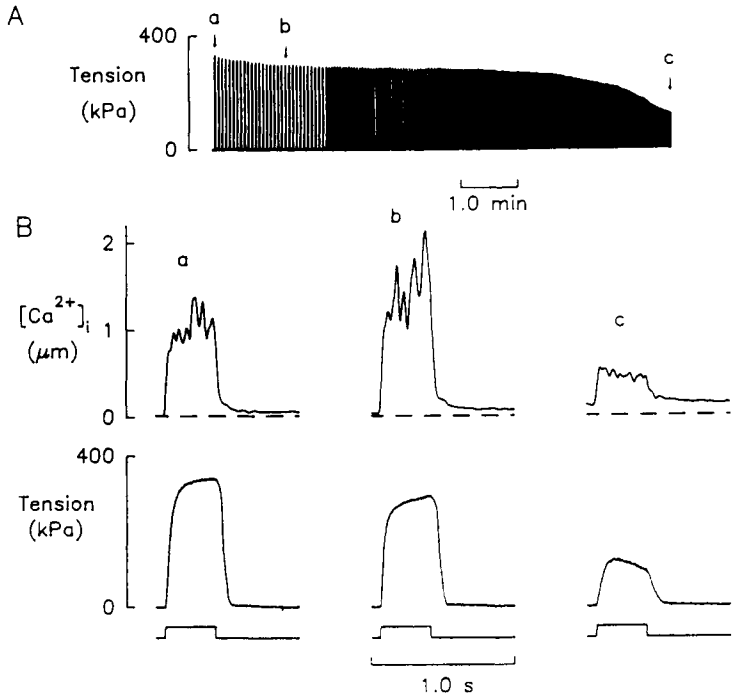


Figure 1. Original records of tension and intracellular free calcium concentration (Ca_i^{2+}) obtained from a single mouse muscle fiber during a fatigue run (modified from Westerblad and Allen, 1991). **A**: continuous tension record in which each vertical line represents a tetanus. **B**: $[Ca^{2+}]_i$ (measured with fura-2) and tension records obtained from the individual tetani (a, b, and c) indicated above the record in **A**. Three major features are illustrated: 1.) the initial tension decline is accompanied by an increase in tetanic $[Ca^{2+}]_i$, 2.) late in fatigue the tetanic $[Ca^{2+}]_i$ is reduced, and 3.) the resting $[Ca^{2+}]_i$ increases during fatiguing stimulation (dashed line indicates resting $[Ca^{2+}]_i$ in control). Stimulation periods are shown below tension records in **B**. From Westerblad et al., 1991, with permission from the Amer. Physiol. Society.

binding capacity in the myoplasm, and consequently the reason must be a pronounced decrease in SR Ca^{2+} release. Three mechanisms could be responsible for this decrease: (a) failure of action potential propagation on the sarcolemmal or T-tubular membranes, (b) impaired coupling of T-tubular depolarization to SR Ca^{2+} release, and (c) reduced Ca^{2+} content of the SR (Westerblad et al., 1991).

Electrolyte Changes and Action Potential Propagation

As described above, the propagation of the action potential is sensitive to Na^+-K^+ changes over the membranes. These changes are most pronounced during continu-

ous high frequency stimulation characterized by a rapid fall in force generation and a substantial recovery of the force within 1–2 sec after stimulation.

Such rapid changes in force production could be explained by Na^+-K^+ changes in the T-tubular fluid during the continuous stimulation. Initially the action potential propagation may be decreased due to a K^+ increase and a Na^+ decrease. Following stimulation a fast recovery of impulse propagation may occur via equilibration of K^+ and Na^+ by diffusion into the extracellular fluid space. These movements will be slower in the T-tubules in the center of the fiber compared to those near the surface. It was also shown by Westerblad et al. (1990) that continuous high frequency stimulation resulted in a Ca^{2+} gradient in the myoplasm of the stimulated fiber, with the lowest concentration in the center of the fiber.

With continued high frequency stimulation, propagation of the action potential in the SR becomes increasingly impaired leading to decreased Ca^{2+} release and eventually decreased force generation. During intermittent stimulation with repeated tetani similar changes occur, but to a smaller extent. With a 60% decrease in tension the membrane potential decreased from –90 to –70 mV. There was no marked recovery of tension within the first 2 sec after stimulation and the Ca^{2+} distribution was homogeneous throughout the fiber (Lännergren and Westerblad, 1986; Westerblad et al., 1990). Therefore the tension decline with repeated tetani in the *Xenopus* fiber was not caused by impaired propagation of action potentials.

T-Tubular Depolarization and Ca^{2+} Release From the Sarcoplasmic Reticulum

According to Schneider and Chandler (1973), depolarization of the T-tubules affects sensors which open Ca^{2+} channels in the SR. The sensors are modified Ca^{2+} channels which act as voltage sensors (Tanabe et al., 1987). The signal from the sensor reaches the SR and opens the Ca^{2+} channels with the release of Ca^{2+} to the myoplasm. The Ca^{2+} channels in the SR system are opened by micromolar [Ca^{2+}], mM [ATP], and caffeine but are inhibited by Mg^{2+} (Smith et al., 1986; Rosseau et al., 1988). The channels are closed in resting muscle and are opened when the voltage sensor is activated.

High concentrations of K^+ activate the sensor by depolarizing the T-tubule and releasing SR Ca^{2+}. The same effect is produced by caffeine which has a direct effect on the Ca^{2+} channel. Application of solutions with either a high [K^+] or [caffeine] to fatigued muscle restores the tension to at least 80% of initial, indicating that a major reason for the fatigue is a decrease in myoplasmic [Ca^{2+}]. The remaining decrease in tension (20%) could be explained by a reduction of maximum Ca^{2+} activated tension induced by metabolic changes, i.e., P_i increase.

Possible mechanisms responsible for the decreased Ca^{2+} release are changes in the sensitivity of the voltage sensor in the T-tubular system or in the SR Ca^{2+} channel to the sensor stimulus. A third possibility would be a decreased availability

of Ca^{2+} in the SR. The responsible mechanism should in some way be related to the metabolic changes in fatiguing muscle. There is to our knowledge no such relation described for the sensitivity of the voltage sensor. However, the rate of Ca^{2+} release from the SR is decreased when the ATP concentration falls (Meissner et al., 1986). An increase in $[Mg^{2+}]$ has a similar effect. Decreasing the concentration of the MgATP complex with the release of Mg^{2+} in the fatiguing cell is known to occur, and has been suggested as a fatigue factor. Lamb and Stephensson (1991) showed, however, that a decrease in MgATP from 7 to 2 mM resulted in only a marginal decrease in Ca^{2+} release from the SR in skinned fibers. Further studies are needed to exclude the decrease in [MgATP] as a fatigue factor.

It has been shown that inositol triphosphate (IP_3) is involved in the excitation-contraction coupling in smooth muscle (Vergara et al., 1985), but presently no clear evidence has been reported for a similar involvement in skeletal muscle. If IP_3 functions as a messenger for Ca^{2+} release, it would bridge the gap between muscle metabolic changes and Ca^{2+} release, as ATP is a prerequisite for IP_3 regeneration.

Ca^{2+} Content of the Sarcoplasmic Reticulum

Another possible mechanism for decreased SR Ca^{2+} release is a fall in Ca^{2+} availability in the SR of fatiguing muscle. It is well known that a slowing of muscle relaxation occurs after repeated tetani. As much as a 60% decline in relaxation rate, especially after isometric contractions, has been regarded as an integral part of fatigue (Sahlin et al, 1981; Bigland-Ritchie et al., 1983b; Hultman and Sjöholm, 1983b). The slowing of relaxation has been attributed to a reduced rate of cross-bridge cycling (Edwards et al., 1975), impaired function of the SR-Ca^{2+} pump (Dawson et al., 1980; Allen et al., 1989), or Ca^{2+} binding to myoplasmic Ca^{2+} buffers such as parvalbumin. The two first processes are affected by a pH decrease in fatigued muscle and the second process, the SR pump, is also affected by the affinity of the Ca^{2+}-ATPase interaction. A relationship was observed between the free energy change (ΔG) for the ATPase reaction and the rate of relaxation (Dawson et al., 1980). Direct measurements of Ca^{2+} content during and after tetani revealed a lower rate of decline in fatigued muscle compared to control (Allen et al., 1989; Lee et al., 1991). Accumulation of Ca^{2+} during repeated tetani via Ca^{2+} binding to parvalbumin and possibly to mitochondria could decrease the amount of free Ca^{2+} transported back to the SR. Recent studies with K^+ depolarization and caffeine stimulation of muscle fibers demonstrated a lower Ca^{2+} release from the SR of fatigued fibers compared to the release in control fibers (Westerblad and Allen, 1991). Consequently, there are reasons to believe that decreased Ca^{2+} availability in the SR, due to metabolically induced decreases in Ca^{2+} pump effectiveness and increased myoplasmic Ca^{2+} binding, is an important factor in fatigue.

FATIGUE IN INTACT SKELETAL MUSCLE OF HUMAN SUBJECTS DURING HIGH INTENSITY EXERCISE

High intensity exercise will be defined in this chapter as muscle contractions requiring ~40–100% of the maximal voluntary contraction force. The energy required for these contractions cannot be provided solely by oxidative processes and therefore anaerobic processes must also provide energy. It has been shown in a series of studies using isometric or dynamic high intensity exercise that the ATP resynthesis rate declines at approximately the same rate as the force generation (Boobis et al., 1982; Hultman and Sjöholm, 1986). The interpretation of this is either that the decreased ATP resynthesis rate inhibits force generation and is responsible for fatigue, or that force generation decreases due to other factors, resulting in a lower rate of ATP demand. The two mechanisms would be:

1. An insufficient rate of ATP resynthesis for optimal energy supply for actomyosin crossbridge formation and cycling, or for the additional ATPase reactions, Na^+-K^+ pumping and Ca^{2+} reuptake and/or release by the SR.
2. Inhibition of any of the above processes by products formed in the energy supplying pathways, alterations in intracellular composition or alterations in membrane function induced by repeated contraction.

Muscle Metabolism During Electrically Evoked High Intensity Contraction

Electrical stimulation with surface electrodes on the quadriceps muscle combined with muscle tissue sampling using the needle biopsy technique has been utilized to study the relationship between force generation and energy metabolism. An experiment of this type is presented in Figure 2 (Spriet et al., 1987a,b; Hultman et al., 1990). Intermittent tetanic contractions were evoked by stimulation at 20 Hz. Contractions were 1.6 sec in duration and separated by 1.6 sec rest periods. Leg blood flow was occluded with a tourniquet and the energy for muscle contraction could only be provided anaerobically from PCr degradation and glycolysis. The data in Figure 2 are expressed per kg dry muscle (dm), where 1 kg dm corresponds to ~4.3 kg wet muscle containing 3 L intracellular and 0.3 L extracellular water.

The upper panel in Figure 2 shows the force generation in percent of initial and the muscle pH measured in wet tissue at the five muscle sampling points. It also shows the total ATP turnover rate and the amount of ATP formed via PCr degradation and glycolysis. During the initial 16 contractions, the ATP utilized was derived from both the degradation of PCr and the glycolytic pathway. Only a minor part of the ADP formed was unphosphorylated at the end of this period. With continued contractions the glycolytic ATP formation became the main pathway as the PCr store was nearly depleted. The total ATP utilization rate decreased along with the fall in force generation.

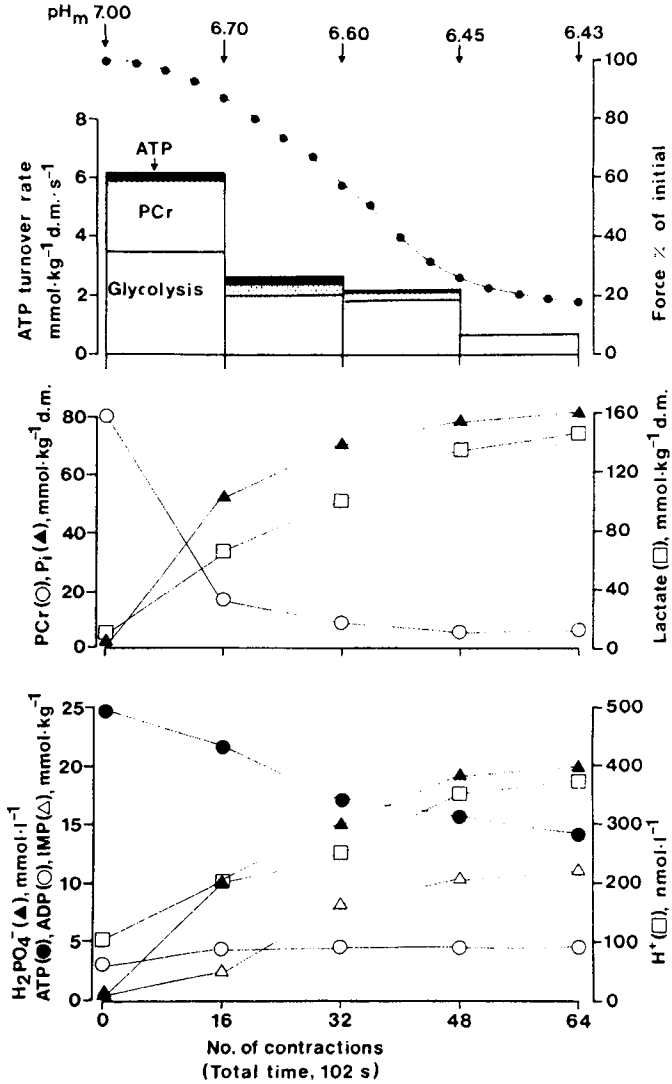


Figure 2. Force generation and energy metabolism in human quadriceps femoris muscle stimulated intermittently at 20 Hz, with 1.6 sec tetani with 1.6 sec rest periods between tetani. The upper panel shows force, ATP turnover rate, and pH; the middle panel, the concentrations of PCr, P_i, and lactate; and the lower panel, ATP, ADP, IMP, H^+, and calculated $H_2PO_4^-$. From Hultman et al. (1990), with permission from Human Kinetics Publishers.

The middle panel shows the fall in PCr concentration in the muscle and the simultaneous increase in P_i released during ATP degradation and resynthesis via the creatine kinase reaction:

$$ATP \rightarrow ADP + H^+ + P_i$$

$$ADP + PCr + H^+ \leftrightarrow ATP + Cr$$

The lower panel shows the decreasing concentration of ATP, to about 60% of resting levels, and the simultaneous equimolar increase in IMP. The fall in ATP started when most of the PCr store was utilized, resulting in a decreased rate of ADP phosphorylation via the creatine kinase reaction. The resultant accumulation of ADP stimulates adenylate kinase activity and subsequently IMP is formed via the AMP deaminase reaction:

$$2\ ADP \rightarrow ATP + AMP$$

$$AMP \rightarrow IMP + NH_3$$

The lower panel also contains H^+ concentration and the calculated $[H_2PO_4^-]$.

Figure 2 shows the expected changes during an intense contraction: an 80% decrease in force generation and a comparable decrease in ATP turnover rate. The ATP content was decreased by about 40% and PCr by 95%. At the same time large increases were observed in H^+ and P_i.

Total Energy Production and Fatigue

The energy produced during the contraction period in Figure 2 corresponded to 300 mmol ATP/kg dm. Of this, 210 mmol were resynthesized by anaerobic glycolysis, 70 by PCr degradation and 20 mmol by degradation of the ATP pool. The PCr store was almost completely utilized and glycolysis resulted in a lactate increase to ~150 mmol/kg dm, a value which is repeatedly observed after intense exercise continued to exhaustion (Harris et al., 1977; McCartney et al., 1986). The limitation is not the availability of glycogen, which is normally ~360 mmoles of glucose units/kg dm prior to exercise. The limitation appears to be due to the H^+ increase with inhibiting effects on the contractile system. The ATP decreased by ~40%, a value which also is repeatedly observed in fatigued muscle after dynamic or isometric exercise induced voluntarily or by electrical stimulation (Hultman et al., 1967; Spriet et al., 1987a). The ATP decrease is comparable to that observed in electrically stimulated frog muscle (Nassar-Gentina et al., 1981). A further ATP decrease occurred only after caffeine-induced contracture in fatigued frog muscle (Nassar-Gentina et al., 1981), or after artificially induced Ca^{2+} release in muscle.

The ATP in fatigued muscle is, however, well above the K_m for actomyosin ATPase activity (Nanninga and Mommaerts, 1960). This indicates that the decrease in force generation is not related to a lack of ATP for crossbridge formation but

rather mainly due to inhibition of force generation by formed products, local changes of ATP/ADP ratios and/or a decrease in Ca^{2+} release. It is obvious from Figure 2 that the fatigue factors discussed above (accumulation of H^+ and P_i) could explain the decrease in force generation, at least in the initial half of the stimulation period. The third fatigue factor, decreased Ca^{2+} release, occurred late during the contraction period. In this experiment, it may have been caused by the decrease in the MgATP complex, which inhibits the SR Ca^{2+} channels. Further, the decreased ATP/ADP ratio at the end of exercise could inhibit ATPase dependent uptake of Ca^{2+} by the SR (Dawson et al., 1980). This would result in a fall in Ca^{2+} content in the SR and eventually decrease the Ca^{2+} release during stimulation.

Fatigue Development and Muscle Fiber Type

It is generally accepted that the two main fiber types have different metabolic profiles with higher activities of Ca^{2+} activated myosin ATPase, creatine kinase,

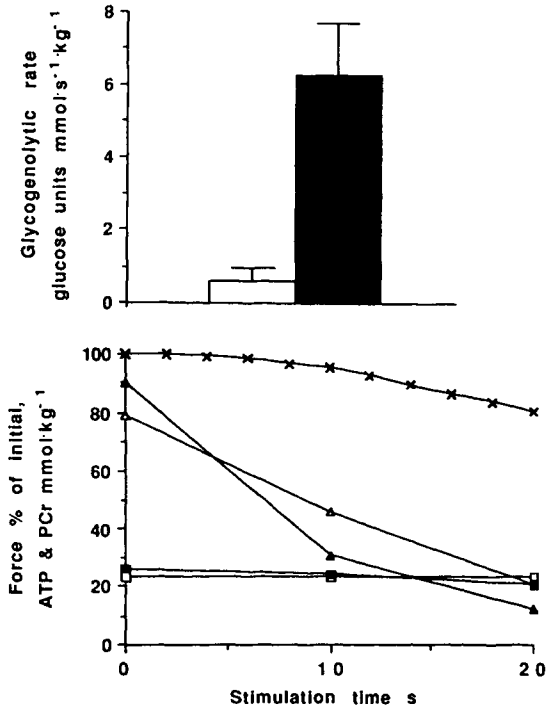


Figure 3. Top panel: Whole muscle force (×) and single fiber PCr (△, ▲) and ATP (□, ■) concentrations at rest and after 10 and 20 sec of intermittent electrical stimulation at 50 Hz. Open symbols denote type I fibers; closed symbols denote type II fibers. Bottom panel: Glycogenolytic rates in type I and II fibers during the 20 sec stimulation period. The open bar denotes type I fibers; the closed bar denotes type II fibers.

and glycogen phosphorylase in type II fibers. Force generation and fatiguability are also higher in type II fibers while type I fibers are fatigue resistant. Direct force measurements in type II motor units of human muscle showed a high force generation initially, but also a rapid loss of force while type I motor units produced much lower initial forces but maintained the force constant during a longer stimulation period (Faulkner et al., 1986). In all human skeletal muscles the two fiber types are mixed and for metabolic studies of individual fiber types, the fibers have to be separated from the mixed biopsy sample, characterized, and analyzed.

A study of this type is shown in Figure 3. The quadriceps muscle was stimulated intermittently for 20 sec. The fall in force generation was most pronounced after 10 sec of stimulation. The initial rate of PCr degradation was 5.8 mmol/sec/kg dm in type II fibers (0–10 sec) and decreased thereafter to 2.3 mmol/sec/kg during the 10–20 sec period. The rate in type I fibers was lower and essentially equal during the two periods. The rate of glycogen degradation was very high in type II fibers while no significant glycogen degradation was observed in type I fibers (Hultman et al., 1991). It has also been shown that the glycolytic rate in muscle during this type of stimulation is maximal during the initial 10 sec with no further increase during the 10–20 sec period (Hultman et al., 1991). This means that there was no compensation for the decrease in ATP resynthesis from PCr in type II fibers during the 10–20 sec period. The only possible way to keep the balance between ATP utilization and formation would be to decrease the force generation. The conclusion was that one limitation for energy production and force generation during this type of contraction may be the size of the PCr store, especially in type II fibers.

Phosphocreatine and Fatigue

Depletion of the PCr store is observed after short-lasting intense contractions evoked by electrical stimulation (Figures 2 and 3) during voluntary isometric contraction (Katz et al., 1986b), and during dynamic exercise with high work loads (Hultman et al., 1967; Karlsson and Saltin, 1970). During this type of exercise there is also an accumulation of H^+ and P_i in the fatiguing muscle and these other fatigue factors could explain the decrease in force generation. The specific effect of PCr could be studied by repeating an exercise with the same subject with different concentrations of PCr before the start of exercise.

Increased creatine and PCr concentrations have been achieved by giving creatine orally. The creatine dose used was 20 g daily for 4–5 days, administered in repeated 5 g doses. Muscle creatine increased by 20–40% above the preadministration content and the PCr increased by 8–15% (Harris et al., 1992). Torque generation by the knee extensor muscles was measured during five repeated bouts of 30 isokinetic MVCs, with 2 min rest periods between bouts. The exercise protocol was performed four times, before and after placebo or creatine supplementation. The total peak torque production increased significantly in all five bouts of maximal

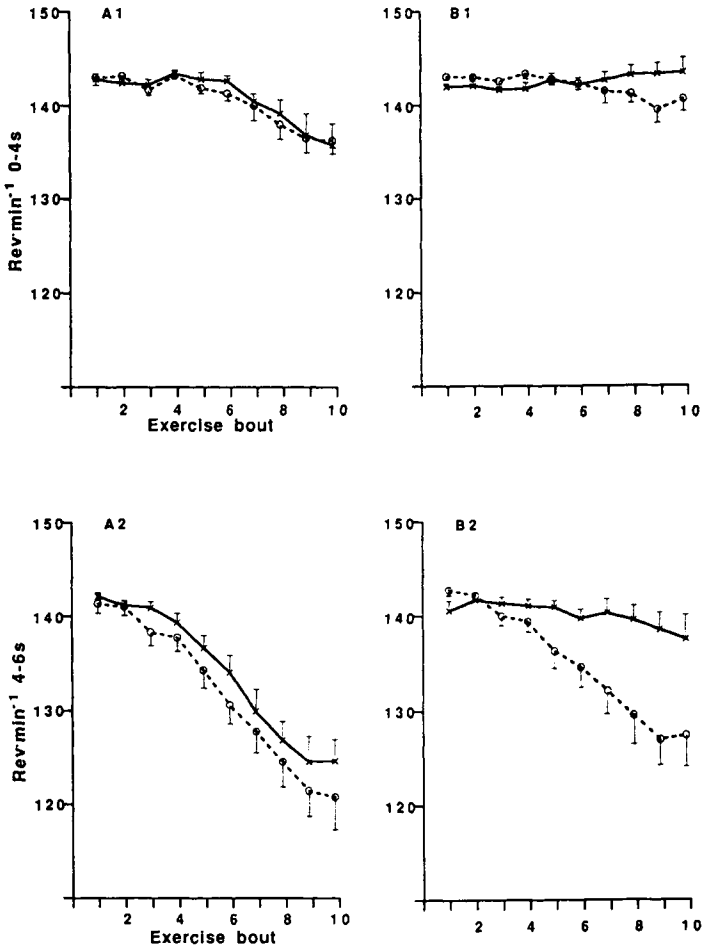


Figure 4. Ten bouts of high intensity cycle exercise, each with a duration of 6 sec and with 30 sec rest periods between bouts were performed on a given day. The entire exercise protocol was performed on four separate days by each subject (n = 8), before and after placebo and creatine supplementation. Mean power output was 882 watts. Upper panels (A1 and B1) demonstrate mean revolutions per min during the 0–4 sec period of each bout and the lower panel (A2 and B2) during 4–6 sec of each bout. The left panels (A1 and A2) show results before (○) and after (×) placebo and the right panels (B1 and B2) before (○) and after (×) creatine supplementation. The study was performed in a double blind manner.

contraction (5% mean increase) following creatine supplementation, while placebo produced no significant change (Greenhaff et al., 1993b). In a later study by Balsom et al. (1993), high intensity exercise was performed before and after supplementation of creatine or placebo. Each exercise protocol consisted of ten 6 sec bouts of

high intensity cycling with rest periods of 30 sec between bouts. The results (Figure 4A1,B1) demonstrate the relatively well preserved force during the initial 4 sec of each bout followed by a progressive force decline during the last 2 sec (4–6 sec) of bouts 1–10 before and after placebo, as well as before creatine supplementation (Figure 4A2,B2). After creatine supplementation the force was essentially unchanged throughout the entire exercise period (Figure 4B1,B2).

There is evidence to suggest that the difference in torque/force generation in both of these experiments was due to two factors; higher initial PCr content in muscle after creatine supplementation, and a creatine induced acceleration of PCr resynthesis during the rest periods between bouts. The higher PCr availability seems to have delayed fatigue. The mechanism for this delay could be a better maintenance of the required rate of ATP resynthesis during contraction, either in the contractile apparatus or in other ATPase reactions such as Na^+-K^+ pumping or Ca^{2+} transport and release. These reactions are dependent on the ATP/ADP ratio, which is directly related to the PCr store via the creatine kinase reaction. Indications of a better preserved ATP/ADP ratio after creatine supplementation compared to placebo were lower blood concentrations of ammonia (Greenhaff et al., 1993b) and hypoxanthine (Balsom et al., 1993) after exercise. This was interpreted as a sign of lower mean ADP during the exercise bouts resulting in less AMP formation, which is the substrate for the deaminase reaction and the subsequent formation of hypoxanthine. Blood lactate accumulation was also lower during high intensity cycling exercise following creatine supplementation compared to control, in spite of a higher force output (Balsom et al., 1993).

Creatine supplementation has been shown to increase the rate of PCr resynthesis in the recovery period after ischemic exercise (Greenhaff et al., 1993a). This could be attributed to an acceleration of oxidative phosphorylation by increased free creatine content available to the mitochondrial fraction of the creatine kinase enzyme, as previously suggested (Bessman and Fonyo, 1966).

As discussed earlier, a decrease in muscle MgATP could inhibit Ca^{2+} release from the SR, by lowering ATP content and increasing the free Mg^{2+}. A fall in the ATP/ADP ratio may also inhibit Ca^{2+} release by slowing Ca^{2+} reuptake into the SR. A better preserved ATP/ADP ratio in exercising muscle as a result of increased creatine content could counteract the inhibition of Ca^{2+} kinetics and delay fatigue.

Fatigue and Glycolytic Rate

A decreased glycolytic rate has been proposed as a cause of muscle fatigue and related to pH inhibition of glycolytic enzymes. Decreasing pH inhibits both phosphorylase kinase and phosphofructokinase (PFK) activities. PFK is rate determining for glycolytic flux and therefore must be precisely matched to the rate of ATP expenditure. The essential characteristic of PFK control is allosteric inhibition by ATP. This inhibition is increased by H^+ and PCr (Storey and Hochachka, 1974;

Bock and Frieden, 1976). Removal of this inhibition provides the primary mechanism by which PFK responds to increased energy expenditure with the onset of contraction and maintains activity in the face of decreasing pH. The factors contributing to this are decreased contents of ATP and PCr and increased contents of metabolites, which either reduce the inhibitory effects of ATP or appear to activate the enzyme directly; ADP, AMP, P_i, NH_3, fructose-6-phosphate, and possibly the hexose bisphosphates (Dobson et al., 1986; Spriet et al., 1987b). As illustrated in Figure 2, glycolysis in intact human muscle continues at an unchanged rate during the period 16–48 contractions, despite a drop in pH from 6.70 to 6.45. Consequently the pH inhibition of PFK must be almost completely released. During this period a large decrease in force generation is observed which cannot be explained by lack of glycolytic ATP generation.

Inhibition of the Contractile Processes by Accumulation of H^+ and P_i

The lactate accumulation during intense anaerobic exercise was implicated as a possible fatigue factor as early as the beginning of this century (Fletcher and Hopkins, 1907). A detailed presentation of the role of H^+ and P_i in the contractile process has been presented in the section on fatigue in isolated fibers.

Several studies of intact human muscle stimulated electrically have shown good relationships between the force decrease and accumulations of lactate (H^+) and P_i. Force generation measured in the recovery period after stimulation shows a slightly different picture. Figure 5 illustrates the relationships between force, P_i, and lactate, calculated from the experiment in Figure 2 and data from a series of 1.6 sec stimulations (20 sec in total) at 1, 2, and 3 min of recovery (Hultman et al., 1990). The calculated concentration of $H_2PO_4^-$ was used instead of P_i. The results show a linear relationship between force generation and the accumulation of lactate and $H_2PO_4^-$ during the stimulation. The simultaneous formation of P_i from PCr degradation and H^+ from glycolysis makes it difficult to separate these effects from each other or from other effects of the stimulation. In the recovery period, PCr resynthesis with the removal of P_i is faster than the disappearance of lactate, and the effects of the two factors on force generation can be at least partly separated. A closer relationship between force and $[H^+]$ indicates an additional effect of pH on force generation above that from $H_2PO_4^-$. The results support the findings of Nosek et al. (1987), suggesting that two separate mechanisms are responsible for development of fatigue: H^+ increase and accumulation of $H_2PO_4^-$. However, they also show that none of these factors are solely responsible for fatigue.

Fatigue Unrelated to Metabolic Changes

As discussed above, pronounced metabolic changes are associated with the decrease in force during intense exercise. Reasonably good correlations are ob-

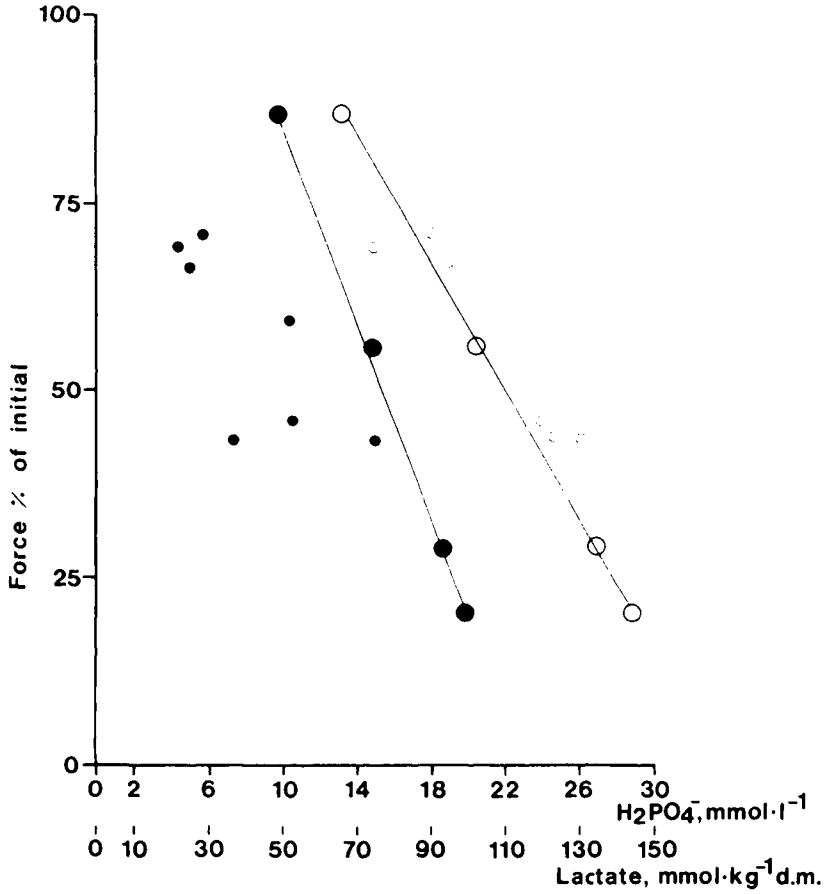


Figure 5. The relationships between force generation and calculated concentration of $H_2PO_4^-$ (closed circles) and between force and lactate content (open circles) in the stimulated muscle presented in Figure 2. Corresponding relationships are also presented for the muscle during a 3 min recovery period (small closed and open circles). In the recovery period, the muscle was stimulated at 20 Hz for 1.6 sec at 30 sec and at one, two, and three min after the fatiguing contraction. From Hultman et al. (1990), with permission from Human Kinetics Publishers.

served between the decrease in force, the fall in available high energy compounds in the cell, and the accumulation of degradation products such as P_i and lactate. However, in some situations, there are dissociations between metabolic changes and fatigue. If the intermittent stimulation shown in Figure 2 is repeated with an open circulation instead of occluded, the stimulation can be continued. The force decreased by 50% within the first 90 sec and thereafter continued to fall, reaching

40% after 5 min and 20% of initial after 20 min stimulation (Hultman and Spriet, 1986). At the same time, ATP and PCr concentrations were the lowest after 40 sec of contraction and increased thereafter. The lactate concentration peaked at 60 sec and fell thereafter. At the end of the stimulation (45 min) the force had decreased to 18% of initial while the ATP, PCr, and lactate concentrations had essentially returned to resting levels. Similar results have been observed in intact rat muscle studied by NMR (Le Rumeur et al., 1989). Apparently factors other than metabolic changes are responsible for this type of fatigue. As already discussed, a pronounced fall in Ca^{2+} release from the SR could explain the fall in force generation, either specifically in type II fibers or in both fiber types. The mechanism for this fall in Ca^{2+} release, when the muscle energy status is preserved, is not obvious. It may be due to Ca^{2+} binding to parvalbumin and mitochondria, with a resulting decrease in Ca^{2+} availability in the SR.

FATIGUE IN INTACT SKELETAL MUSCLE OF HUMAN SUBJECTS DURING SUBMAXIMAL EXERCISE

Introduction to Submaximal Exercise

Most of the activities that humans normally engage in do not fall into the category of high intensity exercise. The majority of routine daily activities and many forms of exercise are submaximal and require contractions that are also submaximal. When studying submaximal exercise there is a greater emphasis on whole body exercise, where several muscles or muscle groups act in concert to accomplish the required movement, and less on isolated muscle or muscle group preparations. While the various muscles used in any task may not be working at the same intensity, a standardized method is used to measure the submaximal intensity of various forms of whole body activity. This method expresses the intensity of a given task as a percentage of the exercise intensity that requires the body to use oxygen at its maximal rate. A person's oxygen uptake will increase linearly as the intensity of the exercise increases, until the maximum oxygen uptake (VO_2 max) is reached. Humans who are moderately active will generate ~200–300 Watts (W) of power before reaching their VO_2 max, while aerobically trained individuals will reach power outputs that are ~10–25% higher at VO_2 max. This system for classifying whole body submaximal exercise as a percentage of VO_2 max is valuable. For example, knowing that a person is cycling at 50% VO_2 max provides information about the intensity of the contractions required by the muscles responsible for the cycling movement.

This categorization of relative whole body intensity is also valuable in distinguishing between high intensity and submaximal exercise. Exercise intensities or power outputs at or above 100% VO_2 max are generally considered high intensity. These intensities require muscle contractions that generate ~40–100% of the force

produced during a maximal voluntary contraction and are often referred to as maximal or near-maximal. For example, truly maximal muscle contractions during cycling will generate 100% of the MVC force and 750–900 W of external power, but can only be sustained for a few seconds. These high external powers are ~3 times the power output required to elicit 100% VO_2 max (300% VO_2 max)! On the other hand, exercising at 125% VO_2 max requires muscle contractions that are only ~50% of MVC. While this muscle contraction intensity is much lower and can be maintained for 2–3 min, they are still considered high intensity and near-maximal. As mentioned earlier, high intensity contractions rely heavily on anaerobic energy to produce high forces. Because of the limited capacity of the anaerobic processes, exercise intensities between 100–300% VO_2 max cannot be sustained for more than a few seconds to a few minutes.

Exercise intensities below 100% VO_2 max are considered submaximal and require muscles to contract with forces less than ~40% of MVC. The energy required to produce forces at or below 40% of MVC can generally be sustained by oxidative metabolism, with small contributions from anaerobic metabolism. Oxidative metabolism produces ATP at a lower rate than the anaerobic processes but has a much greater capacity. Therefore, depending on the intensity, submaximal muscle contractions can be maintained from minutes to several hours. This assumes that the delivery of oxygen to the working muscles is not impaired and that the substrates for oxidative metabolism (fat and carbohydrate) are not depleted. Well-motivated individuals can exercise for 5–7 min at 100% VO_2 max, 30–60 min at 85% VO_2 max, over two hours at 70%, 3–6 hours at 50% and eight or more hours if the intensity is at or below 40% VO_2 max. Aerobically trained subjects will exercise longer at any given intensity before fatigue occurs while the opposite is true for people unaccustomed to exercise.

Factors That Limit Maximal Oxygen Uptake (VO_2 max)

Paramount to any discussion of fatigue in muscles exercising at submaximal intensities is an understanding of the factors that set the upper limit for a muscle to produce oxidative energy. The ability of a muscle to sustain a submaximal intensity is a function of the capacity of the oxidative processes to produce ATP, which in turn are dependent on adequate availability of oxygen and oxidizable substrates. A less than adequate provision of oxygen or substrates for oxidation will force the muscle to rely more heavily on anaerobic energy processes. Anaerobic metabolism is associated with the buildup of various byproducts that are associated with fatigue and a need to shift from aerobic to anaerobic energy processes will rapidly lead to muscle fatigue. Therefore, while fatigue in many submaximal exercise situations ultimately occurs in muscle, the cause of the fatigue may originate outside the muscle.

The respiratory and cardiovascular systems are responsible for transporting oxygen from the environment to the working muscles. The potential sites for limitations in the transport and utilization of oxygen by working muscles during whole body exercise are often categorized as: (a) the respiratory system, (b) the central circulation, (c) the peripheral circulation, and (d) the oxidative capacity of the muscle. In moderately active people and well trained endurance athletes, the factor which sets the upper limit for VO_2 max appears to be the central circulation. When the amount of oxygen that is transported to the working muscles is increased during exercise at ~110–120% VO_2 max, the VO_2 max increases (Thomson et al., 1982; Spriet et al., 1986). This suggests that the ability of the working muscles to utilize oxygen is greater than the amount of oxygen it receives in normal situations. However, there are individuals where systems other than the central circulation appear to limit VO_2 max. In one example, the respiratory system limits VO_2 max in a select group of highly trained runners (Powers et al., 1989). During exercise at or near VO_2 max, these individuals do not fully saturate their blood with oxygen as it passes through the lungs, resulting in less than optimal amounts of oxygen delivery to the working muscles. When this desaturation is corrected, VO_2 max increases. In a second example, the peripheral circulation or the oxidative capacity of the muscles limits VO_2 max in people who are very unaccustomed to exercise. In these individuals, few capillaries surround the muscle fibers making it difficult to deliver oxygen and the muscle fibers also have very low oxidative rates, such that some of the delivered oxygen cannot be used. In this case the ability of the central circulation to deliver oxygen to the muscles is adequate but local delivery and utilization are limiting the capacity of the muscles to produce oxidative energy.

In the discussions that follow, it is important to remember that the ability to deliver and utilize oxygen is very different between individuals. Two individuals may be working at 80% of VO_2 max, where one is producing 250 W of power and the other is producing only 175 W. Their absolute power outputs are quite different but their relative intensities are the same. Therefore, expressing exercise intensity as a percentage of the VO_2 max permits comparisons between people of differing aerobic potentials. It is equally important to remember that the provision of oxidizable substrate may also limit muscle performance and be the cause of muscle fatigue as discussed in a following section.

Muscle Fatigue During Exercise at 85–100% of VO_2 max

In terms of the working muscles, exercise in this range of whole body intensities can be looked upon as the transition zone between exercise that is clearly high intensity or clearly submaximal. At the onset of exercise, the oxygen uptake by the working muscles will increase linearly with time until VO_2 max is reached within 60–90 sec. This is the time required for the cardiovascular system to increase the delivery of oxygen to the working muscles, by increasing the cardiac output and

diverting blood flow from nonexercising tissues to the working muscles. The lag in oxidative energy production necessitates the utilization of anaerobic processes to meet the energy requirements of this submaximal but intense power output. The amount of anaerobic energy required during this period can be up to 50–70 ml of oxygen/kg body weight or 3.5–5.0 L of oxygen in total (Medbo et al., 1988; Scott et al., 1991).

A second factor also contributes to muscle fatigue at 85–100% VO_2 max, even when the maximal or near maximal oxygen uptake has been reached. The muscles used to accomplish these whole body exercise intensities will not all be working at the same level. Some muscles will be working at intensities that are actually high intensity and will require an anaerobic energy contribution, while others will be at submaximal intensities that can be sustained aerobically. Muscles in the former group may continue to use anaerobic energy processes throughout exercise, until these processes are exhausted (PCr degradation) or inhibited (glycolysis).

The requirement to use anaerobic energy processes during whole body exercise at 85–100% VO_2 max causes the same changes in muscle as described for high intensity exercise. PCr is degraded with the resultant release of P_i and high anaerobic glycolytic activity produces large amounts of lactate and the associated increase in H^+. Some of the produced lactate can escape the muscle, leading to high plasma lactate concentrations, but the muscle lactate production rate is often so high that the muscles are acidotic at the point of fatigue. If the intensity of the exercise is 100% VO_2 max, and fatigue occurs in ~5–7 min, muscle metabolite contents closely mimic those measured at the fatigue point during high intensity exercise (Karlsson and Saltin, 1970). If the exercise intensity is 85% VO_2 max, where fatigue may occur in 30–60 min, the muscles often do not mimic the profile of high intensity exercise (Hultman et al., 1967) and fatigue occurs for other reasons, as discussed in the next section. Therefore, exercise intensities between 100% and 85% VO_2 max produce muscle characteristics at the point of fatigue that follow a continuum between high intensity and submaximal exercise.

Many factors will determine how long each individual will be able to exercise at a given whole body intensity, even though muscle fatigue may be the common problem. For example, a well trained competitive athlete may be able to exercise for 30–60 min at 90% VO_2 max, while a very untrained subject may fatigue in 2–5 min. In other words, the systems that provide oxygen to the working muscles and the metabolic pathways that use the oxygen to provide ATP are maximized in the athlete through a combination of genetic endowment and environmental adaptation to training and minimized in the untrained individual.

Muscle Fatigue During Exercise at 60–85% of VO_2 max

People who perform aerobic exercise on a regular basis will most often exercise at an intensity that falls between 60–85% VO_2 max. While exercise at 80–85% VO_2

max will produce muscle fatigue in 30–90 min, exercise at 60% VO_2 max may be continued for several hours. In this type of exercise adequate oxygen delivery to the working muscles can be sustained and does not limit muscle performance. Muscle fatigue appears to be more related to the ability to provide adequate amounts of oxidizable fuel to the muscles.

Substrate Supply

Carbohydrate (CHO), fat, ketone bodies, and protein can all be oxidized by working muscle. However, in the healthy well-fed individual, CHO and fat provide greater than 90% of the oxidizable substrate during this range of intensities. Ketone bodies are not present in significant amounts in the plasma of well-fed subjects and only become a significant fuel source during fasting and starvation. The quantitative role of protein as an oxidizable fuel is also minor in the well-fed state and may contribute at most 5–7% of the total required ATP during prolonged submaximal exercise.

The energy for exercising at submaximal intensities is provided through the combustion of both fat and CHO. The major sources of oxidizable CHO are glycogen stored in the muscles and glucose taken up from the blood. The blood glucose is maintained by the liver, through the degradation of glycogen and to a smaller extent by gluconeogenesis. Fat is available from local stores of muscle triacylglycerol (TG) and via the uptake of plasma free fatty acids (FFAs). Hydrolysis of TG in the adipose tissue stores of the body releases FFAs into the blood. During prolonged exercise at intensities below 50% VO_2 max, CHO in the form of blood glucose and muscle glycogen provide less than one-half of the total substrate used, while plasma FFAs contribute most of the fat substrate. Above 50% VO_2 max, CHO provides the largest amount of substrate. As the exercise intensity increases above 50% VO_2 max, muscle glycogen becomes the dominant substrate and glucose uptake is quantitatively less important.

The importance of muscle glycogen for submaximal exercise is demonstrated by the fact that depletion of muscle glycogen stores results in muscle fatigue and therefore limits prolonged exercise at intensities between 60% and 85% VO_2 max. In one of the first studies examining the relationship between muscle glycogen and endurance performance, subjects cycled intermittently for as long as possible at ~75–80% VO_2 max in a sequence of 15 min of work and 15 min of rest (Bergström and Hultman, 1967). Muscle glycogen utilization was the greatest during the initial 15 min and then decreased with successive 15 min bouts such that glycogen content plotted against time decreased in a semilogarithmic fashion (Figure 6). Most importantly, muscle and whole body fatigue coincided with the depletion of glycogen in the working muscles. Measured rates of muscle glycogen degradation during submaximal exercise are illustrated in Figure 7. The rate increases exponentially with increasing submaximal intensity as studied in vastus lateralis, soleus, and gastrocnemius muscles during running and cycling. There is an absolute need

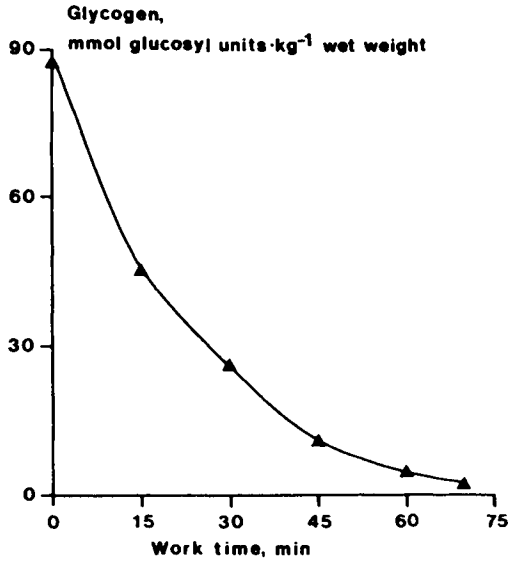


Figure 6. Glycogen content in the vastus lateralis muscle as a function of cycling time at 75–80% of maximal oxygen uptake (VO_2 max). Data points are mean values from 10 subjects. For each subject, exercise was performed repeatedly in periods of 15 min separated by 15 min rest periods. At the point of exhaustion and muscle fatigue, muscle glycogen stores were depleted. From Bergström and Hultman (1967) with permission from the publisher.

for CHO in the form of muscle glycogen at intensities above 65–70% VO_2 max, as the rate of muscle glucose uptake from blood is much lower than can be provided from muscle glycogen (Katz et al., 1986a).

While it is clear that muscle glycogen is the dominant oxidative fuel during exercise at 60–85% VO_2 max, and muscle fatigue results when it is depleted, why is fat not capable of replacing CHO as the major substrate when glycogen depletion occurs? There are at least three reasons why CHO is a more versatile fuel than fat. The most important difference is that CHO, in the form of muscle glycogen, can be provided as a substrate for aerobic metabolism at a rate which is ~1.5–2 times that of fat (McGilvery, 1975). In other words, reliance on fat oxidation alone provides only enough ATP to sustain exercise at an intensity of up to ~60–70% VO_2 max. Therefore, CHO oxidation is required at or above this submaximal intensity in order to meet the energy demand of the task. A second advantage to using CHO aerobically is that it is ~10% more efficient than the combustion of fat in terms of oxygen utilized per kilojoule of energy released. A third advantage is that only CHO can be used anaerobically. While this is not a major concern at these submaximal intensities, it is important during transition periods such as from rest to exercise or from one intensity to a higher intensity. It may also be useful in

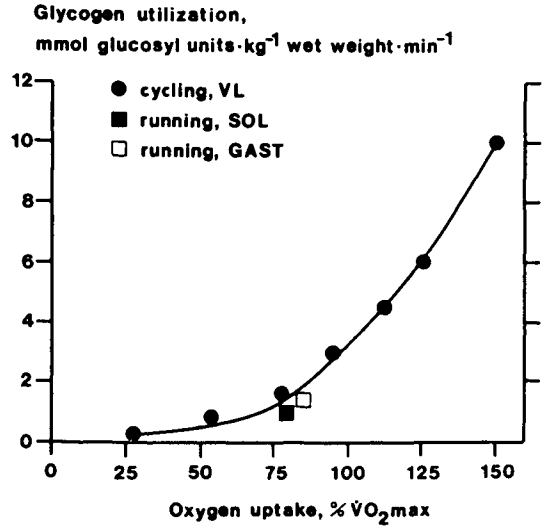


Figure 7. Muscle glycogen utilization rates at various exercise intensities expressed as a percentage of VO_2 max. VL, vastus lateralis; SOL, soleus; GAST, gastrocnemius. Redrawn from Costill et al. (1971), Hermansen et al. (1967), Saltin and Karlsson (1971), and Sherman et al. (1981).

exercising at 80–85% VO_2 max where small amounts of anaerobic energy are periodically required.

There is, however, one major disadvantage to the continued combustion of CHO and specifically when the source is muscle glycogen. The capacity for energy production from CHO is finite and much less than from fat (Table 1). Continued

Table 1. Energy Stores in a Well-Fed Man[1]

Energy Source	*Content (mmol/kg wet muscle)*	*Energy per Mole (kJ)*	*Energy Stored in Body (kJ)*
Carbohydrate			
Muscle glycogen	80–100	2,850	6,400–8,000
Liver glycogen	300–500	2,850	1,550–2,600
Plasma glucose[2]	5	3,150	50
Fat			
Adipose tissue TG	—	30,500	275,000
Intramuscular TG	10–15	30,500	8,500–12,800
Plasma FFAs	0.3–0.6	10,150	9–18

Notes: [1]Assuming a moderately active 70 kg man with 40% of body weight as muscle (28 kg), a liver weight of 1.8 kg, a plasma volume of 3 L, and 9 kg of adipose tissue. Endurance trained individuals store ~125–150 mmol/kg wet muscle of glycogen in muscle and 400–700 mmol/kg wet tissue of glycogen in liver.
[2]mmol/l.
TG, triacylglycerol; FFAs, free fatty acids.

use of the muscle glycogen store will cause depletion and muscle fatigue, limiting the amount of exercise that can be performed. Therefore, working muscles are faced with a situation where fat stored in adipose tissue and muscle is an almost unlimited source of energy for submaximal exercise, but it can only be used at a rate that will sustain 60–70% VO_2 max. On the other hand, CHO, mainly in the form of muscle glycogen, can be used to sustain exercise up to and beyond 85% VO_2 max, but the CHO reserves are very limited and quickly depleted (Table 1). This limited storage capacity for muscle CHO is the reason why dietary intake of CHO prior to and during exercise is very important. Proper nutritional practices will ensure that body CHO stores are maximized and muscle fatigue is prevented or at least delayed during submaximal exercise.

Diet, Exercise, and Muscle Glycogen Content

Christensen and Hanson (1939, as cited by Asmussen, 1971) were the first to demonstrate that varying the availability of CHO in the diet influenced submaximal endurance performance. A high-fat, low-CHO diet in the days prior to exercise decreased exercise time to fatigue, while CHO-rich food increased exercise endurance. These early experiments were unable to identify the mechanisms involved in the relationship between CHO supply and exercise endurance at the level of the muscle. The reintroduction of the needle biopsy technique for sampling human skeletal muscle by Bergström and Hultman in the 1960s made it possible to directly examine the importance of dietary CHO supply and muscle glycogen stores during exercise. The administration of a CHO-rich diet, a CHO-poor diet, or even total starvation without accompanying exercise had little effect on muscle glycogen stores (Hultman and Bergström, 1967) (Figure 8A). If the dietary manipulations were preceded by glycogen depleting exercise, remarkable differences in the rate of glycogen resynthesis and final glycogen content attained were observed. Administration of a CHO-poor diet or complete starvation following the glycogen depleting exercise produced slow rates of glycogen resynthesis over a 1–7 day period (Figure 8B). When this period was followed by 1–2 days of a CHO-rich diet, muscle glycogen stores were quickly replenished to values in excess of the normal resting levels. To further investigate the phenomenon of repleting glycogen stores to higher than normal levels (supercompensation), the same investigators had subjects exercise with only one leg to deplete muscle glycogen in the working leg but not in the rested leg (Bergström and Hultman, 1966). Following exercise, two subjects ate a CHO-rich diet for 3 days and increased glycogen content in the depleted muscles to 120%, 193%, and 230% of normal resting levels after one, two, and three days, respectively, while only small increases occurred in the muscles of the rested leg (Figure 9A). Two other subjects fasted for two days following exercise and then consumed a CHO-rich diet for four days (Hultman and Bergström, 1967). In the exercised leg, glycogen resynthesis was low during fasting

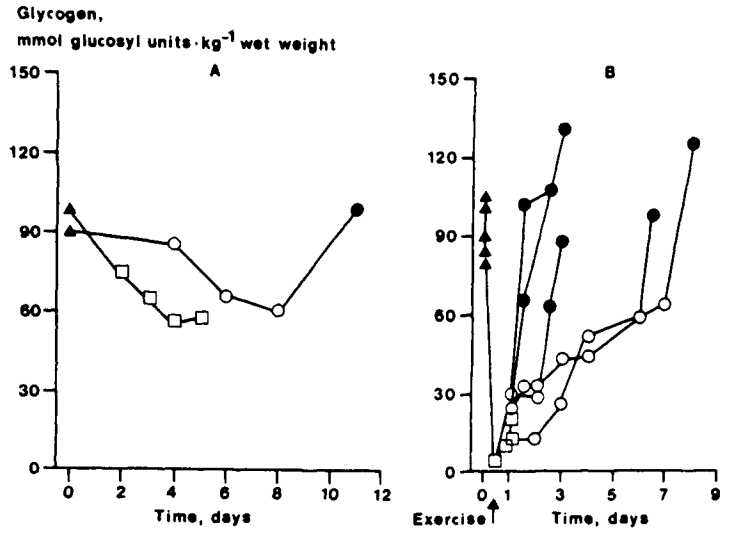


Figure 8. **A.** Glycogen content in the vastus lateralis muscle after a mixed diet (▲) and during 5 days of total starvation (□) in one subject and eight days of carbohydrate-poor diet (○) followed by a carbohydrate-rich diet (●) in a second subject. **B.** Muscle glycogen content before and after exercise. Before exercise the diet was mixed (▲) and in the following days was either total starvation (□) or carbohydrate-poor (○) and finally followed by 1–2 days of a carbohydrate-rich diet (●). Note the slow rate of glycogen resynthesis when the diet is carbohydrate-poor compared to the rate when the diet is carbohydrate-rich. Redrawn from Hultman and Bergström (1967).

but quickly increased to about double the normal value on the CHO-rich diet (Figure 9B). In the rested leg, glycogen decreased to 60% of the postexercise level during the fast, and increased to only 140% of the postexercise level with the high-CHO diet. Similar results were obtained in exercised and nonexercised muscles of two subjects following three days of a CHO-poor diet followed by three days of a CHO-rich diet (Figure 9C). These studies clearly demonstrated that this supercompensation of muscle glycogen levels was a local phenomenon, restricted to the exercised muscles. The next phase of this research was to confirm that supercompensated muscles could contract longer at a submaximal intensity before fatigue occurred, as originally reported by Christensen and Hanson (1939).

Diet, Exercise, and Muscle Performance

Bergström et al. (1967) examined the relationship between initial muscle glycogen content and the capacity for prolonged submaximal exercise. Subjects cycled to exhaustion at 75% VO_2 max on three occasions, each separated by three days.

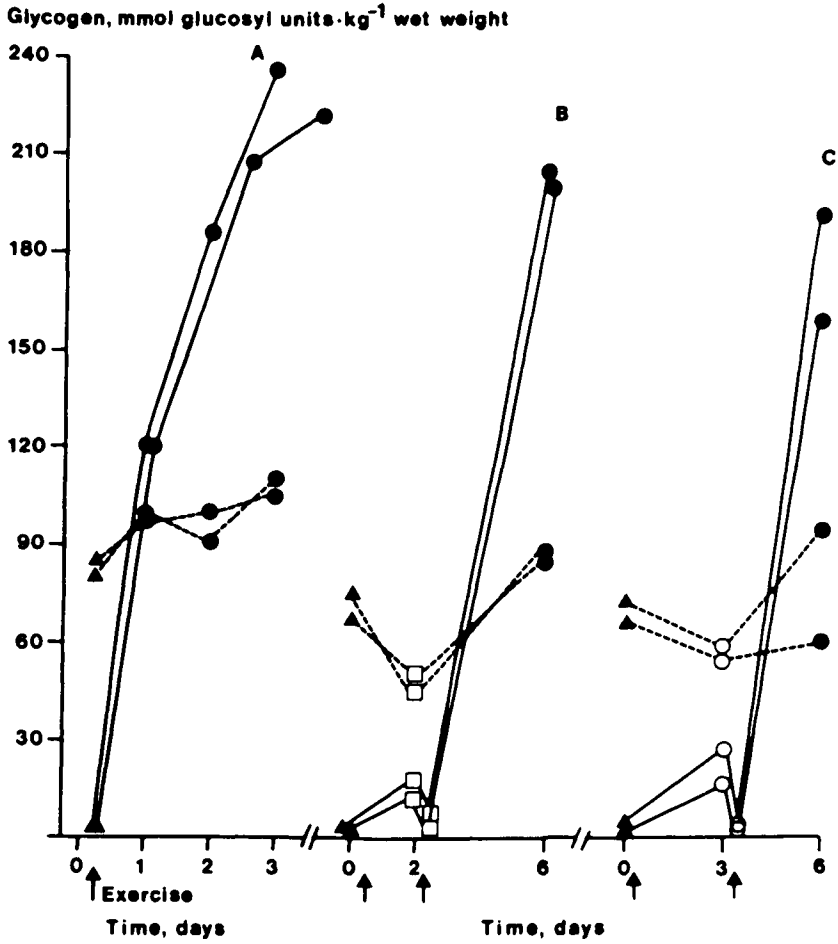


Figure 9. One-legged exercise studies showing the muscle glycogen content of the exercised (—) and rested legs (- - -) in two subjects. **A**. Muscle biopsy samples were obtained immediately after exercise (▲) and during three days when fed a carbohydrate-rich diet (●). **B** and **C**. The diet was total starvation (□) for two days following exercise (**B**) or carbohydrate-poor (○) for three days following exercise (**C**). This was followed by a second one-leg exercise bout (↑) and a carbohydrate-rich diet (●). Redrawn from Bergström and Hultman (1966) in panel A, and from Hultman and Bergström (1967) in panels B and C.

A normal mixed diet was given prior to the first ride, a CHO-poor diet prior to the second, and a CHO-rich diet before the third. The mixed, CHO-poor, and CHO-rich diets produced mean preexercise concentrations of 118, 42, and 227 mmol/kg wet muscle, respectively. The corresponding exercise times were 126, 59, and 189 min. An excellent correlation existed between preexercise glycogen content and cycling

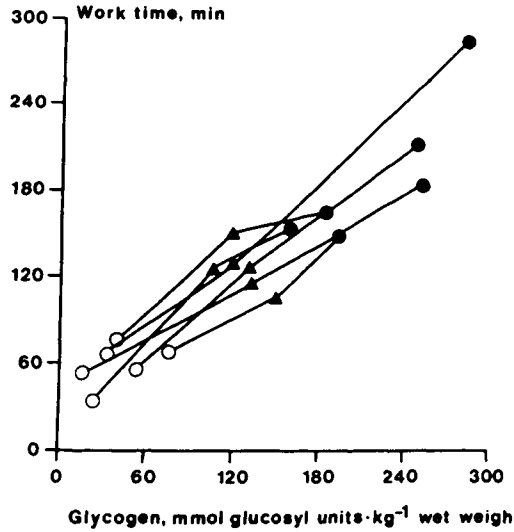


Figure 10. The relationship between the initial glycogen content in vastus lateralis muscle and work time in six subjects who cycled to exhaustion at 75% VO_2 max. Each subject cycled to exhaustion on three occasions. The first experiment was preceded by a mixed diet (▲), the second by a carbohydrate-poor diet (○), and the third by a carbohydrate-rich diet (●). The energy contents of the diets were identical. In all experiments depletion of the muscle glycogen store coincided with exhaustion and muscle fatigue. From Bergström et al. (1967) with permission from the publisher.

time to exhaustion (Figure 10). A most important finding was that exhaustion coincided with glycogen depletion in the working muscles regardless of the preceding dietary regimen and preexercise glycogen content. A field study produced similar results. Subjects ran a 30 km cross-country race on two occasions separated by three weeks, after consuming a mixed diet on one occasion and a CHO-rich diet on the other (Karlsson and Saltin, 1971). Preexercise mean glycogen contents were 109 and 216 mmol/kg wet muscle for the mixed and CHO-rich diets, respectively. Corresponding race times were 143 and 135.3 min. Higher preexercise glycogen content was associated with a faster running speed during the last half of the race.

Liver Carbohydrate and Muscle Performance

The glycogen content in the liver of a well-fed individual is 3–5 times the level found in muscle (Table 1). However, since the mass of the liver is small, the total amount of energy that can be provided from liver CHO is much less than muscle. The liver's primary responsibility with respect to CHO metabolism is to maintain

blood glucose levels during rest and exercise, as certain tissues in the body rely exclusively on CHO for fuel. While muscle glycogen is the main source of CHO during submaximal exercise at 60–85% VO_2 max, the working muscles do take up and oxidize significant amounts of blood glucose. This loss of blood CHO must be replaced by the liver. Experiments have shown that liver glycogen content decreases at a rate of 2–2.5 mmol/kg wet tissue/min during one hour of exercise at 70–75% VO_2 max (Hultman and Nilsson, 1971). This rate of liver glycogen breakdown is sufficient to maintain blood glucose at resting levels of ~5 mM for at least 2.5 hours. In addition, the liver uses substrates that are delivered via the blood (amino acids, lactate, glycerol) for the *de novo* synthesis of glucose during exercise. However, none of this *de novo* glucose is stored as glycogen during exercise and the capacity of this process is much smaller than the provision of glucose from liver glycogen. Therefore, in well-fed individuals the blood glucose should be maintained for over three hours during intense submaximal exercise.

When well-fed subjects are fasted and not allowed to exercise, liver glycogen content decreases while muscle glycogen stores are essentially unchanged. The few liver measurements that exist suggest that a 12–16 hour fast will decrease the glycogen store by 50% and a full day fast will almost deplete it (Nilsson and Hultman, 1973). When subjects are then asked to perform prolonged submaximal exercise, they fatigue not because of muscle glycogen depletion and fatigue but due to low blood glucose (< 2.5 mM) and the associated CNS dysfunction. Therefore, it is important that fasting does not occur prior to prolonged submaximal exercise. If CHO is consumed 2–4 hours prior to exercise, the liver glycogen stores will be full and the cause of fatigue during submaximal exercise should shift back to the working muscles. Liquid feedings of CHO up to one hour prior to exercise will also ensure optimal liver glycogen content if a preexercise meal is not desired. During the final 30–60 min prior to exercise, liquid CHO feedings are not recommended because of the associated increase in plasma insulin in some individuals. A high plasma insulin at the onset of exercise can inhibit the release of FFAs from adipose tissue to blood and may contribute to hypoglycemia, presumably by augmenting the insulinlike effect of exercise on muscle glucose uptake (Costill et al., 1977).

Carbohydrate Intake During Submaximal Exercise

The previous sections clearly indicate that proper dietary intake during the days and hours prior to exercise, ending with a CHO-rich meal 2–4 hours prior to exercise, will ensure that muscle and liver glycogen levels are optimal. Adherence to these dietary principles will delay the onset of muscle fatigue during prolonged submaximal exercise at 60–85% VO_2 max. However, there is a limit to these procedures and while muscle fatigue will be delayed it cannot be prevented if exercise continues. Eventually the extended exercise will lead to hypoglycemia if

the liver glycogen store becomes depleted or muscle fatigue if the muscle glycogen stores are depleted first. In an attempt to delay or prevent these events from occurring, CHO supplements are routinely administered during exercise when CHO depletion may occur (e.g., ~1–2 hours at 70–85% VO_2 max or >2 hours at 60–70% VO_2 max).

Studies examining the effects of CHO feedings (150–250 mL of water containing 140–420 mmol/L CHO (25–75 g/L) every 15–20 min) during submaximal exercise reported little effect on the rate of muscle glycogen utilization. For example, Coyle et al. (1986) reported no glycogen sparing effect of CHO feedings during three hours of cycling at 71% VO_2 max in trained cyclists. This is not surprising since it is known that blood glucose only provides 12% of the total energy required during prolonged exercise at 70% VO_2 max (Costill et al., 1973). However, the major effect of the CHO feedings in the study by Coyle et al. (1986) was to prevent hypoglycemia, such that exercise could be continued for an additional hour when given CHO. Other studies suggest that some individuals do improve their exercise performance by slowing the rate of muscle glycogen utilization and delaying muscle fatigue, rather than delaying or preventing hypoglycemia-induced central fatigue (Coyle et al., 1983).

Other Potential Causes of Muscle Fatigue at 60–85% VO_2 max

Muscle fatigue during this intensity range may be due to other factors besides lack of CHO substrate. The potential mechanisms include dehydration, electrolyte imbalances, and the onset of muscle injury and soreness. As these potential fatigue factors are common to the range of intensities in this section and exercise below 60% VO_2 max, they will be discussed in the following section.

Muscle Fatigue During Exercise Below 60% of VO_2 max

Muscles are capable of contracting for very long periods of time when the exercise intensity is below 60% VO_2 max. For example, endurance times of 3–5 hours at 50% and over eight hours at or below 40% VO_2 max are reported. There are many instances in the workplace where individuals work at intensities averaging 20–40% VO_2 max for an eight hour shift.

At an exercise intensity of 50% VO_2 max, the substrate used for oxidative metabolism is equally split between CHO and fat. Above this intensity CHO becomes the dominant fuel, while fat is the dominant fuel below 50% VO_2 max. There is an increasing reliance on plasma glucose and FFAs (extramuscular sources) for meeting the substrate needs of the muscle at lower exercise intensities. During the course of prolonged low intensity exercise there is also a progressive shift to greater reliance on fat and less on CHO. The rates of glucose and FFA uptake

and oxidation during low intensity submaximal exercise can provide a large percentage of the total energy requirements of the working muscles. This means that very low rates of muscle glycogen and triacylglycerol (TG) are needed to supplement the exogenous provision of substrate.

Maintenance of blood glucose concentration is a concern during prolonged low intensity exercise. At 70% VO_2 max, plasma CHO contributed 12% of the total CHO oxidized and provided ~7% of the total energy required. At 45% VO_2 max, plasma glucose provided 55% of the metabolized CHO and 24% of the total energy expenditure. Since exercise at 45% VO_2 max can be sustained for several hours, the potential for a decrease in blood glucose is very real. This could occur in spite of a shift from CHO to fat oxidation as the exercise continues. A lowered blood glucose will either cause an increased reliance on muscle glycogen and subsequent muscle fatigue when muscle glycogen is depleted or hypoglycemia and the associated central fatigue. This problem can easily be avoided through regular ingestion of meals and liquid feedings containing CHO-rich foods.

Dehydration and Muscle Performance

Most forms of exercise are relatively inefficient as only ~20–25% of the energy released from metabolism is ultimately used to perform measurable external work. This means that 75–80% of the energy metabolism is converted to heat. The heat is dissipated through the process of sweating and the evaporation of the water from the surface of the body. The fluid loss associated with cooling the body during exercise is substantial. A 70 kg subject working at 80% VO_2 max will lose 1.6, 2.0, and 2.4 L of water/hour at ambient temperatures of 10, 20, and 30 °C, respectively (Saltin and Costill, 1988). Dehydration will occur if the lost water is not replaced. When exercise-induced dehydration reaches 4–6% of total body water, 50% (1.5–2 L) of the lost water has come from the intracellular compartment. This impacts negatively on exercising muscle cells as they comprise a large portion of the cells in the body. All processes in the muscle cell and especially aerobic metabolism must then function in the face of a shrinking volume of intracellular water.

Decrements in muscle and exercise performance are measurable during prolonged work when the body fluid loss is as little as 1%. Muscular work and physical labor of long duration decrease by 20–30% if the body water loss is 4–5% (Saltin, 1964). Performance during high intensity submaximal exercise for shorter duration will be affected less, but race performances by runners decreased by 3–7% during races of 1,500, 5,000, and 10,000 m (Armstrong et al., 1985). Therefore, while not all of the negative effects of dehydration can be attributed specifically to the working muscles, the evidence suggests that muscle ability to contract and provide the required rate of ATP during exercise is compromised.

Electrolyte Balance and Muscle Performance

Discussion regarding the possibility that electrolyte disturbances contribute to fatigue during prolonged submaximal exercise has centered on two possibilities: Na^+ loss from the body in sweat, and loss of K^+ from muscle to extracellular sites. Since large amounts of Na^+ can be lost from the body in sweat, the plasma and total body stores of Na^+ can decrease during prolonged submaximal exercise. This problem can be heightened if only water is ingested as it serves to further dilute the remaining Na^+ in the plasma. The onset of hyponatremia affects the CNS more than muscle. There have been reports of runners who had plasma Na^+ values of 118–123 mM at the point of disorientation and collapse, instead of the usual 140 mM (Saltin and Costill, 1988). Heat-acclimated and endurance trained individuals lose less Na^+ in their sweat and ingesting small amounts of Na^+ is recommended if exercise is to be continued for 2–6 hours.

A more serious electrolyte change may be the redistribution of K^+ that occurs during exercise. K^+ leaves muscle fibers and contributes to an increase in plasma $[K^+]$ during exercise in direct proportion to the intensity of muscular contraction (Lindinger and Sjögaard, 1991). It has been proposed that these K^+ changes contribute to decreases in muscle activation at the level of the muscle membrane (Lindinger and Sjögaard, 1991). When exercise is completed, intra- and extracellular $[K^+]$ are quickly restored. It is not clear whether the loss of muscle K^+ continues during prolonged submaximal exercise as suggested in an earlier study (Sjögaard et al., 1985), as a more recent investigation reported that continued muscle K^+ efflux did not occur (Lindinger et al., 1993). Thus the contribution of K^+ shifts to muscle fatigue during submaximal exercise remains to be determined.

Muscle Injury and Soreness

Skeletal muscles are designed to produce force during contractions while the muscle is either shortening (concentric), not changing its length (isometric), or lengthening (eccentric). Delayed onset muscle soreness occurs within 24–48 hours after three main types of exercise situations: exercise where many eccentric contractions are required (e.g., downhill running), any form of prolonged exercise by a person that is unaccustomed to exercise, and ultra-endurance exercise in individuals who are accustomed to shorter duration exercise (referred to as overuse syndrome). While it is difficult to understand how muscle could be injured during the activity it was designed for, symptoms include weakness, stiffness, and extreme soreness to touch and during contraction. Muscle soreness is also characterized by an inability to generate and maintain high forces (> 50–100% MVC).

The high active tension and/or high active strain that occurs in muscle during lengthening contractions is believed to cause mechanical disruption of muscle fibers and connective tissue (Armstrong, 1984; Lieber and Friden, 1993). Activa-

tion of proteolytic enzymes following the injury degrade lipid and protein structures most notably in membranes and along the Z-band in muscle. This permits the diffusion of fluids and electrolytes between the plasma and muscle cell compartments leading to inflammation and abnormally high Ca^{2+} levels in the cells. It is believed that sensory fibers in the muscle are sensitive to this increased pressure in the cell and may be responsible for the sensation of the pain associated with muscle soreness. The best course to prevent these problems is to avoid unusually strenuous amounts of activity and build up to increased amounts of submaximal activity slowly. Recent investigations have demonstrated that antiinflammatory drugs are useful either before or after exercise that leads to muscle soreness (Hasson et al., 1993). They decrease the sensation of pain and improve muscle performance in situations where athletes are unwilling or unable to avoid exercise that provokes muscle soreness.

SUMMARY

Fatigue in Isolated Skeletal Muscle Fibers

This section examined small muscle preparations stimulated electrically or skinned muscle fibers maintained in superfused baths with different substance concentrations. The electrical stimulation was either continuous or intermittent for relatively short durations.

High frequency fatigue or fatigue during continuous high frequency stimulation seems to be mainly due to impaired propagation of the stimulating impulse in the T-tubular system.

Fatigue during intermittent tetanic stimulation is caused by three mechanisms: (a) decreased maximum Ca^{2+} activated tension, (b) decreased Ca^{2+} sensitivity of the contractile proteins, and (c) decreased Ca^{2+} release from the SR.

The first two mechanisms are probably due to the combined effects of myoplasmic accumulation of P_i and H^+ formed when PCr and glycogen are degraded in muscle when the rate of oxidative ATP formation is insufficient. Crossbridge movements are impaired due to the P_i increase and crossbridge formation is decreased when Ca^{2+} binding sites on troponin are occupied by H^+. The third fatigue factor, reduced Ca^{2+} release from the SR, is related to metabolism and the decrease in MgATP following the depletion of the PCr store. The MgATP decrease may inhibit Ca^{2+} release from the SR due to the increased free $[Mg^{2+}]$ and the decrease in ATP. A decreased [ATP] could affect action potential propagation, IP_3 turnover rate and the activity of the SR Ca^{2+} pump. A fall in SR Ca^{2+} pump activity together with increased Ca^{2+} binding in the myoplasm of fatiguing muscle would decrease the SR Ca^{2+} availability.

During intense contractions P_i and H^+ accumulation begins early, resulting in a decrease in maximum Ca^{2+} activated tension and an increase in the Ca^{2+} concen-

tration necessary to produce tension. Both phenomena are potentiated when the Ca^{2+} release from the SR falls during continued stimulation, resulting in nearly complete loss of tension development.

Fatigue in Intact Skeletal Muscle of Human Subjects During High Intensity Exercise

High intensity exercise requires skeletal muscles to contract with forces that represent 40–100% of the maximal voluntary contraction force. Fatigue factors in human subjects during high intensity exercise include the accumulation of H^+ and P_i, with effects on Ca^{2+} sensitivity and maximum Ca^{2+} induced tension, and the availability of PCr and total creatine. The size of the PCr store is directly related to the MgATP/ADP ratio in the myoplasm of exercising muscle, while free creatine may have an impact on the rate of mitochondrial ATP and PCr formation, thereby adding to the regulation of the ratio during exercise. The MgATP/ADP ratio is involved in several processes regulating SR Ca^{2+} release.

Fatigue in Intact Skeletal Muscle of Human Subjects During Submaximal Exercise

Submaximal exercise in human subjects requires skeletal muscles to contract with forces that are below 40% of the maximal voluntary contraction force. Submaximal exercise intensities are expressed as a percentage of the power output required to elicit maximal oxygen uptake (VO_2 max). During exercise at 85–100% VO_2 max, anaerobic energy is required to supplement oxidative energy at the onset of exercise and in some muscles that cannot meet the energy demand aerobically. This produces accumulations of P_i and H^+ and produces fatigue as described in the high intensity exercise sections. At 100% VO_2 max, muscle fatigue will occur in 5–7 min or less, and at 85% VO_2 max in ~30 min.

Muscle fatigue during exercise at 60–85% VO_2 max may occur following 30 min at 85% VO_2 or after several hours at 60% VO_2 max. Fatigue is related to the ability to provide adequate amounts of oxidizable fuel to the muscles. Carbohydrate in the form of muscle glycogen is the required substrate and muscle fatigue coincides with glycogen depletion. Diet and exercise manipulations increase the storage of muscle glycogen thereby delaying muscle fatigue. Hypoglycemia can also occur during exercise when liver glycogen stores are depleted, leading to fatigue caused by central nervous system dysfunction. Ingestion of CHO during exercise can delay fatigue either by preventing hypoglycemia or slowing the rate of muscle glycogen use.

During exercise below 60% VO_2 max, there is greater reliance on fuel sources from outside the muscle (plasma glucose and FFAs). Since exercise can be sustained for 3–5 hours at 50% VO_2 max and over eight hours at or below 40%

VO_2 max, there is a strong possibility that dehydration and/or hypoglycemia will occur. Routine ingestion of meals and liquid feedings rich in CHO throughout exercise prevent these problems from occurring. Muscle injury and soreness will occur when exercise is performed by an individual unaccustomed to exercise or during exercise requiring a large number of eccentric contractions (e.g., downhill running). The affected muscles will be unable to generate normal forces while injured. Prevention of muscle injury through gradual increases in the amount of activity performed is recommended, but antiinflammatory agents decrease the soreness and improve force generation when muscle injury has occurred.

REFERENCES

Allen, D.G., Lee, J.A., & Westerblad, H. (1988). Intracellular calcium during fatigue in single fibers from *Xenopus* toe muscles. J. Physiol. 407, 75 (Abstract).

Allen, D.G., Lee, J.A., & Westerblad, H. (1989). Intracellular calcium and tension in isolated single muscle fibers from *Xenopus*. J. Physiol. 415, 433–458.

Armstrong, L.E., Costill, D.L., & Fink, W.J. (1985). Influence of diuretic-induced dehydration on competitive running performance. Med. Sci. Sports Exerc. 17, 456–461.

Armstrong, R.B. (1984). Mechanisms of exercise-induced delayed onset muscular soreness: A brief review. Med. Sci. Sports Exerc. 16, 529–538.

Asmussen, E. (1971). Muscle metabolism during exercise in man: A historical survey. In: Muscle Metabolism During Exercise (Pernow, B. & Saltin, B., eds.), pp. 1–12, Plenum Press, New York.

Balsom, P.D., Ekblom, B., Söderlund, K., Sjödin, B., & Hultman, E. (1993). Creatine supplementation and dynamic high intensity intermittent exercise. Scand. J. Med. Sci. Sports (in press).

Bergström, J. & Hultman, E. (1966). Muscle glycogen synthesis after exercise: An enhancing factor localized to the muscle cells in man. Nature 210, 309–310.

Bergström, J. & Hultman, E. (1967). A study of the glycogen metabolism during exercise in man. Scand. J. Clin. Lab. Invest. 19, 218–228.

Bergström, J., Hermansen, L., Hultman, E., & Saltin, B. (1967). Diet, muscle glycogen and physical performance. Acta Physiol. Scand. 71, 140–150.

Bergström, M. & Hultman, E. (1990). Contraction characteristics of the human quardriceps muscle during percutaneous electrical stimulation. Pflügers Arch. 417, 136–141.

Bessman, S.P. & Fonyo, A. (1966). The possible role of mitochondrial bound creatine kinase in regulation of mitochondrial respiration. Biochem. Biophys. Res. Commun. 22, 597–602.

Bigland, B. & Lippold, O.C.J. (1954). Motor unit activity in the voluntary contractions of human muscle. J. Physiol. 125, 322–335.

Bigland-Ritchie, B. (1981). EMG and fatigue of human voluntary and stimulated contractions. In: Human Muscle Fatigue: Physiological Mechanisms (Porter, R. & Whelan, J., eds.), pp. 130–156, Pitman, London.

Bigland-Ritchie, B. & Woods, J.J. (1984). Changes in muscle contractile properties and neural control during human muscular fatigue. Muscle Nerve 7, 691–699.

Bigland-Ritchie, B., Jones, D.A., Hosking, G.P., & Edwards, R.H.T. (1978). Central and peripheral fatigue in sustained maximum voluntary contractions of human quardriceps muscle. Clin. Sci. Mol. Med. 54, 609–614.

Bigland-Ritchie, B., Bellemare, F., & Woods, J.J. (1983a). Central and peripheral fatigue in intermittent submaximal contractions. Neurosci. Abst. 9, 631 (Abstract).

Bigland-Ritchie, B., Johansson, R., Lippold, O.C.J., & Woods, J.J. (1983b). Contractile speed and EMG changes during fatigue of sustained maximal voluntary contractions. J. Neurophys. 50, 313–324.

Bock, P.E. & Frieden, C. (1976). Phosphofructokinase: I. Mechanism of pH-dependent inactivation and reactivation of the rabbit muscle enzyme. II. Role of ligands in pH-dependent structural changes of the rabbit muscle enzyme. J. Biol. Chem. 251, 5630–5643.

Boobis, L., Williams, C., & Wootton, S.A. (1982). Human muscle metabolism during brief maximal exercise. J. Physiol. 338, 21–22 (Abstract).

Costill, D.L., Sparks, K., Gregor, R., & Turner, C. (1971). Muscle glycogen utilization during exhaustive running. J. Appl. Physiol. 31, 353–356.

Costill, D.L., Gollnick, P.D., Jansson, E.D., Saltin, B., & Stein, E.M. (1973). Glycogen depletion pattern in human muscle fibers during distance running. Acta Physiol. Scand. 89, 374–383.

Costill, D.L., Coyle, E., Dalsky, G., Evans, W., Fink, W., & Hoopes, D. (1977). Effects of elevated plasma FFA and insulin on muscle glycogen usage during exercise. J. Appl. Physiol. 43, 695–699.

Coyle, E.F., Hagberg, J.M., Hurley, B.F., Martin, W.H., Ehsani, E.E., & Holloszy, J.O. (1983). Carbohydrate feeding during prolonged exercise can delay fatigue. J. Appl. Physiol. 55, 230–235.

Coyle, E.F., Coggan, A.R., Hemmert, M.K., & Ivy, J.L. (1986). Muscle glycogen utilization during prolonged strenuous exercise when fed carbohydrates. J. Appl. Physiol. 61, 165–172.

Dawson, M.J., Gadian, D.G., & Wilkie, D.R. (1978). Muscular fatigue investigated by phosphorus nuclear magnetic resonance. Nature 274, 861–866.

Dawson, M.J., Gadian, D.G., & Wilkie, D.R. (1980). Mechanical relaxation rate and metabolism studied in fatiguing muscle by phosphorus nuclear magnetic resonance. J. Physiol. 299, 465–484.

Dobson, G.P., Yamamoto, E., & Hochachka, P.W. (1986). Phosphofructokinase control in muscle: Nature and reversal of pH-dependent ATP inhibition. Am. J. Physiol. 250, R71–R76.

Edwards, R.H.T. (1981). Human muscle function and fatigue. In: Human Muscle Fatigue: Physiological Mechanisms (Porter, R. & Whelan, J., eds.), pp. 1–18, Pitman, London.

Edwards, R.H.T., Hill, D.K., & Jones, D.A. (1975). Metabolic changes associated with the slowing of relaxation in fatigued mouse muscle. J. Physiol. 251, 287–301.

Fabiato, A. & Fabiato, F. (1978). Effects of pH on the myofilaments and the sarcoplasmic reticulum of skinned cells from cardiac and skeletal muscle. J. Physiol. 276, 233–255.

Faulkner, J.A., Claflin, D.R., & McCully, K.K. (1986). Power output of fast and slow fibers from human skeletal muscles. In: Human Muscle Power (Jones, N. L., McCartney, N., & McComas, A.J., eds.), pp. 81–94, Human Kinetics, Champaign, IL.

Fletcher, W.M. & Hopkins, F.G. (1907). Lactic acid in amphibian muscle. J. Physiol. 35, 247–303.

Godt, R.E. (1981). A simple electrostatic model can explain the effect of pH upon the force-pCa relation of skinned muscle fibers. Biophys. J. 35, 385–392.

Godt, R.E. & Nosek, T.M. (1989). Changes of intracellular milieu with fatigue of hypoxia depress contraction of skinned rabbit skeletal and cardiac muscle. J. Physiol. 412, 155–180.

Greenhaff, P.L., Bodin, K., Harris, R.C., Hultman, E., Jones, D.A., McIntyre, D.B., Söderlund, K., & Turner, D.L. (1993a). The influence of oral creatine supplementation on muscle PCr resynthesis following muscle contraction in man. J. Physiol. 467, 75 (Abstract).

Greenhaff, P.L., Casey, A., Short, A.H., Harris, R.C., Söderlund, K., & Hultman, E. (1993b). Influence of oral creatine supplementation on muscle torque during repeated bouts of maximal voluntary exercise in man. Clin. Sci. 84, 565–571.

Harris, R.C., Sahlin, K., & Hultman, E. (1977). Phosphagen and lactate contents of m. quadriceps femoris of man after exercise. J. Appl. Physiol. 43, 852–857.

Harris, R.C., Söderlund, K., & Hultman, E. (1992). Elevation of creatine in resting and exercised muscle of normal subjects by creatine supplementation. Clin. Sci. 83, 367–374.

Hasson, S.M., Daniels, J.C., Divine, J.C., Niebuhr, B.R., Richmond, S., Stein, P.G., & Williams, J.H. (1993). Effect of ibuprofen use on muscle soreness, damage, and performance: A preliminary investigation. Med. Sci. Sports Exerc. 25, 9–17.

Hermansen, L., Hultman, E., & Saltin, E. (1967). Muscle glycogen during prolonged severe exercise. Acta Physiol. Scand. 71, 129–139.

Hultman, E. & Bergstrom, J. (1967). Muscle glycogen synthesis in relation to diet studied in normal subjects. Acta Med. Scand. 182, 109–117.

Hultman, E. & Nilsson, L.H. (1971). Liver glycogen in man. Effect of different diet and muscular exercise. In: Muscle Metabolism During Exercise (Pernow, B. & Saltin, B., eds.), pp. 143–151, Plenum Press, New York.

Hultman, E. & Sjöholm, H. (1983a). Substrate availability. In: Biochemistry of Exercise V, Int. Series Sport Sci. (Vol. 13) (Knuttgen, H.G., Vogel, J.A., & Poortmans, J., eds.), pp. 63–75, Human Kinetics, Champaign, IL.

Hultman, E. & Sjöholm, H. (1983b). Electromyogram, force and relaxation time during and after continuous electrical stimulation of human skeletal muscle in situ. J. Physiol. 339, 33–40.

Hultman, E. & Sjöholm, H. (1986). Biochemical causes of fatigue. In: Human Muscle Power (Jones, N.L., McCartney, N., & McComas, A.J., eds.), pp. 215–238, Human Kinetics, Champaign, IL.

Hultman, E. & Spriet, L.L. (1986). Skeletal muscle metabolism, contraction force and glycogen utilization during prolonged electrical stimulation in humans. J. Physiol. 374, 493–501.

Hultman, E., Bergström, J., & McLennan-Andersson, N. (1967). Breakdown and resynthesis of phosphorylcreatine and adenosine triphosphate in connection with muscular work. Scand. J. Clin. Lab. Invest. 16, 56–66.

Hultman, E., Bergström, M., Spriet, L.L., & Söderlund, K. (1990). Energy metabolism and fatigue. In: Biochemistry of Exercise VII, Int. Series Sport Sci. (Vol. 21) (Taylor, A.W., Gollnick, P.D., Green, H.J., Ianuzzo, C.D., Noble, E.G., Mètevier, G., & Sutton, J.R., eds.), pp. 73–92, Human Kinetics, Champaign, IL.

Hultman, E., Greenhaff, P.L., Ren, J.M., & Söderlund, K. (1991). Energy metabolism and fatigue during intense muscle contraction. Biochem. Soc. Transact. 19, 347–353.

Juel, C. (1988). Intracellular pH recovery and lactate efflux in mouse soleus muscles stimulated in vitro: The involvement of sodium/proton exchange and a lactate carrier. Acta Physiol. Scand. 132, 363–371.

Karlsson, J. & Saltin, B. (1970). Lactate, ATP, and CP in working muscles during exhaustive exercise in man. J. Appl. Physiol. 29, 598–602.

Karlsson, J. & Saltin, B. (1971). Diet, muscle glycogen, and endurance performance. J. Appl. Physiol. 31, 203–206.

Katz, A., Broberg, S., Sahlin, K., & Wahren, J. (1986a). Leg glucose uptake during maximal dynamic exercise in humans. Am. J. Physiol. 251, E65–E70.

Katz, A., Sahlin, K., & Henriksson, J. (1986b). Muscle ATP turnover rate during isometric contraction in humans. J. Appl. Physiol. 60, 1839–1842.

Lamb, G.D. & Stephenson, D.G. (1991). Effect of Mg^{2+} on the control of Ca^{2+} release in skeletal muscle fibers of the toad. J. Physiol. 434, 507–528.

Länngergren, J. & Westerblad, H. (1986). Force and membrane potential during and after fatiguing, continuous high-frequency stimulation of single *Xenopus* muscle fibers. Acta Physiol. Scand. 128, 359–368.

Lee, J.A., Westerblad, H., & Allen, D.G. (1991). Changes in tetanic and resting $[Ca^{2+}]_i$ during fatigue and recovery of single muscle fibers from *Xenopus laevis*. J. Physiol. 433, 307–326.

Lehman, C. F. (1850). Lehrbuch der physiologischen Chemische 2nd ed., Cavendish Society, London.

Le Rumeur, E., Le Moyec, L., Chagneau, F., Levasseur, M., Toulouse, P., Le Bars, R., & De Certaines, J. (1989). Phosphocreatine and pH recovery without restoration of mechanical function during prolonged activity of rat gastrocnemius muscle: An in vivo ^{31}P NMR study. Arch. Int. Physiol. Biochem. 97, 381–388.

Lieber, R.L. & Friden, J. (1993). Muscle damage is not a function of muscle force but active strain. J. Appl. Physiol. 74, 520–526.

Lindinger, M.I. & Heigenhauser, G.J.F. (1988). Ion fluxes during tetanic stimulation in isolated perfused rat hindlimb. Am. J. Physiol. 254, R117–R126.

Lindinger, M.I. & Sjögaard, G. (1991). Potassium regulation during exercise and recovery. Sports Med. 11, 382–401.

Lindinger, M.I., Spriet, L.L., Hultman, E., Putnam, C.T., McKelvie, R.S., Lands, L.C., Jones, N.L., & Heigenhauser, G.J.F. (1993). Plasma volume and ion regulation during exercise after low- and high-carbohydrate diets. Am. J. Physiol. (in press).

McCartney, N., Spriet, L.L., Heigenhauser, G.J.F., Kowalchuk, J.M., Sutton, J.R., & Jones, N.L. (1986). Muscle power and metabolism in maximal intermittent exercise. J. Appl. Physiol. 60, 1164–1169.

McGilvery, R.W. (1975). The use of fuels for muscular work. In: Metabolic Adaptations to Prolonged Physical Exercise (Howald, H. & Poortmans, J.R., eds.), pp. 12–30, Birkhauser Verlag, Basel.

Medbo, J.I., Mohn, A.-C., Tabata, I., Bahr, R., Vaage, O., & Sejersted, O.M. (1988). Anaerobic capacity determined by maximal accumulated O_2 deficit. J. Appl. Physiol. 64, 50–60.

Meissner, G., Darling, E., & Eveleth, J. (1986). Kinetics of rapid Ca^{2+} release by sarcoplasmic reticulum. Effects of Ca^{2+}, Mg^{2+}, and adenine nucleotides. Biochem. 25, 236–244.

Merton, P.A. (1954). Voluntary strength and fatigue. J. Physiol. 128, 553–564.

Merton, P.A., Hill, D.K., & Morton, H.B. (1981). Indirect and direct stimulation of fatigued human muscle. In: Human Muscle Fatigue: Physiological Mechanisms (Porter, R. & Whelan, J., eds.), pp. 120–129, Pitman, London.

Nanninga, L.B. & Mommaerts, C.R. (1960). Studies on the formation of an enzyme substrate complex between myosin and adenosine-triphosphate. Proc. Natl. Acad. Sci. USA 46, 1155–1166.

Nassar-Gentina, V., Passonneau, J.V., & Rapoport, S.I. (1981). Fatigue and metabolism of frog muscle fibers during stimulation and in response to caffeine. Am. J. Physiol. 241, C160–C166.

Nassar-Gentina, V., Passonneau, J.V., Vergara, J.L., & Ropoport, S.J. (1978). Metabolic correlates of fatigue and of recovery from fatigue in single frog muscle fibers. J. Gen. Physiol. 72, 593–606.

Nilsson, L.H. & Hultman, E. (1973). Liver glycogen in man. The effect of total starvation or a carbohydrate-poor diet followed by carbohydrate refeeding. Scand. J. Clin. Lab. Invest. 32, 325–330.

Nosek, T.M., Fender, K.Y., & Godt, R.E. (1987). It is diprotonated inorganic phosphate that depresses force in skinned skeletal muscle fibers. Science 236, 191–193.

Powers, S.K., Lawler, J.A., Dempsey, S., Dodd, S., & Landry, G. (1989). Effects of incomplete pulmonary gas exchange on VO_2 max. J. Appl. Physiol. 66, 2491–2495.

Robertson, S.P. & Kerrick, W.G.L. (1979). The effects of pH on Ca^{2+}-activated force in frog skeletal muscle fibers. Pfluegers Arch. 380, 41–45.

Rousseau, E., LaDine, J., Liu, Q.-Y., & Meissner, G. (1988). Activation of the Ca^{2+} release channel of skeletal muscle sarcoplasmic reticulum by caffeine and related compounds. Arch. Biochem. Biophys. 267, 75–86.

Sahlin, K., Alvestrand, A., Brandt, R., & Hultman, E. (1978). Intracellular pH and bicarbonate concentration in human muscle during recovery from exercise. J. Appl. Physiol. 45, 474–480.

Sahlin, K., Edström, L., Sjöholm, H., & Hultman, E. (1981). Effects of lactic acid accumulation and ATP decrease on muscle tension and relaxation. Am. J. Physiol. 240, C121–C126.

Saltin, B. (1964). Aerobic and anaerobic work capacity after dehydration. J. Appl. Physiol. 19, 1114–1118.

Saltin, B. & Karlsson, J. (1971). Muscle glycogen utilization during work of different intensities. In: Muscle Metabolism During Exercise (Pernow, B. & Saltin, B., eds.), pp. 289–299, Plenum Press, New York.

Saltin, B. & Costill, D.L. (1988). Fluid and electrolyte balance during prolonged exercise. In: Exercise, Nutrition, and Energy Metabolism (Horton, E.S. & Terjung, R.L., ed), pp. 150–158, MacMillan. New York.

Schneider, M.F. & Chandler, W.K. (1973). Voltage dependent charge movement in skeletal muscle: A possible step in excitation-contraction coupling. Nature 242, 244–246.

Scott, C.B., Roby, F.B., Lohman, T.G., & Bunt, J.C. (1991). The maximally accumulated oxygen deficit as an indicator of anaerobic capacity. Med. Sci. Sports Exerc. 23, 618–624.

Sherman, W.M., Costill, D.L., Fink, W.J., & Miller, J.M. (1981). Effect of exercise-diet manipulation on muscle glycogen and its subsequent ultilization during performance. Int. J. Sports Med. 2, 114–118.

Simonson, E. (1971). In: Physiology of Work Capacity and Fatigue (Simonson, E., ed.), pp. xi–xviii, Charles C. Thomas, Springfield, IL.

Sjögaard, G., Adams, R.P., & Saltin, B. (1985). Water and ion shifts in skeletal muscle of humans with intense dynamic knee extension. Am. J. Physiol. 248, R190–R196.

Smith, J.S., Coronado, R., & Meissner, G. (1986). Single channel measurements of the calcium release channel from skeletal muscle sarcoplasmic reticulum. J. Gen. Physiol. 88, 573–588.

Spriet, L.L., Gledhill, N., Froese, A.B., & Meyers, E.C. (1986). Effect of graded erythrocythemia on cardiovascular and metabolic responses to exercise. J. Appl. Physiol. 61, 1942–1948.

Spriet, L.L., Soderlund, K., Bergstrom, M., & Hultman, E. (1987a). Anaerobic energy release in skeletal muscle during electrical stimulation in men. J. Appl. Physiol. 62, 611–615.

Spriet, L.L., Soderlund, K., Bergstrom, M., & Hultman, E. (1987b). Skeletal muscle glycogenolysis, glycolysis, and pH during electrical stimulation in men. J. Appl. Physiol. 62, 616–621.

Storey, K.B. & Hochachka, P.W. (1974). Activation of muscle glycolysis: A role for creatine phosphate in phosphofructokinase regulation. FEBS Lett. 46, 337–339.

Tanabe, T., Takeshima, H., Mikami, A., Flockerzi, V., Takahashi, H., Kangawa, K., Kojima, M., Matsuo, H., Hirose, T., & Numa, S. (1987). Primary structure of the receptor for calcium channel blockers from skeletal muscle. Nature 328, 313–318.

Thomson, J.M., Stone, J.A., Ginsburg, A.D., & Hamilton, P. (1982). O_2 transport during exercise following blood reinfusion. J. Appl. Physiol. 50, 864–868.

Van Der Laarse, W.J., Lännergren, J., & Diegenbach, P.C. (1991). Resistance to fatigue of single muscle fibers from *Xenopus* related to succinate dehydrogenase and myofibrillar ATPase activities. Exp. Physiol. 76, 589–596.

Vergara, J., Tsien, R.Y., & Delay, M. (1985). Inositol 1, 4, 5-trisphosphate: Possible chemical link in excitation-contraction coupling in muscle. Proc. Natl. Acad. Sci. USA 82, 6352–6356.

Westerblad, H. & Lännergren, J. (1988). The relation between force and intracellular pH in fatigued, single *Xenopus* muscle fibers. Acta Physiol. Scand. 133, 83–89.

Westerblad, H. & Allen, D.G. (1991). Changes of myoplasmic calcium during fatigue in single mouse muscle fibers. J. Gen. Physiol. 98, 615–635.

Westerblad, H., Lee, J.A., Lamb, A.G., Bolsover, S.R., & Allen, D.G. (1990). Spatial gradients of intracellular calcium in skeletal muscle during fatigue. Pfluegers Arch. 415, 734–740.

Westerblad, H., Lee, J.A., Lännergren, J., & Allen, D.G. (1991). Cellular mechanisms of fatigue in skeletal muscle. Am. J. Physiol. 261, C195–C209.

Wiles, C.M. & Edwards, R.H.T. (1982). The effect of temperature, ischemia and contractile activity on the relaxation rate of human muscle. Clin. Physiol. 2, 485–497.

RECOMMENDED READINGS

Bergström, J., Hermansen, L., Hultman, E., & Saltin, B. (1967). Diet, muscle glycogen and physical performance. Acta Physiol. Scand. 71, 140–150.

Bigland-Ritchie, B. (1981). EMG and fatigue of human voluntary and stimulated contractions. In: Human Muscle Fatigue: Physiological Mechanisms (Porter, R. & Whelan, J., eds.), pp. 130–156, Pitman, London.

Coyle, E.F., Hagberg, J.M., Hurley, B.F., Martin, W.H., Ehsani, E.E., & Holloszy, J.O. (1983). Carbohydrate feeding during prolonged exercise can delay fatigue. J. Appl. Physiol. 55, 230–235.

Edwards, R.H.T. (1981). Human muscle function and fatigue. In: Human Muscle Fatigue: Physiological Mechanisms (Porter, R. & Whelan, J., eds.), pp. 1–18, Pitman, London.

Greenhaff, P.L., Casey, A., Short, A.H., Harris, R.C., Söderlund, K., & Hultman, E. (1993b). Influence of oral creatine supplementation on muscle torque during repeated bouts of maximal voluntary exercise in man. Clin. Sci. 84, 565–571.

Hultman, E. & Nilsson, L.H. (1971). Liver glycogen in man. Effect of different diet and muscular exercise. In: Muscle Metabolism During Exercise (Pernow, B. & Saltin, B., eds.), pp. 143–151, Plenum Press, New York.

Hultman, E., Bergström, M., Spriet, L.L., & Söderlund, K. (1990). Energy metabolism and fatigue. In: Biochemistry of Exercise VII, Int. Series Sport Sci. (Vol. 21) (Taylor, A.W., Gollnick, P.D., Green, H.J., Ianuzzo, C.D., Noble, E.G., Mètevier, G., & Sutton, J.R., eds.), pp. 73–92, Human Kinetics, Champaign, IL.

Lännergren, J. & Westerblad, H. (1986). Force and membrane potential during and after fatiguing, continuous high-frequency stimulation of single *Xenopus* muscle fibers. Acta Physiol. Scand. 128, 359–368.

McCartney, N., Spriet, L.L., Heigenhauser, G.J.F., Kowalchuk, J.M., Sutton, J.R., & Jones, N.L. (1986). Muscle power and metabolism in maximal intermittent exercise. J. Appl. Physiol. 60, 1164–1169.

Saltin, B. & Karlsson, J. (1971). Muscle glycogen utilization during work of different intensities. In: Muscle Metabolism During Exercise (Pernow, B. & Saltin, B., eds.), pp. 289–299, Plenum Press, New York.

Saltin, B. & Costill, D.L. (1988). Fluid and electrolyte balance during prolonged exercise. In: Exercise, Nutrition, and Energy Metabolism (Horton, E.S. & Terjung, R.L., eds.), pp. 150–158, MacMillan, New York.

Westerblad, H., Lee, J.A., Lännergren, J., & Allen, D.G. (1991). Cellular mechanisms of fatigue in skeletal muscle. Am. J. Physiol. 261, C195–C209.

Chapter 9

Skeletal Muscle Disorders

MARGARET A. JOHNSON, KATE M.D. BUSHBY, LOUISE V.B. ANDERSON, and JOHN B. HARRIS

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 281–354.

ISBN: 1-55938-808-0

SKELETAL MUSCLE DISORDERS: AN INTRODUCTION

Skeletal muscle is typically considered a remarkably stable tissue, capable of responding to disease, injury, or assault in a strictly limited way. It is also commonly thought that disorders of skeletal muscle are, with a few notable exceptions, benign, giving rise to minor discomfort or disability and self-limiting because of the highly efficient regenerative capacity of the tissue. Most of these views possess an element of truth but conceal the fact that large numbers of patients suffering quite severe muscle dysfunction adopt a lifestyle that minimizes the observable effects of the disorder; the dysfunction may, as a result, remain undetected for many years.

Diseases affecting skeletal muscle are not always primary diseases of muscle, and it is necessary first to determine whether a particular disorder is a primary disease of muscle, is neurogenic in origin, is an inflammatory disorder, or results from vascular insufficiency. A primary disease of muscle is one in which the skeletal muscle fibers are the primary target of the disease. Neurogenic disorders are those in which weakness, atrophy, or abnormal activity arises as a result of pathological processes in the peripheral or central nervous system. Inflammatory disorders may result in T-cell mediated muscle damage and are often associated with viral infections. Vascular insufficiency as a result of occlusion in any part of the muscle vasculature can cause severe disorders of muscle, especially in terms of pain, metabolic insufficiency, and weakness.

Primary myopathies fall into a number of discrete groups: the inherited diseases of muscle, the metabolic myopathies, the neurogenic disorders, and the acquired disorders of muscle.

The inherited diseases of muscle in adults are highly variable. They may be X-linked, autosomal dominant, or autosomal recessive. They may result from germline mosaicism, from a genetically determined predisposition, or from an abnormality in mitochondrial DNA. As a result, these diseases are also variable in age of onset, in the severity of expression of disease, and in the management of the disease.

The congenital diseases are also inherited, but they are characteristically expressed at birth or in very early infancy. Victims of congenital muscle disease are often hypotonic and weak at birth. The severity of the congenital diseases of muscle is also highly variable; the diseases ranging from the slowly progressive, compatible with an essentially normal lifestyle, to the rapidly progressive and fatal in very early life.

The metabolic myopathies are exceptionally complex. Mitochondrial disorders are usually multisystem disorders, in which metabolic dysfunction affects muscle, liver, CNS, and special senses (especially vision) in almost any combination. There is evidence that some forms of mitochondrial disease are inherited, and the preponderance of maternal rather than paternal inheritance is consistent with an abnormality in the mitochondrial genome because almost all (and perhaps all) mitochondria are derived from the ovum.

The simplest of the neurogenic disorders of muscle to consider is muscle fiber atrophy that results from damage to or disease affecting the motor neuron (either cell body, axon, or both). The first response to such damage is inactivity, followed by atrophy. Physiological abnormalities such as fibrillation rapidly appear. If the damage to the motor neuron is rapidly repaired, muscle function may be equally rapidly restored. Longer term damage, involving randomly affected motor neurons results in a much more complex pattern of histological and physiological change.

Acquired disease of muscle is more common than is generally appreciated. It may result from the use of drugs—prescription or nonprescription—that have a recognized capacity to compromise the structure or function of skeletal muscle. Drugs particularly well recognized as myotoxic include clofibrate and its derivatives, anabolic steroids, penicillamine, and emetine. Many nonprescription drugs, including alcohol and laxatives, are directly or indirectly myotoxic. Other forms of acquired myopathies include the acute myopathic conditions caused by the bites of many snakes.

Muscle dysfunction may also present as part of multisystem disease, often involving vascular insufficiency or autoimmune disease. By far the most important of the multisystem and autoimmune diseases are seen as subclasses of the inflammatory disorders.

To cover these various disorders in an orderly and comprehensive manner, the following sections are devoted, respectively, to the muscular dystrophies; the congenital myopathies; the metabolic myopathies; the myotonias, periodic paralyses, and malignant hyperpyrexia; the neurogenic disorders; the inflammatory muscle disorders; the endocrine myopathies; and the drug-induced and toxic myopathies.

Each section comprises a description of the clinical aspects of the diseases being discussed, and then follow brief discussions of genetics, histopathology, biochemistry, and pathophysiology. The chapter concludes with a brief section on the management of muscle disease. We conclude by offering our personal views on future prospects for improving the accuracy of diagnosis and prognosis, the management of patients, and the search for effective therapeutic strategies.

THE MUSCULAR DYSTROPHIES

"Muscular dystrophy" is a term applied to a group of inherited muscle diseases which vary in their clinical presentation, prognosis, and inheritance patterns. Precise guidance about prognosis and genetic implications relies upon a correct diagnosis; careful clinical examination to establish the pattern of muscle involvement is a vital part of the diagnostic procedure. The value of investigations in making the diagnosis varies from findings generally supportive of myopathy but not diagnostic of a particular type of muscular dystrophy, to very specific tests based upon a knowledge of the molecular biology of the different conditions. These tests are likely to be available for an increasing number of muscular dystrophies as the molecular bases of the latter are elucidated. In the following section, each of the major forms of muscular dystrophy is considered in turn, with special attention given to the pattern of disease, mode of inheritance, and the availability of specific genetic and diagnostic tests.

The Xp21-Linked Muscular Dystrophies

This group of related conditions is now known to be due to defects in the dystrophin gene at position Xp21 on the X chromosome, and varying degrees of dystrophin deficiency in muscle (Hoffman and Kunkel, 1989). The clinically recognized entities in this group are known as Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD), and it is also recognized that some individuals may present with a clinical course intermediate between these two extremes. Because these conditions are caused by a gene on the X chromosome, affected individuals are almost exclusively male. A small number of female carriers of mutations within the dystrophin gene do, however, have clinical problems which may vary in their effects from very mild to as severe as a male with DMD.

The main feature in common between all of the disorders in this group is the very specific pattern of muscle involvement, with early involvement of the iliop-

soas, quadriceps, and gluteal muscles in the lower limb. In the upper limbs, the costal origin of pectoralis major, latissimus dorsi, biceps, triceps, and brachioradialis are involved first. Hypertrophy of other muscle groups is commonly seen, especially the calf muscles. The muscles of mastication and deltoid and occasionally other muscles may also be hypertrophied. In all of the Xp21-linked muscular dystrophies progression and complications are invariable, but the rate at which these develop is the major discriminant between the clinical entities within the group.

Duchenne Muscular Dystrophy (DMD)

Presentation is typically in early childhood. Affected boys often walk late and never run properly. Other developmental milestones may also be delayed. Early physical signs include reduced spring in the step, inability to jump with the feet together, and a characteristic tendency to rise from the floor by "climbing up the legs," known as a Gower's maneuver. Progression of the disease leads to slowing of performance and an increasing tendency to fall. Independent mobility is most commonly lost between the ages of 8 and 12 years, though a small number of cases have even more severe disease progression than this, and some (approximately 10% in most series) have a milder disease course intermediate between that seen in DMD and BMD. In addition to the physical manifestations of the disease, the overall mean IQ for populations of boys with DMD is about one standard deviation below the normal mean. This intellectual impairment is nonprogressive and often particularly involves verbal IQ. Once the boys are wheelchair-bound complications arise in the form of contractures, scoliosis, and respiratory muscle weakness leading to a decrease in forced vital capacity, reduced ability to clear secretions, and increased susceptibility to respiratory infections. Cardiac involvement is almost invariable. Affected boys typically die by their late teens or early 20s from respiratory or cardiac failure, or a combination of the two (Emery, 1993).

Becker Muscular Dystrophy (BMD)

BMD is recognized as a slow-motion version of DMD. Thus the pattern of muscle involvement in the two conditions is almost identical, but the course of the disease in BMD is less severe than DMD in every respect. The mean age of reporting first symptoms is 11 years, with difficulty running and climbing stairs prominent early symptoms, though approximately 30% of BMD patients give a history of some problems with mobility in early childhood. Eighty percent experience calf pains at some stage of their disease, most in the teenage years, and in some patients muscle pain may be the only symptom for many years. By definition, BMD patients maintain independent mobility beyond the age of 16 years. A subgroup of BMD patients have relatively severe disease with confinement to a wheelchair

before the age of 40 years and often a reduced lifespan. The majority of BMD patients, however, have milder progression, with continued ambulation into their 40s or beyond. Lifespan may be normal in this group. Physical and intellectual complications are seen in BMD patients, but with a lower incidence and usually to a less severe degree than in DMD (Bushby and Gardner-Medwin, 1993).

Genetics of Xp21-Linked Muscular Dystrophy

The muscular dystrophies in this group are inherited in an X-linked recessive fashion. The dystrophin gene is on the X chromosome at position Xp21. It is the largest gene yet identified, covering over 2.5 mega bases of genomic DNA with about 79 exons. Most mutations (60–85%) in the gene are submicroscopic deletions, and the type of deletion with respect to the effect it has on the reading frame of the gene has been shown to have a major influence on the severity of the muscle disease. Thus deletions which disrupt the reading frame of the gene are usually associated with low or undetectable levels of dystrophin at the protein level, and tend to result in a severe DMD phenotype. At the other end of the spectrum, patients with BMD most often have deletions which maintain the open reading frame of the gene, resulting in the production of an internally deleted and semifunctional protein produced at relatively high abundance (Figure 1). Patients with intermediate clinical symptoms tend to have a variety of genetic and protein abnormalities, and additional factors are most likely to be responsible for their highly variable progression (Nicholson et al., 1993). Genetic counseling is a vital part of the management of the families of these patients. Given the X-linked recessive pattern of inheritance, women may be carriers of the faulty gene but do not usually manifest symptoms. An affected male may therefore have a number of apparently normal female relatives who are at risk of being carriers. In families with a clearcut X-linked family history, genetic counseling is relatively straightforward. There is however a very high rate of spontaneous mutations in the gene (higher than for any other known gene) and, in apparently sporadic cases, the situation in genetic counseling is further complicated by the high incidence of germline mosaicism. The establishment of the risk of carrier status may be based on a number of indirect tests, including serum creatine kinase estimation. The results of indirect tests such as these can be incorporated with pedigree information into a final estimate of the probability of carrier status. A series of genetic tests are now available which can also help to clarify carrier risks, for example, by using probes within the dystrophin gene to trace the X chromosome at risk within a family. Where a deletion is known in a particular family, specific tests may be able to establish carrier status definitively. Known carriers have a one in four risk of having an affected boy in any pregnancy. Many potential or proven carriers opt for prenatal diagnosis, which can be offered by DNA analysis in appropriate families.

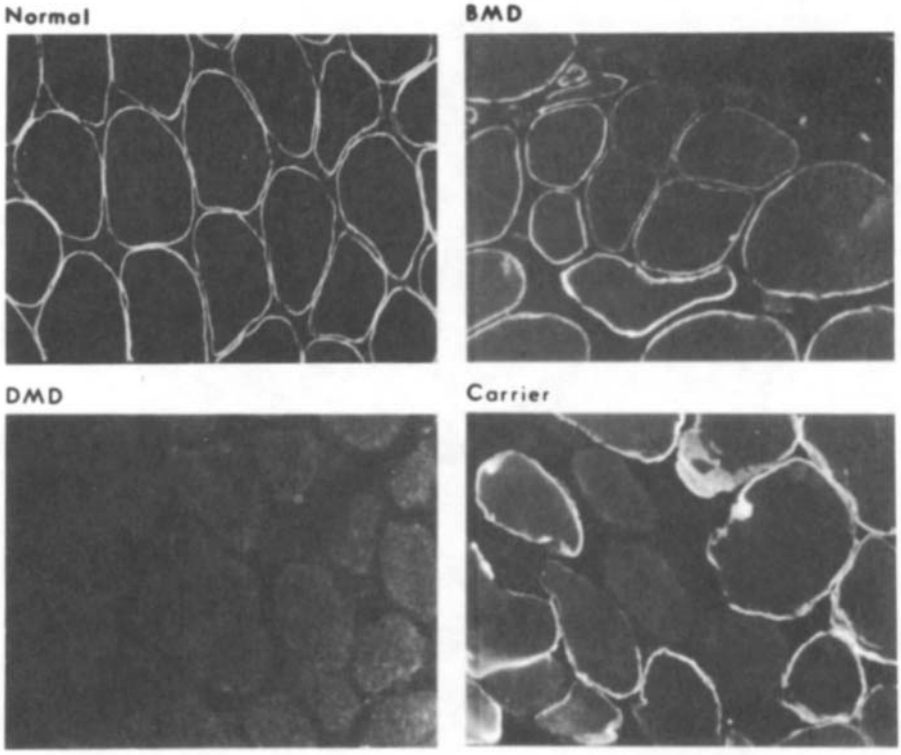


Figure 1. Immunofluorescent labeling of dystrophin in the Xp21 muscular dystrophies. In normal muscle, clear uniform labeling is present at the membrane of each muscle fiber. In Becker muscular dystrophy (BMD), there is inter- and intrafiber variation in labeling intensity. In Duchenne muscular dystrophy (DMD), most fibers are devoid of labeling (note, however, that in most biopsies occasional fibers exhibit weak labeling). In the biopsy from a manifesting carrier, some fibers show normal labeling and others are negative. In the former, the normal X-chromosome is active while in the latter the abnormal X-chromosome is active.

Diagnosis of Xp21-Linked Muscular Dystrophy

Initial suspicion of the diagnosis can be readily substantiated by the finding of a massively elevated serum creatine kinase, which is elevated even in the presymptomatic stages of the disease. Muscle biopsy shows severe dystrophic features, with great variation in fiber size. Fibers often show splitting and central nuclei, and are gradually replaced by proliferation of fibrous tissue and replacement of muscle tissue by fat (Figure 2). The essential findings on biopsy are similar in DMD and BMD, with again the major difference being the severity of the changes seen. A completely specific diagnosis of Xp21-linked muscular dystrophy may be made on

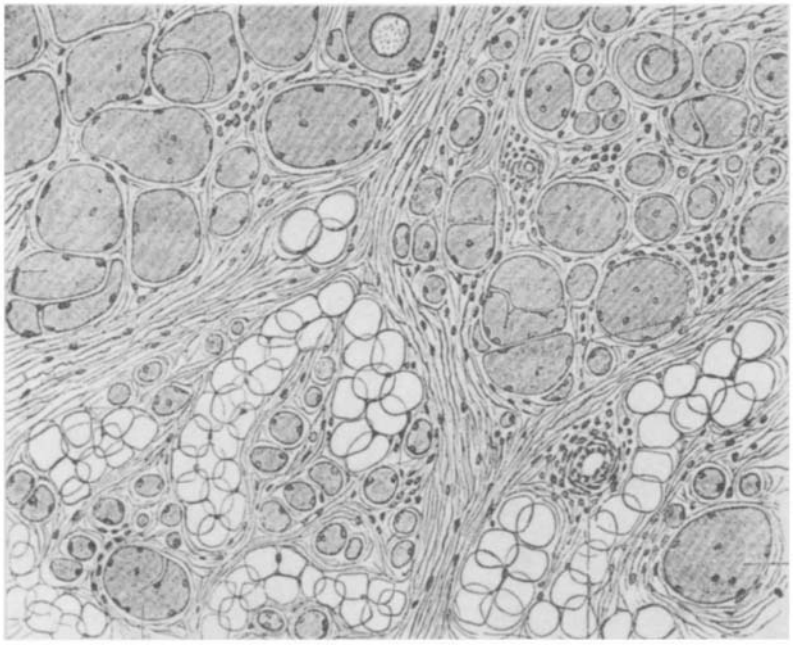


Figure 2. Erb's illustration of the pathology of muscle from patients with Duchenne muscular dystrophy. Note the variation in muscle fiber diameter, fiber-splitting, deposition of fat and infiltration of connective tissue. Drawing from several biopsies produced during final decade of 19th century.

the basis either of the discovery of a dystrophin gene deletion or other mutation, or by the finding of abnormal expression of dystrophin in a muscle biopsy. Immunocytochemical labeling shows an absence or severe reduction of dystrophin positive fibers in DMD, while in BMD most of the muscle section shows some labeling, but there is variation between or within fibers (Figure 1). The abundance of dystrophin detected on Western blotting of a muscle biopsy specimen may be of prognostic as well as diagnostic significance.

Emery-Dreifuss Muscular Dystrophy (EDMD)

EDMD is another X-linked muscular dystrophy, clinically and genetically completely distinct from DMD and BMD. Affected boys usually have onset in childhood of contractures (especially involving the Achilles tendons, elbows, and spinal muscles), humeroperoneal muscle weakness, and cardiac conduction defects, which tend to be mostly a problem in adult life and may necessitate insertion of a pacemaker. The gene for EDMD is known to map to Xq28, but this localization is

only useful in large families, as the faulty gene itself has not yet been isolated (Emery, 1989).

Facioscapulohumeral Muscular Dystrophy (FSH)

FSH is a muscular dystrophy of autosomal dominant inheritance, caused in apparently almost all cases by a defect in a gene localized to chromosome 4q. The clinical pattern is extremely variable between affected individuals, even from the same family. The most characteristic presentation is with facial weakness and weakness and wasting around the shoulder girdle. Winging of the scapulae is seen in most cases. Progression to involve the abdominal muscles, foot muscles, and pelvic girdle muscles is very variable. Approximately one third of patients with FSH remain mildly affected throughout their lives, while about 20% require a wheelchair by the age of 40 years. A DNA rearrangement on chromosome 4q seems to be associated with the disease in some cases, but the nature of the genetic defect in this condition is not yet completely understood. Other autosomally inherited muscular dystrophies may be clinically very similar to FSH (for example, forms of "limb-girdle" muscular dystrophy with scapular involvement but limited facial weakness, and other scapulohumeral syndromes) and may turn out to be related to it at the molecular level (Lunt, 1994).

The "Limb-Girdle" Muscular Dystrophies (LGD)

This group of disorders includes a number of clinically and genetically distinct conditions, the unifying factors among which are (a) muscle weakness predominantly and initially involving the pelvic and/or shoulder girdle musculature, and (b) results of investigations supportive of myopathy, and excluding alternative diagnoses (e.g., Xp21-linked disease, metabolic or mitochondrial disease). A number of fairly disparate groups can now be recognized which fulfill these criteria.

Severe Childhood Autosomal Recessive Muscular Dystrophy of Childhood (SCARMD)

This condition may be clinically very similar to DMD, or a little milder. It can be distinguished from the Xp21-linked muscular dystrophies by the finding of normal dystrophin. Its incidence is highest in the countries of North Africa, though it is also seen elsewhere at a lower frequency. A gene for SCARMD has been localized to chromosome 13. In some patients with this condition, a deficiency of the 50 kDa dystrophin associated glycoprotein has been demonstrated. The relationship of these two findings is not yet entirely clear, and it is likely that SCARMD is itself a heterogeneous group.

Chromosome 15-Linked LGD

LGD linked to chromosome 15 is of recessive inheritance, with the onset of disease at around the age of eight years and confinement to a wheelchair around the age of 30 years. The underlying genetic defect in this group is still unclear.

Chromosome 2-Linked LGD

The disease in this group is also of recessive inheritance, and seems to follow a yet milder course again, with onset of muscle weakness in the late teens and usually slow progression.

Other Autosomal Recessive Forms of LGD

The classification of the different types of LGD has been possible only through the techniques of linkage analysis applied to large enough families. Smaller families and sporadic cases will only be confirmed with certainty as belonging to one of the above groups once direct genetic or protein diagnosis is available. The total number of genes involved in the production of an autosomal recessive LGD phenotype is not yet known.

Autosomal Dominant LGD

While it is probably rarer than recessively inherited disease, families with autosomal dominant LGD have been described, and in these families the disease is usually relatively mild. In one large family with dominant disease and the unusual additional feature of dysarthria, a locus has been identified on chromosome 5. The genes responsible for this condition in other families have not yet been localized.

CONGENITAL MYOPATHIES

The congenital myopathies are a group of genetically-determined disorders in which muscle weakness is typically present from birth or is first manifest in early childhood. A typical congenital myopathy is also often regarded as a condition in which the muscle weakness is only slowly progressive. However, within this group of disorders are to be found conditions which are fatal in infancy and, at the other end of the spectrum, disorders which evolve so slowly so as to be compatible with a normal life-style and to be inapparent throughout childhood and early adult life (Fardeau, 1992; Goebel and Lenard, 1992).

Although clinical examination provides important clues to diagnosis of congenital myopathies, ultrastructural and histochemical examination of muscle biopsies provides the key to definitive identification. Most of the congenital myopathies

were first described in the 1960s when electron microscopy and histochemistry were first used systematically in the investigation of neuromuscular disorders.

Central Core Disease

This disorder takes its name from the abnormal corelike structures which occur in affected fibers (Figure 3). The clinical pattern is less variable in central core disease (CCD) than in other congenital myopathies. Muscle weakness is generally apparent in early childhood but severe hypotonia is rarely seen. Motor milestones may be somewhat delayed. Lower limb weakness predominates and proximal muscles tend to be more affected than distal though foot-drop is seen in some patients. Slight involvement of facial and upper limb-girdle muscles may be seen but extraocular muscles are never affected. Central core disease may be virtually nonprogressive and its benign clinical course seems at variance with the severe structural abnormalities which are apparent on histopathological examination.

Transmission is autosomal dominant, and the relevant gene has been mapped to the same locus on chromosome 19q as that associated with susceptibility to malignant hyperthermia (pages 318–319). It is not established that the conditions are allelic. Almost all patients with CCD whose muscle has been tested *in vitro* for

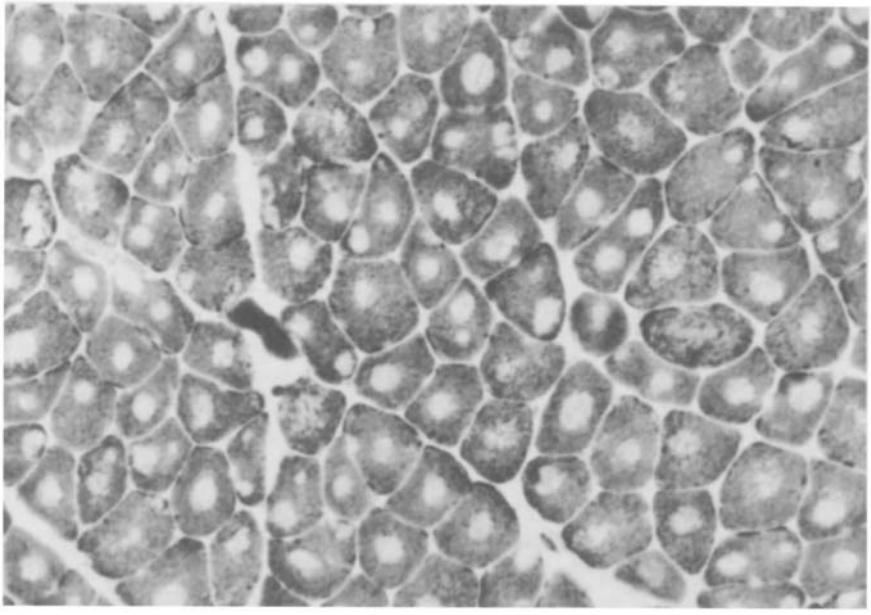


Figure 3. Central core disease: muscle fibers show corelike areas devoid of mitochondrial oxidative enzyme activity.

MH-susceptibility have proved positive and it would seem prudent to regard all CCD patients as MH-susceptible.

Histopathological examination shows the typical corelike lesions in a high proportion of muscle fibers; in older patients this may amount to 100%. Most typically the cores are large and centrally-placed, but multiple cores may occur in the same fiber cross section. Most older patients show a striking predominance of type 1 (slow twitch oxidative) fibers and virtually all fibers with cores are type 1. Sometimes younger family members have more normal proportions of type 1 and type 2 fibers but, again, the cores are confined to the type 1 fibers. It is well established that muscle fiber types can interconvert due to altered physiological demands, and it is likely that fibers with cores convert to a basically slow twitch-oxidative metabolism to compensate for the fact that up to 50% of their cross sectional area may be devoid of mitochondria.

Multi-Core Myopathy (Mini-Core Disease)

This disorder is characterized by multiple small corelike lesions which often extend for only 2–6 sarcomeres, in contrast to the large cores of CCD which can extend for long distances within affected fibers. The disorder is fairly benign, with neonatal hypotonia, delayed motor milestones and a generally slender appearance of most of the musculature. There is a tendency for upper limbs to be more severely involved than lower limbs and facial weakness is common. Ptosis and ophthalmoplegia have been reported. Contractures may occur and tendon reflexes are generally absent or reduced. Not all cases show the usual nonprogressive or slowly-progressive course and cardiomyopathy is sometimes also encountered in less benign cases. Many cases of multicore disease are sporadic but some families have several affected members with a pattern suggestive of autosomal recessive transmission.

Nemaline Myopathy

This disorder takes its name from the nemaline rods (Gk. nema = thread) which occur in large numbers in affected muscle fibers. The name rod myopathy is also used. The condition shows great clinical variability and it has been estimated from published case-studies that about one in five patients die during the first six years of life. In its most severe form the disorder is characterized by profound neonatal hypotonia, and feeding and respiratory difficulties. Death may occur as a result of recurrent pneumonia. In children surviving their first few years, diffuse muscle weakness and general slenderness of muscles is noted. Facial muscle weakness is often apparent and is associated with a high-arched palate and an open mouth. Scoliosis is common. In patients who present in adult life, most show mild diffuse weakness though occasional severe late-onset cases have been recorded. Ptosis may

occur in adult patients whereas extraocular muscles are generally unaffected in juvenile patients.

Genetic transmission in nemaline myopathy is the subject of some uncertainty. A Japanese study of 50 pedigrees came to the conclusion that autosomal dominant with reduced penetrance was the most probable mode. However a Finnish study presented evidence for autosomal recessive transmission. There is no evidence that severe and mild forms are genetically distinct and several pedigrees contain members showing widely differing clinical severity. A candidate gene for autosomal dominant nemaline myopathy has been localized to chromosome 1q 21–23.

Histopathological examination is necessary for confirmation of a clinical diagnosis. The rods often occur in large clusters in severely-affected fibers and these areas show loss of myofibrillar ATPase reactivity. Ultrastructural examination shows Z-line streaming and over-production of Z-line material, which is similar in electron density to the nemaline rods themselves (Figure 4). Occasionally rods are seen in direct continuity with the Z-lines of the adjacent myofibrils. Rods up to 5 μm in length can occur and their width varies from about 0.2 μm to 1.0 μm. Immunocytochemical studies reveal that a major component is α-actinin with some tropomyosin and actin also associated with it. The rods thus have strong similarities with normal Z-line material.

Figure 4. Nemaline myopathy: electron micrograph shows nemaline rods (arrows) lying between disrupted myofibrils.

In severe neonatal nemaline myopathy virtually every muscle fiber shows multiple rods and all muscle fiber types are affected. However in juvenile cases, two different patterns of fiber type involvement are seen. In one there is a clear size difference between type 1 fibers, which are abnormally small (hypotrophic or atrophic) and which contain numerous nemaline rods, and type 2 fibers, which are either of normal diameter or hypertrophic and contain few, if any, nemaline rods. Other patients show a gross predominance of type 1 muscle fibers, again with rods virtually confined to this fiber type. These findings may be explicable in terms of the involvement of isoforms of α-actinin specific to slow and fast muscle fiber types.

Centronuclear Myopathies

As the name suggests, centronuclear myopathies are characterized by the presence of high numbers of internally-placed muscle nuclei, in contrast to the usual subsarcolemmal position. This group of disorders is clinically and genetically heterogeneous. Early-onset and late-onset forms are seen, and one variant is associated with significant type 1 fiber hypotrophy. These three forms are regarded by most investigators as showing an autosomal dominant inheritance with variable expression, though some have argued in favor of an autosomal recessive mode of transmission for both early onset and type 1 hypotrophy-associated forms. Yet another centronuclear myopathy is a severe X-linked condition and has been mapped to Xq 28.

The early-onset centronuclear myopathies were the first to be described in the 1960s. Marked neonatal hypotonia is seen and, although delay in motor development is not severe, a waddling gait is common and there is pronounced weakness of foot flexion. Scapular winging, facial weakness with extraocular muscle involvement and generalized areflexia are typical of the disorder. Progression is fairly gradual in most cases but patients may be severely disabled early in adult life.

Histopathological features are dominated by the large number of centrally-placed muscle nuclei, sometimes affecting more than 90% of muscle fibers. The nuclei form long chains in the middle of the fiber and are surrounded by cytoplasm, which contains mitochondria and membranous vesicles, but no myofibrils. This morphological appearance has prompted comparison with myotubes, and in fact centronuclear myopathies are sometimes referred to as myotubular myopathies. This is a misnomer, however, since although the affected fibers retain some of the structural features of myotubes, and maturational arrest may play a role in their formation, the vast majority of such fibers are fully differentiated histochemically into either type 1 or type 2.

Late-onset centronuclear myopathies show a predominance of limb-girdle and truncal muscle weakness with only rare facial or eye involvement. Although this group is classified with the congenital myopathies, weakness only becomes appar-

ent in adult life, sometimes even in patients past middle age. The histological features are broadly similar to those described for the early-onset variant. Findings common to both variants include the presence of central nuclei in both type 1 and type 2 fibers with equal frequency and a tendency to an overall type 1 fiber predominance.

Centronuclear myopathy with type 1 fiber hypotrophy is sometimes regarded as a separate entity because many cases show central nuclei only in the hypotrophic type 1 fibers, while the type 2 fibers are morphologically normal. Affected type 1 fibers are even more myotubelike than in other variants of the disorder, with the exception of the severe X-linked form, due to the persistence of a mitochondria-rich core within a peripheral ring of myofibrils. These features are clearly demonstrable using histochemical methods for the localization of SDH activity and myofibrillar ATPase, respectively.

The severe X-linked form of centronuclear myopathy is often associated with reduced fetal movement and hydramnios, and may be fatal in the neonatal period due to respiratory failure. Children may survive for several years but often only with assisted ventilation. In only a few reported cases has the condition allowed any form of active life. Female relatives may show a carrier state characterized by the presence of some small myotubelike type 1 fibers in an otherwise normal muscle fiber population.

Congenital Fiber Type Disproportion (CFTD)

The syndrome of congenital fiber type disproportion (CFTD) is generally considered a congenital myopathy because the clinical picture conforms to that of a fairly benign neuromuscular condition with possible familial incidence. However, it differs radically from the conditions so far described in that no morphological abnormalities are detectable in muscle fibers and the only significant biopsy finding is that of an abnormal size difference between type 1 and type 2 fibers. It has frequently been argued that CFTD should be regarded as a defect arising from impaired maturation of the motor unit as a whole, rather than as a true primary myopathy.

Clinical features include neonatal hypotonia, a tendency toward congenital hip dislocation and diffuse muscle weakness. Later on children are frequently of short stature and low body weight and often have long thin faces and high-arched palates. Respiratory difficulties, where present, occur early on and tend to improve with time. In others a virtually static clinical picture is seen.

From the practical viewpoint it is important to be able to distinguish infants and children with this condition from less benign disorders such as the spinal muscular atrophies. Careful histochemical assessment of muscle biopsies with histographic analysis is recommended. Most biopsies from CFTD patients show type 1 fibers which are small in relation to type 2 fibers. A revised definition of CFTD states that

this size difference should exceed 45%. Not all CFTD cases conform to this definition and some show the reverse size difference, i.e., type 2 fibers are significantly smaller than type 1 fibers. In either case, the specificity of the process of atrophy or hypotrophy for one or other of the fiber types allows the condition to be distinguished from SMA, where fibers of all types are affected by atrophy.

METABOLIC MYOPATHIES

Defects in the utilization of carbohydrates and lipids as metabolic fuels are collectively responsible for a large number of myopathic disorders (see Di Mauro et al, 1992). Defects of the mitochondrial respiratory chain also comprise an important group of metabolic disorders, normally referred to as mitochondrial myopathies (Morgan-Hughes, 1992). The myopathies which are associated with disorders of lipid metabolism have an equal claim to be included in this category since the enzymes involved are located in mitochondria. Many metabolic myopathies show histopathological evidence of the underlying abnormality, hence the terms glycogen storage disease and lipid storage disease. However almost as many metabolic myopathies are enzymopathies in which the defective enzymes are responsible for functional impairment but cause little or no morphological damage. Clearly, if a muscle-specific isoenzyme is defective the result will be a pure myopathy, whereas if the affected enzyme is present in all tissues as a single isoform then a multisystem disorder may result. This outcome is not inevitable, however, because deficiencies of an enzyme may be critical in some tissues but not in others. Molecular genetic analysis is beginning to make an important contribution to understanding the complex relationship between biochemical defect and clinical syndrome (see review by Poulton and Land, 1990).

Glycogen Storage Disease

Eleven defects in the metabolism of glycogen have been reported; nine of them affect skeletal muscle directly (see Figure 5), but only glycogenosis type II (acid maltase deficiency) and glycogenesis type V (myophosphorylase deficiency) are reasonably common; the rest are rare and some have been recorded in isolated case studies only.

Glycogenosis type I (glucose-6-phosphatase deficiency) does not affect skeletal muscle directly because the enzyme involved is found in liver, kidney, and small intestine but not muscle itself. However severe hypotonia can occur in affected infants since the defect in liver prevents the release of the glucose formed from glycogenolysis or gluconeogenesis. Fasting hypoglycemia and lactic acidosis are seen, though the severity of these features usually abates with age. The disorder shows autosomal recessive transmission and a hemorrhagic tendency is present in many patients.

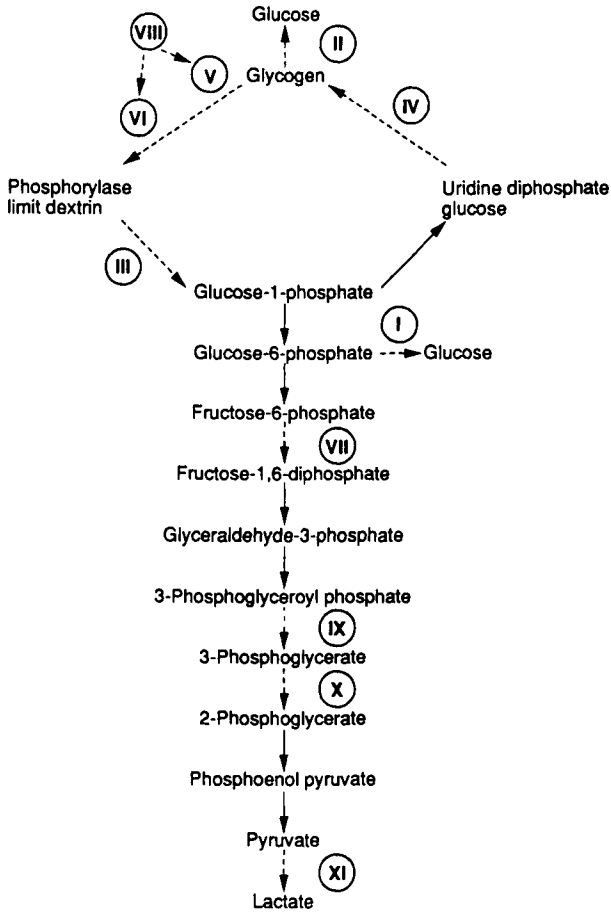


Figure 5. Glycogen metabolism and glycolysis. Dotted lines indicate sites of metabolic defects involving enzymes I–XI.

Glycogenosis type II (acid maltase deficiency) is the only glycogen storage disorder which involves a lysosomal enzyme. Acid maltase (AM) has both α-1,4-glucosidase and α-1,6-glucosidase activities and can therefore hydrolyze both straight and branched-chain moieties of glycogen to form glucose, which can then cross the lysosomal membrane. In acid maltase deficiency, marked glycogen storage is apparent within lysosomes and also in the cytosol (Figure 6). The disorder is responsible for three major clinical syndromes. Pompe's disease is characterized by severe neonatal hypotonia and cardiomegaly with less prominent hepatomegaly. Death occurs within two years, often as a result of pulmonary or cardiac failure. Childhood acid maltase deficiency is distinguishable from Pompe's disease by the

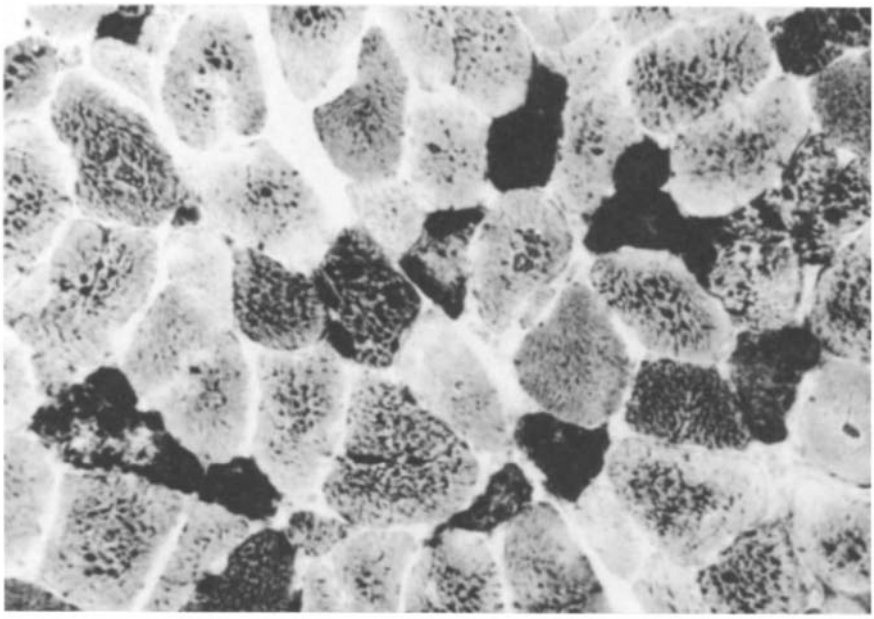


Figure 6. Glycogen storage in acid maltase (AM) deficiency: in this late-onset case not all muscle fibers are affected.

lack of cardiomegaly and has a relatively benign clinical course. Motor milestones may be delayed and there is a tendency for a waddling gait and toe-walking. As in Pompe's disease, respiratory muscles are particularly severely involved and ventilatory insufficiency is a common cause of death, which usually occurs during the second or third decade. Late-onset acid maltase deficiency may present in the fourth or fifth decade, with the mildest forms becoming apparent even later than that. Heart muscle is not affected but respiratory muscle weakness is almost always a prominent feature and is significantly more severe than weakness of limb muscles. Patients are therefore prone to pulmonary infections and pulmonary failure.

As in almost all glycogeneses, the mode of transmission in acid maltase deficiency (AMD) is autosomal recessive. The gene for AM has been assigned to chromosome 17. With only rare exceptions the severity of the disorder is fairly constant within families. In patients with Pompe's disease there is either complete lack of enzyme synthesis, or defective processing of 110 kDa precursor protein leading to absence of the mature 76 kDa enzyme protein, or to its being present but inactive. At the molecular level there is often no detectable AM-specific mRNA in Pompe's disease, while in adult-onset variants the corresponding mRNA is present, though it may be truncated. At present it appears that the major differentiating factor between Pompe's disease and the more benign variants of AMD is the persistence

of residual enzyme activity, albeit in low concentrations, in the latter disorders. Since there are no tissue-specific isoenzymes of AM, glycogen storage would be expected to occur in all tissues in AMD. Low levels of AM activity may be sufficient to prevent this in all tissues, except skeletal and cardiac muscle, in which glycogen metabolism is particularly prominent but which, unlike liver and many other tissues, contain only a poorly developed lysosomal system.

Histopathological examination of muscle biopsies is an important part of the diagnostic process in the investigation of AMD. In Pompe's disease, the glycogen storage is so severe that the muscle fibers are virtually unrecognizable because they are so distorted by vacuolation. Glycogen content can be as high as 10% in some cases. Proliferation of lysosomes gives rise to greatly increased acid phosphatase activity which can also be detected histochemically. The combination of glycogen storage and high lysosomal acid phosphatase allows a diagnosis of AMD to be made at this stage, although confirmation of the enzyme defect should be made by biochemical assay. Prenatal diagnosis can be made by assaying AM activity in cultured amniocytes. Glycogen storage is particularly severe in cardiac muscle and anterior horn cells of the spinal cord, and neurons of brainstem nuclei also show marked glycogen accumulation.

In childhood and adult-onset forms of AM, more moderate glycogen storage and vacuolation of muscle are seen and not all fibers are affected. Although cardiomegaly is not apparent in childhood AMD, glycogen storage is detectable histologically in heart muscle.

Glycogenosis type III (debranching enzyme deficiency) causes a hepatopathy in infancy and childhood but the same enzyme is found in muscle as well as liver, and myopathy may appear in later life after liver disease has largely disappeared. The enzyme is bifunctional; in addition to hydrolyzing glycogen branch points (amylo-1,6-glucosidase activity), it also acts as an oligo-1,4→7 1,4-glucan transferase. In this function, its substrate is phosphorylase-limit-dextrin (PLD) from which it removes a maltotriosyl moiety leaving a glucosyl unit in an α-1,6-glucoside linkage, which is thus susceptible to its debranching activity.

The debranching enzyme gene has been assigned to chromosome 1 and transmission is recessive. There is a tendency for male patients to manifest myopathy to a greater degree than female patients. Muscle weakness is rarely seen before the third decade and is mainly distal in distribution; peripheral neuropathy may be present as a complicating feature. Muscle biopsy shows a severe vacuolar myopathy with associated glycogen accumulation. The accumulation of glycogen is not membrane-bound, in contrast to the situation in acid maltase deficiency where a large component is intralysosomal.

Glycogenosis type IV (branching enzyme deficiency) results in the formation of a variant of glycogen, characterized by abnormally long inner and outer glucosyl chains and fewer branch points than normal. The abnormal variant is stored in sufficient amounts to cause some vacuolation. The clinical manifestations of this

defect are variable. The disorder may be associated with severe liver disease in infancy and childhood with an early fatal outcome. Muscle involvement is a minor feature but another group of patients has been reported in whom liver disease was moderate and cardiomyopathy was the major finding. Affected tissues contain intensely PAS-positive material that unlike normal glycogen is resistant to diastase digestion. Ultrastructural studies show that it consists of finely filamentous and granular material and is found in association with normal glycogen particles. The existence of normal glycogen in these patients would seem to indicate the existence of an alternative mechanism for producing glycogen branch-points.

Glycogenosis type V (myophosphorylase deficiency, McArdle's disease) is a disorder which is generally characterized by exercise intolerance, cramps, and early fatigue. Many patients experience episodes of myoglobinuria after exercise and, if this is severe, it may lead to renal failure. Although the disorder is transmitted as an autosomal recessive trait, there is a preponderance of males among diagnosed cases. Onset in males is typically during the second decade, but there is a tendency for female patients to be diagnosed late, i.e., during third or fourth decade or even later. The difference may reflect different exercise patterns in males and females.

The gene for myophosphorylase has been assigned to chromosome 11q13. The enzyme is a dimer of two identical 97 kDa subunits and is the sole isoform present in skeletal muscle. Heart and brain also contain this isoform in addition to a distinct brain isoenzyme and a hybrid muscle/brain isoform. Smooth muscle also contains a phosphorylase isoform distinct from the muscle isoenzyme. If regenerating muscle fibers are present they also contain phosphorylase activity due to the presence, in fetal and developing muscle, of an isoform said to be identical with brain phosphorylase.

The laboratory diagnosis of McArdle's disease involves assessment of blood lactate levels after ischemic exercise. Normal subjects show a three-to-fivefold increase in lactate after one minute of forearm exercise with arterial blood flow occluded by means of a blood-pressure cuff. In myophosphorylase-deficient muscle, since anaerobic glycolysis is blocked at the level of glycogenolysis, the increase in lactate is either nonexistent or minimal. Histochemical demonstration of the lack of myophosphorylase activity in tissue sections is used as a confirmatory test (Figure 7). Glycogen storage is very variable from patient to patient but rarely exceeds more than about 3% muscle weight. Routine histological examination may show subsarcolemmal vacuolation due to the glycogen storage, and ultrastructural studies also reveal excess glycogen between individual myofibrils. The myophosphorylase protein can be shown, by immunoblotting, to be absent from the muscle of most McArdle's disease patients. However Northern blotting has demonstrated that, whereas myophosphorylase-specific mRNA is missing in a majority of patients, various other situations exist including the production of truncated mRNA or production of normal-length message in decreased amounts.

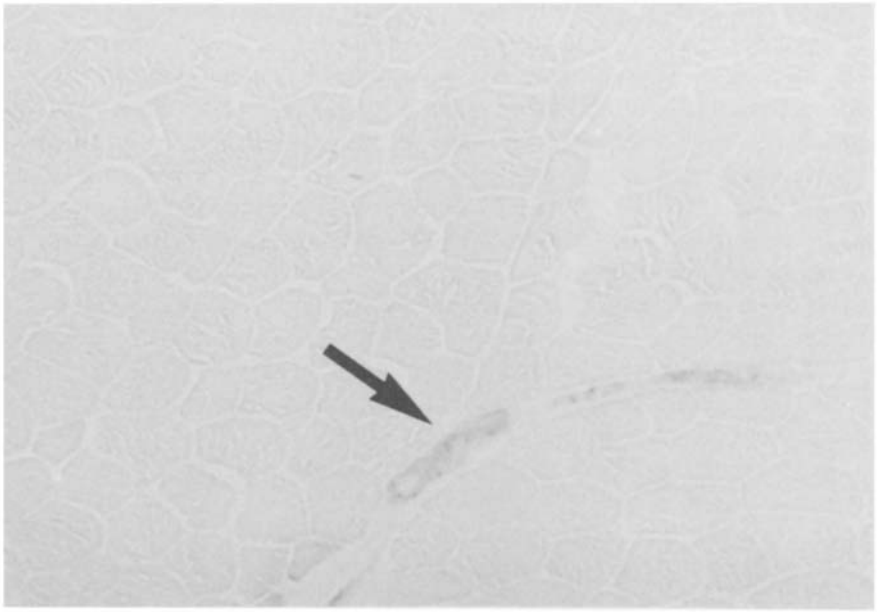


Figure 7. Myophosphorylase deficiency (McArdle's disease): enzyme is absent from muscle fibers but present in smooth muscle cells of blood vessel (arrow).

Glycogenosis type VI (liver myophosphorylase deficiency) gives rise to hepatomegaly and hypoglycemia in childhood. The enzyme involved is under separate genetic control from the muscle isoform and has been assigned to chromosome 14.

Glycogenosis type VII (phosphofructokinase deficiency) causes a very similar pattern of symptoms to that seen in McArdle's disease patients, including adverse reactions to vigorous exercise often relieved by rest. The disorder is rare in the general population and virtually all recorded cases have been of Ashkenazi Jewish descent. Phosphofructokinase (PFK) protein is a tetramer which, in muscle, is present as a homotetramer of M subunits. The gene encoding this M subunit has been assigned to chromosome 1, while those for P (platelet) and L (liver) subunits are found on chromosomes 10 and 21, respectively. Erythrocytes contain both M_4 and L_4 homotetramers as well as the three heterotetramers, M_3L_1, M_2L_2, and M_1L_3. PFK deficiency has been shown, in some patients, to involve a 75 base in-frame deletion in the gene coding for the M subunit. As might be expected, in addition to myopathy, some patients show blood disorders including hemolytic anemia and jaundice as a result of mild hemolysis. Since the M subunit is a major component of brain and heart PFK, some involvement of these tissues might be expected, but cardiomyopathy has not been recorded and encephalopathy has occurred only within an atypical syndrome of severe neonatal myopathy.

Differentiation from McArdle's disease can be made histochemically since there are specific methods for demonstrating the activity both of myophosphorylase and of PFK. Muscle biopsy sections also show accumulation of abnormal glycogen. Electron microscopic examination shows it to be similar to the filamentous and granular deposits described in branching enzyme deficiency (glycogenosis type IV). Investigation of anaerobic glycolysis *in vitro* shows that no increase in lactate occurs when PFK-deficient muscle is incubated with glycogen, glucose-1-phosphate, glucose-6-phosphate, or fructose-6-phosphate as substrates but that normal lactate production occurred with fructose-1,6-diphosphate as substrate (see Figure 5). Because PFK-deficient muscle cannot utilize either glycogen or glucose, it is largely dependent on free fatty acids (FFA) and ketones as alternative metabolic fuels. Whereas a high glucose intake may be helpful in McArdle's disease patients in avoiding adverse effects of exercise, glucose actually has a negative effect in PFK deficiency since it tends to lower levels of FFA and ketones.

Glycogenosis type VIII (phosphorylase b kinase deficiency) gives rise to myopathy and liver disease, either singly or in combination. Phosphorylase b kinase (PBK) converts the inactive b form of both muscle and liver phosphorylases to the active a forms of the enzymes. The ischemic lactate test sometimes shows a flat result as in McArdle's disease, but is more likely to be normal. Histochemical demonstration of myophosphorylase activity in tissue sections shows a near-normal reaction due to the presence of phosphorylase a. Accumulation of glycogen is modest and found mainly in type 2 (fast-twitch glycolytic) muscle fibers.

The molecular genetics of PBK is a complex subject since the enzyme consists of four different subunits α, β, γ, and δ which are present in the configuration $(\alpha\beta\gamma\delta)_4$, a hexadecamer. The genes encoding α and β have been assigned to chromosomes Xq12-q13 and 16q12-q13, respectively, and the gene for the γ subunit to chromosome 7. Of the various clinical syndromes which have been recorded, myopathy occurring alone is usually transmitted as an autosomal recessive trait, as is a combination of liver disease and a static myopathy. However a benign syndrome of liver disease in infancy, characterized by hepatomegaly and fasting hypoglycaemia, in association with growth retardation and delayed motor milestones, appears to be inherited in an X-linked pattern.

Glycogenosis type IX (phosphoglycerate kinase deficiency) may be symptomless or may cause syndromes of exercise intolerance, cramps and myoglobinuria, or of hemolytic anemia with seizures and mental retardation. Phosphoglycerate kinase (PGK) deficiency is inherited as an X-linked trait and the PGK polypeptide is expressed in a wide range of tissues with no tissue-specific isoforms. In patients with myopathy, biopsy findings in skeletal muscle are not indicative of glycogen storage and, on biochemical analysis, glycogen content has proved to be within normal limits. Decreased lactate production *in vitro* with glycogen and its four hexose phosphate metabolites showed that the defect was distal to the PFK reaction (see Figure 5).

Glycogenosis type X (phosphoglycerate mutase deficiency) is another rare disorder of terminal glycolysis and is associated with cramps, myoglobinuria, and exercise intolerance. Phosphoglycerate mutase (PGAM) is a dimer and about 95% of the muscle enzyme is in an MM homodimeric form. A brain (BB) isoform predominates in tissues other than muscle and in patients with PGAM deficiency, the small amount of residual activity in muscle is due to the BB isoform. This suggests that the defect involves the M subunit only. The gene for this subunit has been assigned to chromosome 7.

Glycogenosis type XI (lactate dehydrogenase deficiency) is even more rare than deficiencies of PGK or PGAM and has been recorded in only two individuals presenting with exercise intolerance and myoglobinuria. However the chance identification of the same defect in an asymptomatic woman may mean that lactate dehydrogenase (LDH) deficiencies could be more widespread but clinically silent. In all three patients an ischemic lactate test showed little increase in venous lactate but an abnormally high increase in pyruvate, indicating the site of the metabolic block. LDH is a tetrameter and may consist of all muscle (M) subunits, all heart (H) subunits, or conform to one of the three heterotetrameric configurations. Inheritance is autosomal recessive and the gene for the M subunit which is implicated in the LDH defect has been assigned to chromosome 11.

Disorders of Lipid Metabolism

Alternative names for this class of metabolic defect include lipid storage disorders and lipidoses, but not all lipid defects cause actual accumulation of lipid in affected tissues. As a general rule, defects involving lipid transport give rise to insignificant lipid storage, whereas disorders of lipid utilization are accompanied by moderate to severe lipid accumulation (Figure 8). In skeletal muscle, long-chain fatty acids (LCFAs) are major substrates for oxidation, both at rest and under conditions of glycogen depletion resulting from prolonged exercise. LCFAs cross the muscle plasma membrane from the circulatory system where they are bound to albumin. The precise mechanism of transport into muscle is not known, though cytosolic fatty acid-binding proteins are found in skeletal and cardiac muscle. Within muscle, LCFAs may be esterified and stored as neutral lipid droplets, consisting mainly of triglycerides, which are located between the myofibrils in close proximity to mitochondria.

The steps in the subsequent utilization of muscle LCFAs may be summarized as follows. The free fatty acids, liberated from triglycerides by a neutral triglyceride lipase, are activated to form acyl CoAs by the mediation of LCFAcyl-CoA synthetase which is situated on the outer mitochondrial membrane. The next step involves carnitine palmitoyl transferase I (CPT I, see Figure 9) which is also located on the outer mitochondrial membrane and catalyzes the transfer of LCFAcyl residues from CoA to carnitine (γ-trimethyl-amino-β-hydroxybutyrate). LCFAcyl

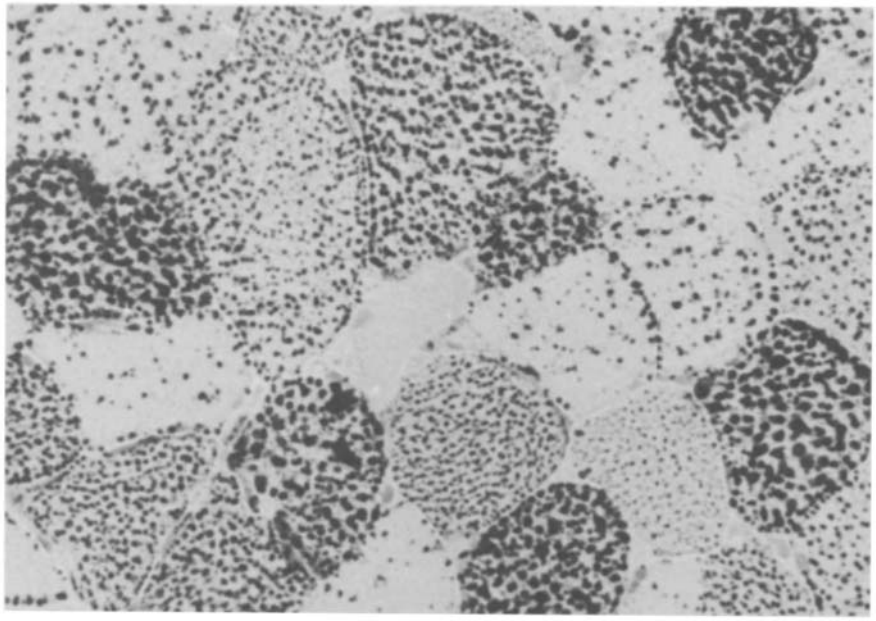


Figure 8. Lipid storage in muscle from a patient with short chain acyl CoA dehydrogenase (SCAD) deficiency.

carnitines are exchanged for free carnitine across the inner mitochondrial membrane by carnitine:acylcarnitine translocase, and another carnitine palmitoyl transferase (CPT II), situated on the inner (matrix) side of the inner mitochondrial membrane, reconverts LCFAcyl carnitines to LCFAcyl-CoAs.

The next sequence of reactions involves β-oxidation, in which fatty acyl-CoA is shortened by two carbon atoms with each repeated sequence of four enzyme reactions: (a) acyl-CoA dehydrogenase, (b) enoyl-CoA hydratase, (c) 3-hydroxyacyl-CoA dehydrogenase, and (d) 3-oxoacyl CoA thiolase. The acyl-CoA dehydrogenases present in the mitochondrial matrix show specificities for long-chain, medium-chain or short-chain fatty acyl-CoAs, although the specificities are not absolute and overlap to some extent. These enzymes (LCAD, MCAD, SCAD) are all homotetramers with each subunit containing FAD. Electrons are transferred from substrate via this FAD moiety to another flavoprotein, electron transfer flavoprotein (ETF), and via ETF dehydrogenase to the respiratory chain. Defects at many of the steps in this complicated series of reactions have been reported and are responsible for a wide range of myopathic syndromes (Harris and Turnbull, 1990).

Carnitine palmitoyl transferase (CPT) deficiencies are commonly associated with myoglobinuria after prolonged exercise; typically patients are young men and

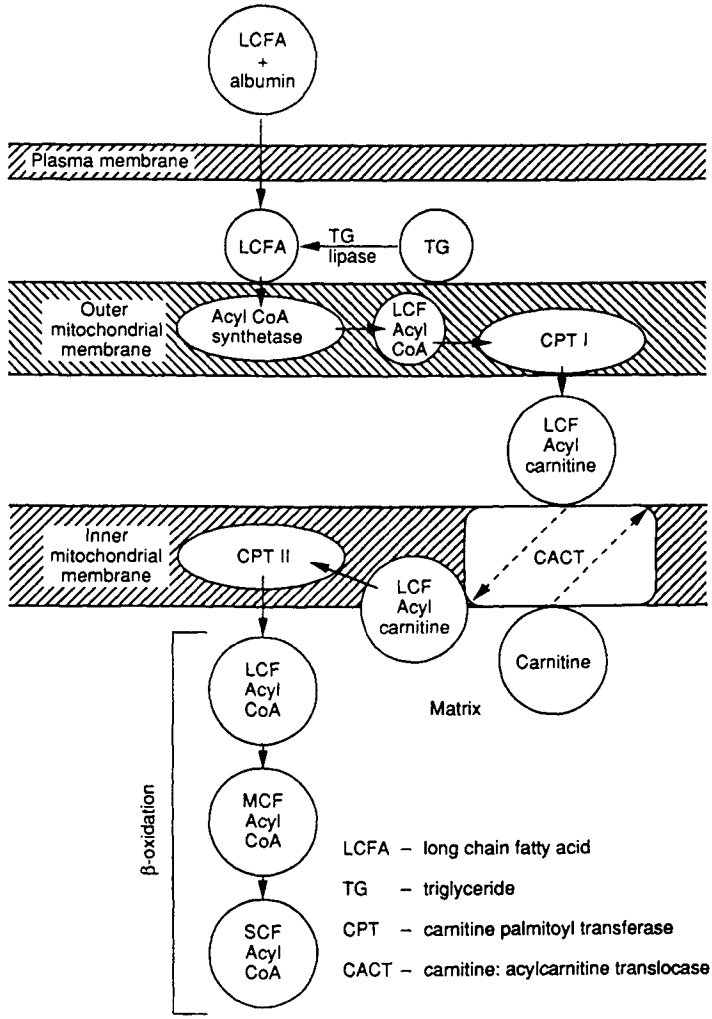


Figure 9. Fatty acid metabolism in skeletal muscle.

myoglobinuria may cause renal failure (Di Mauro and Papadimitriou, 1986). Muscle biopsy, if undertaken between attacks, generally shows little or no excess of lipid droplets. The transmission of the CPT defect is said to be autosomal recessive but there is continuing controversy as to the involvement of CPT I and CPT II in reported cases. According to one school of thought, CPT I is deficient only in a rare hepatic form of CPT deficiency which presents in infancy with hypoketosis, hypoglycemia, and coma precipitated by fasting. In contrast, the myopathic form of CPT deficiency is said to involve CPT II. Evidence leading to

these conclusions was derived from studies of cultured fibroblasts and myoblasts from patients with hepatic or myopathic forms of the disorder. Other studies have challenged the concept of separate CPT I and CPT II defects and have suggested that myopathy may result from defective or altered regulatory properties rather than from loss of catalytic activity.

Long-chain acyl-CoA dehydrogenase (LCAD) deficiency is typical of defects of lipid substrate utilization in that excessive fat storage in affected tissues is a characteristic feature. It is an autosomal recessive disorder and LCAD assays in fibroblasts taken from parents of affected children have shown enzyme levels intermediate between those of normal controls and the homozygous patients. LCAD deficiency primarily affects liver with hypoketotic hypoglycemic coma, but long chain acyl carnitines are also increased in muscle tissue. A small number of patients who survive the infantile syndrome of liver disease subsequently develop chronic myopathy and myoglobinuria under stress, which is similar to the findings in myopathic CPT deficiency.

Medium-chain acyl-CoA dehydrogenase (MCAD) deficiency is the most common of the acyl-CoA dehydrogenase defects and occurs in up to one in 5,000 live births. Presentation is in infancy with hypoketosis, hypoglycemia and encephalopathy, hepatomegaly with fatty degeneration and metabolic crises reminiscent of Reye's syndrome. The defect can be demonstrated in liver, muscle, fibroblasts, and leukocytes. It is likely that MCAD deficiency is genetically heterogeneous. Investigation of MCAD protein in several patients has shown that biosynthesis and processing appears normal and that the protein is present in normal abundance. This suggests that the mutated enzyme may show either decreased catalytic activity and/or decreased substrate binding. It is known that the mutation in one isolated case involved the signal peptide of MCAD precursor protein and thus prevented its import into mitochondria. The gene for MCAD has been assigned to chromosome 1.

Short-chain acyl-CoA dehydrogenase (SCAD) deficiency has been recorded in only a few patients and these show wide variation in clinical presentation. The defect has been seen in infants with a syndrome of psychomotor retardation and failure to thrive. These infants showed abnormal organic aciduria, and drastically decreased SCAD activity was demonstrable in cultured fibroblasts. Muscle symptoms were only part of a wider syndrome in all infants and children so far reported to have SCAD deficiency, but were the sole presenting feature in two adult patients, in whom lipid storage was demonstrable in skeletal muscle. The gene encoding for human SCAD has been mapped to chromosome 12.

Multiple acyl-CoA dehydrogenase (MAD) deficiency occurs due to abnormalities in the common pathway linking the FAD-dependent acyl-CoA dehydrogenases with the respiratory chain. The linking steps involve the electron transfer flavoprotein (ETF) and its corresponding dehydrogenase (ETF-DH). The most severely affected patients present soon after birth with glutaric aciduria type II (excretion of

glutaric, ethylmalonic, isovaleric, and isobutyric dicarboxylic acids). A later-onset form is referred to as ethyl-malonic-adipic aciduria and is said to involve only ETF, while the more severe syndrome involves both ETF and ETF-DH. The more benign syndromes tend to present with myopathy, show severe lipid storage in muscle and to be responsive to riboflavin. ETF protein has been shown to be dimeric, consisting of α and β subunits, and the gene for the α subunit has recently been mapped to chromosome 15.

In LCAD, MCAD, SCAD, and MAD, systemic carnitine deficiency may occur as a secondary feature of the disorder, and administration of oral carnitine is usually recommended as part of the treatment for patients. In older literature, dating from a time prior to the recognition of the various acyl-CoA dehydrogenase defects as separate entities, systemic carnitine deficiency is accorded the status of a primary disorder. However there is only one form of carnitine deficiency which can be considered to be a primary genetic defect; this disorder is characterized by a progressive cardiomyopathy, which may be fatal, and a demonstrable defect of high-affinity carnitine uptake. In contrast, secondary carnitine deficiency, occurring in any of the β-oxidation defects, has a totally different basis. Accumulation of excessive amounts of acyl-CoAs can only be regulated by a corresponding overproduction of acyl-carnitines. This results in excessive excretion of acyl-carnitines and a net loss of carnitine itself from many tissues, including skeletal muscle.

Respiratory Chain Disorders

This group of disorders is often referred to under the general heading of mitochondrial myopathies but since disorders of lipid metabolism involve mitochondrial enzymes the term respiratory chain disorder is to be preferred. The five enzyme complexes which comprise the respiratory chain are located within the inner mitochondrial membrane, and reducing equivalents (electrons) generated by substrate oxidation flow along this electron transport chain with the formation of a proton gradient across the membrane. This gradient is harnessed by complex V (ATP synthase) in the generation of ATP by phosphorylation of ADP. The electron transport function of the respiratory chain thus links substrate oxidation with ATP production (oxidative phosphorylation). Defects within the complexes of the respiratory chain will therefore compromise cellular ATP concentrations, and tissues which are particularly dependent upon aerobic metabolism, such as skeletal muscle, heart, and brain, are especially susceptible. When oxidative phosphorylation is impaired reduced enzyme cofactors, NADH in particular, will accumulate in the cytosol. The regeneration of NAD^+, the oxidized form of the cofactor, is possible through the catalytic action of lactate dehydrogenase (LDH) which links the reduction of pyruvate to lactate with the oxidation of NADH. It is not surprising,

therefore, to find that excessive lactate production is a common hallmark of respiratory chain disorders.

Mitochondria are unique organelles in that they contain their own DNA (mtDNA), which, in addition to ribosomal RNA (rRNA) and transfer RNA (tRNA)-coding sequences, also encodes 13 polypeptides which are components of complexes I, III, IV, and V (Anderson et al., 1981). This fact has important implications for both the genetics and the etiology of the respiratory chain disorders. Since mtDNA is maternally-inherited, a defect of a respiratory complex due to a mtDNA deletion would be expected to show a pattern of maternal transmission. However the situation is complicated by the fact that the majority of the polypeptide subunits of complexes I, III, IV, and V, and all subunits of complex II, are encoded by nuclear DNA. A defect in a nuclear-coded subunit of one of the respiratory complexes would be expected to show classic Mendelian inheritance. A further complication exists in that it is now established that some respiratory chain disorders result from defects of communication between nuclear and mitochondrial genomes (Zeviani et al., 1989). Since many mitochondrial proteins are synthesized in the cytosol and require a sophisticated system of posttranslational processing for transport and assembly, it is apparent that a diversity of genetic errors is to be expected.

Defects of nuclear DNA (nDNA) will normally involve only a single respiratory complex and cause a monoenzymopathy (single enzyme defect) though this may be tissue-specific or systemic depending on the affected gene. In contrast, large-scale deletions of mitochondrial DNA (mtDNA) or point mutations involving mitochondrial tRNA genes are likely to affect all the respiratory complexes which have mitochondrially-coded subunits (multienzymopathy). Defects involving single respiratory complexes are less common than multicomplex disorders but each of the individual complexes appears to be involved in specific monoenzymopathies.

Complex I (NADH-Ubiquinone Reductase)

This complex consists of at least 25 separate polypeptides, seven of which are encoded by mtDNA. Its catalytic action is to transfer electrons from NADH to ubiquinone, thus replenishing NAD^+ concentrations. Complex I deficiency has been described in myopathic syndromes, characterized by exercise intolerance and lactic acidemia. In at least some patients it has been demonstrated that the defect is tissue specific and a defect in nuclear DNA is assumed. Muscle biopsy findings in these patients are typical of those in many respiratory chain abnormalities. Instead of the even distribution of mitochondria seen in normal muscle fibers, mitochondria are seen in dense clusters, especially at the fiber periphery, giving rise to the ragged-red fiber (Figure 10). This appearance is a hallmark of many mitochondrial myopathies.

Patients with complex I deficiency may also present with severe congenital lactic acidosis, hypotonia, weakness, cardiomyopathy, and cardiorespiratory failure caus-

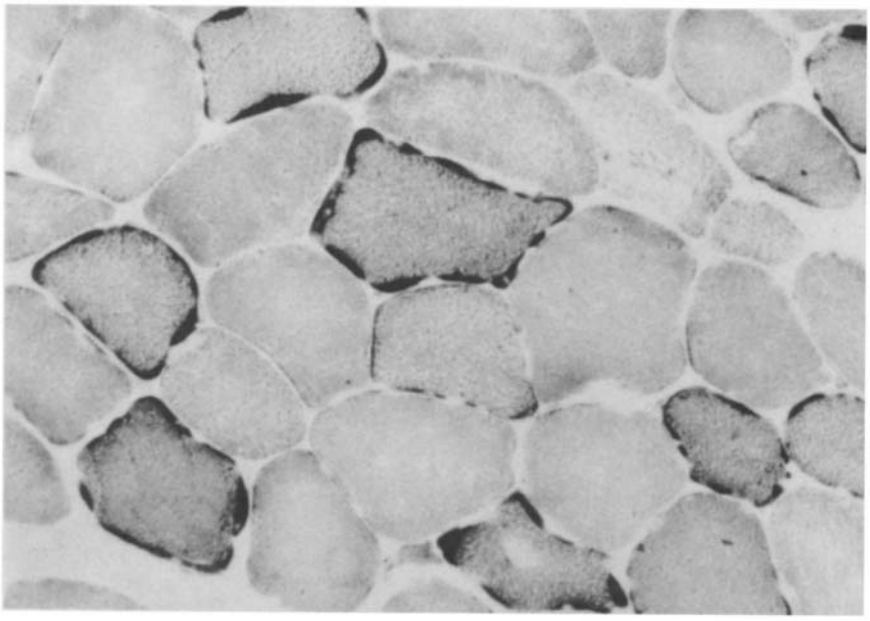


Figure 10. Ragged-red fibers with peripheral accumulations of mitochondria in muscle from a patient with Kearns-Sayre syndrome (KSS).

ing death in infancy. The defect thus appears to involve both skeletal and cardiac muscle and was documented in other tissues as well. The absence of a 20 kDa nuclear-coded subunit of complex I has been detected by immunoblotting in one patient, but in view of the large number of subunits in complex I, it is likely that this disorder will prove to have several molecular etiologies.

Complex II (Succinate-Ubiquinone Reductase)

This complex consists of four subunits, all of which are encoded on nuclear DNA, synthesized on cytosolic ribosomes, and transported into mitochondria. The succinate dehydrogenase (SDH) component of the complex oxidizes succinate to fumarate with transfer of electrons via its prosthetic group, FAD, to ubiquinone. It is unique in that it participates both in the respiratory chain and in the tricarboxylic acid (TCA) cycle. Defects of complex II are rare and only about 10 cases have been reported to date. Clinical syndromes include myopathy, but the major presenting features are often encephalopathy, with seizures and psychomotor retardation. Succinate oxidation is severely impaired (Figure 11).

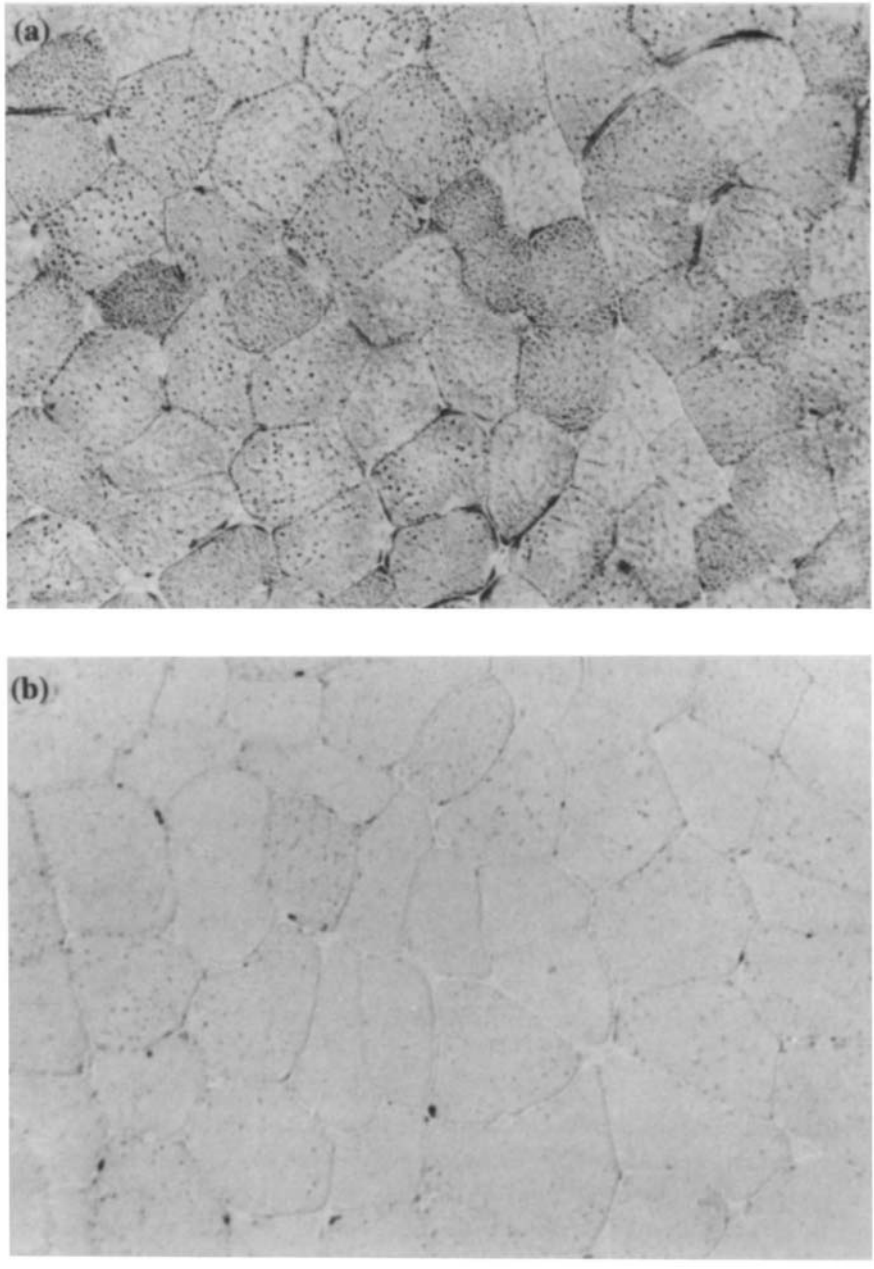


Figure 11. **(a)** Succinate dehydrogenase activity in normal skeletal muscle. **(b)** Muscle from patient with complex II deficiency showing severely decreased succinate dehydrogenase activity.

Complex III (Ubiquinol-Cytochrome c Reductase)

This complex contains 11 polypeptide subunits of which only one is encoded by mtDNA. Defects of complex III are relatively uncommon and clinical presentations vary. Fatal infantile encephalomyopathies have been described in which severe neonatal lactic acidosis and hypotonia are present along with generalized amino aciduria, a Fanconi syndrome of renal insufficiency and eventual coma and death. Muscle biopsy findings may be uninformative since abnormal mitochondrial distribution is not seen, i.e., there are no ragged-red fibers. Other patients present with pure myopathy in later life and the existence of tissue-specific subunits in complex III has been suggested since one of these patients was shown to have normal complex III activity in lymphocytes and fibroblasts.

Complex IV (Cytochrome c Oxidase)

This respiratory complex consists of 13 subunits, of which the three largest are encoded on mtDNA and contain the redox centers. Complex IV is involved in a greater diversity of defects affecting human skeletal muscle than any other respiratory complex.

Fatal infantile cytochrome c oxidase (CCO) deficiency is characterized by total absence of catalytic activity in skeletal muscle. This often occurs within the context of the Fanconi syndrome, or less commonly in association with a cardiomyopathy. Although the deficiency is global in skeletal muscle, with all fibers affected, only isolated scattered fibers show abnormal aggregations of mitochondria (ragged-red fibers). Multiple affected siblings within one family are frequently encountered and suggest autosomal recessive inheritance. The condition normally proves fatal before the age of six months and is characterized by worsening intractable lactic acidemia.

Benign infantile cytochrome c oxidase deficiency presents in the neonatal period with severe hypotonia and feeding difficulties, but unlike the fatal form of CCO deficiency, affected skeletal muscle shows a gradual increase in CCO activity, so that at three years, children show virtually normal catalytic activity. However, some residual motor weakness may remain. This syndrome was originally labeled mitochondria-lipid-glycogen disease (MLG) disease, due to the abnormal accumulation of both lipid and glycogen in the enzyme-deficient muscle. It is quite common to encounter excessive amounts of triglyceride and/or glycogen in respiratory chain defects, due to a generalized impairment of intermediary metabolism in affected tissues. The benign CCO deficiency must be carefully monitored in the first year of life in order to differentiate it from the fatal form of the disorder.

Leigh's syndrome (subacute necrotizing encephalomyelopathy) is characterized by a variable combination of clinical abnormalities including cerebellar ataxia, developmental delay, mental regression, deafness, optic atrophy, hypotonia, and

peripheral neuropathy. Muscle biopsies from patients with Leigh's syndrome and CCO deficiency show few mitochondrial abnormalities at light microscope level (ragged-red fibers are typically absent), but electron microscopic examination may reveal ultrastructural abnormalities such as concentric cristae and intramitochondrial inclusions.

Complex V (Proton Translocase, ATP Synthase)

Complex V catalyzes the phosphorylation of ADP to ATP. Deficiencies of complex V are rare and clinical presentation is usually nonprogressive muscle weakness. One subject included growth retardation, sensorineural deafness, and involvement of basal ganglia, which are occasional features of various other respiratory chain abnormalities.

Mitochondrial DNA Abnormalities

In addition to the conditions described above, which involve deficiencies of individual respiratory complexes, there is another important group of mitochondrial disorders which are associated with defects of multiple respiratory complexes. The underlying abnormalities in these disorders are to be found within the mitochondrial genome, which encodes some subunits of all the respiratory complexes except complex II (Figure 12).

Kearns-Sayre syndrome (KSS) is characterized by skeletal muscle weakness, cerebellar ataxia, cardiac conduction defects, deafness, and pigmentary retinal degeneration. It is regarded by many as a more severe manifestation of the syndrome of chronic progressive external ophthalmoplegia (CPEO) and myopathy, which is of more frequent occurrence and more benign clinical course. Together,

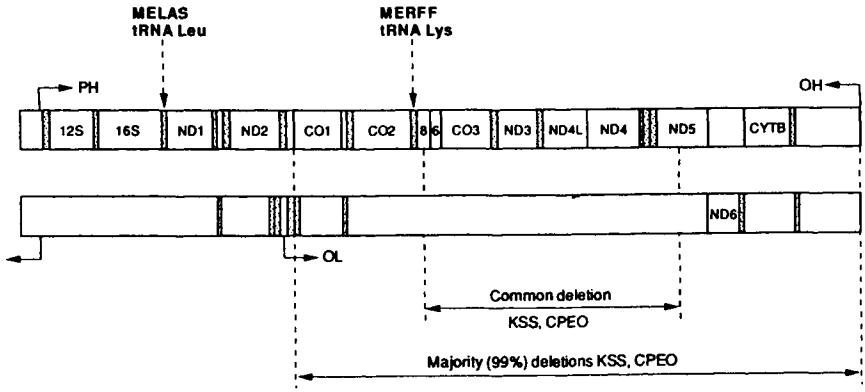


Figure 12. Schematic diagram of mitochondrial DNA (linearized).

these syndromes account for a large number of patients with mitochondrial myopathies. Muscle biopsies show the typical features of this type of disorder: ragged-red fibers (RRF) with pronounced peripheral aggregates of abnormal mitochondria (Figure 10), intramitochondrial paracrystalline inclusions, and abnormalities of mitochondrial cristae. Cytochrome c oxidase (CCO) activity is absent or severely decreased in most of the fibers showing typical RRF changes. Most patients with KSS or CPEO have large-scale deletions of mtDNA and these deletions are heteroplasmic (Figure 13). This means that both deleted and wild-type mtDNA are present in affected skeletal muscle, with a tendency for the deleted mtDNA to segregate in individual muscle fibers, with the result that these fibers become respiration-deficient. It is common for large-scale deletions of mtDNA to occur involving up to 8 kb of the total 16.5 kb mtDNA. Many regions encoding the polypeptide subunits of complexes I, III, IV, and V may be deleted, but it is likely that this is of less importance than the involvement of multiple transfer RNA (tRNA)-encoding sequences which will effectively impair the translation of all the polypeptides of respiratory complexes encoded by the mitochondrial genome (Poulton and Land, 1990).

Multiple deletions of mtDNA have been recorded in several families, where members show an autosomal dominant mode of inheritance of a syndrome involv-

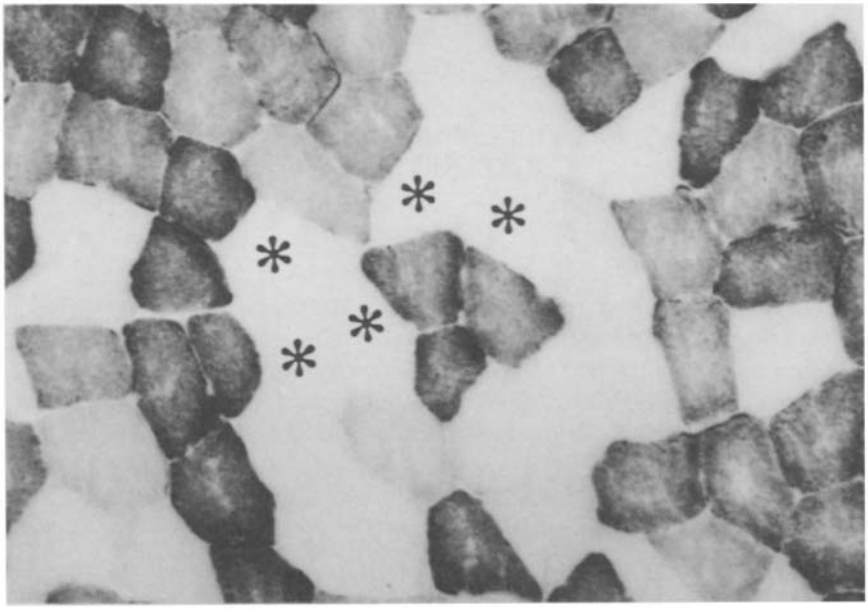


Figure 13. Mosaic of cytochrome oxidase-deficient muscle fibers (asterisks) in a patient with KSS and a heteroplasmic mtDNA deletion.

ing CPEO, myopathy, cataract and decreased life span. The mtDNA deletions tended to be different in different family members and appear to arise *de novo* in each individual by a common mechanism. This suggests mutation in a nuclear gene which acts on mtDNA replication (a *trans*-acting factor) and is a prime example of defective communication between nuclear and mitochondrial genomes (Zeviani et al, 1989).

Myoclonic epilepsy with ragged-red fibers (MERFF) is a rare syndrome which shows clear maternal inheritance and a variable clinical pattern including progressive myoclonus, cerebellar ataxia, dementia, and muscle weakness. It is associated with an A-to-G transition at position 8344 of the tRNA Lys gene in the mtDNA. The mutation is heteroplasmic and produces similar multicomplex deficiencies as are seen in KSS.

Mitochondrial myopathy, encephalopathy lactic acidosis, and strokelike episodes (MELAS) constitute another rare syndrome characterized by a single point mutation of mtDNA, this time an A-to-G transition at position 3243 in the tRNA Leu (UUR) gene. The site of this mutation is such that it may interfere with the termination of mtDNA heavy strand (H strand) transcription. MELAS can be differentiated clinically from KSS and MERRF since all three syndromes contain distinguishing features. In the case of KSS these are ophthalmoplegia, pigmentary retinopathy and heart block; in MERRF, myoclonus, optic atrophy, and sensory neuropathy; and in MELAS, the presence of seizures and recurrent cerebrovascular incidents. Since all three syndromes have mtDNA abnormalities affecting tRNA genes, it is common to find involvement of complexes I, III, and IV coexisting in individual patients.

MYOTONIAS, PERIODIC PARALYSES, AND MALIGNANT HYPERPYREXIA

Myotonias

This group of conditions, in which myotonia (the failure of voluntary muscle to relax following contraction) may be a feature, can now be classified according to the primary molecular defect responsible for the relevant condition. The clinical features of the different conditions within this group can show some significant differences, even among the diseases now known to be due to mutations within the same genes (Table 1).

Steinert's Disease (Myotonic Dystrophy)

The only one of these conditions which causes progressive and degenerative disease of muscle is myotonic dystrophy. Inheritance is autosomal dominant, but expression is very variable. The condition also illustrates the phenomenon of

Table 1. The Myotonias and Periodic Paralyses

Disease	*Mode of Inheritance*	*Gene Location*	*Gene Product*
Myotonic dystrophy	AD	19q13	Myotonin protein kinase
Dominant myotonia congenita (Thomsen's disease)	AD	7q35	Muscle chloride channel
Recessive generalized myotonia (Becker)	AR	7q35	Muscle chloride channel
Hyperkalemic periodic paralysis	AD	17q13	Sodium channel alpha subunit
Hypokalemic periodic paralysis	AD	1q31	Not known
Paramyotonia congenita	AD	17q13	Sodium channel alpha subunit

Note: AD = autosomal dominant; AR = autosomal recessive.

anticipation, with a tendency to increasing severity with successive generation. Among mildly affected individuals, the only manifestations of the disease may be frontal balding or cataracts, and these people and their families may be completely unaware of their condition. In other obligate gene carriers, the condition may only be detected by electromyography or slit lamp examination. The typical features of the condition in more classically affected individuals include myotonia seen mainly in the small hand muscles, forearms, and tongue. Muscle weakness is seen predominantly in facial and sternomastoid muscles and the distal limb musculature. In the most severe form of the condition, congenital myotonic dystrophy, babies are born with hypotonia which may be severe enough to cause feeding and respiratory problems. Severe facial weakness is present from birth, but myotonia is not seen in the first two years of life. These children are usually globally retarded. Children with congenital myotonic dystrophy are born only to affected mothers, but the mothers may themselves have been so mildly affected as to have been previously unaware of their condition.

The increasing severity of the condition, which is seen in successive generations, can now be explained by the dynamic nature of the gene mutation responsible for myotonic dystrophy (Harper et al., 1992). The disease is caused by the expansion of a CTG repeat in the 3′ untranslated region of a protein kinase gene on chromosome 19q13. Unaffected individuals have approximately 5–30 copies of the repeat, while affected individuals have an increased number. In general, the number of repeats correlates with the clinical severity of the condition, with congenitally affected children having the largest number of repeats, up to several thousand, and minimally affected individuals having repeats just outside the normal range.

The histopathological features of muscle samples from patients with myotonic dystrophy are not particularly distinctive. Early changes appear to be a selective atrophy of type 1 fibers, and hypertrophy of type 2 fibers, but the biochemical and/or physiological basis of these possibly related phenomena is not known. The incidence of degenerating fibers increases with age, although the presence of internally nucleated muscle fibers in early stages of the disease suggests that the muscle retains

a strong capacity to regenerate. Muscle weakness is a significant problem in these patients.

Thomsen's Disease (Dominant Myotonia Congenita)

The predominant feature of Thomsen's disease is severe myotonia, worse in the cold and early morning than later in the day or in the warmth, unassociated with significant muscle weakness, degeneration, or other severe pathology. Myotonia is particularly common in ocular muscles, and muscles of the arms and legs. Onset is typically during childhood.

Paramyotonia Congenita

This condition is rare, and difficult to understand. Its primary characteristic is that the inability to relax following a movement of mechanical activity lasts for a prolonged period in the absence of electrical activity. The condition is also unlike Thomsen's disease in that exercise does not relieve the problem, but makes it worse. It is also exacerbated by cold. Some patients exhibit hyperkalemia during an attack, and this has led to considerable problems of diagnosis (see below). It is of interest that the defect is in chromosome 17q13 in both this condition and in primary hyperkalemic paralysis. The histopathology of the muscle of patients with Paramyotonia congenita is variable, but generally involves an apparent lack of differentiation between type 1 and type 2 fibers, central nucleation, and highly variable fiber diameter. This may suggest a basic immaturity of the muscle, but the condition is so rare that definitive studies are unavailable.

Congenital Myotonic Dystrophy

Congenital myotonic dystrophy is a relatively rare condition in which myotonia (defined electrically) is mostly absent in the affected newborn infant, but becomes apparent in the older infant. Histopathology shows a consistent feature of arrested development and maturation of muscle fibers, but there is, currently, no adequate explanation for this phenomenon. Patients with congenital myotonic dystrophy rarely survive without aggressive ventilatory support, and survivors, without exception, are severely multiply handicapped.

There is no single underlying cause for the myotonia seen in the muscles of myotonic patients. The typical myotonic response is a train of action potentials generated in a muscle fiber in response to a single stimulus. Experimental work has shown that such a response can be generated in normal muscle fibers in which chloride conductance is suppressed, and this may be the cause of the myotonia of Thomsen's disease (see Barchi, 1988 for examples). It is almost certainly not the cause of myotonia in myotonic dystrophy in which there is an associated fall in

muscle fiber membrane potential (itself not a consequence of a reduction in Cl conductance). In this case, the gene product appears to be a protein kinase. Neither is it the cause of the electrically-silent myotonia of Paramyotonia congenita, in which the alpha-subunit of the fast Na^+ channel is involved. The physiological basis of myotonia in these conditions requires much more detailed evaluation.

The Periodic Paralyses

This is another group of diseases characterized by abnormalities in muscle fiber excitability. They are all periodic in the sense that periods of normal behavior are interspersed with periods of abnormally depressed excitability. During these latter phases, which may last for anything from a few hours to several days, there is a characteristic muscular weakness. The conditions are usually subdivided on the basis of serum K^+ levels during paralytic episodes, and are thus described as hyperkalemic, normokalemic, or hypokalemic.

Hyperkalemic Periodic Paralysis

Primary hyperkalemic periodic paralysis is usually first manifest in childhood. Attacks may last for a period of a few hours to several days, and the degree of muscle damage associated with the condition appears to increase with age and frequency of attacks. Vacuolation and dilatation of the SR is the most obvious form of damage, and it increases with age.

It is clear that paramyotonia congenita (p. 316) and primary hyperkalemic periodic paralysis may share many common features (autosomal dominant inheritance, myotonia, weakness exacerbated by cold and exercise, involvement of locus on 17q13 encoding the alpha-subunit of the fast Na^+ channel), but there are sufficient numbers of exceptions to suggest that these are two discrete conditions with overlapping symptomology. Engel (1988) has succinctly discussed the similarities and differences between the two conditions. Secondary hyperkalemic periodic paralysis may be precipitated by a large number of factors that cause the elevation of serum K^+ but renal failure is the most common.

Normokalemic Periodic Paralysis

This is not a distinct clinical entity. The most frequent form of the condition is primarily associated with severe cardiac dysfunction often in combination with dysmorphy. It is confusing, however, because a paralytic attack is accompanied by extrasystoles and tachycardia. Serum K^+ may be high, low or normal. A familial form of the disorder is exacerbated by cold and high K^+ but is relieved by Na^+ loading. It is, however, not associated with any specific changes in serum K^+.

Hypokalemic Periodic Paralysis

A fall in serum K^+ is commonly associated with hypokalemic periodic paralysis. Primary hypokalemic paralysis is usually first expressed in children and young adults. Paralytic attacks may fluctuate with remarkable frequency, and there is a common diurnal variation in severity, with weakness especially bad in the morning and evening. The condition has an autosomal dominant pattern of inheritance caused by an abnormality in or close to locus 1q13. The gene product is unknown.

Secondary hypokalemic paralysis can be caused by almost any event that causes excessive K^+ loss. One particularly interesting secondary condition, however, is thyrotoxic periodic paralysis, seen most commonly in Chinese patients.

The pathophysiology of the periodic paralyses, as a group of diseases, is not well understood. The biggest difficulty is that although it is clear that the paralytic phase may be associated with consistent shifts of major electrolytes (principally K^+ but also Na^+ and Cl^-), the fluctuating nature of the conditions is unexplained; neither is it known whether the changes in electrolyte distribution are causally or casually related to what is often a very sudden onset of weakness. It does seem, however, that the description of the diseases as hyper-, normo-, or hypokalemic has concentrated excessive numbers of investigations on the role of K^+. In fact, it is more likely that changes in serum K^+ trigger abnormal behavior of the Na^+ channels in excitable membrane. For example, in patients with hyperkalemic periodic paralysis, small increases in serum K^+ cause a small depolarization of muscle fiber membranes. This would not normally present a problem, but in these patients the fast Na^+ channels appear to be blocked (i.e., inactivated) at abnormally high membrane potentials. Similarly, in hypokalemic periodic paralysis there is a very high resting permeability to Na^+ and this results in an abnormal depolarization when serum K^+ falls, rather than a more typical hyperpolarization. As in hyperkalemic periodic paralysis, the Na^+ channels also appear to block at remarkably high membrane potentials. The pathophysiology of the conditions has been reviewed by Rüdel and Ricker (1985).

Malignant Hyperpyrexia

Malignant hyperpyrexia was not formally recognized as a discrete syndrome until the 1960s. The classical response is a rapidly developing rise in body temperature, cyanosis, tachycardia and/or dysrythmia, and muscle rigidity. The reaction is triggered in susceptible patients by a range of anesthetic agents and muscle relaxants. Halothane and succinylcholine (suxamethonium) are particularly commonly implicated. In its most extreme expression, body temperature can exceed 40 °C, and the combination of extreme hyperpyrexia, acidosis, hyperkalemia, and myoglobinuria can be fatal.

Malignant hyperpyrexia is inherited as an autosomal dominant condition, but penetrance is variable. The gene locus has been identified (19q, 12–13–2). There is no standardized diagnostic test to predict susceptibility, but as many as 70% of susceptible patients have an elevated serum creatine kinase, and muscle biopsies are liable to exhibit contractures when exposed to known triggering agents. There is no definitive histopathology. Malignant hyperpyrexias tend to exhibit mild myopathic signs such as focal loss of myofibrils, irregular fiber size, regenerating fibers, and central nucleation, but there is a rather loose association between susceptibility to malignant hyperpyrexia and Duchenne muscular dystrophy, myotonia congenita, a family history of sudden cot death, and particularly central core disease which shares the same genetic locus (see pp. 291, 316). The association with central core disease is particularly interesting, since the gene locus is identical. Whether the two conditions are allelic has not been established.

The cause of malignant hyperpyrexia is instability of intracellular Ca^{2+} homeostasis. It appears that triggering agents are capable of causing the abnormal release of Ca^{2+} from the sarcoplasmic reticulum. The rising intracellular free Ca^{2+} then triggers hypercontraction, mitochondrial overload, and the breakdown of normal muscle fiber metabolism. The incidence of this condition is probably higher than often appreciated. It has been calculated to be as high as 1:15,000 administrations of general anaesthetic to children, and should be considered a potential adverse reaction of great severity. It follows that a resuscitative strategy, that can be rapidly deployed in the event of hyperpyrexia, should form part of normal surgical provision. The high fatality rates suggest that this is often not the case.

NEUROGENIC MUSCLE DISORDERS

Pathological Features

Motor units depend on the functional integrity of their neural component for the maintenance of normal muscle fiber size, and fiber atrophy is a hallmark of a neurogenic muscle disorder. Muscle fibers will also atrophy as a result of simple disuse and in some primary myopathies, but neurogenic atrophy has its own particular distinctive features. The atrophy may be disseminated in distribution, i.e., affected fibers are scattered among fibers of normal size; this occurs when a relatively small number of motor units are affected. Since the territory of an individual normal motor unit is diffuse and component fibers are intermixed with those of other motor units, the disseminated distribution of atrophied fibers reflects this organizational pattern. Small group atrophy occurs when several adjacent motor units are affected and under these conditions clusters of 10–20 atrophied fibers may be seen. Large group atrophy results when the majority of motor units in a given area are affected; this may involve whole muscle fascicles and hundreds or even thousands of contiguous muscle fibers. The electrophysiological conse-

quence of denervation, however caused, is fibrillation and a detailed EMG recording exhibits characteristic spontaneous electrical firing of the denervated muscle fibers (Barwick and Fawcett, 1988).

The severe weakness which occurs as a result of denervation atrophy may be offset if effective reinnervation takes place. This process involves healthy surviving motor neurons enlarging their territory (the number of muscle fibers which they innervate) by means of axonal sprouting. This mechanism results in the incorporation of previously denervated fibers within the territory of a larger remodeled motor unit. Since the component muscle fibers of any one unit are metabolically uniform, reinnervation may result in the formation of large groups of fibers with the same metabolic characteristics. This phenomenon is referred to as uniform fiber type grouping and contrasts sharply with the random mosaic pattern of fiber types seen in normal muscle (Figure 14). The demonstration of fiber type grouping using histochemical techniques, for example, myofibrillar ATPase, provides the histopathologist with a means of recognizing reinnervated muscle (Figure 15). However the absence of uniform fiber grouping cannot be construed as evidence that reinnervation has not occurred. Two or more motor units of differing metabolic type may each participate in the reinnervation of a group of denervated fibers, resulting in a random mixture of fibers of different types which is indistinguishable from normal muscle.

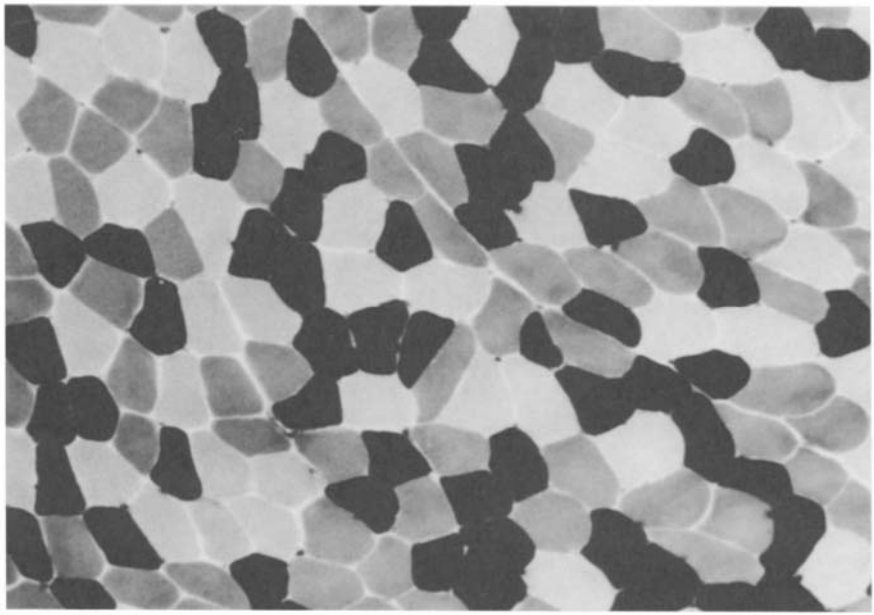


Figure 14. Normal skeletal muscle showing random distribution of type 1 (dark), type 2A (pale) and type 2B (intermediate) fibers; myofibrillar ATPase after pH 4.6 preincubation.

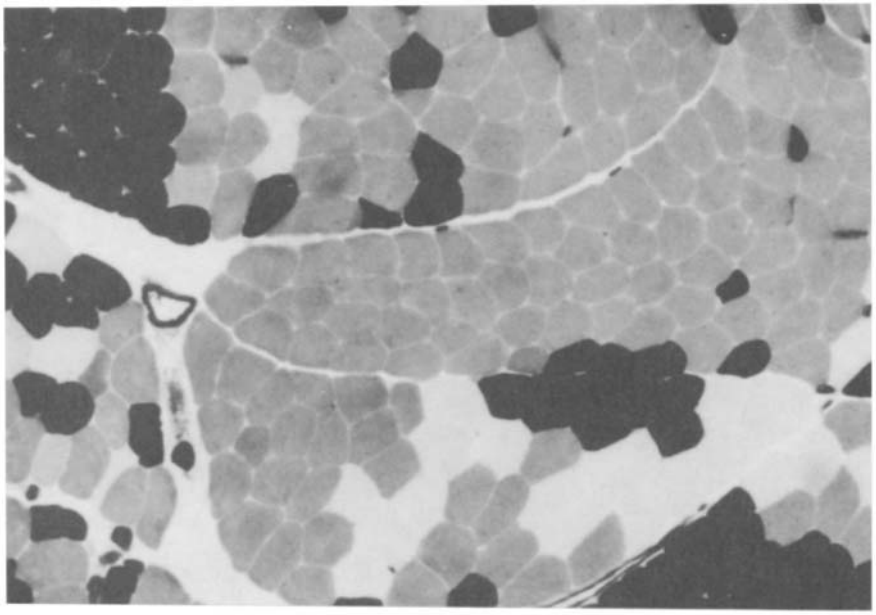


Figure 15. Grouping of muscle fibers of uniform histochemical type indicating reinnervation in a patient with SMA.

The clinical and histopathological features of individual neurogenic muscle disorders are determined to a large extent by the balance of the opposing processes of denervation and reinnervation, at least as far as the muscle dysfunction is concerned. The involvement of upper motor neurons in some of these disorders is outside the scope of this chapter and will only be referred to in passing.

Spinal Muscular Atrophy (SMA)

This group of neurogenic muscle disorders are progressive inherited diseases specifically affecting the lower motor neuron and have an infantile or juvenile onset. Additional involvement of thalamic neurons and dorsal root ganglia is seen in the severe infantile forms of SMA rather than in those which are of later onset (Gomez, 1986). The most severe form of infantile SMA is Werdnig-Hoffman disease (SMA type I) in which profound hypotonia is present at birth. Decreased fetal movement in the last trimester of pregnancy is frequently reported. The symptoms are present by 3–6 months at the latest and death occurs before the age of 18 months. Histopathological examination of muscle biopsies shows widespread denervation atrophy, often affecting over 90% of the muscle fiber population. The muscle fibers of surviving innervated motor units may show marked hypertrophy

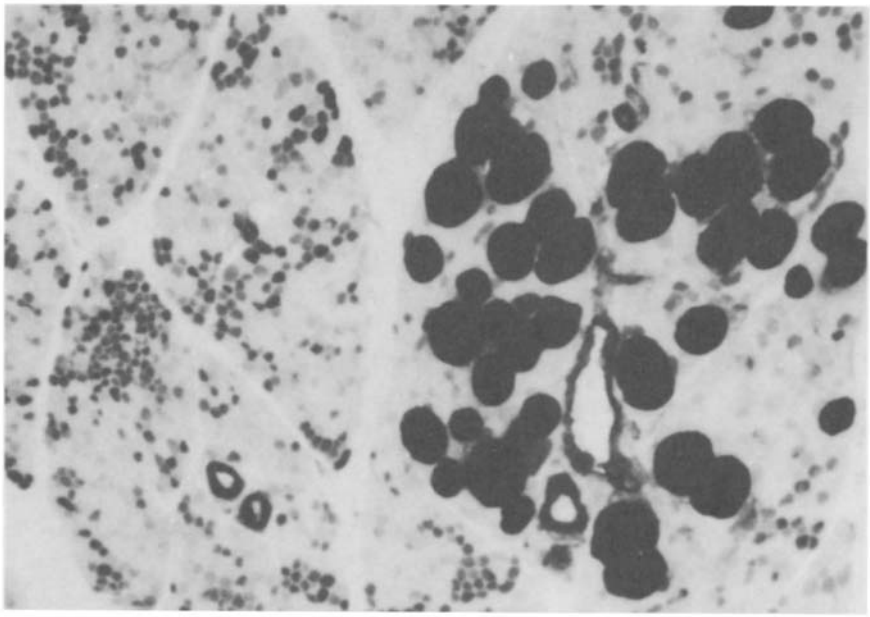


Figure 16. Werdnig-Hoffman disease: most muscle fibers show severe atrophy with some type 1 (dark) fibers showing hypertrophy.

(up to five or six times normal diameter) in an attempt to compensate for loss of functioning muscle (Figure 16). The territory of a surviving hypertrophic motor unit is the same as that of a normal motor unit, i.e., its component fibers are intermixed with those of other motor units, which, in Werdnig-Hoffman disease are likely to be denervated. The compact motor unit territory characteristic of reinnervated muscle is not seen and there is no evidence to suggest that reinnervation occurs in this type of SMA.

In contrast, infants with SMA type II (intermediate type SMA) may show a limited capacity for muscle reinnervation. This disorder is more benign in its clinical course than Werdnig-Hoffmann disease; although hypotonia and weakness may be present in the neonatal period, they may not become apparent until 6–12 months of age. Survival into the second decade is sometimes seen, even in some patients whose early biopsy findings were virtually as severe as those in Werdnig-Hoffman disease. Other infants with intermediate SMA show clear signs of reinnervation in that uniform fiber type grouping is commonly present. Limited postural control is often achieved in these infants and children and some may manage a few steps with appropriate walking aids.

Kugelberg-Welander disease (type III SMA) is characterized by juvenile onset and prolonged survival. In early stages of the disease the process of reinnervation

seems to keep pace with that of denervation and few atrophied fibers may be seen in biopsy material. On the other hand, extensive remodeling of motor units due to reinnervation may be seen and uniform fiber type grouping may be widespread. Later on reinnervated motor units may undergo further denervation; this gives rise to groups of atrophied fibers of uniform fiber type and is characteristic of secondary denervation. In this chronic form of SMA, muscle fibers may undergo repeated cycles of denervation and reinnervation. However denervation gradually outstrips reinnervation, and excess fibrous connective tissue and fat replace lost muscle tissue.

All three forms of spinal muscular atrophy are inherited as autosomal recessive disorders, linked to chromosome 5q. Prenatal diagnosis using closely linked markers is now available. A rare, autosomal dominant form of juvenile SMA is similar in expression to the recessive forms, but 5q is not involved.

Adult motor neuron disease (MND) is broadly synonymous with the syndrome of amyotrophic lateral sclerosis (ALS) and includes the clinical features of progressive muscular atrophy, progressive bulbar palsy and, in about 10% cases, primary lateral sclerosis, a purely upper motor neuron disorder (Mulder, 1986). ALS comprises both upper and lower motor neuron disease and is the most common manifestation of the disease complex. Progressive weakness of all limb, bulbar, and respiratory muscles occurs and respiratory complications are the most usual cause of death. At its most severe, motor neuron disease can be fatal within three years of onset. A muscle biopsy is most likely to show a picture of disseminated fiber atrophy with little evidence of reinnervation. Such evidence, i.e., uniform fiber type grouping, is more common in milder forms, such as those presenting with predominantly lower motor neuron disease and progressive muscular atrophy. Clinically, care must be taken to differentiate such forms of MND from late-onset forms of SMA. Histopathological examination of muscle, while of immense value in differentiating between the various infantile and juvenile spinal muscular atrophies, is notoriously unable to assist the clinician in the differential diagnosis of late-onset motor neuron disorders. The same is also true of muscle biopsy findings in the many and various forms of peripheral neuropathy, which fall outside the scope of this chapter, and which give rise to muscle fiber atrophy of neurogenic pattern but with no indication as to precise type.

INFLAMMATORY MUSCLE DISORDERS

These disorders are all acquired conditions with no evidence of an hereditary basis. Most of them involve inflammation of the skeletal muscle itself (myositis) (Figure 17), though this may sometimes occur because of initial targeting of the muscle vasculature or connective tissue. Many instances of myositis are classed as idiopathic disorders, in that the precise mechanisms of muscle degeneration are not known, but is widely accepted that these syndromes are associated with abnormal function of the immune system. The syndromes of polymyositis (PM) and derma-

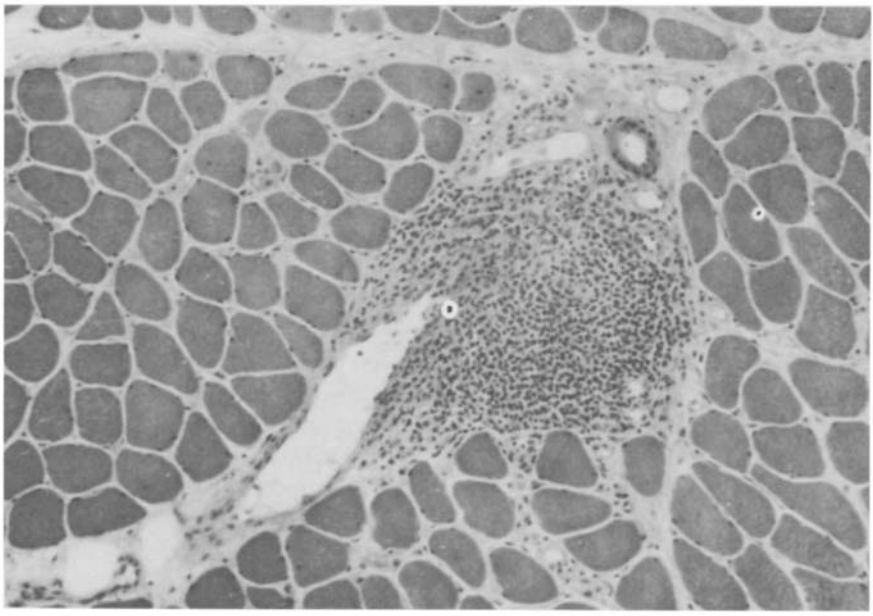


Figure 17. Inflammatory cell infiltrate in a muscle biopsy from a patient with dermatomyositis; note compact nature of infiltrate and perivascular location.

tomyositis (DM) are, numerically, the most important members of this idiopathic group. They are differentiated clinically by the presence of characteristic skin changes in the DM syndrome and can often be shown to involve separate dysimmune mechanisms. Both PM and DM can affect adults and children. The adult forms of both disorders are also prone to occur in conjunction with various connective tissue disorders in the so-called overlap syndromes. Disorders occurring most frequently in association with PM or DM include rheumatoid arthritis, scleroderma, and systemic lupus erythematosus. PM and DM can also occur in conjunction with various malignancies.

Myositis may also have an infective basis. Viral myositis has been recorded in association with influenza and picornavirus infections, particularly those due to viruses of the Coxsackie group, and HIV infection is an increasingly common cause of myositis seen in routine practice. Fungal, bacterial, and parasitic myositis is seen much more rarely in North America and Europe than in tropical parts of the world, but in these regions these forms of infective myositis constitute a significant problem. In any survey of inflammatory muscle disorders, it is also necessary to consider other inflammatory conditions which affect muscle indirectly, but do not cause myositis in the strict sense of the word. In this group are to be found various forms of arteritis and fascitis and granulomatous conditions such as sarcoidosis.

Polymyositis and Dermatomyositis Syndromes

The incidence of these syndromes seems to vary according to geographical area and ethnic background but is about 0.2–0.3 per 100,000 population (mean annual incidence rate). However incidence rates calculated for adult populations are up to three times higher. The fifth and sixth decades show peak incidence rates and there is also clear biomodality across the full age spectrum due to the existence of a juvenile form of dermatomyositis (JDM) which is pathogenetically distinct. Polymyositis, uncomplicated by skin changes, can also occur as a juvenile condition.

The vast majority of cases of PM and DM are sporadic, though other family members appear to show a higher than normal incidence of other disorders with an autoimmune or putatively autoimmune basis. Various studies have also shown a tendency for PM and DM patients to express preferentially particular human leukocyte antigens (HLA). HLA-B8 antigen has been recorded in both childhood and adult PM and DM with an incidence significantly higher than that manifest in the population as a whole. An increased incidence of HLA-B14 was also found in adult dermatomyositis patients showing overlap syndromes of rheumatoid arthritis or other connective tissue disorders. However the available evidence is still too preliminary to attempt any generalizations regarding HLA typing and particular forms of PM or DM.

Juvenile dermatomyositis (JDM) is perhaps the most uniform, in terms of clinical and histopathological features, of the whole PM/DM disease complex. Presentation may be before 5 years of age with peak incidence between 8 and 12 years. The disease may remit and recur until well into young adult life. The skin lesions include a facial rash in butterfly distribution across nose and cheeks. Erythematous skin changes are seen over extensor surfaces of joints, especially knees, knuckles and elbows. Muscle involvement is generally evident some time later and takes the form of weakness and stiffness, particularly affecting shoulder and pelvic musculature. Proximal muscles are often worse affected than distal muscles and extensors worse than flexors. In the absence of prompt and effective treatment contractures may occur at elbows, ankles, knees, and hips. Subcutaneous calcification and skin ulceration may be found; calcification of deeper-lying connective tissue may be apparent on X-ray.

In severe cases, or those refractory to treatment, truncal and limb weakness may be accompanied by involvement of masticatory, bulbar, and respiratory muscles. However the most life-threatening clinical manifestations are those affecting the gastrointestinal tract, since stomach ulceration can occur and death from perforation and peritonitis are not unknown. Medication with steroidal antiinflammatory agents is necessary but weakens the childrens' resistance to infection, so that systemic spread of usually self-limiting disorders, such as candidiasis, may occasionally occur.

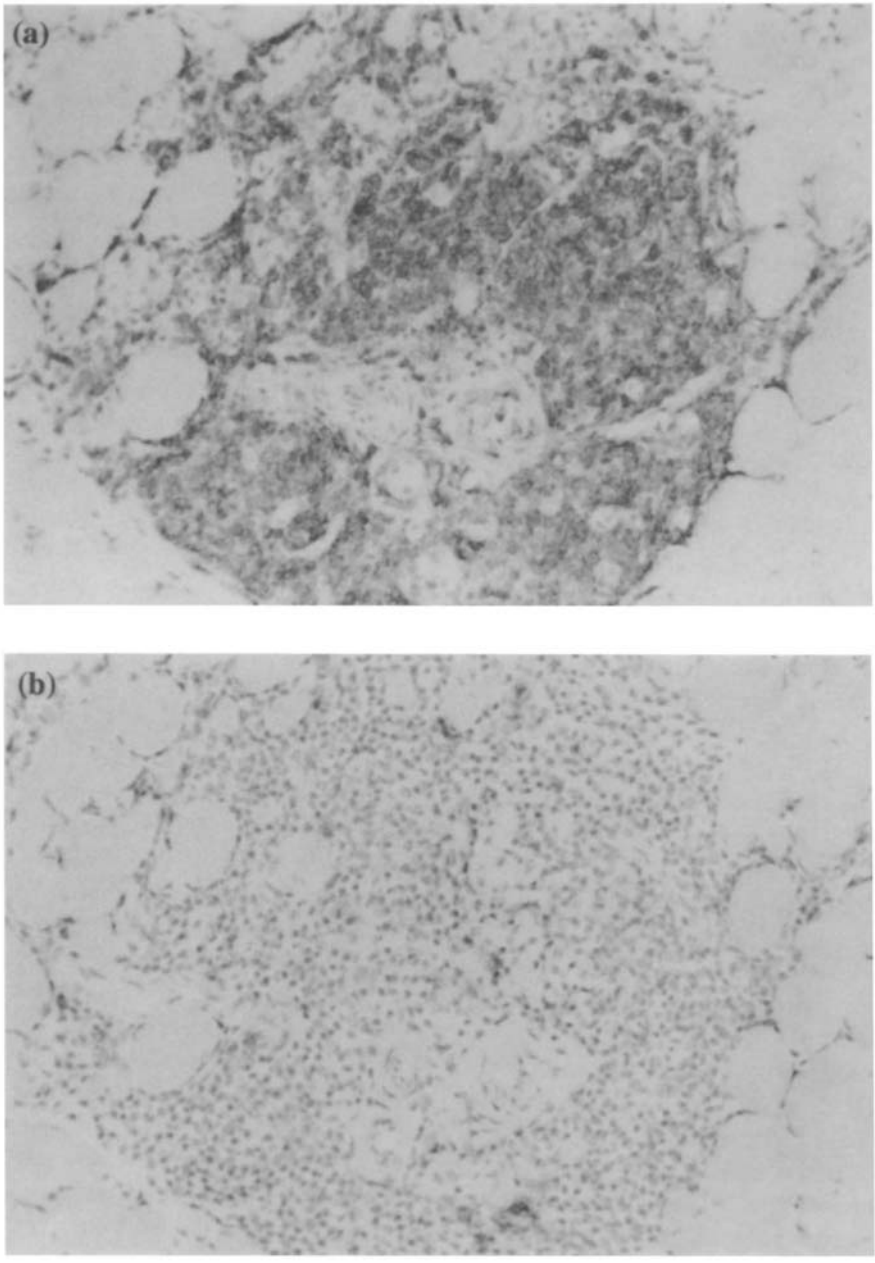


Figure 18. **(a)** Immunocytochemical labeling of B-lymphocytes in muscle from a patient with juvenile dermatomyositis. **(b)** Immunocytochemical labeling of T8 lymphocytes showing that the infiltrate contains very few cells of this type (serial section).

Muscle biopsy is usually undertaken to confirm the provisional clinical diagnosis. Because the skin lesions normally precede those in muscle, biopsies of muscle taken early may show little abnormality. Inflammatory foci may be scanty or absent and muscle fiber diameters may be normal. However typical biopsies show discrete foci of inflammatory cells, with a predominance of B-lymphocytes (see Figure 18). These cells are situated in perimysial connective tissue rather than in the endomysium and are often also perivascular in location. Muscle fiber necrosis occurs in JDM but muscle fibers do not appear to be the primary target of the disordered immune process. Rather, it is the microvasculature of the muscle which appears to degenerate first and muscle necrosis is preceded by capillary necrosis, detectable at the ultrastructural level.

The trigger for endothelial cell degeneration is probably linked to the predominant cell type found in JDM, the B-lymphocyte. It is postulated that antibody secreted by B-lymphocyte-derived plasma cells may react with antigen on the surface of, or associated with, endothelial cells and that this may give rise to antibody-dependent cytotoxic reactions. Certainly immunoglobulins are found in association with the microvasculature in JDM and the membrane attack complex, consisting of elements C5b-9 of the complement cascade, has also been detected. Alternatively if antibody reacts with soluble, circulating antigen, immune complexes may be formed. Deposition of such complexes, for example in small vessels, usually leads to complement activation and tissue damage. It is fairly well established that the muscle necrosis in JDM is secondary to the damage to the microvasculature. The topography of muscle damage is also interesting, in that most muscle necrosis, and the muscle atrophy which occurs subsequently, shows a perifascicular distribution (Figure 19). This may be due to either or both of the following mechanisms.

1. Perifascicular capillaries are closer to aggregates of antibody-secreting cells (B-lymphocytes) situated in perimysial connective tissue and therefore are most severely affected by antibody-dependent cytotoxic reactions.
2. Immune-complex deposition occurs at a higher level in the vascular tree (i.e., at arteriolar level) and this may cause fluctuations in perfusion pressure. Perifascicular capillaries are most distal from the head of vascular pressure and therefore most likely to suffer from periodic anoxia.

In addition to marked perifascicular atrophy, infarctlike areas are sometimes seen, and are also consistent with a microangiopathy. Muscle fibers which appear normal morphologically may show loss of myofibrillar ATPase activity from the center of the fibers; this is also characteristic of muscle subject to ischemia. Such changes may be reversible, but more prolonged ischemia undoubtedly causes irreversible muscle necrosis.

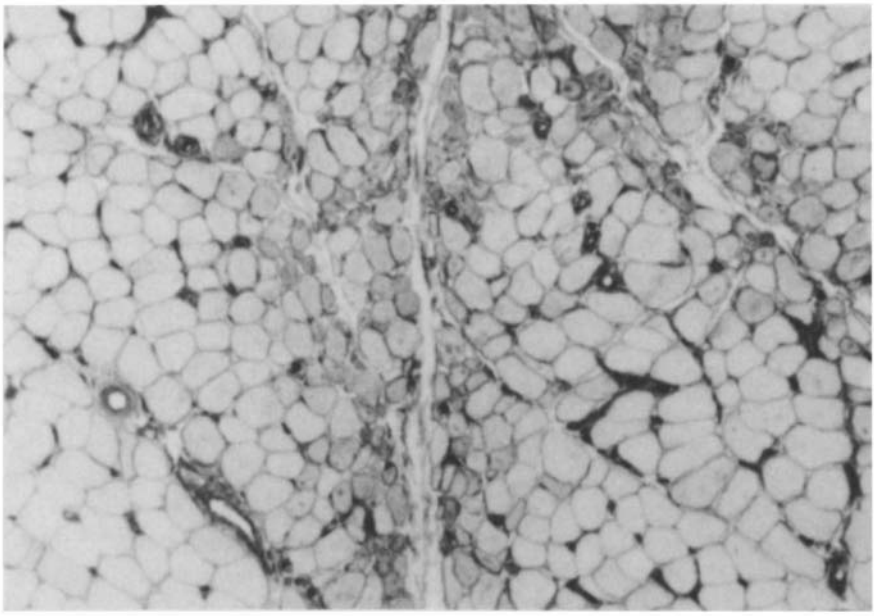


Figure 19. Juvenile dermatomyositis, showing atrophy of perifascicular fibers and strong reactivity for MHC-I antigen in these fibers.

Adult dermatomyositis (ADM) shows the same peak incidence in the fifth and sixth decades as does adult PM; in each condition there are at least twice as many women affected as men. Muscular weakness is generally apparent after the skin changes have become noticeable. The facial rash is most prominent on the cheeks and around the eyes. Neck and upper chest are often affected along with both proximal and distal regions of the arms. Elbows, fingers, knees, and ankles are sometimes affected by thicker patches of dermatitis with induration and scaling. The bases of fingernails often show hyperemia and the tips of the fingers may be reddened. Muscle weakness is proximal rather than distal in at least 75% cases. Neck flexors are most often involved but general shoulder and pelvic girdle weakness may develop. Weakness is symmetrical in distribution and is accompanied by myalgia in about two-thirds of the cases. Dysphagia and respiratory muscle involvement may occur. Heart abnormalities are well documented, both by ECG evidence of atrial arrhythmias, atrioventricular conduction defects, and bundle-branch block, and also by autopsy evidence of myocardial myositis.

Muscle biopsy shows some features in common with JDM; for example, perifascicular atrophy resulting from preferential involvement of peripheral muscle fibers is seen in some cases of ADM whereas it is never seen in adult PM. Since

ADM may evolve over several years, the extent of fiber atrophy provides an important indication of the chronicity of muscle degeneration. Acute muscle necrosis and phagocytosis give some indication as to how active the disease is at the time of biopsy. In most biopsies from ADM patients, the inflammatory cell foci are perivascular and perimysial rather than endomysial and are dominated by B-lymphocytes. The ratio of $T4^+$ lymphocytes (helper cells) to $T8^+$ lymphocytes (cytotoxic) generally indicates a predominance of the former. As in JDM, this is consistent with humoral mechanisms of cell damage, and vascular involvement is also apparent in the form of capillary endothelial cell abnormalities (tubular arrays) and duplication of basal lamina. Loss of myofibrillar ATPase from the central portions of fibers is a common prelude to muscle necrosis.

In common with PM and JDM, adult DM is a patchy process and care must be taken to maximize the chances of a muscle biopsy yielding positive diagnostic information. Electromyography (EMG) is helpful in this regard since sampling of multiple muscle sites is possible. Characteristic EMG findings include the presence of spontaneous activity, fibrillation and increased insertional activity. Polyphasic potentials of small amplitude and short duration are also seen as are high-frequency discharges after mechanical stimulation of the muscle. In order to avoid needle-tracks in a subsequent muscle biopsy it is best to select a corresponding contralateral muscle.

Polymyositis (PM) is a disorder in which adult patients present with myositis without skin involvement. It is a disorder twice as common as classical ADM. As in DM proximal muscle weakness is an almost invariable feature. Severity of symptoms at presentation varies greatly and, on this basis, it is usual to classify cases of PM as acute, sub-acute, or chronic. At its worst, acute PM is characterized by pain and tenderness of muscles, profound weakness, and associated evidence of muscle damage such as grossly elevated creatine kinase (CK) levels and myoglobinuria. Such patients may show high fever and if the myoglobinuria persists it may lead to oliguric renal failure. Sub-acute PM shows a more gradual onset, evolving over weeks or months and symptoms are less dramatic. Chronic PM has an onset which can only be described as insidious, with slowly progressive weakness and atrophy.

The histopathological features of PM may be radically different from those of JDM and ADM. There is little, if any, evidence of involvement of the microvasculature and the muscle necrosis which occurs appears to be the direct result of targeting of individual muscle fibers. In the dermatomyositis syndromes, antibody-dependent humoral mechanisms are predominant and B-lymphocytes are seen to be the most abundant cell type in almost all JDM cases and a substantial proportion of ADM cases. In contrast, most muscle biopsies from PM patients show evidence of inflammation in which $T8^+$ (cytotoxic) lymphocytes predominate (Figure 20). Moreover, the distribution of inflammatory cell infiltrates tends to be different. Instead of the mainly perifascicular location of lymphocytes in JDM/ADM, there

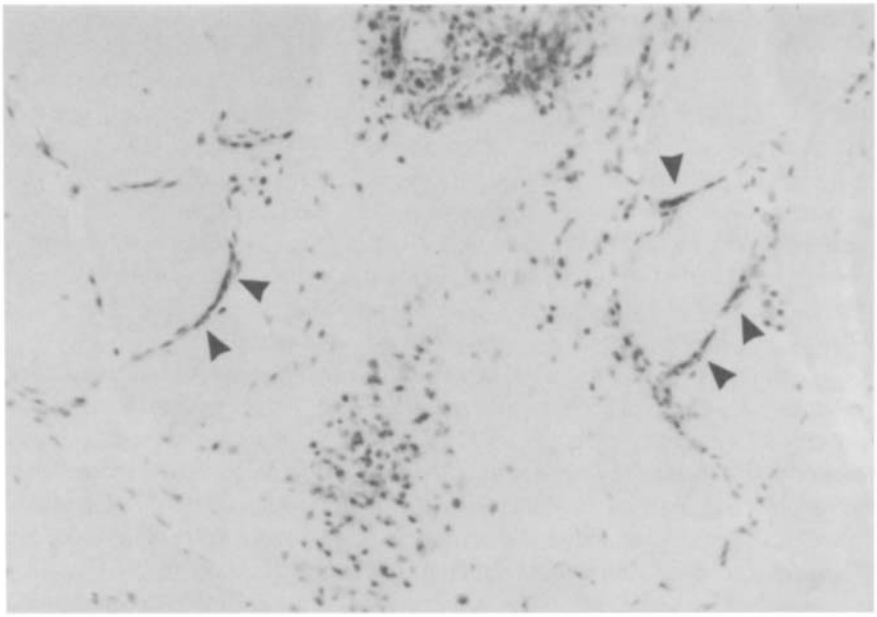


Figure 20. Immunocytochemical labeling of T8 lymphocytes showing some cells in close contact with muscle fiber plasma membranes (arrowheads).

is a much greater tendency for these cells to be found in the endomysium, often in close apposition to individual muscle fibers. This is consistent with the function of $T8^+$ lymphocytes as cytotoxic cells.

$T8^+$ lymphocytes which have been previously sensitized to an antigen (presumed in this context to be a muscle surface antigen), will recognize that antigen only if it occurs on a histocompatible target. Thus target recognition by $T8^+$ cytotoxic cells is restricted by major histocompatibility complex (MHC) gene products. Expression of MHC Class I (MHC-I) antigen is a necessary prerequisite for an individual muscle fiber to become a possible target for $T8^+$ cells. In normal skeletal muscle the expression of MHC-I antigen is confined to endothelial cells of capillaries and larger vessels, although macrophages, which can function as antigen-presenting cells, also express MHC-I. In PM a variable proportion of muscle fibers are seen to express MHC-I on their surfaces with variable intensity (Figure 21). It can also be shown that $T8^+$ cells only occur within or in close proximity to muscle fibers which express MHC-I. The cytotoxic lymphocytes bearing a receptor for the muscle-specific antigen will be stimulated by contact with it so that they either invade the target cell or otherwise damage its surface leading to the release of lymphokines. This process is likely to result in recruitment of phagocytic cells

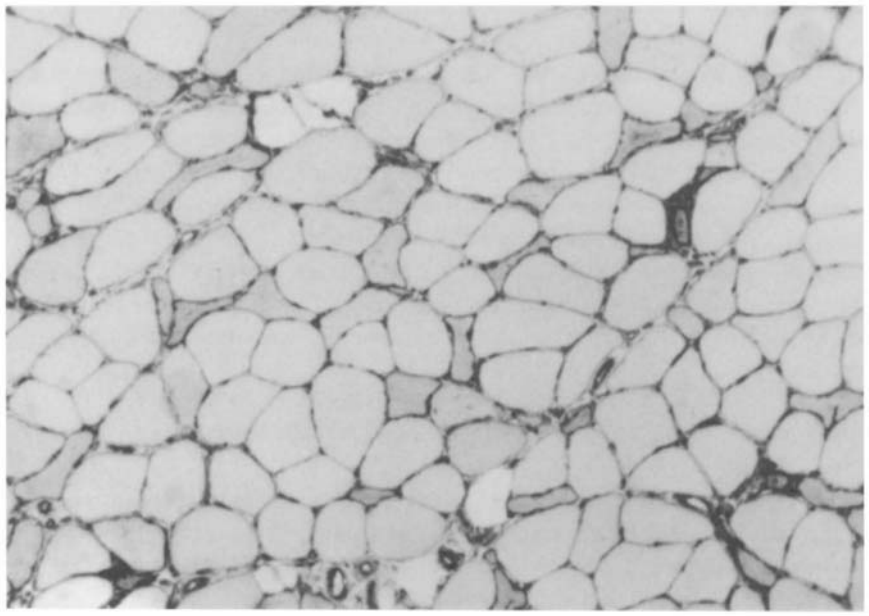


Figure 21. MHC-I antigen expression in muscle biopsy from polymyositis patient, showing random distribution of fibers with strong reactivity.

which will remove muscle debris and of more cytotoxic lymphocytes which may target adjacent muscle fibers. The distribution of targeted muscle fibers, and hence of muscle necrosis, in PM is essentially random and there is no preferential involvement of perifascicular fibers. This scattered distribution of affected fibers is in contrast with the situation in JDM and many cases of ADM where the underlying abnormality of the microvasculature gives rise to a markedly clustered distribution of necrotic muscle fibers and also to selective perifascicular involvement.

Muscle regeneration is usually remarkably effective in PM and DM syndromes, provided that prompt treatment is given to prevent recurrence of muscle damage. This is not surprising, since the underlying abnormalities in PM and DM reside in the immune system and not in skeletal muscle itself. In contrast to primary degenerative muscle disorders, such as the muscular dystrophies, in which regenerative capacity is impaired, there is no a priori reason why muscle in PM/DM patients should not regenerate fully. Nevertheless, in cases refractory to treatment, muscle fiber atrophy and replacement by fat and fibrous connective tissue may occur as irreversible features.

Overlap Syndromes

The definition of an overlap syndrome dictates that the criteria for diagnosis of both disorders (in the present context, of PM/DM and of some other connective tissue disorder), are fulfilled. It is not unexpected that those syndromes which overlap with PM/DM are also either known autoimmune conditions or ones in which an autoimmune basis is strongly suspected. The association of these disorders with PM/DM syndromes may not materially alter the basic histopathological features expected in PM/DM but some differences may be identifiable. The disorders most frequently associated with an overlap syndrome are rheumatoid arthritis, systemic lupus erythematosis, scleroderma, and mixed connective tissue disease.

Polymyositis and Dermatomyositis Associated with Malignancy

On review of all cases of PM/DM in any large center for the investigation of neuromuscular diseases, an incidence of associated malignancy varying from about 7% to over 20% is to be expected. In contrast to PM/DM alone where the male:female ratio is about 1:2, approximately equal involvement of the sexes is seen in PM/DM with malignancy. In patients over 40 years of age the association with DM greatly exceeds that with PM.

Inclusion Body Myositis (IBM)

This form of myositis stands apart from the classical PM/DM syndromes on account of its distinctive clinical and histopathological features. There is no clear difference in incidence between males and females and the disorder is typically one of middle or old age. In the majority of cases, progression is slow and skin involvement is not seen, so that the main question of differential diagnosis is its distinction from chronic PM. Unlike classic PM, weakness involves distal muscles as frequently as proximal muscles. CK levels are usually only moderately raised. A common finding which leads to the correct diagnosis of this condition is its nonresponsiveness to steroid treatment or other forms of immunosuppression.

Muscle biopsy with full histochemical and ultrastructural investigation is necessary for the confirmation of a diagnosis of IBM. The inclusions which are the hallmark of this disorder are to be found in three locations: (a) basophilic granular inclusions are found at the periphery of vacuoles within the cytoplasm of muscle fibers; (b) eosinophilic hyaline inclusions are also found in the cytoplasm but are not associated with vacuoles; and (c) intranuclear inclusions consisting of aggregates of filamentous microtubules are found in a variable percentage of muscle nuclei. Inclusions of the first two types are visible at light microscope level, whereas the third type is detectable at the electron microscope level only. Ultrastructural

examination reveals that the basophilic inclusions are lamellated membranous structures (myeloid bodies).

Infective Myositis

Myositis can occur concomitantly with various virus infections and in at least some cases there is good evidence that the viral infection is causative. There is also a continuing controversy regarding the etiological link between previous virus infections and some of the idiopathic PM/DM syndromes. Bacterial myositis is not common and nor are cases of myositis due to fungal infections, though both are more prevalent in tropical areas than in temperate latitudes. Major protozoan causes of myositis include toxoplasmosis, sarcosporidiosis, and African and American forms of trypanosomiosis. Nematode infections (trichinosis and toxocariasis) and cestode infections (cysticercosis and echinococcosis) are also important causes of parasitic myositis.

Viral Infections

Acute myositis with rhabdomyolysis and myoglobinuria is a rare complication of some viral infections. Influenza A virus has been implicated most often in this regard but influenza B, parainfluenza B, and the picornaviruses, Echovirus 9 and Coxsackievirus B2, B3, and B5, are also said to have caused similar syndromes. Biopsy findings generally show scattered muscle fiber necrosis with regeneration and only sparse inflammatory infiltration. Viral agents have been isolated from the muscle only on very rare occasions and this argues against direct virus invasion being responsible for the muscle damage. HIV infection can give rise to an inflammatory myopathy which resembles polymyositis both clinically and histopathologically. Most of the inflammatory cells are $T8^+$ lymphocytes, and whereas it is likely that the condition is immunologically-mediated, it is unlikely that muscle fibers are invaded by the virus since *in situ* hybridization studies have consistently failed to demonstrate this. The persistence of viral genomes or genomic fragments in idiopathic inflammatory myopathies is still hotly debated. Mumps virus has been reported in inclusion body myositis (IBM) but attempts to substantiate this finding by seeking evidence of mumps nucleocapsid gene using *in situ* hybridization were not successful. Some workers have reported the existence of picornavirus genome in PM/DM whereas other workers using similar *in situ* hybridization, PCR, and dot-blot techniques have obtained negative results.

Bacterial and Fungal Myositis

Bacterial agents, particularly *Staphylococcus aureus*, β-hemolytic streptococci, and *Escherichia coli* may give rise to pyomyositis, a form of acute suppurative

myositis. This may follow penetrating injuries but may also occur apparently without any associated trauma. Biopsy findings show interstitial edema with a mixed mononuclear cell and polymorphonuclear leukocyte infiltrate. Muscle fibers become necrotic and in later stages the abscess cavity becomes surrounded by granulation tissue. Permanent scarring and loss of muscle tissue may occur. Clostridial myositis (gas gangrene) is an increasingly rare condition but may occur in military personnel due to contamination of open wounds. It can also be encountered following infection of ulcerated ischemic limbs, or after intestinal surgery. The main organisms involved are *Clostridium welchii*, *Cl. edematiens*, and *Cl. septicum*. Rapid muscle damage is caused by the collagenase and hyaluronidase secreted by the bacterial cells. There is a massive polymorphonuclear cell response but lymphocytes may be sparse or absent. Of the fungal agents causing myositis, *Candida albicans* has been reported in patients with acute leukemia where the muscle inflammation may develop as part of a disseminated candidiasis. Actinomycosis and coccidiomycosis have also been implicated in muscle abscess or sinus formation; these are essentially focal disorders and do not appear to become generalized.

Protozoan Infections

Toxoplasmosis (causative agent *Toxoplasma gondii*) is a ubiquitous condition and up to one-third of the adult population world-wide is estimated to have antibodies to the parasite. However the condition is most often asymptomatic. Encysted bradyzoites are ingested with infected meat or through contact with cat feces and are liberated into the gastrointestinal tract. After a proliferative tachyzoite stage they may re-encyst in skeletal muscle of the human host. The cysts measure 10–100 μm in diameter and are easily identifiable by a strongly PAS-positive capsule and the thousands of bradyzoites in each cyst. Free tachyzoites can be identified in tissue sections by immunoperoxidase techniques. Infection by *Sarcocystis* in humans is much less common than *Toxoplasma* infections and can be differentiated from the latter histologically because of the nature of its cysts. These are divided into septa and are much larger than those in toxoplasmosis, measuring up to 2 mm in length. Often, the presence of sarcocysts in muscle appears to elicit no marked inflammatory response but patients may complain of local pain and tenderness. Most human *Sarcocystis* infections have occurred in Southeast Asia. The condition can be differentiated from toxoplasmosis by complement fixation tests.

African trypanosomiasis (sleeping sickness) and American trypanosomiasis (Chagas' disease) are caused by *Trypanosoma brucei* and *Trypanosoma cruzi*, respectively. Sleeping sickness results from being bitten by the insect vector, the tsetse fly. At first there is only local lymphadenitis but about a month later generalized malaise, fever, and systemic disease involving skeletal muscle is seen.

Symptoms next include dementia, tremor, and ataxia progressing most often to coma and death; survivors show severe chronic meningoencephalitis. The American form of trypanosomiasis (Chagas' disease) is a common health problem in Central and Southern America and is transmitted by reduviid bugs. Polymyositis, myocarditis and encephalomyelitis are constant features of the disorder but may be accompanied by more widespread visceral involvement. Infection can be confirmed by complement fixation tests but antibodies do not occur in the blood until about two months after infection.

Nematode and Cestode Infections

Infection with the nematode *Trichinella spiralis* used to be common in Europe and North America at the beginning of this century and is still not completely eradicated. Larvae are ingested with inadequately cooked or processed pork, penetrate the mucosa of the small intestine and reach skeletal muscle via the bloodstream. Myalgia and tenderness may be apparent from one to six weeks after infection. Ocular muscles may be particularly affected and a macular or petechial rash may occur. An eosinophil leukocytosis is present as an almost invariable feature. Muscle biopsy will usually show the nematode at various stages of development: larvae grow from about 100 μm to about 1 mm within muscle fibers which are grossly distended. A pseudocyst is formed, consisting of altered host tissue which becomes calcified, and this elicits a chronic inflammatory response. The larvae of the nematodes *Toxocara canis* or *Toxocara cati* can be ingested by children playing on lawns contaminated by dog or cat feces. In mild infections, persistent eosinophilia may be the only sign of the disorder, but symptomatic cases show myalgia, fever, hepatosplenomegaly, and pneumonitis. Eye involvement is not uncommon. The muscle lesions are visible to the naked eye as white patches 5–10 mm in diameter; microscopically they consist of intense granulomatous infiltration of necrotic muscle fibers with lymphocytes, eosinophils, and giant cells. Larvae, measuring about 400 μm × 20 μm can be found within the granulomata, but may be elusive. The cestode parasite *Taenia solium* causes cysticercosis, which is a major health problem in Africa, Asia, and Central and South America. *T. solium* is a hermaphrodite tapeworm which can grow to more than 20 ft in length. Ingestion of *Taenia* cysticerci in uncooked pork can lead to colonization of the jejunum by the tapeworm stage, but the larval cysticercus form can invade many tissues, including skeletal muscle, heart, central nervous system, and liver. The muscle involvement gives rise to striking pseudohypertrophy which is symmetrical and affects limbs and limb girdles. Biopsy shows numerous cysts up to 1 cm in diameter, each containing a single invaginated scolex complete with suckers and hooklets. Dead larvae behave as foreign bodies and elicit a granulomatous response and calcification. Cestodes of *Echinococcus* are prevalent in sheep-rearing areas, especially Australia and New Zealand. Dogs are the usual definitive host and humans

(the intermediate host) acquire infection with the larvae or hydatid stages by contamination of food with dog excreta. Both unilocular and multilocular cysts are found in infected muscle and provoke a marked granulomatous reaction in the zone between cyst and muscle. In these areas muscle fascicles are destroyed and replaced by fibrous connective tissue.

Other Inflammatory Muscle Disorders

It is important to bear in mind that disorders other than myositis, in the strict sense of the word, can give rise to very similar signs and symptoms. Most of these disorders involve muscle-associated connective tissue or vasculature and hence possible confusion with true myositis is not surprising. For example, *polyarteritis nodosa* is a disease of small and medium-sized arteries and their adjacent veins, but presentation may be with muscle weakness and myalgia. Muscle biopsy will reveal that an inflammatory reaction is certainly present but this is centered on arterioles and not on muscle fibers. Inflammatory cells include a predominance of neutrophil leukocytes and all zones of the vessel walls are involved. Fibrinoid necrosis is common and splitting or disruption of the internal elastic lamina may occur adjacent to inflammatory foci. The muscle tissue may show areas of infarction or may be morphologically normal apart from the perivascular inflammatory reaction. Wegener's granulomatosis is a necrotizing vasculitis involving small vessels and, as the name suggests, is characterized by granulomatous infiltration. In early stages of the disorder, neutrophils predominate but are then succeeded by mononuclear cells of the macrophage lineage, and some giant cells may be found. This latter feature is much more prevalent in temporal arteritis which is often referred to as giant cell arteritis. The disorder targets large and medium-sized arteries and, in addition to temporal arteries, the carotid, aortic, vertebral, and mesenteric arteries are vulnerable. Muscle arterioles may also be involved, with the inflammatory reaction extending into the adjacent perimysial connective tissue. Affected vessels show intimal hyperplasia, disruption of the internal elastic lamina and a granulomatous reaction with numerous Langhans-type giant cells.

Various syndromes associated with hypereosinophilia involve skeletal muscle. There is a rare form of polymyositis which is characterized by this feature (defined as exceeding 1,500 eosinophils/mm^3 for at least six months). Clinical presentation includes skin changes, heart and lung involvement, and peripheral neuropathy as well as proximal myopathy. The condition must be distinguished from trichinosis and other parasitic infections associated with hypereosinophilia. Muscle biopsy findings are interstitial and perivascular infiltrates in which eosinophils predominate but are accompanied by lymphocytes and plasma cells, and occasional muscle fiber necrosis. Fascitis may also be associated with hypereosinophilia (Shulman's syndrome). This condition is characterized by painful swelling of skin and soft tissues of trunk and extremities and weakness of limb muscles. Biopsy of muscle

and overlying fascia shows collagenous hypertrophy and granulomatous infiltration with lymphocytes, macrophages, and some eosinophils. The underlying muscle shows perifascicular atrophy. Some areas of fascia may show fibrinoid necrosis and fascial blood vessels usually involved in the obliteration of their lumina are often seen. A form of eosinophilic fascitis is also seen in patients medicated with L-tryptophan who present with a sclerodermalike syndrome.

Sarcoidosis is a disorder with worldwide distribution but blacks are about 10 times more likely to have the disease than whites. Many organ systems are involved and initial symptoms tend to originate in pulmonary and lymphatic systems. Muscle biopsy is often undertaken to substantiate the diagnosis and it is estimated that about half of all patients with sarcoidosis will show sarcoid granulomata in muscle. However clinical myopathy is present in less than one in 100 patients. In such cases swelling or hypertrophy of calf muscles may be noted, or there may be a proximal myopathy with tenderness, weakness, and wasting. Muscle biopsy findings are striking giant cell and granuloma formation in the perimysium. This may sometimes extend between perifascicular muscle fibers, which often atrophy as a result, but acute muscle necrosis is not seen. In addition to giant cells and other cells of the macrophage lineage, epithelioid cells are common, and sparsely distributed lymphocytes and polymorphonuclear leukocytes may also be present.

ENDOCRINE MYOPATHIES

Skeletal muscle represents the largest protein store in the human body and is important in the process of gluconeogenesis. Many endocrine disturbances accelerate protein catabolism, and in addition, can interfere with neural electrolyte concentrations and cause impairment of excitation-contraction coupling. Early descriptions of endocrine myopathies highlighted severe debilitating effects but, as prompt diagnosis and treatment have become the accepted norm, there is a tendency for mild cases to comprise a large majority of those seen in routine practice (Kissel and Mendell, 1992).

Thyroid Disorders

The terms hyperthyroidism and thyrotoxicosis are often used almost interchangeably but hyperthyroidism is more correctly reserved for conditions arising from true hyperfunction of the thyroid whereas thyrotoxicosis can result from diverse causes including secretion of excess TSH by the pituitary, toxic adenoma, and Graves' disease, an autoimmune condition of uncertain etiology. The clinical features of thyrotoxic myopathy are proximal muscle weakness with wasting of shoulder girdle muscles and scapular winging. Although thyrotoxicosis is more common in women, thyrotoxic myopathy shows an equal sex distribution suggesting that males are more susceptible. Rare cases show a more severe acute illness characterized by dysphagia, dysphonia, bulbar palsy, and profound generalized

muscle weakness but these patients may be cases of acute myasthenia gravis with accompanying thyrotoxicosis, since the prevalence of myasthenia gravis in thyrotoxic patients is about 30 times that seen in the population as a whole. Biopsy findings in thyrotoxic myopathy are often nonspecific with disseminated muscle fiber atrophy affecting both fiber types and a tendency for interstitial fibrous connective tissue to be increased. Minor mitochondrial changes may be seen but current evidence does not support previous suggestions that oxidative phosphorylation is uncoupled since respiratory chain function is normal. It is fairly well established, however, that excess thyroid hormone impairs the anabolic effect of insulin on the metabolism of amino acids and proteins and therefore results in decreased protein synthesis.

Thyroid-associated ophthalmopathy (TAO) is present in 90% of patients with the classical triad of Graves' disease (goiter, ophthalmopathy, dermopathy) but these features may follow independent courses and successful control of the hyperthyroidism improves TAO in less than 5% cases. Immunosuppression has been used since theories of the etiology of TAO include the presence of circulating antibodies to both thyroid and ocular muscle fibers, and of thyroglobulin-antithyroglobulin complexes with high affinity for extraocular muscles.

Thyrotoxic periodic paralysis (TPP) is another well-recognized complication of thyrotoxicosis, occurring predominantly in Asians. Susceptibility to TPP seems to be transmitted as an autosomal dominant trait and is associated with various HLA haplotypes (A2, BW22, AW19, B17, DRW8). Clinical features are indistinguishable from familial hypokalemic periodic paralysis (see p. 318). Muscle biopsy findings may be noncontributory though mild vacuolation may be seen. An abnormally high proportion of type 2 (fast-twitch) muscle fibers may occur, analogous to that seen in rats rendered hyperthyroid by excess T3 administration.

Hypothyroid myopathy occurs in about 30% of patients with hypothyroidism irrespective of its cause. Muscle pain, cramps, and stiffness may be seen, and are often exacerbated by cold weather. Pseudomyotonic features of delayed muscle contraction and relaxation are common. Myoedema (the mounding phenomenon) is due to the painless, electrically silent contracture produced on direct percussion. Muscle biopsy often shows a predominance of type 1 (slow-twitch) fibers, again analogous to that seen in experimental hypothyroidism (Figure 22). Muscle hypertrophy with weakness and slowness of movement occurs in the Debre-Semelaigne syndrome seen in severely hypothyroid children, and Hoffman's syndrome is a similar condition seen in adults with hypothyroidism, but is also accompanied by painful spasms.

Adrenal Disorders

Glucocorticoids are secreted by the cells of the zona fasciculata in the adrenal cortex. Hydrocortisone is the most important, and regulates intermediary metabo-

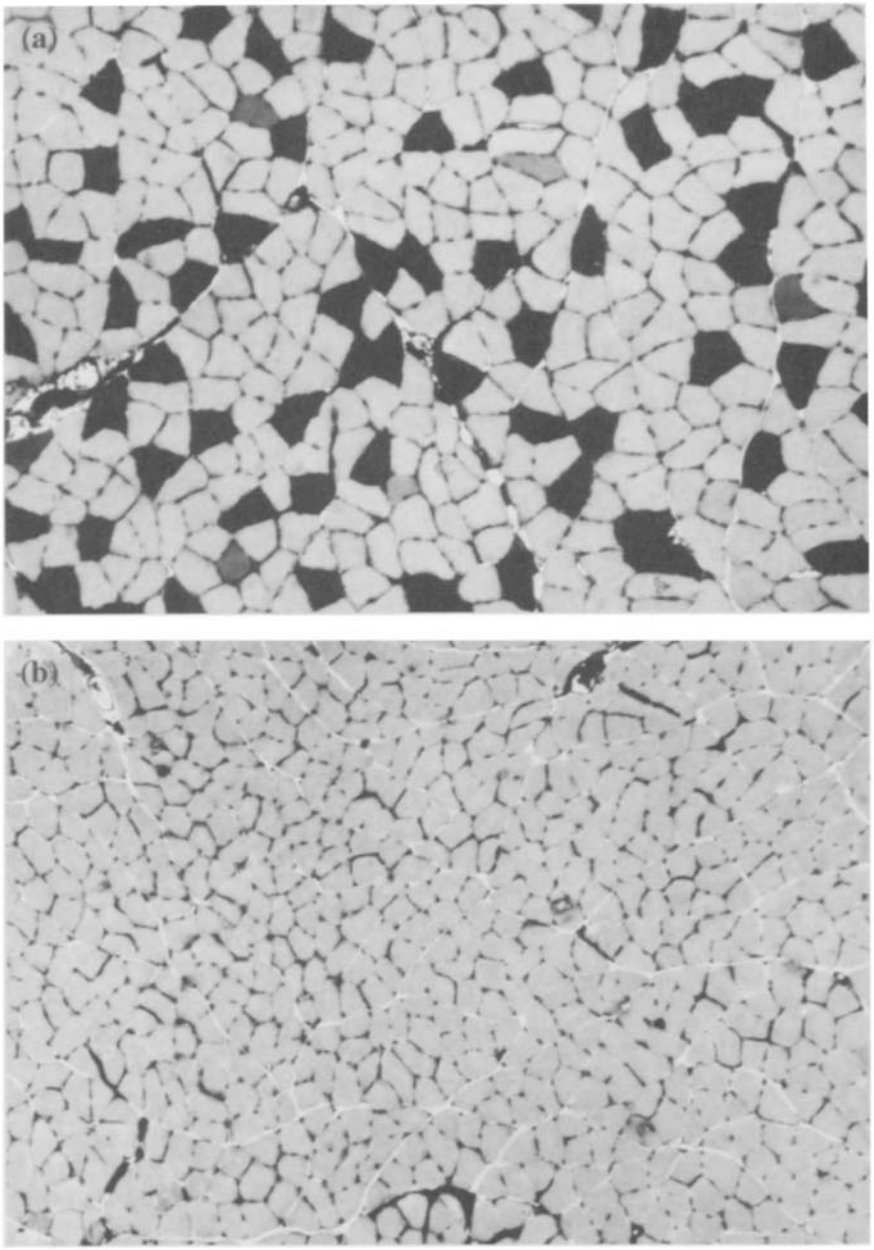


Figure 22. **(a)** Euthyroid (normal) rat muscle showing mixture of type 1 (slow-twitch) and type 2 (fast-twitch) fibers. **(b)** Hypothyroid rat muscle showing uniformly type 1 (slow-twitch) histochemical profile; myofibrillar ATPase after alkaline preincubation.

lism; its actions are mainly catabolic so that excessive production is often associated with myopathy whereas deficiency rarely is. Cushing's syndrome is characterized by glucocorticoid excess as a result of pituitary adenomas, and concomitant overproduction of ACTH; muscle weakness is an almost invariable feature of this syndrome. Proximal muscles are worse affected than distal and facial muscles are spared. Women are much more likely to develop myopathic features than men. This is also true of iatrogenic steroid myopathy which may result from treatment with any glucocorticoid but is more prevalent in patients who have been given fluorinated forms such as dexamethasone or triamcinolone. It has been shown that patients on alternate day regimes of prednisone (nonfluorinated) or on daily dosages of less than 30 mg are less likely to develop steroid myopathy.

Biopsy findings show disseminated muscle fiber atrophy which is confined to type 2 fibers, in many instances with type 2B (glycolytic) fibers most affected (Figure 23). Muscle necrosis is not seen, though at ultrastructural level focal myofibrillar disruption and myofilament loss may be evident. The muscle atrophy seems to be due to decreased protein synthesis, and at high doses, to increased catabolism. The reason for the selective effect on phasic, glycolytic fibers is not clear since, although steroids interfere with carbohydrate metabolism and oxidative capacity, there seems to be no overall effect on ATP levels. Nevertheless it has been

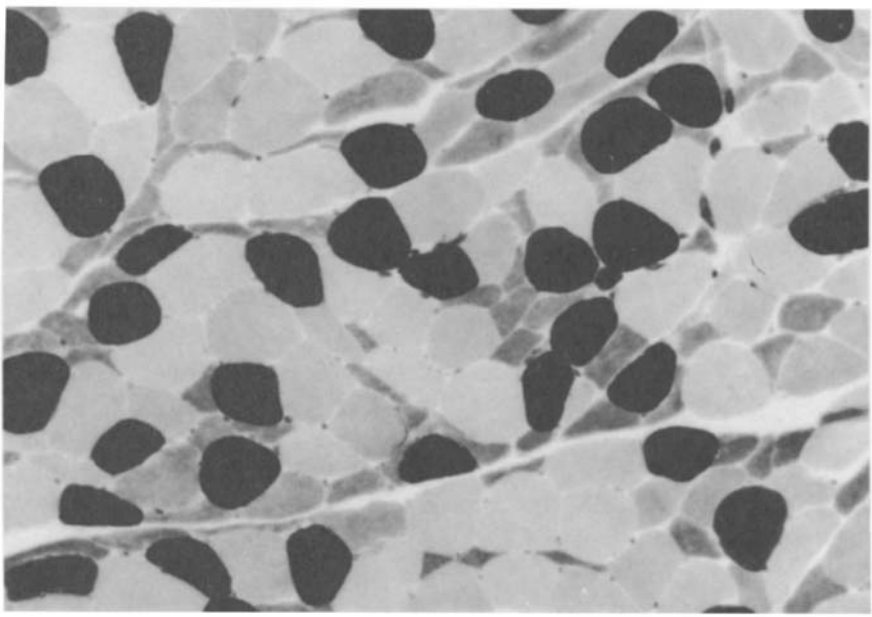


Figure 23. Steroid-induced atrophy affecting type 2B fibers selectively; these fibers show intermediate ATPase reactivity after pH 4.6 preincubation.

suggested that the resistance of type 1 muscle fibers to steroid atrophy may be linked to their ability to oxidize substrates derived from free fatty acids as alternative energy sources, whereas type 2 fibers are solely dependent on carbohydrate substrates. Androgens partly antagonize the catabolic effects of glucocorticoids which may explain the greater susceptibility of women to steroid-induced atrophy.

The major mineralocorticoid, aldosterone, is secreted by cells of the zona glomerulosa. Primary hyperaldosteronism (Conn's syndrome) is associated with potassium depletion which is, in turn, responsible for the observed neuromuscular abnormalities seen in the disorder. These are similar to those seen in hypokalemic periodic paralysis (PP), with episodic and severe exacerbations of fixed muscle weakness. Muscle biopsy shows occasional muscle necrosis and vacuoles; often these features are accompanied by tubular aggregates as in hypokalemic PP. All these changes can be attributed to the hypokalemia and not to excess aldosterone production per se.

Pituitary Disorders

The effects of ACTH excess are difficult to define and separate from accompanying glucocorticoid excess. However, in patients with bilateral adrenalectomy for Cushing's syndrome, muscle fatiguability is reported (Nelson's syndrome). This effect has been attributed to the extra-adrenal action of the high ACTH levels which become even higher after adrenalectomy. Fatiguability has also been recorded in acromegaly, due to excess growth hormone, though in early phases of the disorder increased muscle bulk is generally associated with increased strength. Muscle fiber diameter measurements are consistent with hypertrophy in early stages, followed by atrophy in later stages. The action of human growth hormone (GH) is still poorly understood. Some work has suggested that amino acid transport is stimulated by reduction in cAMP or by elevation in cGMP activity. The action of GH is also mediated via somatomedins which enhance amino acid and glucose transport and protein synthesis. Hypopituitarism in children causes dwarfing and poor muscle development unless corrected by administration of exogenous human GH. Patients show a gross reduction in muscle fiber number, but this deficit is reversed by GH treatment. Pituitary failure in adults (Simmond disease) causes weakness and fatiguability but muscle mass is preserved. Many of the problems in these patients can be attributed to the associated loss of thyroid and adrenal hormone function. Isolated loss of growth hormone in adults is rare but gives rise to fasting hypoglycemia and glucose intolerance. The latter finding may be due to loss of GH-stimulated glucose uptake, increased hepatic sensitivity to glucagon and decreased insulin secretion. Fasting hypoglycemia occurs in spite of low insulin levels and suggests that supersensitivity to insulin develops.

Parathyroid Disorders

Hypoparathyroidism is most often encountered as a result of surgical removal of the parathyroid glands although it can be associated with thymic agenesis. Whatever the cause of the hypoparathyroidism, it is characterized by hypocalcemia, hyperphosphatemia, and sometimes also by hypomagnesemia. Overt myopathy due to the hypocalcemia is rare but parethesias due to hyperexcitability of nerve fibers and generalized tetany are more common. Axonal hyperexcitability is linked to decreased serum calcium and magnesium since a deficit of these ions lowers the action potential threshold. Parathyroid hormone (PTH) regulates blood calcium levels by inducing bone resorption, augmenting renal calcium absorption and phosphorus excretion and by mediating in the interconversion of vitamin D (cholecalciferol) metabolites. As a result hyperparathyroidism is associated with disorders such as osteomalacia and renal failure. Whereas the major target tissues of PTH are bone and kidney, it has been suggested that excessive PTH could elevate cytoplasmic calcium levels in skeletal muscle to a concentration which would result in the activation of neutral protease. PTH also increases cAMP levels in skeletal muscle and it has been postulated that this may reduce the calcium sensitivity of the contractile system since cAMP-dependent phosphorylation of troponin-I diminishes the affinity for calcium ions of the calcium sensor troponin-C.

Primary hyperparathyroidism occurs as a result of hyperplasia or the occurrence of adenoma. Secondary hyperparathyroidism may result from renal failure because of the associated phosphate retention, resistance to the metabolic actions of PTH, or impaired vitamin D metabolism. The last-mentioned factor is primarily responsible for the development of osteomalacia. Muscle symptoms are much more common in patients with osteomalacia than in primary hyperparathyroidism. Muscle biopsy has revealed disseminated atrophy, sometimes confined to type 2 fibers, but in other cases involving both fiber types. Clinical features of osteomalacic myopathy are proximal limb weakness and associated bone pain; the condition responds well to treatment with vitamin D.

Pancreatic Disorders

The neuromuscular complications of diabetes mellitus are most often neuropathic in origin, with distal sensorimotor polyneuropathies being the most common. In addition, ischemic infarction of skeletal muscle may occur due to occlusive vascular disease, with small and medium-sized arterioles particularly affected. This occurs in poorly-controlled diabetes and affects thigh muscles in most cases. In acute stages, muscle biopsy findings are those of widespread muscle necrosis, edema, and phagocytic cell infiltration. Muscle regeneration may be incomplete and increased fibrous connective tissue may replace lost muscle tissue.

Insulin is a powerful anabolic hormone but it is unlikely that insulin deficiency causes skeletal muscle atrophy by direct action on muscle fibers (as opposed to neurogenic atrophy) except in chronic untreated cases. There is however a close parallel between the catabolic states induced by glucocorticoid excess and by insulin deficiency. Moreover, impaired insulin action is implicated in other endocrine myopathies as a contributory cause of muscle wasting. Both acromegaly and thyrotoxicosis are associated with insulin resistance due to a postreceptor defect, and secondary hyperparathyroidism due to hypophosphatemia also gives rise to insulin insensitivity.

DRUG-INDUCED AND TOXIC MYOPATHIES

Drug-induced and toxic myopathies are probably more common than is generally realized, but the distinctive feature of these conditions is that muscle damage is usually resolved rapidly once the causal agent is removed. The specific causes of damage vary considerably as does the presence of pain and discomfort.

Drug-induced myopathies may arise as a result of the normal use of therapeutic agents or the abuse of drugs such as steroids and narcotics, and the damage may be focal or disseminated. Focal muscle damage arises as the result of the direct inoculation of myotoxic agents into the skeletal muscle. The process of inoculation itself is probably of no consequence; experimental studies have shown that apart from occasional needle-track damage caused by the insertion of the needle, sterile vehicle (water, 0.9% w/v NaCl for example) causes no untoward effect even after repeated inoculation. Some drugs that may routinely be injected intramuscularly are known to cause focal necrosis, and particularly problematic are chloroquine, opiates, and chlorpromazine. In normal practice, there is little cause to use this route of administration over long periods of time, and the focal necrosis is of little long-term consequence. Long-term abuse of anabolic steroids, opiates and antibiotics, however, can lead to the formation of local scar tissue and disabling contractures and abscess formation. The drugs themselves, the repeated introduction of low grade infection, occasional intramuscular hemorrhage and repeated needle track damage all probably contribute to the problem.

Systemic muscle damage, often associated with pain and discomfort, is a well known problem associated with specific drugs such as epsilon amino caproic acid (EACA), clofibrate, emetine, vincristine, chloroquine, D-penicillamine, and anabolic steroids. Notes on each of these drugs follow, but for a detailed discussion of drug-induced muscle damage refer to Argov and Mastaglia (1988) and Harris and Blain (1990).

EACA: muscle damage is sudden in onset and though rare is well recognized. It is uncommon when treatment involves the use of low doses below 18 g day^{-1}. It usually occurs several weeks after the commencement of treatment. The drug can cause a severe necrotizing myopathy. Capillary occlusion and fibrin deposition in

the small blood vessels may also be pathogenic. Muscle damage can lead to a severe rhabdomyolysis. Recovery is rapid when treatment is stopped.

Clofibrate: causes a necrotizing myopathy, particularly in patients with renal failure, nephrotic syndrome or hypothyroidism. The myopathy is painful and myokymia of unknown origin is sometimes present. The mechanism of damage is not known, but p-chlorophenol is a major metabolite of clofibrate and p-chlorophenol is a particularly potent uncoupler of cellular oxidative phosphorylation and disrupts the fluidity of lipid membranes. Muscle damage is repaired rapidly on the cessation of treatment.

Emetine: emetine is still used at high doses for the treatment of patients with severe amebiasis. Muscle damage is uncommon but when it does occur can be a severe generalized necrotizing myopathy. The outcome is, at times, fatal, especially when an emetine-induced cardiomyopathy is also present. Despite the suggestion that there may be neuritic changes, there is no evidence that emetine damages peripheral nerve. The myopathy is usually painful but reversible. The mechanism of action of emetine is unknown.

Vincristine: a severe polyneuropathy is a particularly common side effect, but in some patients a painful autophagic myopathy is also present. Abnormal lysosomal complexes are also formed which function as virtually indigestible lipid accumulations.

Chloroquine: causes a vacuolar myopathy with the formation of lysosomal complexes. There is conflicting evidence on the selective involvement of particular fiber-types. The myopathy is reversible.

Corticosteroids: a chronic painless myopathy associated with the long-term use of corticosteroids is a particularly common example of drug-induced muscle disorder. It is almost certain that mild cases are overlooked because steroids are so frequently used to treat inflammatory myopathies such as polymyositis. Fluorinated steroids are particularly frequently implicated, and the incidence of drug-induced muscle disease is dose and time-related. The presence of muscle weakness can even complicate topical steroid therapy. Corticosteroid-induced myopathy is mediated via intramuscular cytosolic steroid receptors. The steroid-receptor complexes inhibit protein synthesis and interfere with oxidative phosphorylation. The myopathy is associated with vacuolar changes in muscle, and the accumulation of cytoplasmic glycogen and mitochondrial aggregations.

D-penicillamine: a well-recognized complication of the use of D-penicillamine is a myasthenic syndrome. Most reported cases of myasthenic syndrome are women and the ocular muscles appear to be particularly commonly affected at early stages of the abnormality. The onset of the myasthenic syndrome can range between a few months and many years, and is not always reversible. There is a clear immunological basis to this problem, because it is associated with raised anti-acetylcholine receptor antibodies circulating in the serum. Antibody levels fall when the drug is withdrawn, and this, in turn, is associated with clinical improvement. How the

immunological reaction is stimulated is not clear. There is no specific association, in affected patients, with thymoma or HLA status and it has been suggested that the drug is particularly effective at unmasking subclinical myasthenia gravis. Use of D-penicillamine is also associated with the onset of drug-induced polymyositis or dermatomyositis, and several other muscle diseases of immunological origins. Like the myasthenic syndrome, penicillamine-induced polymyositis is usually reversed with the withdrawal of the drug, but this is not always the case, and aggressive therapy of the problem may be required.

The inclusion of the drugs listed above is not intended to be complete, but to give an indication of the extent of the problem, the range of drugs involved and the variety of muscle damage caused. It must be emphasized however, that a very large number of drugs have the capacity to cause muscle problems and that at times the problems may be severe. Acute rhabdomyolysis, for example, is rare but can be fatal because of the potential complication of renal failure. Drugs implicated in rhabdomyolysis include alcohol, heroin, methadone, barbiturates, phencyclidine, vasopressin, and amphotericin. The list is not exhaustive.

Neuromuscular paralysis can also be a significant problem with a number of antibiotics, and can be sufficiently severe as to be life-threatening. As a general guideline, there is sufficient evidence to suggest that when an acute muscular problem of unknown origin is detected, a drug-induced disorder should always be considered.

Toxic myopathies are more common than is often supposed. The most severe and common of the myopathies caused by natural toxins and xenobiotics are probably those caused by snake-bite. Reid (1961) provided the first detailed study of venom induced muscle damage when he described a painful necrotizing myopathy in victims of the sea-snake *Enhydrina schistosa* in Malaysia. Other snakes known to cause such problems are the Australian elapid snakes *Notechis scutatus*, *Oxyuranus scutellatus*, and *Oxyuranus microlepidotus*, the cobra *Naja naja* (esp. subspecies *kaouthia* and *leucodira*), the spitting cobra *Naja nigricollis*, the rattlesnakes *Crotalus durissus terrificus* and *Crotalus horridus atricaudatus*, and the Russell's viper, *Daboia russelli* (esp. subspecies *pulchella*) (Harris and Cullen, 1991).

Laboratory studies have shown that numerous other venomous snakes have the capacity to cause muscle damage, but probably most are only rarely myopathic in man. Similarly, the venoms of many spiders, scorpions, bees, wasps, hornets, and ants cause muscle damage at high doses, but are very rarely implicated in such complications in man. The most myotoxic of the venoms are usually rich in phospholipases A_2 (PLA_2). These enzymes are not necessarily toxic, but many possess very high myotoxic potential. The best studied of the myopathic PLA_2 enzymes is notexin, isolated from the venom of the Australian tiger snake (see Harris, 1990). The toxins, and venoms containing them, cause a rapid necrotizing myopathy that may be localized to the site of inoculation or bite, but may be

widespread, leading to severe weakness, rhabdomyolysis, and kidney failure. The mechanism of action appears to involve the hydrolysis of the plasma membrane and the destruction of key cytoskeletal proteins leading to the breakdown of the myofibrils. Provided there is no extensive hemorrhage, regeneration of the affected tissue is usually rapid and successful.

A second group of myotoxic toxins, found almost exclusively in the venoms of cobras, are the cytotoxins (often called cobratoxins, cytolysins, cardiotoxins, or direct lytic factors). These, rather than phospholipases, are almost certainly the primary cause of muscle damage following bites by cobras. Their mechanism of action is not properly known, but it is certainly the case that their action is potentiated by the presence of phospholipases in the venom, even if the phospholipases concerned are not, themselves, myotoxic. The cytotoxins of cobra venom possess no hydrolytic activity of any kind.

A third group of myotoxic factors are very short polypeptides, devoid of hydrolytic activity. These toxins, found in the venom of a few species of North American rattlesnakes, cause a dilatation of sarcoplasmic reticulum and can cause severe muscle damage.

Although it is possible to identify a group of particularly toxic components in venom, it should be noted that the venoms are complex mixtures of components, many of which are synergistic. Muscle damage is particularly severe if myotoxic activity is combined with hemorrhagic activity. In this case, muscle regeneration is impaired, because the regenerating tissue is rendered anoxic at a time of intense metabolic activity.

Bacterial and viral myositis is well recognized as a clinical entity by muscle pathologists. The viruses most commonly involved appear to be the Coxsackie viruses, the arboviruses, influenza virus, and HIV, but the mechanism whereby the viral infection gives rise to the myositic syndrome is not known. A detailed discussion of such problems is presented later on pages 333–334.

MANAGEMENT OF MUSCLE DISEASE

The first step in the management of muscle disease is achieving a completely accurate diagnosis, as described in the previous sections. Where an iatrogenic or primary condition may be identified, then removal or cure of the primary condition is an obvious part of the management, and may be very successful in relieving the problem. Drug treatment may be appropriate in the management of inflammatory myopathies, and steroids and immunosuppressants are used with varying success in their management. Supplementation with carnitine or other cofactors in the respiratory chain may be used in various metabolic muscle diseases. The drugs of choice in myasthenia gravis are the anticholinesterases, sometimes with the addition of steroids and immunosuppressants. Thymectomy and plasmapharesis are other treatment options in this condition. However, the overall role of drugs or other specific therapeutic measures in the treatment of muscle disease is limited.

Among the disorders discussed in this section, many are inherited and progressive, lead to significant disability, and no specific treatment is available. The management of these conditions depends very much on a holistic and multidisciplinary approach. For example, Duchenne muscular dystrophy (DMD) is a severe and progressive disorder, commencing with problems in early childhood. Despite attempts to find a satisfactory drug treatment for DMD, none has gained widespread acceptance, and prospects for therapy based on the gene or protein itself are still poor (see pp. 349–352). The management of DMD can be divided into several phases depending on the stage of the disease as follows:

1. Ambulant with few problems with mobility. This is the stage which boys with DMD typically pass through between the ages of about two and six years. At this time, the boys will be able to walk and attempt to run and climb stairs, but will be noticeably slower and less proficient than their peers. There are relatively few physical needs at this stage. Physiotherapy is limited to the introduction of simple exercises to maintain a full range of movement at the ankles and hips, with the introduction of ankle-foot orthoses to wear at night towards the end of this stage. Attention to diet may prevent future problems with obesity, and early advice about the need to keep weight under control is appropriate. The major needs of the family are, therefore, to get support through the time of the diagnosis and the realization of its implications. It has been shown that the way the diagnosis is handled can have a major bearing on the future adjustment of the family, and psychological back-up from specialized support staff, such as the Family Care Officers funded in Britain by the Muscular Dystrophy Group, can help considerably with the family's ability to cope with the devastation which the diagnosis inevitably brings. Communication with other families in a similar situation, either informally or through support groups, may also help with adjustment. Genetic issues need to be addressed, and counseling offered to all family members of child-bearing age who may be at risk.
2. Increasing difficulty with ambulation. This is the stage which boys pass through where their physical problems begin to impinge much more on their ability to maintain a normal lifestyle (on average between the ages of six and 8–9 years). Progressive weakness leads to an increasing tendency to fall, more pronounced slowing, and the inability to climb stairs or rise from the floor without assistance. The effect of increasing weakness is often compounded by the development of contractures at the hips and ankles. Regular physiotherapy is vital at this stage to minimize contractures, and the use of nocturnal ankle splinting is also crucial if tenotomies are to be avoided. Some centers advocate fairly widespread surgery on contractures at the hips, ankles, and iliotibial bands to prolong independent mobility, but this approach does not have universal acceptance.

3. Loss of independent mobility. Once independent walking becomes intolerably difficult, it is possible to prolong ambulation by the provision of knee-ankle-foot orthoses, or long-leg calipers. These calipers act by allowing the leg to be maintained in full extension, reducing the tendency to fall, and permitting ambulation by using the residual strength in the trunk muscles to throw each leg forward one at a time. In order for calipers to be successful, contractures must be minimal, so tenotomies at the ankles may be necessary to allow calipers to be used, and the avoidance of such surgery is a major impetus for regular physiotherapy and the use of nocturnal splinting of the ankles. The use of calipers itself requires a considerable commitment from the boy, his parents, and carers (especially physiotherapists) and some families may feel unable to accept it, especially if the boy is of limited intelligence. The advantages of allowing some form of walking to continue are in continuing to prevent contractures and the development of scoliosis. There may also be a psychological benefit to the patient and his parents in being able to have a gradual transition from fully independent mobility to the stage of being completely wheelchair-dependent. During this stage of the disease, the family will have to cope with the visible evidence that the deterioration predicted for their child is really going to happen. Practical issues, such as the suitability of current housing for a wheelchair user, access to schools, and the prospect of opting for special schooling also need to be addressed at this time, and the family helped through each of these decisions. Liaison with social workers, occupational therapists, and medical and paramedical staff needs to be closely maintained at this stage to ensure that all of the current and future needs of the family are addressed.
4. Wheelchair dependent but healthy. The time for which calipers can usefully be managed (if used at all) is variable, depending on a number of factors including the degree of muscle weakness and rate of deterioration, the motivation of the family, and intelligence of the child. Usually within 1–2 years mobility with calipers becomes impossible, and the boy becomes confined to a wheelchair, initially manual, but ultimately electrically powered. Maintaining standing posture for an hour or so a day is a useful physiotherapy measure at this stage, either by standing in long-leg calipers, or through the use of some sort of standing frame. Physiotherapy continues to be a mainstay of management at this stage, because of the tendency to develop contractures around all of the major joints, and hydrotherapy may be an important part of this physiotherapy input. For a period of maybe a couple of years after confinement to a wheelchair, the disease may appear to be fairly stable, and management should revolve around the prediction of problems and minimizing them wherever possible. Monitoring of scoliosis and respiratory function are the other major needs at this stage.

5. Wheelchair-dependent with complications. A number of major decisions face the patient and his parents as the disease progresses. Scoliosis is an almost invariable complication of DMD in the late stages, and can be managed by either surgery, with spinal instrumentation, or bracing which can act to delay the progression of scoliosis but does not prevent it altogether. Respiratory failure is another almost inevitable feature of the late stages of DMD, and the decision of whether to opt for assisted ventilation, at home, is another important option which needs to be considered, and the decision, whichever way it is made, fully supported. Providing comfortable seating and beds can become more difficult as contractures worsen and may require the input of specialized occupational therapists. Preparation for the inevitable death of the patient is a painful process for all concerned.

From this description of the clinical management of patients with DMD a number of general principles emerge, which are applicable to patients with other primary muscle diseases. First, there should be a multidisciplinary approach to the management of these patients. Second, psychological support of the family is a major factor in the well-being of the patient. Third, where a disease is progressive and that progression is predictable, problems should be anticipated and planned for so that their impact is lessened as much as possible. Fourth, where these diseases are inherited, genetic counseling is an essential part of their management, and this is an ongoing commitment, as new family members approach childbearing age, and as potentially new tests can be applied to different disorders. For these reasons, genetic registers are often kept of families with inherited disorders such as the muscular dystrophies, so that at-risk individuals can easily be traced when new developments are made. An extension of this principle is that DNA information should always be stored from patients with inherited muscle diseases, even if only limited information on the genetic basis of the particular disorder is available at the time of making the diagnosis, as relatives may present seeking genetic advice at a later date, by which stage direct DNA analysis may have become available.

FUTURE PROSPECTS FOR THE TREATMENT OF INHERITED MUSCLE DISEASES

Successful treatment for any condition depends, to a very large extent, on an understanding of exactly what is going wrong and why. As the genes which are responsible for inherited disorders of muscle are identified, our understanding of the roles that defective gene and protein expression play in the development of the disease increases. Possible therapeutic approaches may then suggest themselves, but the problem of generating premature expectations and excitement about a "cure" for these conditions must be avoided.

Isolation of genes. On the basis of "prevention is better than cure," the process of identifying and isolating the affected gene in any inherited disorder almost

always provides an improvement in genetic counseling. As soon as evidence of linkage to a particular locus is established, adjacent polymorphic DNA markers can be used to improve the identification of affected family members and carriers. This is currently the case for facioscapulohumeral MD and some of the forms of limb girdle MD and SCARMD. The possibility of recombination places a limit on the accuracy of this manner of gene tracking but the use of two markers which map on either side of the disease gene locus (flanking markers) improves the accuracy. Recent developments in the field of molecular genetics have speeded up the process of isolating the cDNA from the disease gene once the approximate location has been mapped.

Characterization of gene expression and improved diagnostic techniques. Analysis of DNA from patients indicates what type of gene defect is associated with the expression of a particular disease. This might be an exonic deletion or duplication (which would directly affect the molecular structure of the protein product), an expansion or contraction of DNA within an untranslated region (indirectly affecting the regulation of gene transcription and translation), or a point mutation changing one base for another (which can alter an individual amino acid, alter a splice site or generate a new stop codon). When the characteristic genetic lesion is known, screening for this specific defect by PCR analysis or Southern blotting provides an immediate improvement in the accuracy of diagnosis. This is now the case for Duchenne, Becker, and myotonic dystrophy. Altered gene expression usually results in an obvious alteration in the protein expressed (size or abundance) and this may also be used to improve the accuracy of diagnosis.

Changing the genes present. If a disease is caused by the inheritance of a single faulty gene then the only true cure would be to replace the faulty gene with a good copy before the action of the gene comes into effect. The ethical problems that surround human genetic engineering are enormous, however, since what starts as a method to avoid diseases could so easily be abused and turned into a method to generate tailored human beings with characteristics which are perceived to be advantageous (beauty, intelligence, strength, etc.). Nevertheless, a great deal of scientific activity currently surrounds the genetic approach to treating several major inherited human diseases including cystic fibrosis, severe immunodeficiency diseases, and hypercholesterolemia. Most of this work is in animal models of the human diseases, but the pressure to commence human trials is enormous and hierarchies of committees have been established to consider the issues raised by protocols for human gene therapy.

Certain characteristics of skeletal muscle fibers are particularly relevant when considering the future of gene therapy in muscle diseases. Skeletal muscle fibers are large syncytia containing thousands of postmitotic myonuclei, each of which expresses the same set of genes. The postmitotic nature of the myonuclei implies that in mature muscle fibers, cell division cannot play a role in spreading the transferred gene to required locations (i.e., whole muscles). On the other hand, once

a transferred gene is stably established in a myonucleus, it cannot be diluted out by subsequent mitosis. If, however, new genes are to be transferred in the nuclei of donor myoblasts (muscle precursor cells), these can only be introduced into mature fibers following necrosis and regeneration. During this time the host myoblasts vigorously proliferate so that the transferred genes could be diluted out as long as the myoblasts proliferate. Since gene products tend to spread in a limited manner from myonuclei (nuclear domains), if a gene encodes a structural protein like dystrophin, almost every myonucleus would need to acquire a copy of the transferred gene for optimum therapeutic results. At the current time DMD is the only inherited muscle disease for which forms of gene therapy are being actively pursued. Two approaches have been taken to the replacing of defective dystrophin genes with healthy ones: the transfer of intact healthy myoblasts into dystrophin-deficient muscle, and the introduction of the isolated gene or part of the gene directly into muscle fibers (reviewed in Karpati and Acsadi, 1993).

Myoblast transfer therapy. One of the early claims made in support of myoblast transfer therapy was that it would provide a treatment for all diseases of muscle, regardless of whether the gene responsible for individual disorders had been identified or not. Since the healthy myoblasts contain good copies of all the genes required for normal muscle function, it was argued that this approach would be suitable for any condition. In reality, however, this dream has not come true. Early studies looked quite promising in that the introduction of healthy myoblasts into the dystrophin-deficient *mdx* mouse showed that donor myoblasts fused and dystrophin-positive fibers which were a mosaic of donor and host cells could be seen around the site of injection. Unfortunately the myoblasts did not migrate efficiently and immunorejection was a problem. Nevertheless, the transfer of myoblasts into human subjects was undertaken in a limited number of centers. The reports of these studies are highly variable. One center claims a significant improvement in isometric force generation, but another four centers reported more disappointing results: no improvement in muscle strength, limited dystrophin synthesis, and adverse immunological reactions. It is now generally considered that this approach has not proved to be as successful as once hoped (Morgan and Watt, 1993) and it is doubtful that cell-based therapies have a real future.

Gene therapy. Healthy dystrophin genes have been introduced into *mdx* muscle in three ways: direct injection of DNA, or via retroviral or adenoviral vectors which infect the host muscle cells. Again the main problem has been one of inefficient incorporation. Direct injection of pure plasmid DNA resulted in only 1% of the surrounding fibers showing any dystrophin labeling. One of the hurdles associated with viral vectors was the huge size of the dystrophin gene and to counteract that a minigene version has been employed which is approximately half the size of the full length version. The rationale for thinking that the protein expressed by the minigene had all the domains required for a reasonable level of function was that the exons missing were deleted in a BMD patient who was quite mildly affected

by his disease. Transfer of the minigene with a retroviral vector proved to be successful in that dystrophin expression could be observed up to nine months after a single injection. Unfortunately, the level of incorporation was still low with only 6% of fibers showing dystrophin labeling. Retroviruses can only replicate in actively dividing cells which effectively limits their usefulness to regenerating fibers. In contrast, adenoviral vectors are effective in nondividing cells, with transfer of the minigene to *mdx* mice resulting in dystrophin labeling in 5–50% of the fibers in injected muscles, up to six months after a single injection. The expression of dystrophin also appeared to limit the degenerative cycles which characterize dystrophin-deficient muscle. The topic is discussed in Dickson and Dunckley (1993).

Upregulation of alternative genes. Once the amino acid sequence of the affected protein is known, it is possible to compare it with the sequences of other known proteins on computer data bases, to try and locate a protein or family of proteins with which the newly identified protein has homology. This provides clues to the possible function of the protein in normal muscle. This, in turn, may suggest a therapeutic approach, since the function of the affected protein might be provided in a different form. For example, utrophin is homologous to dystrophin and they both bind to the same, or similar, complex of integral membrane glycoproteins. In normal muscle utrophin is restricted to the region of the sarcolemma which underlies neuromuscular junctions, but in dystrophin-deficient muscle utrophin is expressed along the extrajunctional membrane in a similar manner to dystrophin. Studies are now in progress to see if utrophin could substitute for dystrophin in a functional sense. If the utrophin gene could be upregulated in a controlled manner this might provide a novel approach to therapeutic strategies in the future.

Conclusions. For the inherited muscle diseases as a group the future prospects for treatment are not wildly exciting, in that the tracking down and cloning individual disease genes is the first priority. An improvement in diagnosis and counseling is a more achievable aim than novel therapeutic methods. For DMD the possibilities for therapy are slightly greater but caution must be exercised, and expectations of gene transfer therapy must be realistic. The problems of getting persistent, widespread expression of dystrophin are still not completely solved and may well require a lot more work before human trials could be contemplated.

ACKNOWLEDGMENTS

The authors acknowledge the work of our colleagues, clinical and nonclinical, whose data have shaped our views. Our work is funded by grants from the Muscular Dystrophy Group of Great Britain, Action Research, the Wellcome Trust, the MRC, and the Royal Society.

REFERENCES

Anderson, S., Bankier, A.T., Barrel, B.G., et al. (1981). Sequence and organization of the human mitochondrial genome. Nature (Lond.) 290, 457–465.

Argov, Z., & Mastaglia, F.L. (1988). Drug-induced neuromuscular disorders in man. In: Disorders of Voluntary Muscle (Walton, J.N., ed.), pp. 981–1014, Churchill-Livingstone, Edinburgh.

Barchi, R.L. (1988). The pathophysiology of excitation in skeletal muscle. In: Disorders of Voluntary Muscle (Walton, J.N., ed.), pp. 372–389, Churchill-Livingstone, Edinburgh.

Barwick, D., & Fawcett, P.R.W. (1988). The clinical physiology of neuromuscular disease. In: Disorders of Voluntary Muscle (Walton, J.N., ed.), pp. 1015–1080, Churchill-Livingstone, Edinburgh.

Bushby, K., & Gardner-Medwin, D. (1993). The clinical, genetic and dystrophin characteristics of Becker muscular dystrophy. 1. Natural History. J. Neurol. 240, 98–104.

Dickson, G., & Dunckley, M. (1993). Human dystrophin gene transfer: Genetic correction of dystrophin deficiency. In: Molecular and Cell Biology of Muscular Dystrophy (Partridge, T., ed.), pp. 283–302, Chapman & Hall, London.

Di Mauro, S., Bonilla, E., Hays, A.P., & Ricci, E. (1992). Skeletal muscle storage diseases resulting from errors in carbohydrate and fatty acid metabolism. In: Skeletal Muscle Pathology (Mastaglia, F.L., & Walton, J.N., eds.), pp. 425–451, Churchill-Livingstone, Edinburgh.

Di Mauro, S., & Papadimitriou, A. (1986). Carnitine palmitoyl transferase (CPT) deficiency. In: Myology (Engel, A.G., & Banker, B.Q., eds.), pp. 1697–1708, McGraw-Hill, New York.

Emery, A.E.H. (1989). Emery-Dreifuss muscular dystrophy and other related disorders. Br. Med. Bull. 45, 772–787.

Emery, A.E.H. (1993). Duchenne Muscular Dystrophy. Oxford University Press, Oxford.

Engel, A.G. (1988). Metabolic and endocrine myopathies. In: Disorders of Voluntary Muscle (Walton, J.N., ed.), pp. 811–868, Churchill-Livingstone, Edinburgh.

Fardeau, M. (1992). Congenital myopathies. In: Skeletal Muscle Pathology (Mastaglia, F. L., & Walton, J.N., eds.), pp. 237–281, Churchill-Livingstone, Edinburgh.

Goebel, H.H., & Lenard, H.G. (1992). Congenital myopathies. In: Handbook of Clinical Neurology, Vol. 18 (62): Myopathies (Rowland, L.P., & Di Mauro, S., eds.), pp. 331–367, Elsevier Science Publishers, New York.

Gomez, M.R. (1986). Motor neuron diseases in children. In: Myology Volume II (Engel, A.G., & Barker, B.Q., eds.), pp. 1993–1012, McGraw Hill, New York.

Harper, P.S., Harley, H., Reardon, W., & Shaw, D.J. (1992). Review Article: Anticipation in myotonic dystrophy: New light on an old problem. Am. J. Hum. Genet. 51, 10–16.

Harris, J.B. (1990). Phospholipases in snake venoms and their effects on nerve and muscle. In: Snake Toxins (Harvey, A.L., ed.), pp. 91–129, Pergamon Press, Oxford.

Harris, J.B., & Blain, P.G. (1990). Introduction to the toxicology of muscle. In: Muscle Metabolism (Harris, J.B., & Turnbull, D.M., eds.), pp. 665–686, Balliere-Tindall, London.

Harris, J.B., & Cullen, M.J. (1991). Muscle necrosis caused by snake venoms and toxins. Electron Microsc. Rev. 3, 183–211.

Harris, J.B., & Turnbull, D.M. (eds.). (1990), Muscle Metabolism. Balliere-Tindall, London.

Hoffman, E.P., & Kunkel, L.M. (1989). Dystrophin abnormalities in Duchenne/Becker muscular dystrophy. Neuron 2, 1019–1029.

Karpati, G., & Ascadi, G. (1993). The potential for gene therapy in Duchenne muscular dystrophy and other genetic muscle diseases. Muscle Nerve 16, 1141–1153.

Kissell, J.T., & Mendell, J.R. (1992). The endocrine myopathies. In: Handbook of Clinical Neurology, Vol. 18 (62): Myopathies (Rowland, L.P., & Di Mauro, S. eds.), pp. 527–551, Elsevier Science Publishers, New York.

Lunt, P. (1994). Report of the sixth international workshop on facioscapulohumeral muscular dystrophy; San Francisco 11 November 1992; and current guidelines for clinical application of DNA rearrangements at locus D4S810. Neuromusc. Disord. 4, 83–86.

Morgan, J.E., & Watt, D.J. (1993). Myoblast transplantation in inherited myopathies. In: Molecular and Cell Biology of Muscular Dystrophy (Partidge, T., ed.), pp. 303–332, Chapman & Hall, London.

Morgan-Hughes, J.A. (1992). Mitochondrial diseases. In: Skeletal Muscle Pathology (Mastaglia, F.L., & Walton, J.N., eds.), pp. 367–424, Churchill-Livingstone, Edinburgh.

Mulder, D.W. (1986). Motor neuron disease in adults. In: Myology Volume II (Engel, A.G., & Banker, B.Q., eds.), pp. 2013–2029, McGraw Hill, New York.

Nicholson, L.V.B., Johnson, M.A., Bushby, K.M.D., Gardner-Medwin, D., Curtis, A., Ginjaar, I.B., Den Dunnen, J.T., Welch, J., Butler, T.J., Bakker, E,. Van Ommen, G.J.B., & Harris, J.B. (1993). Integrated study of 100 patients with Xp21-linked muscular dystrophy using clinical, genetic, immunochemical and histopathological data. J. Med. Genet. 30, 728–775.

Poulton, J., & Land, J. (1990). The genetics of metabolic disorders of muscle. In: Muscle Metabolism (Harris, J.B., & Turnbull, D.M., eds.), pp. 621–664, Balliere-Tindall, London.

Reid, H.A. (1961). Myoglobinuria and sea-snake poisoning. Br. Med. J. 1, 1284–1289.

Rüdel, R., & Ricker, R. (1985). The primary periodic paralysis. Trends in Neurosci. 8, 467–470.

Zeviani, M., Servidei, S., Gellera, C., Bertini, E., Di Mauro, S., & Di Donato, S. (1989). An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop. Nature (Lond.) 339, 309–311.

Chapter 10

Principles of Medical Cryobiology: The Freezing of Living Cells, Tissues, and Organs

PETER MAZUR

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 355–384.

ISBN: 1-55938-808-0

INTRODUCTION

Cryobiology is concerned with the responses of living cells to freezing, and medical cryobiology is primarily concerned with the use of low temperature to preserve cells and tissues for clinical applications. The applications and implications are varied. Many relate to transplantation. Thus, efforts are under way to establish fetal tissue banks to provide sources of cells to treat diabetes, Parkinson's disease, and liver failure (Cohen and Jonsen, 1993). A hospital in Philadelphia is offering a service to freeze umbilical cord blood to provide a potential source of stem cells should the newborn require them some time in the future. Another area is assisted reproduction. Cryopreservation of sperm cells is now required to ensure that the donor is HIV free before the sperm are used in artificial insemination (Henry et al., 1993). In embryo transfer procedures, more embryos are often collected than can be immediately transferred. The extras can be frozen for potential use at a later time using procedures originally developed for mouse embryos (Whittingham et al., 1972; Trounson and Mohr, 1983). Other forms of assisted reproduction involve the use of unfertilized ova and, here again, the ability to cryopreserve the cells would have important logistic advantages. Immunological rejection continues to be a major limitation in the transplantation of cells, tissues, and organs. Cryopreservation would ameliorate the problem by permitting the establishment of banks of a wide variety of immunologically typed tissues and organs. The problem is that most organized tissues and organs cannot now be successfully frozen (Karow and Pegg, 1981), although some, like skin and cornea, are candidates for success in the near future (Kearney et al., 1990; Rich and Armitage, 1992; Madden et al., 1993).

There are indirect implications as well. Malaria remains one of the world's scourges striking some 200,000,000 people per year and killing some 2,000,000. Unfortunately, the *Anopheles* mosquito has become increasingly resistant to permissible insecticides and the malaria parasite increasingly resistant to antibiotics. It has been discovered, however, that a number of steps in the development and transmission of the *Plasmodium* parasite are under genetic control of the mosquito. Further, the belief is growing that transgenic mosquitos can be created that block one or more of these steps and thus block transmission, and that when these mutants are introduced into the wild, the inability to transmit will spread throughout the indigenous population (Aldous, 1993). Creating the desired transgenic forms, however, requires the creation and maintenance of hundreds of mutant lines, a problem that will be exceedingly difficult and expensive without a way to cryopreserve mosquito embryos. *Anopheles* embryos cannot currently be cryopreserved, but recent success in the cryopreservation of embryos of the fruit fly *Drosophila* (Mazur et al., 1992b) provides optimism that success will be achieved.

Some of these applications are available now because procedures exist for cryobiologically preserving the cells and tissues involved. Others remain potentials for the future because the cells and tissues currently do not survive attempts to freeze them. Understanding the difference requires an understanding of the major

physical-chemical events that occur in and around cells during exposure to subzero temperatures and the responses of the cells to these events. A discussion of these underlying fundamentals is the chief purpose of this chapter.

STOPPING BIOLOGICAL TIME

Contrary to the usual impression, the challenge to cells during freezing is not their ability to endure storage at very low temperatures; rather it is the lethality of an intermediate zone of temperature (~ –15 °C to –60 °C) that a cell must traverse twice: once during cooling and once during warming. No thermally driven reactions occur in aqueous systems at liquid nitrogen temperatures (–196 °C), the refrigerant commonly used for low temperature storage. And few, if any, occur below –130 °C because the viscosity of water rises to enormous values by that temperature. The only reactions that can occur in frozen aqueous systems at –196 °C are photophysical events such as the formation of free radicals and the production of breaks in macromolecules as a direct result of "hits" by background ionizing radiation or cosmic rays. But damage from this source only becomes significant if cells are stored for centuries (Mazur, 1984)

The implications and applications of frozen cells stem, therefore, from the fact that storage temperatures of –130 °C to –196 °C effectively stop biological time. The problem is how to get cells down to these temperatures and later return them to normal temperatures without killing them.

LOW TEMPERATURES VS. FREEZING

Lowered temperature is a self-evident concomitant of freezing, thus raising the question of the extent to which freezing injury is a result of reduced temperature rather than of the ice formation that accompanies reduced temperatures. The evidence is that in most cells lowered temperature per se has little relevance to freezing injury. The two can be experimentally separated down to about –15 °C to –20 °C by the ability to keep cells and their surrounding medium unfrozen (supercooled) to these temperatures. This is not to say that there are not adverse effects of lowered temperature, but they generally occur over a much longer time scale than is involved in freezing. There are exceptions, however, some of which have been reviewed by Morris (1987). Pig sperm, for example, is highly sensitive to chilling and may be totally inactivated by cooling to 0 °C even before freezing is initiated (Watson and Plummer, 1985). *Drosophila* embryos also become increasingly sensitive to lowered temperature in the absence of ice formation (Mazur et al., 1992a). Chilling injury of this sort, however, has a very different genesis than freezing injury, and appears to involve membrane lipid phase transformations and the impairment of active transport pumps. It will not be treated further in this chapter (see Chapter 11).

THE FATE OF INTRACELLULAR WATER DURING FREEZING

To survive freezing, a cell must be cooled in such a way that it contains little or no freezable water by the time it reaches the temperature at which internal ice formation becomes possible. Above that temperature, the plasma membrane is a barrier to the movement of ice crystals into the cytoplasm. The critical factor is the cooling rate. Even in the presence of external ice, most cells remain unfrozen, and hence, supercooled, 10 to 30 degrees below their actual freezing point (–0.5 °C in mammalian cells). Supercooled cell water has a higher chemical potential than that of the water and ice in the external medium, and as a consequence, it tends to flow out of the cells osmotically and freeze externally (Figure 1).

If cooling is sufficiently slow, the cell is able to lose water rapidly enough by exosmosis to concentrate intracellular solutes sufficiently to eliminate supercooling and maintain the chemical potential of intracellular water in equilibrium with that of extracellular water. The result is that the cell dehydrates and does not freeze intracellularly. But, if the cell is cooled too rapidly, it is not able to lose water fast enough to maintain equilibrium; it becomes increasingly supercooled and eventually attains equilibrium by freezing intracellularly. Whether the cell undergoes osmotic dehydration during slow cooling or intracellular freezing during rapid

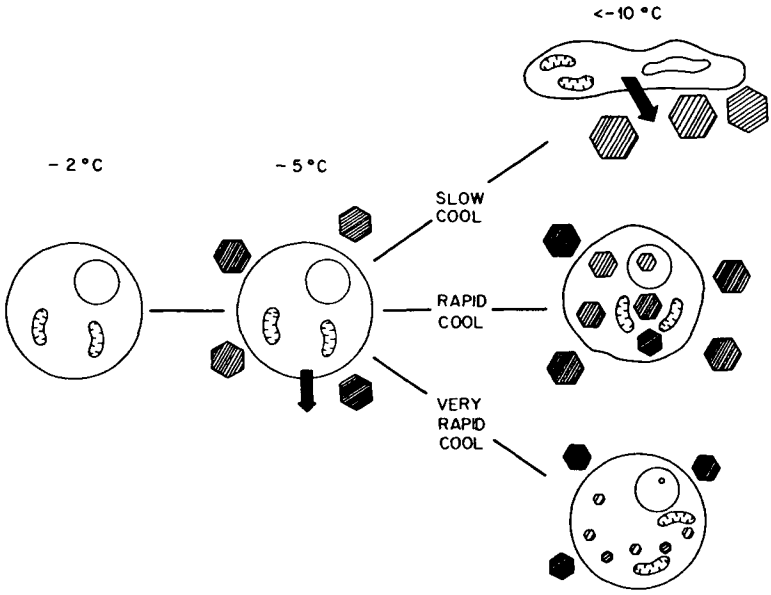


Figure 1. Schematic of physical events in cells during freezing. The cross-hatched hexagons represent ice crystals. (From Mazur, 1977a.)

cooling is critical to the cell's survival. In general, intracellular freezing is lethal, apparently because it mechanically disrupts cellular membranes.

The dependence of the rate of exosmosis of water on cooling rate and the likelihood of intracellular freezing is described by four coupled equations (Mazur, 1963; Mazur et al., 1984). The first describes the increase in the difference between the chemical potentials of internal water and external ice that will develop with falling temperature for cells with a given amount of supercooled water. The second equation relates the rate of water efflux to that difference in chemical potential. The important cell parameter in that equation is the water permeability (L_p) of the plasma membrane. The third equation describes the temperature dependence of L_P in terms of its activation energy (E_a). And the fourth equation relates temperature and time, i.e., it expresses the cooling rate. These equations may be numerically solved to compute the extent of supercooling in cells as a function of cooling rate provided one knows or can estimate L_p, E_a, and the surface-to-volume ratio of the cell or cell system. The results of such calculations are commonly expressed as plots of the water content of a cell as a function of temperature relative to the normal or isotonic water volume. Figure 2, for example, shows computed water loss curves for mouse ova. The calculated extent to which a cell becomes supercooled is the number of degrees that any given curve is displaced to the right of the equilibrium curve at a given subzero temperature. In the example shown, embryos cooled

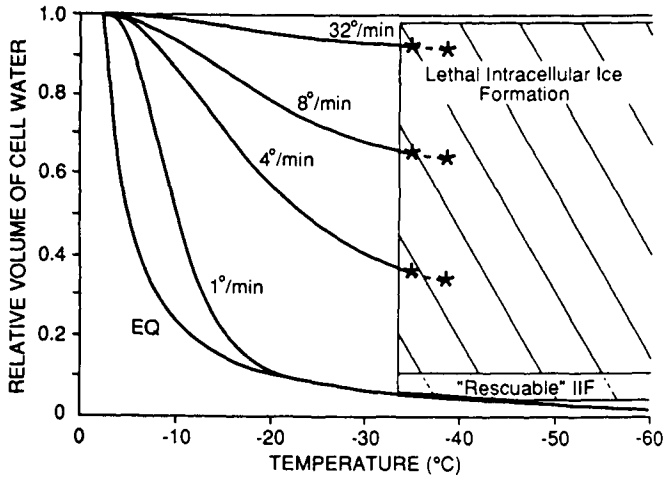


Figure 2. Computed kinetics of water loss from mouse ova cooled at 1 °C to 32 °C/min in 1M DMSO. The curve labeled EQ shows the water content that ova have to maintain to remain in equilibrium with extracellular ice. If ova or embryos contain more than equilibrium amounts of water when they cool to below –30 °C, they will undergo intracellular freezing. Usually such freezing is lethal, but if the quantity of ice is small, some internally frozen cells can be rescued by rapid warming. (From Mazur, 1990.)

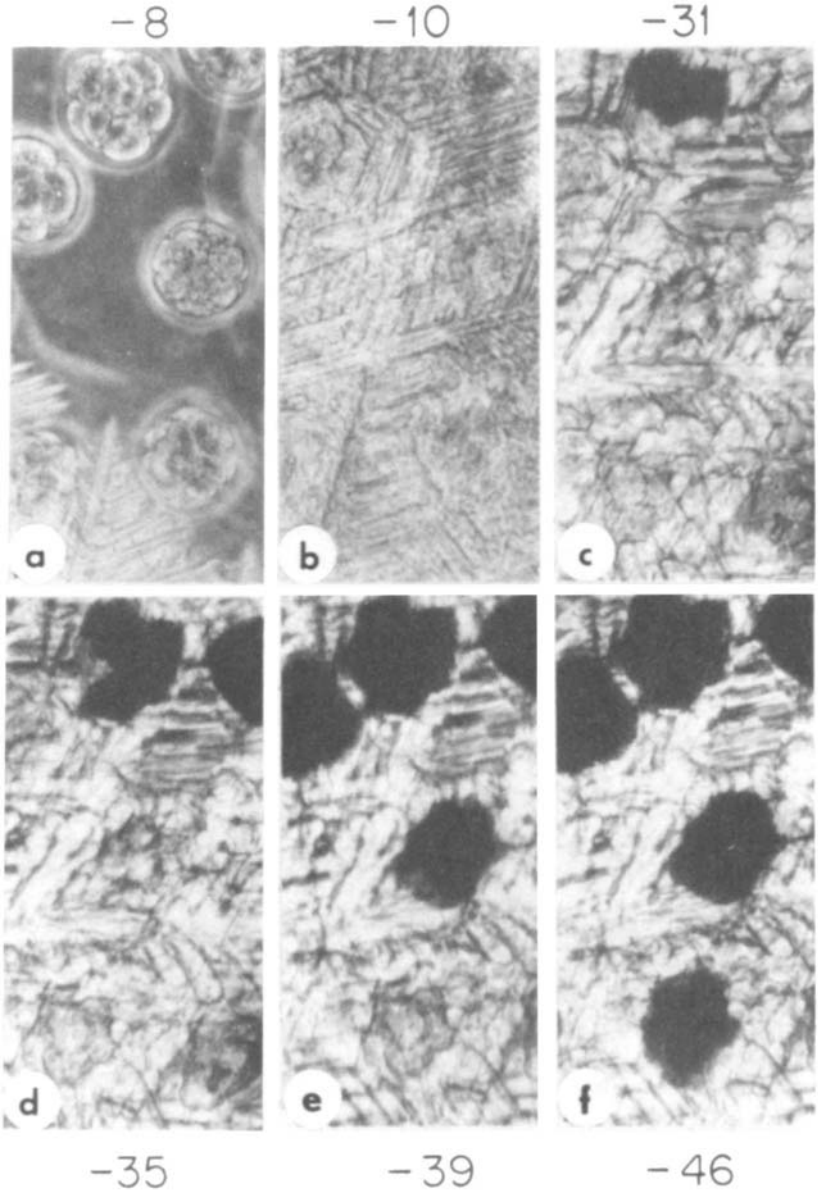


Figure 3. Intracellular freezing of 8-cell mouse embryos cooled at 20 °C/min in 2 M DMSO. The black "flashing" occurring in cells at –31 °C to –46 °C is characteristic of intracellular ice formation, and is caused by the scattering of light by many small highly branched ice crystals. (Modified from Rall et al., 1983.)

at 4 °C/min will have lost about half their water by –20 °C and the remaining water will be supercooled 15 degrees.

Three biological parameters have special influence on the position and shape of these curves; namely, water permeability (L_p) and its activation energy (E_a), and the size of the cell, or more properly its surface to volume ratio (A/V). An increase in L_p produces the same effect as a comparable decrease in cooling rate. The effect of cell size is the opposite. An increase in diameter reduces the cooling rate required to produce a given probability of intracellular freezing. Changes in these two parameters shift the positions of the curves. Changes in the activation energy, E_a, have a major effect on the shape of the curves. As E_a increases, the curves flatten at lower temperatures.

To avoid intracellular freezing, the cell must dehydrate to close to its equilibrium value before it has cooled to the temperature at which the plasma membrane loses its ability to block ice crystal growth. As illustrated in Figure 3, intracellular nucleation occurs between about –33 °C and –45 °C in mouse embryos that are supercooled in one to two molar solutions of the cryobiologically protective solutes dimethyl sulfoxide (DMSO) or glycerol. Superimposing this range of nucleation temperatures on the computed water loss curves in Figure 2 leads to the prediction that embryos cooled at less than 2 °C/min should not freeze internally since they will have dehydrated to equilibrium prior to cooling to the nucleation temperatures. But embryos cooled at 4 °C/min or faster ought to freeze internally since they will still contain appreciable supercooled water as they enter the temperature zone for nucleation. These predictions have been found accurate for mouse and hamster ova (Mazur et al., 1984; Shabana and McGrath, 1988). The set of equations governing the kinetics of water loss have also been found to predict the cooling rate depend-

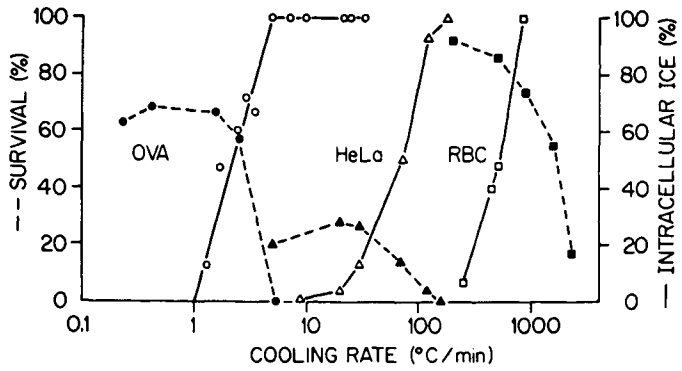


Figure 4. Percentage survival (dashed lines) vs. percentage cells undergoing intracellular freezing (solid lines) in three mammalian cells frozen at various rates to –20 °C (HeLa) or to –78 °C to –196 °C (ova and red blood cells [RBC]). (Modified from Leibo, 1977. Sources of data for individual curves are given there.)

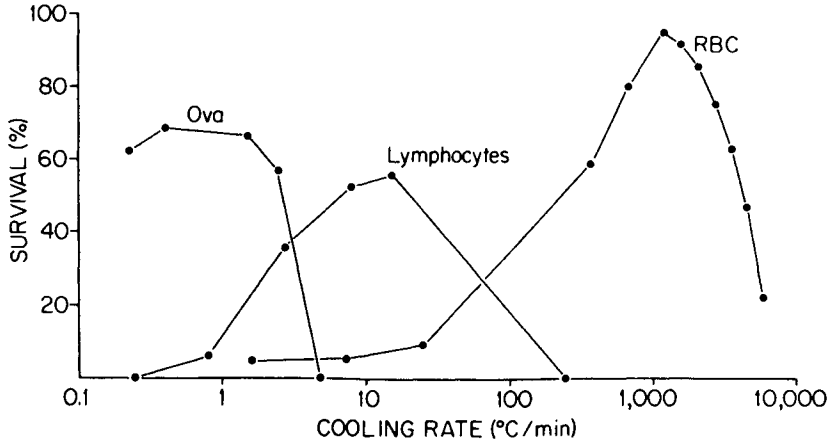


Figure 5. Survival vs. cooling rate of three cell types frozen to −196 °C in 0.7 to 1M DMSO. (From Leibo, 1981.)

ence of intracellular freezing in yeast, HeLa, Chinese hamster tissue culture cells, human lymphocytes, red blood cells, and rye protoplasts (Mazur, 1984).

Why the concern about the relation between cooling rate and intracellular freezing? It is because there is a close correlation between the cooling rates that produce intracellular ice and the cooling rates that produce killing (Figure 4). However, as shown in Figure 4, the numerical value of the critical cooling rate can differ widely in different cells because the L_p, E_a, and surface-to-volume ratio differ widely in different cells.

In cells that can survive freezing, the usual finding is that plots of survival versus cooling rate form inverted "U's" as exemplified in Figure 5. Survival is maximum at some critical rate, a rate that can vary a thousandfold in different cell types. The drop in survival at supraoptimal rates results from intracellular freezing as just discussed. And, with exceptions to be noted later, high survival demands that the cooling rate be low enough to avoid such internal freezing. Although such low rates are necessary for high survival, they are not sufficient, for cooling at suboptimal rates can also be highly injurious.

THE NATURE AND ELIMINATION OF SLOW FREEZING INJURY

In 1950, Audrey Smith and Christopher Polge in Mill Hill, England, discovered that compounds like glycerol in molar concentrations can reduce or eliminate slow freezing injury in sperm and other cells. The story of how they made the discovery is a nice illustration of serendipity. Although repeatedly failing to preserve the

viability of sperm by freezing in sucrose solutions, Smith and Polge made still another attempt. To their astonishment, the viability this time was spectacularly good. What had happened, they discovered, was that the refrigerator containing the sucrose solutions also held a bottle of a glycerol-containing solution used in the histological fixation of sperm. Subsequent analysis showed that that bottle was the source of the solution used in the successful experiment. Apparently, the labels had fallen off both bottles and had been inadvertently switched when someone reapplied them prior to the successful experiment!

Subsequent studies by them and others showed that the effect of solutes like glycerol and dimethyl sulfoxide (DMSO) was on the left arm of the inverted "U" in Figure 3. As illustrated in Figure 6 for human red blood cells and Figure 7 for mouse marrow stem cells, the higher the concentration of glycerol, the less are the damaging effects of freezing at cooling rates below the optimum. In most cases, to be effective, these cryoprotectants (as they came to be termed) had to be present inside the cells as well as outside.

The above findings provided an approach for the cryobiological preservation of a wide variety of single cells; namely, suspend cells in 1–2 M concentrations of glycerol or DMSO and cool them at rates low enough to preclude intracellular freezing. A commonly used cooling rate was, and is, 1 °C/min. Left unanswered, however, was the cause of slow freezing injury and the basis by which glycerol and DMSO protected against that injury.

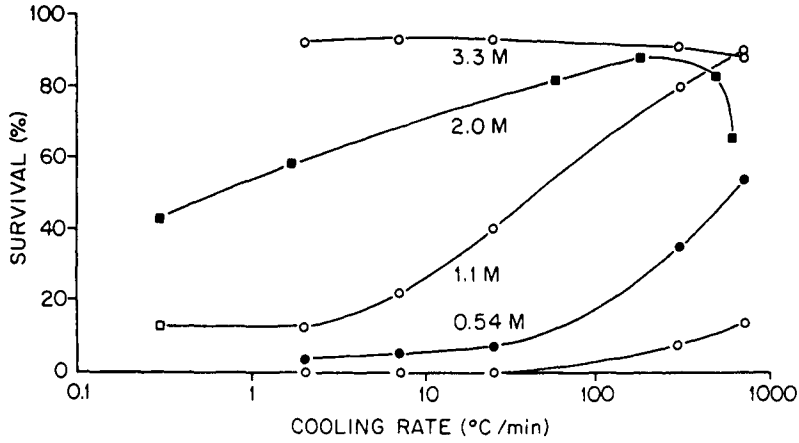


Figure 6. Survival as a function of cooling rate for human red blood cells suspended in physiological saline containing the indicated concentrations of glycerol, cooled to −196 °C, and warmed rapidly. Glycerol was absent in the bottom curve. (The data are from Morris and Farrant (1972) and Mazur and Miller (1976). The figure is modified from Souzu and Mazur (1978).)

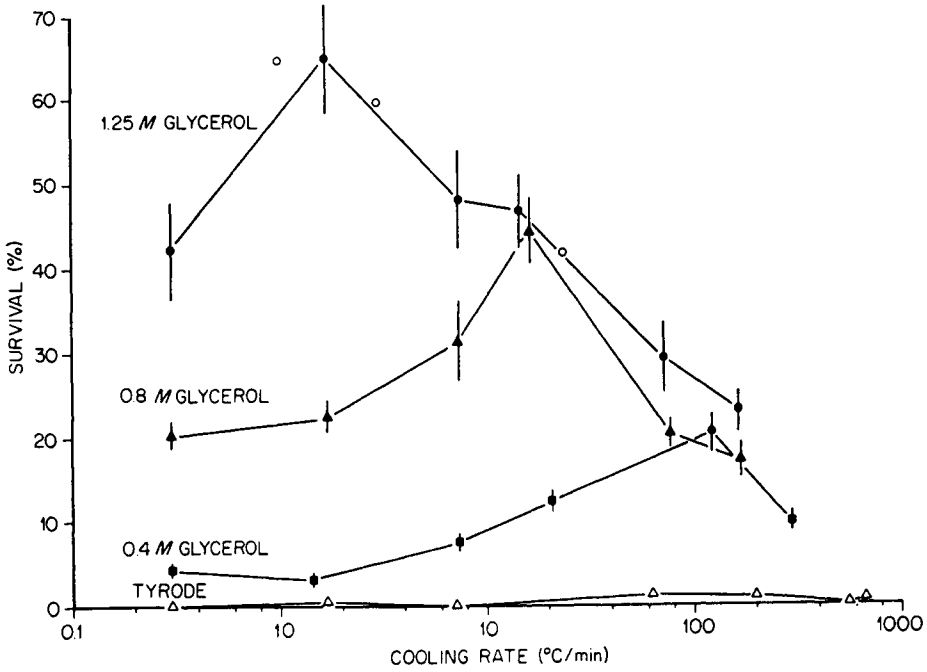


Figure 7. Survival of frozen-thawed mouse marrow stem cells (based on colony-forming activity) as a function of the concentration of glycerol in the tyrode saline suspending medium. (From Leibo et al., 1970.)

Figure 8 depicts a photographic sequence in the slow freezing of frog erythrocytes in saline. The cells are sequestered in unfrozen channels between the growing ice crystals. These crystals grow by pulling pure water out of these channels, and as a consequence, the solute concentration in the channels rises and the channels progressively shrink in size. The rise in solute concentration in turn causes the cells to progressively shrink osmotically as depicted in the upper right branch of Figure 1. Until recently, most cryobiologists have ascribed slow freezing injury to either the rise in solute (electrolyte) concentrations in the channels or to the consequent cell shrinkage. We, however, have proposed that a considerable portion of the damage is due to the reduction in the size of the unfrozen channels and the consequent increase in cell-ice contacts or cell-cell contacts (Mazur, 1984).

Concentration of Extracellular Solutes During Freezing

As ice forms outside the cell, the concentration of extracellular solute in the residual unfrozen medium increases according to the relation $M^e = \phi v m^e = \phi v n_2 / V^e = \Delta T / 1.86$, where M^e is by definition the external osmolality (i.e., the osmotically

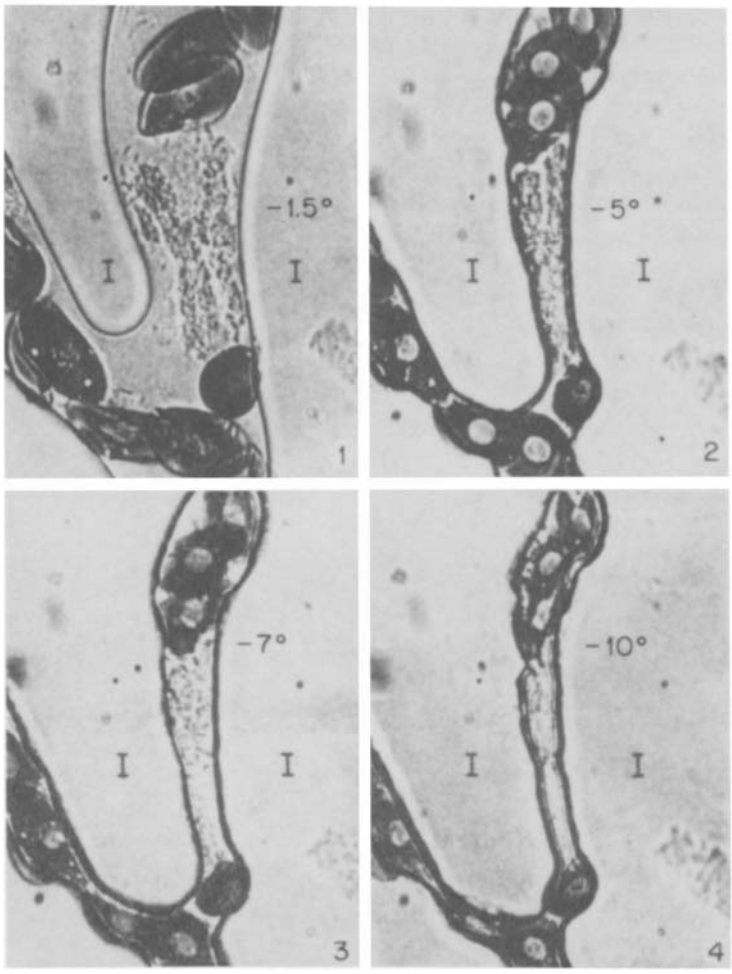


Figure 8. Photomicrographs of frog erythrocytes in serum during the course of slow freezing from –1.5 °C to –10 °C. Note that the cells are confined to the channels of unfrozen solution between the ice crystals (I), that the channels decrease in diameter with decreasing temperature, and that the cells shrink. (From Rapatz et al., 1966.)

effective concentration), ϕ is the osmotic coefficient, ν the number of species into which the solute dissociates, n_2 the moles of solute, m^e the molality (moles/kg.H_2O), V^e the volume of extracellular water, and ΔT is the number of degrees below 0 °C. The value 1.86 is the molal freezing-point depression constant for water. If the solution contains more than one solute, the second and third terms become $\Sigma\phi\nu m^e$ and $\Sigma\phi\nu n_2/V^e$.

In partly frozen solutions, M^e is independent both of the nature of the solutes and of their total concentration prior to freezing. At constant pressure it is dependent only on temperature. For a solution containing a single solute, this is also roughly true of the molality m^e (It is only roughly true because ϕ changes somewhat with concentration.)

A consequence of these considerations is that the total osmolal concentration of solutes in the unfrozen portion of a solution at a given temperature is not influenced by the addition of solutes like glycerol or DMSO. For example, the unfrozen portions of both an isotonic salt solution (0.3 osmolal) and an isotonic salt solution containing 1 M glycerol (1.4 osmol/kg H_2O, total) will have the same total osmolal concentration at −10 °C, namely, 10/1.86 or 5.4 osmol/kg.

The presence of cryoprotectant, however, does reduce the concentrations of salt at the given temperature according to the relation $M^e_{NaCl} = M^{e'}_{NaCl}/(1 + \mathbf{R})$, where

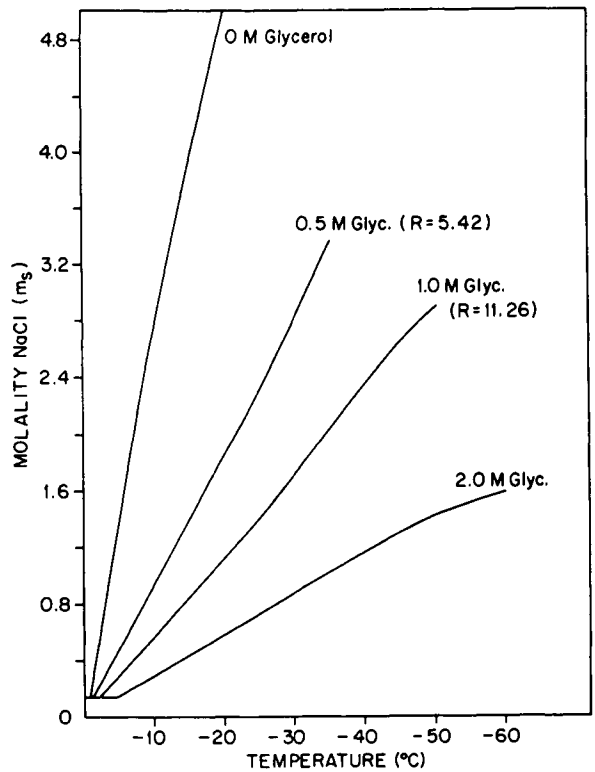


Figure 9. Molality of NaCl (m_s) in unfrozen portions of glycerol-NaCl-H_2O solutions at various subzero temperatures. Curves apply to any glycerol-NaCl solution with stated **R** values, where **R** is %wt glycerol/%wt NaCl. (From Mazur et al., 1981.)

M^{e}_{NaCl} and $M^{e'}_{NaCl}$ are osmolal concentrations of electrolyte in the presence and absence of additive and **R** is osmolal ratio of cryoprotectant to electrolyte.

The changes in osmotic coefficients with temperature and concentration make it difficult to solve the above equations accurately, but accurate determinations of the composition and relative amounts of the concentrated liquid and ice can be made from phase diagrams which are plots of the freezing points of solutions versus their concentration. From these, it is possible to determine the exact NaCl concentration at any temperature. Examples are shown in Figure 9 for solutions of 0 to 2.0 M glycerol in 0.15 M NaCl. This figure nicely illustrates how the presence of glycerol reduces the concentration of NaCl in the residual unfrozen solution.

Effect of Slow Freezing on Cell Survival

The injurious effects of slow freezing are well exemplified by the human red blood cell (Figure 10). First, we see that the survival of slowly frozen cells decreases with decreasing temperature. Second, as the concentration of glycerol is increased from 0 to 1.75 M, the decrease in survival occurs at progressively lower temperatures. Third, as the glycerol concentration rises above 1.75 M, survival remains high even at very low temperatures, a fact that has provided the basis for existing banks of frozen human erythrocytes.

If we combine these data on survival vs. temperature with the data on NaCl concentration vs. temperature in Figure 9, we obtain the results shown in Figure 11A for two of the lower concentrations of glycerol. The NaCl concentration in the

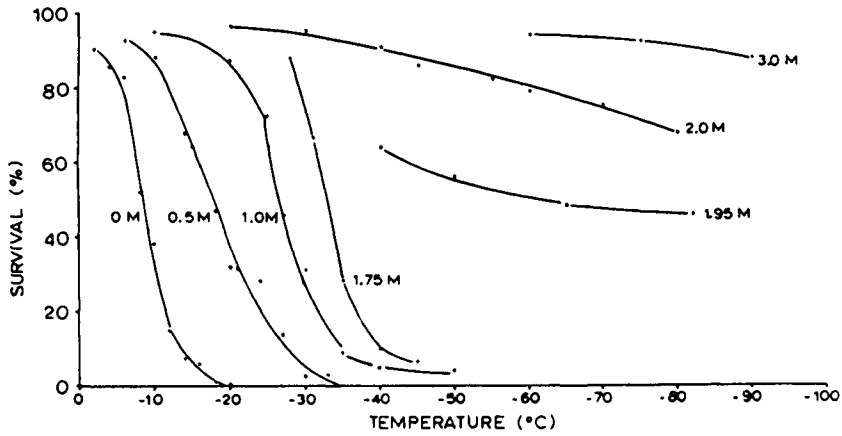


Figure 10. Survival (% unhemolyzed cells) of frozen-thawed human red blood cells as a function of the concentration of glycerol in the medium (buffered saline) and as a function of temperature. Freezing was slow (1.7 °C/min); thawing was rapid. (From Mazur, 1977b, based on data of Souzu and Mazur, 1978.)

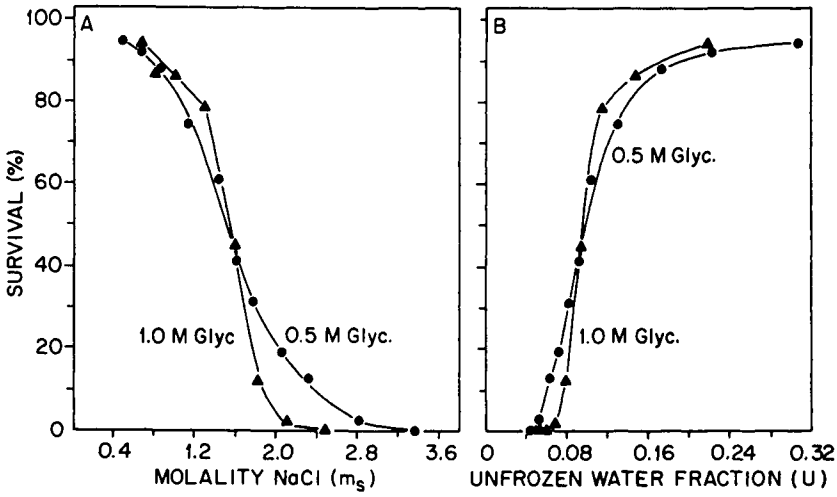


Figure 11. A: Survival of human red blood cells as a function of the molality of NaCl (m_s) to which they are exposed after being frozen at 1.7 °C/min to various subzero temperatures while suspended in solutions of 0.5 (●) or 1.0 (▲) M glycerol in isotonic NaCl. Thawing was rapid. **B**: Survival as a function of the fraction of water that remains unfrozen in the solutions. (From Mazur and Rigopoulos, 1983.)

residual unfrozen solution appears to exert major effects; survivals drop from above 80% for salt concentrations of less than 1 molal to below 20% for salt concentrations of 2 molal or more.

James Lovelock in Mill Hill, England (1953a,b), was the first to show results of the sort indicated in Figure 11A. He also showed that red cells exposed at 0 °C in the absence of freezing to the same concentrations of NaCl to which they are exposed upon freezing to specified temperatures undergo similar extents of hemolysis. From these results he hypothesized that injury during slow freezing is the result of the increase in the concentration of extracellular electrolytes, which in turn causes osmotic dehydration and an increase in the concentration of intracellular electrolytes. He argued that glycerol and other low molecular weight additives protect by virtue of their colligative ability to lower the salt concentration in the manner shown in Figure 9, the term "colligative" referring to effects that depend on the number of ions or molecules, and not their chemical nature.

Cells also shrink osmotically in response to the increasing concentration of extracellular solutes during freezing, and Meryman (1970) proposed an alternative hypothesis that slow-freezing injury is a consequence of that shrinkage, or, to be more precise, it is a consequence of the inability of the cells to shrink to the extent required for osmotic equilibrium without membrane damage. A few years later, Wiest and Steponkus (1978) proposed a variant of this hypothesis for plant protoplasts. Their view is that cells that shrink osmotically during freezing lose

membrane material and consequently undergo damage as they attempt to return to normal isotonic volume during thawing.

Common to all three hypotheses is the assumption that injury is a direct or indirect consequence of the concentration of solutes in the medium surrounding the cells. However, the concentration of solutes during freezing occurs because of the conversion of water to ice; i.e., m^e in the first of the equations on p. 364 increases because V^e decreases. The value of n_2 to a first approximation remains constant. But if we increase the prefreezing value of n_2 by introducing another solute such as glycerol, the required value of m^e during freezing is achieved with less decrease in V^e. That is to say the introduction of glycerol (or any other solute) results in an increase in the unfrozen fraction at any temperature (Figure 12).

If we combine the data on survival of red cells vs. temperature in Figure 10 with the data in Figure 12 on the effect of temperature on the unfrozen fraction, we obtain the results shown in Figure 11B. The curves are mirror images of those in Figure 11A which relate survival to salt concentration. They show that survival drops from above 80% when the unfrozen water fraction is 0.14 or more, to below 20% when

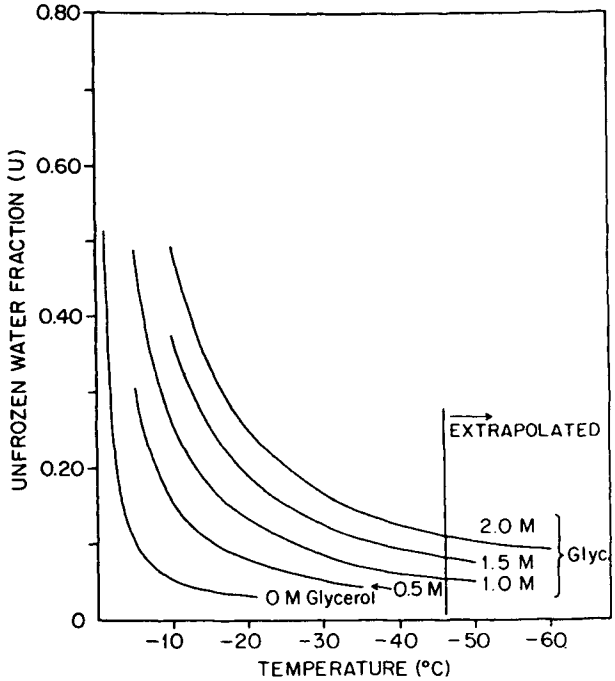


Figure 12. Fractions of water in glycerol-NaCl-H_2O solutions remaining unfrozen at various subzero temperatures. Glycerol molarities refer to the initial unfrozen solutions. The initial NaCl concentration in these solutions was 0.15 M. (From Mazur et al., 1981.)

the unfrozen fraction is 0.07 or less. The question, then, is whether the hemolysis of slowly frozen red blood cells is a result of the attainment of high NaCl concentrations as indicated in Figure 11A or is a result of the attainment of low unfrozen fractions as shown in Figure 11B.

Mazur et al. (1981) and Mazur and Rigopoulos (1983) have shown that the effects of salt concentration and the unfrozen fraction can be separated in part by appropriate manipulation of the starting glycerol and NaCl concentration and the temperatures to which the cells are frozen. We noted in connection with the equations on p. 364 that the total concentration of solute in a partly frozen solution depends on temperature alone and is therefore independent of the starting solute concentration. But that is not the case with the unfrozen fraction. Its magnitude is dependent on both the starting solute concentration and the temperature. It is possible then, by varying the total starting solute concentration (holding the ratio of cryoprotectant to NaCl constant), to vary the fraction unfrozen while maintaining a constant solute concentration at a given temperature in that unfrozen fraction.

For example, if red cells are suspended in solutions containing 0.11 *m* NaCl/0.37 *m* glycerol, 0.15 *m* NaCl/0.51 *m* glycerol, 0.31 *m* NaCl/1.07 *m* glycerol, and 0.7 *m* NaCl/2.42 m glycerol (all of which have the same mole ratio of glycerol to NaCl; i.e., 3.44) and are frozen to −16.8 °C, the concentration of NaCl in the unfrozen portion will be the same in all four cases; namely, 1.6 molal. But the percentage of water remaining unfrozen will be 7%, 9%, 19%, and 44% in the four solutions, respectively. According to Lovelock-type hypotheses, the survival should be the same in all four cases since the composition of the unfrozen portion of the solution is the same (i.e., 1.6 m NaCl, 5.5 m glycerol). But survival is not the same; it varies from about 15% in the first solution to about 95% in the third and fourth solutions. Figure 13 shows the results for a number of such experiments in which human red cells were frozen to temperatures yielding specific salt concentrations (m_s) ranging from 1.0 to 2.8 molal with the unfrozen fractions being allowed to vary for a given constant m_s. Since the Lovelock-type hypotheses include no effect of unfrozen fraction, they would predict a series of horizontal lines in such a graph with the survival line for lower values of m_s lying above the survival lines for higher values of m_s. Clearly, the results are very different. What we see in Figure 13 is that in the range of $U = 0.05$–0.15 (i.e., 85–95% of the water frozen), survival depends entirely on the unfrozen fraction and is independent of the concentration of salt in that fraction. Only at unfrozen fractions above 0.15 is an effect of salt concentration seen. We have found a similar dependence of survival on unfrozen fraction in hamster tissue culture cells and mouse embryos (Schneider and Mazur, 1987).

Figure 14 shows the computed osmotic volume changes the red cells undergo when suspended in the same four test solutions at room temperature (left side) and when slowly frozen (right side). They are responses characteristic of any cell that behaves as an ideal osmometer. Initially the cells are at normal isotonic volume (relative volume 1.0). After transfer to the test solution, however, they shrink in all

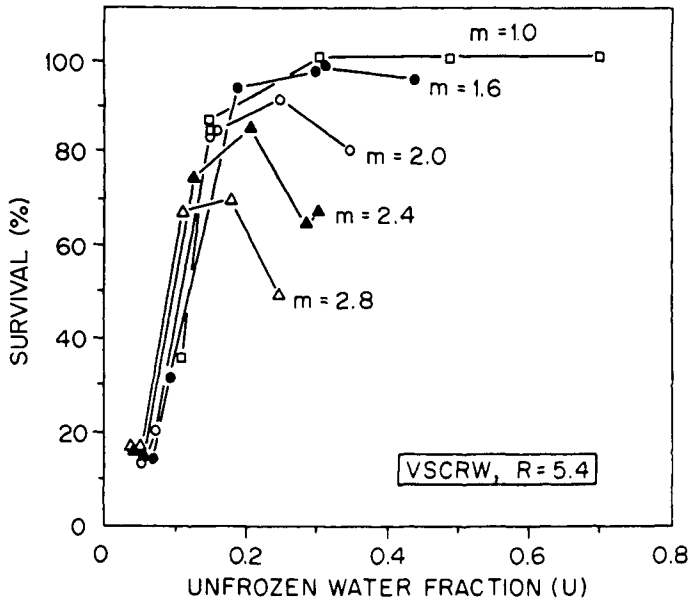


Figure 13. Survival (% unhemolyzed) vs. unfrozen water fraction of human red blood cells frozen very slowly at 0.6 °C/min (VSC) in solutions of glycerol-NaCl-H_2O to temperatures producing the indicated concentrations of NaCl (m_s) in the unfrozen channels. Warming and thawing were rapid (RW). (From Mazur et al., 1981.)

four cases because the osmolal concentrations of extracellular glycerol plus NaCl exceed the isotonic concentration, and water is withdrawn from the cell to restore osmotic balance. But then the glycerol gradually permeates the cell, and, as it does, water moves back into the cells to maintain osmotic balance, an influx that causes the cells to swell gradually. The final volume they attain depends solely on the concentration of the impermeant species (here, NaCl), and since the concentrations of NaCl differ in the four cases, the final equilibrium volumes differ.

The right-hand side of Figure 14 depicts the volume changes when the glycerol-equilibrated cells are then slowly frozen. Note that they all follow a single shrinkage curve. Once the temperature has dropped below the freezing point of the test solutions, all cells at a given temperature have the same volume relative to the initial isotonic volume irrespective of the test solution. Where the cells in the several solutions differ is with respect to the proportion of the shrinkage that occurs during the room temperature equilibration and the proportion that occurs during freezing. For example, the shrinkage of cells in the 4× solutions occurs primarily in the former whereas the shrinkage of cells in the 1× solutions occurs entirely in the latter.

The smaller, harmful, unfrozen fractions are obtained predominantly by the use of the lower tonicity test solutions, in which the cells undergo the greater volume

excursions during freezing. Pegg and Diaper (1988) have argued that this greater volume excursion is the injurious factor rather than the low unfrozen fraction attained. A major reason for their conclusion is that they obtained somewhat similar survivals when the sequence of changes in m_s and cell volume were produced not by freezing and thawing but by dialysis at room temperature. Mazur and Cole (1989) have given counter arguments, based in part on the fact that they, Nei (1981), and Pegg and Diaper have found, as discussed below, that higher cell hematocrits exacerbate the damage to red cells. The matter is unresolved at this time.

One deficiency in both hypotheses is a mechanistic explanation of injury. With respect to Pegg and Diaper's hypothesis, why should cells that start at a given

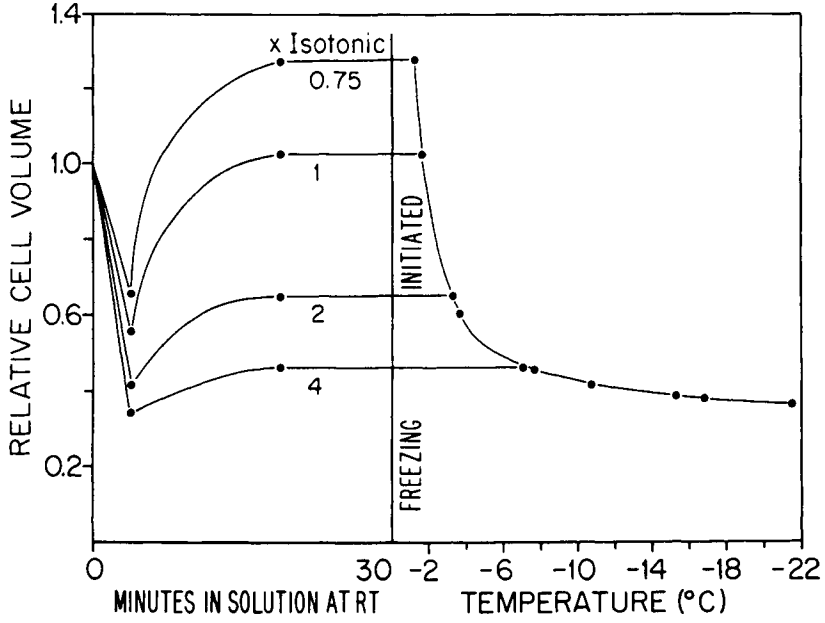


Figure 14. Computed relative volume changes in human red blood cells during (1) equilibration in four glycerol-NaCl-H_2O solutions, and (2) subsequent freezing. The four test solutions contain 0.75, 1, 2, and 4 times the isotonic concentration of salt and corresponding concentrations of glycerol such that the weight ratio of glycerol to NaCl remains 5.42 (i.e., ~0.4, 0.5, 1.0, and 2.0 M glycerol). At zero time, cells at relative volume 1.0 are transferred from isotonic NaCl to the test solutions at room temperature (RT). They initially shrink as a result of exosmosis of water and then swell as glycerol permeates. (Time scale during swelling is approximate.) At equilibrium, their volume is determined by the concentration of the impermeant species, NaCl. After some 30 min, slow freezing is initiated. Cells reshrink as a consequence of exposure to progressively increasing concentration of solutes in the unfrozen extracellular medium. (From Mazur and Rigopoulos, 1983.)

volume (Figure 14, left side) and end at a given volume (Figure 14, right side) care about when in time the intervening volume changes occur, especially in the dialysis experiment where both the left and right sides take place at room temperature?

With respect to the unfrozen fraction hypothesis, why should the size of the unfrozen channels per se influence cell survival? We have suggested that it does so because as the channels narrow, osmotically shrunken and deformed cells have a higher probability of coming in close contact with ice and with each other. Two indirect lines of evidence suggest that cell-cell contacts are a factor in freezing injury, at least in red cells. One, as mentioned, is that hemolysis from freezing and thawing becomes exacerbated at high hematocrits. As shown in Figure 15, the higher the initial hematocrit, the greater is the incidence of contact between osmotically shrunken cells in the unfrozen channels. A second line of evidence is that osmotically shrunken cells are hemolyzed if they are forced together by the centrifugation of unfrozen cell suspensions at –8 °C (Mazur and Cole, 1988). Unshrunken cells are not damaged by centrifugation and shrunken cells are not damaged in the absence of centrifugation.

Two more recent lines of evidence suggest damaging effects from cell-ice and cell-cell interactions. Both involve modifying the ice crystal habit during the freezing of a given solution at a given rate. The first makes use of so-called directional solidification; the second involves the use of antifreeze proteins from antarctic fish. The cooling rate of a sample is determined by the product of the thermal gradient G (units: °C/cm), and the rate of growth of the crystal front V (units: cm/min). A directional solidification stage is a special microscope stage that permits one to observe samples cooled at given rates but at varying G and V. The thermal gradient is established by holding the two ends of the stage at different temperatures. The crystal growth velocity is controlled by allowing the portion of the stage that carries the sample to move at a constant and controllable rate towards

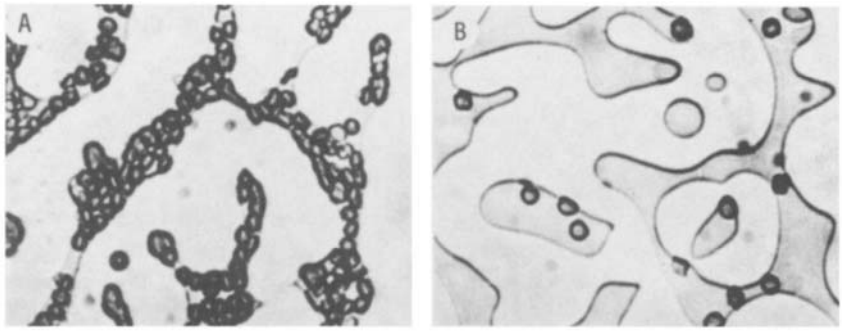


Figure 15. Appearance of a suspension of human red cells when frozen to –4 °C. The initial hematocrit was 20% in (**A**) and 2% in (**B**). Note that as in Figure 8, the cells lie in unfrozen channels between the ice crystals. (From Nei, 1981.)

the cold sink. Changes in G and V affect the morphology and spacing of the ice crystals and the unfrozen spaces between crystals, but they do not materially affect either the solute concentrations in the unfrozen fraction to which the cells are exposed or the duration of the exposure. The latter are determined by the temperature and the cooling rate. Nevertheless, Rubinsky and DeVries (1989), Beckmann et al. (1990), and Arav et al., (1993) have found the survival of human red cells, white cells, and bovine sperm are substantially influenced by changes in G and V at a constant cooling rate. In general those combinations of gradient and velocity that decrease the spacing between crystal dendrites are the more damaging.

A second way to modify ice crystal habits at a constant cooling rate is to introduce low concentrations of fish antifreeze proteins. Ordinarily, ice crystals grow more or less as hexagonal prisms. But in the presence of antifreeze protein, they grow as slender spicules or needles. Rubinsky and DeVries (1989) have found that the survival of human red cells frozen under the two conditions while suspended in 20% glycerol is dramatically different. In the former case, essentially all survive; in the latter case, none survive. Ishiguro and Rubinsky (1994) have observed a major difference in the two cases in the distribution of cells in the partly frozen samples. In the presence of a glycerol solution containing the antifreeze protein, the cells appeared to be crowded into unfrozen spaces between the ice spicules, in a manner analogous to that depicted in Figure 8 for red cells frozen in saline alone. But in the presence of a glycerol solution without antifreeze protein, the cells remained separated and uniformly distributed. It appeared as though they (and presumably pockets of solution) were individually encapsulated in the advancing ice. If confirmed, these findings raise the important possibility that the protective effect of glycerol may arise at least in part by its keeping cells separated during freezing.

Why should contact between shrunken cells be damaging? One possibility is that the contacts induce damaging membrane-membrane fusion, for close contact between shrunken cells has been found to induce membrane fusion above 0 °C. Moreover, freezing has been shown to induce fusion in model membranes (Anchordoguy et al., 1987) and to produce changes in fungal hyphae that are symptomatic of fusion (Fujikawa and Miura, 1986).

WARMING AND THAWING

A cell that has survived cooling to low subzero temperatures still faces the challenges of warming and thawing. The rate of warming can exert effects on survival of a magnitude comparable to those of the rate of cooling. The effects depend on whether the prior rate of cooling has been high enough to induce intracellular freezing or low enough to produce cell dehydration. In the former case, if the cells are not killed outright, many can be rescued by rapid thawing. The explanation is that although high cooling rates produce intracellular ice, the ice crystals tend to be small. Small crystals, because of their small radii of curvature,

have higher surface free energies than large crystals, and because of this higher free energy they tend to recrystallize or fuse together to form larger crystals. The extent of recrystallization depends on temperature and time. Consequently, recrystallization is enhanced when warming is slow and it is suppressed when warming is rapid. The enlargement of the ice crystals by recrystallization is damaging, probably because it disrupts intracellular membranes.

On the other hand, if cells are cooled slowly enough to preclude intracellular freezing, the response to warming rate is highly variable. In some cells (e.g., yeast) warming rate makes little or no difference; in some cells (e.g., boar sperm) rapid warming appears mandatory; in other cells (e.g., mouse embryos, higher plant cells) slow warming may be mandatory. Mechanistic explanations have been offered for these differences, but are not proven (Mazur, 1984).

NONEQUILIBRIUM FREEZING

The term "slow" in the slow freezing we have been discussing means a cooling rate that is low enough to allow the chemical potential of water in the cells to remain in near-equilibrium with that of the extracellular ice and water; hence, such slow freezing is referred to as equilibrium freezing, and it results in the cell water volume remaining close to the equilibrium volume during freezing (e.g, the EQ curve in Figure 2). The first successful freezing of early mouse embryos (2–8 cell) required such equilibrium freezing; i.e., a cooling rate of $\leq$ 1 °C/min to about –70 °C (Wittingham et al., 1972). But more recently, nonequilibrium procedures involving higher cooling rates have been used successfully. These nonequilibrium procedures have two elements in common. One is that the embryos must be partially dehydrated before they have cooled to the temperatures at which intracellular ice can form (the nucleation temperature). The other is that they must be warmed rapidly. The partial dehydration apparently reduces the embryo water content to the point that the subsequent rapid nonequilibrium cooling either results in the formation of intracellular ice crystals that are small enough to be innocuous, or it results in the conversion of the intracellular solution into a noncrystalline glass—a process referred to as vitrification. The subsequent rapid warming apparently ensures that the small innocuous ice crystals cannot recrystallize and that vitrified solutions cannot crystallize or devitrify during the return to normal temperatures (Fahy, 1987; Rall, 1987; Mazur, 1990).

Several methods have been used to achieve the initial dehydration. One is to place the embryos in glycerol-containing media made hypertonic with (nonpermeating) sucrose (e.g., Széll and Shelton, 1986). The result prior to the initiation of rapid freezing is analogous to the 2× isotonic curve in the left side of Figure 14. Another method achieves the same effect by using temperatures near 0 °C to slow the permeation of cryoprotectant and by initiating rapid freezing when the embryo volumes are near the minimums shown in the left side of Figure 14 (Rall, 1987). Also important to success with these nonequilibrium methods is the use of multi-

molar concentrations of the permeating cryoprotectant (often glycerol), as high as 6.5 M. These high concentrations of cryoprotectant facilitate the formation of small intracellular ice crystals and they facilitate vitrification (Fahy, 1987; Rall, 1987; Mazur, 1990).

With most single cells so far examined, the survivals obtained with nonequilibrium freezing procedures and vitrification are no better than those obtainable with equilibrium freezing procedures. But vitrification procedures may be necessary for the successful preservation of tissues and organs, which appear not to be able to tolerate the formation of ice in capillaries and other spaces between the cells.

THE FREEZING OF MULTI-CELLULAR SYSTEMS

When water is pulled osmotically out of individual cells during freezing, it passes to and freezes in the extracellular medium—the true outside. But, as illustrated in Figure 16, when water is osmotically withdrawn from the cells that comprise a tissue or organ, it passes not to the true outside of the tissue or organ but to spaces lying within the organ—namely, the lumens of capillaries or intercellular sinusoids, and other, nonvascular, space—where it is converted to ice. The result is that these intercellular spaces increase in volume. The volume increase during freezing is large: 8-fold in rabbit liver according to studies by Rubinsky and Pegg (1988). The cells from which the water is flowing are thus dehydrating and avoiding intracellular ice formation, but the endothelial cells lining the capillaries and the cells lining the sinusoids into which the water is moving and freezing are subjected to major stretching forces as the endothelial linings attempt to accommodate the large increase in water volume. It is highly doubtful that they can do so without damage. This topological geometric problem may explain in large measure why individual cells isolated from hearts and kidneys survive standard slow freezing techniques reasonably well (Alink et al., 1977), whereas the organs from which they are derived do not survive any freezing procedures so far tried (Armitage, 1981; Karow, 1981).

A previous section pointed out that damage to single cells may arise as a result of their being forced into contact with other shrunken cells during slow freezing. If that proves generally true, it could be a major problem in the cryopreservation of tissues and organs by slow equilibrium freezing because their component cells are in close apposition to begin with, and because the apposed plasma membranes often possess specialized regions such as tight and gap junctions.

Since the large increase in the volumes of intercellular lumens and the dehydration of the cells surrounding those lumens are both consequences of ice formation in the lumens, a solution to both problems would be to eliminate ice formation altogether by vitrifying the tissue or organ. The feasibility of this approach has recently been demonstrated by the successful cryobiological cryopreservation of 13–15 hr embryos of the fruit fly *Drosophila* (Steponkus et al., 1990; Mazur et al., 1992b; Steponkus and Caldwell, 1993). At this stage of development, these insect

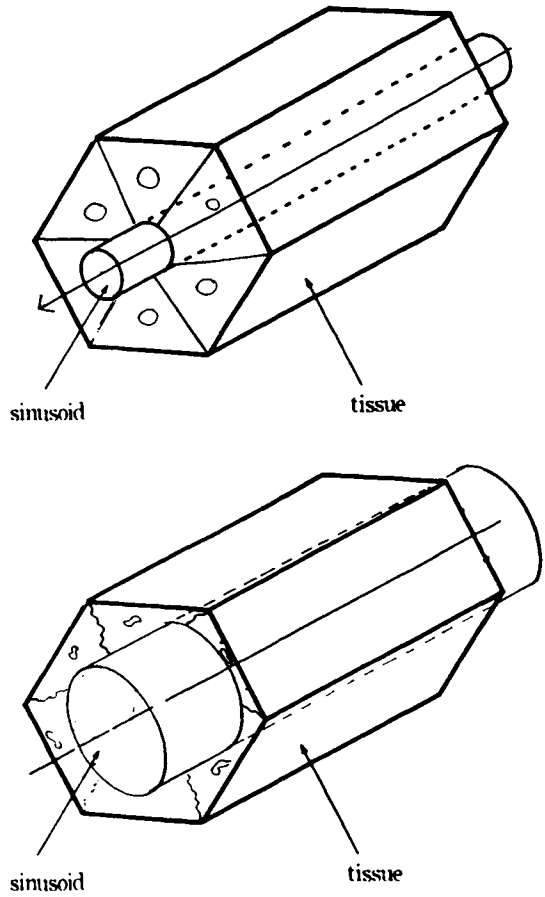


Figure 16. Schematic of the geometry of a section of liver before (top) and after (bottom) being subjected to slow freezing. During slow freezing, ice forms in the sinusoids and water flows osmotically from the cells into the sinusoids. (The upper drawing is slightly modified from Rubinsky and Pegg, 1988.)

embryos contain some 50,000 cells well differentiated into tissue and organ systems. Indeed, at the stage preserved, myoneural junctions are in the process of being formed (Broadie and Bate, 1993). The success demonstrates that complex multicellularity does not per se preclude cryopreservation. Success, however, as with any vitrification procedure, demands that the cells be surrounded by and contain high concentrations of glass-inducing solutes, in this case, 8.5 M ethylene glycol. The problem is that the ability of cells to tolerate such concentrations is at best borderline. Vitrification can be achieved with lower solute concentrations, but

only if the cooling and warming rates are increased dramatically. For example, Fahy (1987) reports data and computations showing that if the solute concentration is reduced five weight percent, the warming rate required to prevent the crystallization of a vitrified sample during warming increases several thousandfold.

A number of vitrification approaches have been and are being pursued in mammalian systems by a number of investigators, including Rall at the Smithsonian National Zoological Park and National Institutes of Health; Fahy at the Naval Medical Research Institute; Pegg at York University, UK; Armitage at Bristol University, UK; and Boutron in Grenoble, France. Rall (1987) showed that mouse embryos were better able to tolerate the required high concentrations of cryoprotectant if they were achieved by a combination of solute permeation and the osmotic removal of intracellular water. Boutron has demonstrated that various glass-inducing solutes differ in the concentrations required to induce vitrification during cooling and to prevent devitrification during warming. Rall, Fahy, and Boutron have found that mixtures of solutes are often more effective in producing stable glasses than are single solutes. Indeed, Boutron and Peyridieu (1993) have reported the successful vitrification of 35 wt % butanediol solutions by the addition of low concentrations of sugars like sucrose. Fahy has found that vitrification is enhanced by the application of hydrostatic pressure, and that the toxicity of high concentrations of glass-inducing solutes can be reduced by the presence of certain compounds (Khirabadi et al., 1993). Armitage and Pegg are applying such findings in attempts to vitrify corneas.

RETURNING THAWED CELLS TO PHYSIOLOGICAL MEDIA

Most cells contain equilibrium concentrations of cryoprotective solutes prior to freezing, and this equilibrium concentration will be reattained upon thawing. Thawed cells that contain appreciable concentrations of cryoprotectants will abruptly swell osmotically if they are abruptly returned to normal physiological media. Unless precautions are taken, this swelling can be extensive enough to cause damage and death. The procedure used for removing cryoprotectants has thus been found to be critical to obtaining high survivals of frozen-thawed rabbit, mouse, and cattle embryos and frozen-thawed erythrocytes, lymphocytes, and hematopoietic cells (see references in Mazur, 1984).

Two procedures can be used to avoid osmotic damage during cryoprotectant removal. One is to reduce the concentration of cryoprotectant in the external medium in steps of appropriate size separated for appropriate times to keep swelling within tolerated limits. This procedure produces transient spikes of increased cell volume at each dilution step. The second procedure ensures that the cell volume never exceeds isotonic. It involves transferring cells to a solution of a nonpermeating solute like sucrose that is about equiosmolal to the cryoprotectant. The intracellular cryoprotectant diffuses out of the cell into the outside medium, and as it does, the cell gradually shrinks because of the presence of extracellular nonper-

meating solute. When sufficient cryoprotectant has left the cell, the cell can then be returned to an isotonic physiological medium in a single step without undergoing a volume increase above that tolerated.

The proper step size, step interval, and temperature in the first procedure depend on the inherent permeability of the cell to the cryoprotectant (P_g) and its temperature coefficient, the size of the cell, and the degree to which it will tolerate swelling above isotonic. These variables interact in such a complex way that their optimum values are difficult to determine empirically. However, if one measures P_g and its temperature coefficient independently, one can compute the osmotic volume changes resulting from a given procedure by the solution to two simultaneous equations (Mazur et al., 1974), and thereby quickly compute a dilution procedure that holds the volume excursions to tolerable limits. One equation gives cell water volume as a function of the intra- and extracellular concentration of permeating and nonpermeating solutes. The other differential equation gives the rate of efflux of permeating additive as a function of its P_g, the cell's surface-to-volume ratio, and the differences in the intracellular and extracellular concentrations of the additive. Gao et al. (1995) have recently used this approach to minimize osmotic injury to human sperm from the addition and removal of hyperosmotic solutions of glycerol.

FREEZING AND MEMBRANES

The Significance and Determination of Permeation Coefficients

As indicated, knowledge of the permeability of a cell to water and to cryoprotectants can be powerful tools in predicting the likely optimum values for the major steps involved in freezing. From the value of the permeability coefficient for the cryoprotectant, one can compute the time required to achieve its complete penetration into the cells prior to initiating freezing, and one can compute the osmotic responses of the cells to various procedures for returning them to isotonic saline, and thereby select the procedure most likely to eliminate osmotic damage. From knowledge of the permeability of the cell to water and its temperature coefficient, one can predict the cooling rate likely to be low enough to preclude lethal intracellular freezing. Several published examples testify as to the accuracy of the latter predictions (Mazur, 1984).

Freezing and Cell Membranes

There are aspects of cell membranes other than their permeability to water and solutes that also play a critical role in the responses of cells to freezing. The structure of the plasma membrane allows cells to supercool and probably determines their ice-nucleation temperature. The nucleation temperature along with the permeability of membranes to water are the chief determinants of whether cells cooled at

given rates will equilibrate by dehydration or by intracellular freezing. Furthermore, surface and internal membranes seem to be the chief targets of injury with both slow and rapid freezing (Heber et al., 1973; Steponkus et al., 1977). Soluble enzymes are far more resistant. There is evidence that changes in cell volume or distortion of cell shape can be damaging at subzero temperatures (Leibo, 1976). Very possibly this reflects changes in the physical properties of membranes induced by low temperatures. One such change in human red cells is that the plasma membrane apparently loses all elasticity by –20 °C (Thom et al., 1988).

Most of this review has considered the cell to be a sack of dilute aqueous solution and has neglected the possible involvement of internal membrane-bounded components or other structures. Despite this gross oversimplification, the predicted responses of many cells are in reasonable accord with observation. But this is not universally so. Armitage and Mazur (1984), for example, have suggested that the inability of human granulocytes to survive even moderately hyperosmotic solutions of both permeating and nonpermeating solutes (glycerol, sucrose, and NaCl) may reflect damage to intracellular organelles rather than the cell surface. Another example is sperm. The water permeability of the plasma membrane of human sperm has been found to be so high that intracellular freezing is predicted to occur only when cooling rates exceed 10,000 °C/min (Noiles et al., 1993). Yet when motility is measured as a function of cooling rate, it decreases when the rates exceed one to 10 °C/min (Henry et al., 1993)—a thousandfold discrepancy if, as in Figure 4, the drop is a consequence of internal ice formation. One explanation of the discrepancy is that the critical factor is not the water permeability of the plasma membrane but the water permeability of intracellular organelles like mitochondria which must remain functional for motility. It may be that loss of motility is not due to ice formation in the cytosol but rather to ice formation in the mitochondria. A third example is the oocyte. Although early embryos from more than a dozen mammalian species can be cryopreserved with high survival, the survival of unfertilized eggs after freezing is poor (Leibo, 1992). One reason may be that such oocytes are locked in meiotic metaphase II, a stage in which the chromosomes lie on the spindle. The spindle is composed of microtubules, and microtubules are known to disassemble at low temperatures. Thus, as cryobiology seeks a deeper understanding of the mechanisms of injury, it will have to become increasingly concerned with the responses of intracellular components such as mitochondria, lysosomes, microtubules, and microfilaments.

SUMMARY

Biological time ceases at the temperature of liquid nitrogen (–196 °C), a fact that permits the long-term preservation of cells, provided that they can withstand the dangerous roundtrip voyage to such low temperatures. The cell is poised between Scylla and Charybdis. If it is cooled too rapidly, its water freezes internally with lethal consequences. If it is cooled too slowly, intracellular freezing is avoided, but

the cell becomes exposed to enormous concentrations of solutes and narrowing channels of liquid as the water in the external medium is converted to ice. Certain cryoprotective solutes such as glycerol and dimethyl sulfoxide have been found to ameliorate this latter damage either by reducing the concentration of electrolytes or by inhibiting cell-ice and cell-cell contacts. In essence, freezing subjects cells to extremes of osmotic pressure. Consequently, their response to freezing can often be satisfactorily described by the physical-chemical relations that govern the normal osmotic movement of water and solutes through plasma membranes. When one knows the responses, one can take measures to avoid or minimize damaging consequences and thereby achieve the stoppage of biological time. This fundamental approach has led to the successful cryobiological preservation of a variety of single cells and simpler multicellular systems. It will be an important step in developing methods for the successful preservation of more complex tissues and organs.

ACKNOWLEDGMENTS

Research supported by the Office of Health and Environmental Research, U.S. Department of Energy under contract DE-AC05-84OR21400 with the Martin Marietta Energy Systems, Inc.

REFERENCES

Aldhous, P. (1993). Malaria: Focus on mosquito genes. Science 261, 546–548.

Alink, G.M., Verheul, C.C., & Offerijns, F.G.J. (1977). Isolation and low temperature preservation of adult rat heart cells. Cryobiol. 14, 399–408.

Anchordoguy, T.J., Rudolph, A.S., Carpenter, J.F., & Crowe, J.H. (1987). Modes of interaction of cryoprotectants with membrane phospholipids during freezing. Cryobiol. 24, 324–331.

Arav, A., Granut, R., Shturman, H., & Rubinsky, B. (1993). Cryopreservation of bovine semen with directional solidification. Cryobiol. 30, 632–633 (abstract).

Armitage, W.J. (1981). Heart. In: Organ Preservation for Transplantation, 2nd ed. (Karow, A.M. Jr. & Pegg, D.E., eds.), pp. 577–597, Dekker, New York.

Armitage, W.J. & Mazur, P. (1984). Toxic and osmotic effects of glycerol on human granulocytes. Am. J. Physiol.: Cell Physiol. 247, (Cell Physiol. 16) C382–C389.

Beckmann, J., Korber, C.H., Rau, G., Hubel, A., & Carvalho, E.G. (1990). Redefining cooling rate in terms of ice front velocity and thermal gradient: First evidence of relevance to freezing injury of lymphocytes. Cryobiol. 27, 279–287.

Boutron, P. & Peyridieu, J.-F. (1993). Absence of toxicity for erythrocytes of buffered solutions containing high concentrations of 2,3-butanediol and sugars. Cryobiol. 30, 658 (abstract).

Broadie, K. & Bate, M. (1993). Innervation directs receptor synthesis and localization in *Drosophila* embryo synaptogenesis. Nature 361, 350–353.

Cohen, C.B. & Jonsen, A.R. (1993). The future of the fetal tissue bank. Science 262, 1663–1665.

Fahy, G.M. (1987). Biological effects of vitrification and devitrification. In: The Biophysics of Organ Cryopreservation (Pegg, D.E. & Karow, A.M., Jr., eds.), pp. 265–297, Plenum Press, New York.

Fujikawa, S. & Miura, K. (1986). Plasma membrane ultrastructural changes caused by mechanical stress in the formation of extracellular ice as a primary cause of slow freezing injury in fruit-bodies of basidiomycetes (*Lyophyllum ulmarium* [Fr.] Kühner). Cryobiol. 23, 371–382.

Gao, D.Y., Liu, J., McGann, L.E., Watson, P.F., Kleinhans, F.W., Mazur, P., Critser, E.S., & Critser, J.K. (1995). Prevention of osmotic injury to human spermatozoa during addition and removal of glycerol. Hum. Reprod. 10, 1109–1122.

Heber, U., Tyankova, L., & Santarius, K.A. (1973). Effects of freezing on biological membranes *in vivo* and *in vitro*. Biochim. Biophys. Acta 291, 23–37.

Henry, M.A., Noiles, E.E., Gao, D., Mazur, P., & Critser, J.K. (1993). Cryopreservation of human spermatozoa. IV. The effects of cooling rate and warming rate on the maintenance of motility, plasma membrane integrity, and mitochondrial function. Fertil. and Steril. 60, 911–918.

Ishiguro, H. & Rubinsky, B. (1994). Mechanical interactions between ice crystals and red blood cells during directional solidification. Crybiol. 31, 483–500,

Karow, A.M., Jr. (1981). Problems of organ preservation. In: Organ Preservation for Transplantation, 2nd ed. (Karow, A.M., Jr., & Pegg, D.E., eds.), pp. 517–552, Dekker, New York.

Karow, A.M., Jr. & Pegg, D.E. (eds.) (1981). Organ Preservation for Transplantation. Marcel Dekker, New York.

Kearney, J.N., Wheldon, L.A., & Gowland, G. (1990). Effects of cryobiological variables on the survival of skin using a defined murine model. Cryobiol. 27, 164–170.

Khirobadi, B.S., Fahy, G.M., Saur, J., & Meryman, H.T. (1993). Perfusion of rabbit kidneys with 8 molar cryoprotectant (V52). Cryobiol. 30, 611–612 (abstract).

Leibo, S.P., Farrant, J., Mazur, P., Hanna, M.G., Jr., & Smith, L.H. (1970). Effects of freezing on marrow stem cell suspensions: Interactions of cooling and warming rates in the presence of PVP, sucrose, or glycerol. Cryobiol. 6, 315–332.

Leibo, S.P. (1976). Freezing damage of bovine erythrocytes: Simulation using glycerol concentration changes at subzero temperatures. Cryobiol. 13, 587–598.

Leibo, S.P. (1977). Preservation of mammalian cells and embryos by freezing. In: Cryoimmunology (Simatos, D., Strong, D.M., & Turc, J.M., eds.), Vol. 62, pp. 311–334, INSERM, Paris.

Leibo, S.P. (1981). Preservation of ova and embryos by freezing. In: New Technologies in Animal Breeding (Brackett, E.G., Seidel, G.E. & Seidel, S.M., eds.), pp. 127–139, Academic Press, New York.

Leibo, S.P. (1992). Techniques for preservation of mammalian germ plasm. Animal Biotech. 3, 139–153.

Lovelock, J.E. (1953a). The haemolysis of human red blood cells by freezing and thawing. Biochim. Biophys. Acta 10, 414–426.

Lovelock, J.E. (1953b). The mechanism of the protective action of glycerol against haemolysis by freezing and thawing. Biochim. Biophys. Acta 11, 28–36.

Madden, P.W., Taylor, M.J., Hunt, C.J., & Pegg, D.E. (1993). The effect of polyvinylpyrrolidone and the cooling rate during corneal cryopreservation. Cryobiol. 30, 135–157.

Mazur, P. (1963). Kinetics of water loss from cells at subzero temperatures and the likelihood of intracellular freezing. J. Gen. Physiol. 47, 347–369.

Mazur, P., Miller, R.H., & Leibo, S.P. (1974). Survival of frozen-thawed bovine red cells as a function of the permeation of glycerol and sucrose. J. Membr. Biol. 15, 137–158.

Mazur, P. & Miller, R.H. (1976). Survival of frozen thawed human red cells as a function of the permeation of glycerol and sucrose. Cryobiol. 13, 523–536.

Mazur, P. (1977a). The role of intracellular freezing in the death of cells cooled at supraoptimal rates. Cryobiol. 14, 251–272.

Mazur, P. (1977b). Slow-freezing injury in mammalian cells. In: The Freezing of Mammalian Embryos. Ciba Foundation Symposium No. 52 (Elliott, K. & Whelan, J., eds.), pp. 19–42, Elsevier, Amsterdam.

Mazur, P., Rall, W.F., & Rigopoulos, N. (1981). Relative contributions of the fraction of unfrozen water and of salt concentration to the survival of slowly frozen human erythrocytes. Biophys. J. 36, 653–675.

Mazur, P. & Rigopoulos, N. (1983). Contributions of unfrozen fraction and of salt concentration to the survival of frozen human erythrocytes: Influence of warming rate. Cryobiol. 20, 274–289.

Mazur, P. (1984). Freezing of living cells: Mechanisms and implications. Amer. J. Physiol. 247 (Cell Physiol. 16) C125–C142.

Mazur, P., Rall, W.F., & Leibo, S.P. (1984). Kinetics of water loss and the likelihood of intracellular freezing in mouse ova: Influence of the method of calculating the temperature dependence of water permeability. Cell Biophy. 6, 197–214.

Mazur, P. & Cole, K.W. (1988). Contact between shrunken red cells as a factor in freezing injury. Cryobiol. 25, 510–511 (abstract).

Mazur, P. & Cole, K.W. (1989). Roles of unfrozen fraction, salt concentration, and changes in cell volume in the survival of frozen human erythrocytes. Cryobiol. 26, 1–29.

Mazur, P. (1990). Equilibrium quasi-equilibrium, and nonequilibrium freezing of mammalian embryos. Cell Biophy. 17, 53–92.

Mazur, P., Schneider, U., & Mahowald, A.P. (1992a). Characteristics and kinetics of subzero chilling injury in *Drosophila* embryos. Cryobiol. 29, 39–68.

Mazur, P., Cole, K.W., Hall, J.W., Schreuders, P.D., & Mahowald, A.P. (1992b). Cryobiological preservation of *Drosophila* embryos. Science 258, 1932–1935.

Meryman, H.T. (1970). The exceeding of a minimum tolerable cell volume in hypertonic suspension as a cause of freezing injury. In: The Frozen Cell, Ciba Foundation Symposium (Wolstenholme, G.E.W. & O'Connor, M., eds.), pp. 51–64, Churchill, London.

Morris, G.J. & Farrant, J. (1972). Interactions of cooling rate and protective additive on the survival of washed human erythrocytes frozen to –196 °C. Cryobiol. 9, 173–181.

Morris, G.J. (1987). Direct chilling injury. In: The Effect of Low Temperatures on Biological Systems (Grout, B.W.W. & Morris, G.J., eds.), pp. 120–146, Arnold, London.

Nei, T. (1981). Mechanism of freezing injury to erythrocytes: Effect of initial cell concentration on the post-thaw hemolysis. Cryobiol. 18, 229–237.

Noiles, E.E., Mazur, P., Watson, P.F., Kleinhans, F.W., & Critser, J.K. (1993). Determination of water permeability coefficient for human spermatozoa and its activation energy. Biol. of Reprod. 48, 99–109.

Pegg, D.E. & Diaper, M.P. (1988). The mechanism of injury to slowly-frozen erythrocytes. Biophys. J. 54, 471–488.

Rall, W.F., Mazur, P., & McGrath, J.J. (1983). Depression of the ice-nucleation temperature of rapidly cooled mouse embryos by glycerol and dimethyl sulfoxide. Biophys. J. 41, 1–12.

Rall, W.F. (1987). Factors affecting the survival of mouse embryos cryopreserved by vitrification. Cryobiol. 24, 387–402.

Rapatz, G.L., Menz, L.J., & Luyet, B.J. (1966). Anatomy of the freezing process in biological materials. In: Cryobiology (Meryman, H.T., ed.), pp. 139–162, Academic Press, New York.

Rich, S.J. & Armitage, W.J. (1992). Corneal tolerance of vitrifiable concentrations of glycerol. Cryobiol. 29, 153–164.

Rubinsky, B. & Pegg, D.E. (1988). A mathematical model for the freezing process in biological tissue. Proc. Roy. Soc. London B 234, 343–358.

Rubinsky, B. & DeVries, A.L. (1989). Effect of ice crystal habit on the viability of red blood cells. Cryobiol. 26, 580 (abstract).

Schneider, U. & Mazur, P. (1987). Relative influence of unfrozen fraction and salt concentration on the survival of slowly frozen eight-cell mouse embryos. Cryobiol. 24, 17–41.

Shabana, M. & McGrath, J.J. (1988). Cryomicroscope investigation and thermodynamic modeling of the freezing of unfertilized hamster ova. Cryobiol. 25, 338–354.

Smith, A.U. & Polge, C. (1950). Survival of spermatozoa at low temperatures. Nature 166, 668–669.

Souzu, H. & Mazur, P. (1978). Temperature dependence of the survival of human erythrocytes frozen slowly in various concentrations of glycerol. Biophys. J. 23, 89–100.

Steponkus, P.L., Garber, M.P., Myers, S.P., & Lineberger, R.D. (1977). Effects of cold acclimation and freezing on structure and function of chloroplast thylakoids. Cryobiol. 14, 303–321.

Steponkus, P.L., Myers, S.P., Lynch, D.V., Gardner, L., Bronshteyn, V., Leibo, S.P., Rall, W.F., Pitt, R.E., Lin, T.-T., & MacIntyre, R.J. (1990). Cryopreservation of *Drosophila melanogaster* embryos. Nature 345, 170–172.

Steponkus, P.L., & Caldwell, S. (1993). An optimized procedure for the cryopreservation of *Drosophila melanogaster* embryos. Cryo-Letters 14, 375–380.

Széll, A. & Shelton, J.N. (1986). Sucrose dilution of glycerol from mouse embryos frozen rapidly in liquid nitrogen vapour. J. Reprod. Fert. 76, 401–408.

Thom, F., Richter, E., & Matthes, G. (1988). Reactions of erythrocyte membranes to forces at subzero temperatures. Cryobiol. 25, 512 (abstract).

Trounson, A. & Mohr, L. (1983). Human pregnancy following cryopreservation, thawing and transfer of an eight-cell embryo. Nature 305, 707–710.

Watson, P.F. & Plummer, J.M. (1985). The responses of boar sperm membranes to cold shock and cooling. In: Deep Freezing of Boar Semen (Johnson, L.A. & Larsson, K., eds.), pp. 113–128, Swedish University of Agricultural Sciences, Uppsala, Sweden.

Whittingham, D.G., Leibo, S.P., & Mazur, P. (1972). Survival of mouse embryos frozen to –196° and –269 °C. Science 178, 411–414.

Wiest, S.C. & Steponkus, P.L. (1978). Freeze-thaw injury to isolated spinach protoplasts and its simulation at above freezing temperatures. Plant Physiol. 62, 699–705.

RECOMMENDED READINGS

Stowell, R.E. (ed.) (1966). Cryobiology. Proceedings March-April. Federation of American Societies for Experimental Biology, Washington, D.C.

Meryman, H.J. (ed.) (1966). Cryobiology. Academic Press, New York.

Wolstenholme, G.E.W. & O'Connor, M. (eds.) (1970). The Frozen Cell. Ciba Foundation Symposium, J. & A. Churchill, London.

Elliott, K. & Whelan, J. (eds.) (1977). The Freezing of Mammalian Embryos. Ciba Foundation Symposium 52 (new series) Elsevier, Amsterdam.

Karow, A.M., Jr. & Pegg, D.E. (eds.) (1981). Organ Preservation for Transplantation, 2nd edition. Dekker, New York.

Pegg, D.E. & Karow, Jr., A.M. (1987). The Biophysics of Organ Cryopreservation, NATO ASL Series A: Life Sciences, Vol. 147, Plenum, New York.

Bowler, K. & Fuller, B.J. (1987). Temperature and Animal Cells. Symposia of the Society for Experimental Biology 41, Company of Biologists, Ltd. Cambridge, U.K.

McGrath, J.J. & Diller, K.R. (1988). Low Temperature Biotechnology: Emerging Applications and Engineering Contributions. Amer. Society of Mechanical Engineers, New York.

Chapter 11

Clinical Applications of Cryobiology

JAMES H. SOUTHARD

INTRODUCTION

Cryobiology covers a broad spectrum of specialties and is a science that involves application of information about how low temperatures affect biological systems. Cryobiologists include those interested in plants (cold hardiness, adaptation), animals (hibernators, frogs and insects that tolerate freezing, fish in antarctic waters), protein chemists, biochemists, physiologists, engineers, and clinicians. The temperatures of interest are also varied and include mild hypothermia (10 °C to 20 °C below normothermia, 37 °C), deep hypothermia (near the freezing point

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 385–397.

ISBN: 1-55938-808-0

of tissues, 0 °C to 4 °C), and extreme temperatures that cause solidification (freezing, –0.5 °C to –196 °C).

Medical applications of cryobiology are related to the use of cold to either preserve tissues, cells, and organs, or to destroy unwanted tissues (cancers) by freezing. Preservation can be obtained for limited periods of time by utilizing hypothermia (above 0 °C) or for extremely long periods of time by lowering the temperature to a point at which practically no biological reaction can occur (about –140 °C). The science of cryopreservation (freezing) has progressed a long way since the original description of glycerol as a cryoprotectant useful for freezing sperm, around 1950. This subject is covered in a number of reviews, as well as in the preceding chapter in this volume. Thus, the clinical applications of freezing for preservation of tissues, cells, and organs will not be discussed in this chapter.

This chapter will focus on the principles of hypothermic preservation of tissues of interest clinically, and discuss some of the applications of cryobiological principles to medicine. Various chapters and reviews that cover specific areas in considerable detail that are related to clinical applications of cryobiology are cited at the end of this chapter.

HOW COLD AFFECTS METABOLISM

Most mammalian cells are not adapted for survival at temperatures that deviate much from the norm (about 37 °C). Therefore, subjecting cells to hypothermia will lead to loss of viability. If hypothermia is to be used effectively to preserve isolated tissues, cells, and organs, then the mechanisms by which cold kills cells must be understood. If we obtain that knowledge, it may be possible to design therapeutic approaches to utilize hypothermia for preservation. Hypothermia is used clinically under two general conditions, either where oxygen is present (aerobic incubation or continuous perfusion) or where oxygen is absent (cold static storage). Thus, in addition to the metabolic effects of cold, there are also the metabolic effects of ischemia (lack of oxygen and perfusion) that will affect metabolism and viability of the cells. Cells from different organs and cells that circulate through the vascular system show different metabolic requirements and different responses to cold and ischemia. Thus, the preservation methods for these different cell types vary.

CELLULAR MEMBRANES

Membranes are composed of phospholipids and proteins. The fatty acid composition of the phospholipids in a membrane influences how it is affected by the cold. In general, as the temperature of a cell is lowered the lipids in the membrane bilayer undergo a phase transition from a liquid crystalline (fluid) state to a gel (more solid) state. The temperature at which this transition takes place is very narrow for phospholipids composed of a simple mixture of fatty acids, but is quite broad for the phospholipids in cellular membranes. It is usually implied from various methods

(Fourier Transform Infra Red Spectroscopy, Arrhenius plots of rate vs. temperature of a membrane-linked phenomenon) that biological membranes from nonhibernating or cold acclimated animals show a phase transition around 12 °C to 17 °C. Thus, at useful cold storage temperatures, it is expected that the plasma membrane and membranes of the cellular organelles will be mostly in a gel or solid state.

The solidification of cellular membranes can have a wide range effect on metabolism. The lack of sufficient fluidity in the plasma membrane and the capability for easy deformability can lead to physical disruption of cells during handling and preparation for storage. The membrane regulates the internal composition of the cell and organelles such as the mitochondria, nucleus, and lysosomes. Solidification and changes in the state of lipids, as well as alterations in the location and flexibility of interdigitated proteins could lead to changes in the permeability properties of the membrane. Increased permeability to calcium has been demonstrated as a result of cooling hepatocytes. Thus, increased permeability will cause an influx of solutes from the external medium and the possible loss of intracellular metabolites that may be critical to survival of cells when returned to a normothermicnormoxic environment. Furthermore, the cell membranes contain enzymes that are involved in critical cellular processes, such as cell water content, ion translocation, intracellular hydrogen ion concentration (pH), energy transduction (ATP synthesis by mitochondria), and signal transduction through various receptors. The activity of enzymes in cellular membranes is decreased substantially by cooling because of inhibition of conformational changes necessary for catalysis or inhibition of substrate-enzyme interaction. Thus, the cell loses control of metabolism.

The importance of adaptation of membranes to the cold as a means of inducing tolerance is illustrated by the changes that take place in cellular membranes of hibernators or other species required to adapt to the cold for survival. In general, membranes of cold adapted species show increased capability to remain in a fluid state even at near freezing temperatures. This is often accomplished by incorporating a greater number of unsaturated fatty acids into the phospholipids. The degree of unsaturation influences the phase transition temperature (greater unsaturation greater fluidity at low temperatures). Furthermore, these organisms are capable of decreasing the leak-pump rate for ions and water, thus helping to conserve energy during periods of cold adaptation (such as hibernation) and regulating intracellular ionic and water concentrations.

For most tissues, cells and organs, the effects of cold on the cellular membrane are fully reversible. Cells cooled to 1 °C to 4 °C for short periods of time (about four hours) can regain normal cellular functions, including membrane-linked functions, when rewarmed. This seems to suggest that the phase transition in the membrane-bound phospholipids is reversible when the temperature is elevated to normothermia.

ENERGY METABOLISM

Hypothermia slows down enzyme catalysis of enzymes in plasma membranes or organelle membranes, as well as enzymes floating around in the cytosol. The primary reason enzyme activity is decreased is related to the decrease in molecular motion by lowering the temperature as expressed in the Arrhenius relationship ($k = Ae^{-Ea/RT}$, where k is the rate constant of the reaction, Ea the activation energy, A the constant for the given reaction, R the gas constant, and T the absolute temperature). An enzyme-catalyzed reaction depends on the collision frequency of the reactants which is dependent upon how fast they move about. The rate of motion of reactants (kinetic energy) is dependent upon temperature. However, each collision does not lead to catalysis, and only those reactants with sufficient energy (activation energy, Ea) react. The effects of temperature on the rate of an enzyme-catalyzed reaction can be readily seen by plotting the natural logarithm of the rate vs. the absolute temperature (Arrhenius Plot) (Figure 1). For soluble enzymes (Plot A), the plot gives a straight line with decreasing catalysis with decreasing tempera-

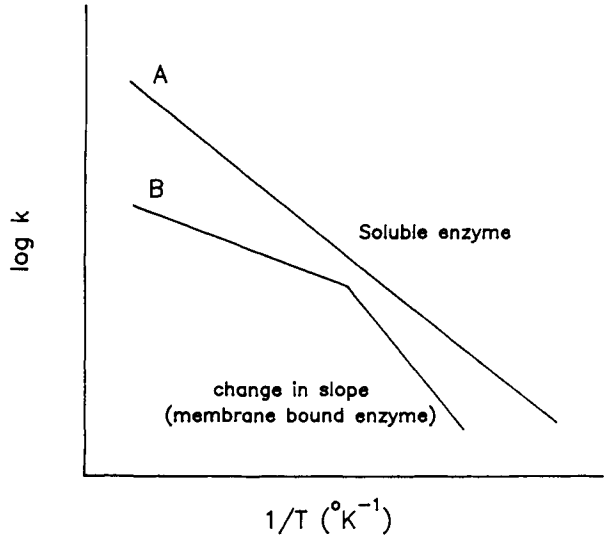


Figure 1. Examples of Arrhenius plots of reaction rate(s) versus temperature. An enzyme that is soluble in the cytosol often shows a linear relationship between the environmental temperature and its activity (Slope A in plot). A membrane-bound enzyme, or a reaction involving several enzymatic steps can show a break or change in slope at a particular temperature ("transition temperature"). This indicates that there has been a change in the rate-limiting enzyme or that there could be a solidification of the membrane which alters the activation energy and rate of reaction (Slope B in plot).

ture. van's Hoff found that for many biological reactions the decrease in rate of activity was about, two for every 10 °C drop in temperature. This finding led him to define the temperature coefficient (Q_{10}) as the degree of change activity for a 10 °C change in temperature. For most biological reactions, the Q_{10} is two to three. Thus, decreasing the temperature from normothermia (37 °C) to about 0 °C would lower the activity of most soluble enzyme reactions by about 12- to 13-fold.

The rate of catalysis of membrane bound enzymes (Plot B, Figure 1) is more greatly affected than soluble enzymes by lowering the temperature. This is due to the effect of low temperatures on the solidification of the membranes. Thus, an Arrhenius plot of the rate of a membrane-bound enzyme as a function of temperature often shows a discontinuity with a sharp break point (transition temperature) and loss of activity at the temperature where the membrane becomes a gel or more solid phase.

In most mammalian cells, energy is generated in the form of ATP by reactions carried out primarily in the mitochondrial inner membrane. Therefore, cooling will have a dramatic effect on the mitochondrial membrane, as well as on the rate and extent of energy generation. The primary effect of cooling on the mitochondria is to lower the rate of ATP synthesis without causing damage to the mechanisms that couple oxygen utilization to phosphorylation of ADP to ATP. The decreased rate of synthesis of ATP appears to be primarily due to an inhibition of the rate of entry of ADP (in exchange for ATP) into the inner mitochondrial space where phosphorylation takes place. Hypothermia decreases the activity of the membrane-bound adenine nucleotide translocase. Thus, in the cold there is a reduction in the rate of ATP production which would be well tolerated if there was a similar reduction in the rate at which ATP was utilized by energy-consuming reactions in the cell. However, in general, the energy-consuming reactions continue at a relatively fast rate and thus deplete ATP content, leaving the cell in an energy compromised state. This can lead directly to cell death due to the failure to regenerate metabolites and due to the failure of energy-dependent membrane-bound ion pumps.

ION MOVEMENTS AND CELL SWELLING

One of the primary ways in which cold causes injury to cells is by the loss of the capability to regulate cellular volume. This occurs because of the decrease in the available energy (ATP) caused by hypothermia and ischemia which is needed by the membrane-bound ion pumps, the increased rate of leakage of ions through the plasma membrane, and the decreased activity of the membrane-bound ion pumps, especially the Na-pump.

Cells are normally kept at osmotic (water activity) equilibrium by the action of the Na-pump. Inhibition of the pump with the specific Na^+-K^+-ATPase inhibitor, ouabain, causes cell swelling as does inhibition of it by hypothermia. The intracellular environment contains a high concentration of K (100 to 120 mM, in most mammalian cells), lower concentrations of Na (about 10 to 30 mM), and high

concentration of molecules that do not easily penetrate the membrane (proteins, phosphorylated compounds). The outside of the cell contains little protein and a high concentration of Na and low concentration of K. The membrane is permeable to Na and K which constantly exchange with each other. The Na that enters the cell, however, is pumped out by the electrogenic Na-pump and thereby the K is pumped back into the cells, the stoichiometry of the exchange being 3:2. The membrane potential retards the entry of chloride (negatively charged), which is in high concentration outside the cell vs. inside, and would move down its chemical potential were there no membrane potential. The pumping of Na out of the cell effectively makes Na an impermeant (osmotically active agent) that contributes to the suppression of water gain by the cell due to the colloidal osmotic pressure inside, which is higher than outside the cell, due to the impermeant proteins and phosphorylated compounds.

When the Na-pump is inhibited by ouabain, hypothermia, or lack of ATP (ischemia), H_2O moves into the cell to compensate for the increased osmolality due to the increased ionic content of the cell. The swelling of the cell is fully reversible if a normoxic or normothermic environment is reestablished in about four hours. However, with longer exposure to a swollen state, injury occurs which is irreversible. The increased water causes stretching of the membrane, swelling of mitochondria and loss of capacity to function properly, swelling of other organelles, and death of the tissue. Suppression of cell swelling under conditions of ischemia or hypothermia can prolong the time of safe storage of many tissues.

LOSS OF CONTROL OF METABOLISM

At hypothermia (0 °C to 4 °C) metabolism is slowed, but it is clear that metabolism continues. This is very evident by the observation that cold stored cells accumulate lactate if stored without oxygen (i.e., anaerobic glycolysis is active) but do not if stored under oxygen (mitochondria generate ATP). Furthermore, ATP disappears indicating that ATPases are active. Glutathione is degraded and because of the lack of energy (ATP) is not readily regenerated. Also, other metabolites are lost and end products accumulate, such as free fatty acids derived from phospholipid or triglyceride hydrolysis. Therefore, the cell continues to carry out metabolic reactions, but in an uncontrolled manner and most of the reactions are catabolic. This continued catabolism results in eventual cell death. Similar events go on in cells exposed to warm ischemia but only much faster. The exact systems that are degraded, at what rate, and how this leads to irreversible injury are subjects of great research interest in both warm and cold ischemic and hypoxic conditions. Despite a great deal of effort to define the mechanisms of ischemic injury, we currently do not know the exact events that lead to irreversible cell injury. This makes the design of therapeutic regimens to prevent cell death very difficult, and consequently, many of the advances in hypothermic medicine, as well as treatment of hypoxic and ischemic injury, are the result of trial and error type experimentation.

CLINICAL APPLICATIONS OF CRYOBIOLOGY

Hypothermia has been used to store a number of tissues, organs, and cells for transplantation or infusion in medicine, and some clinical applications are listed in Table 1. A most obvious application, and one that has been used for the longest period of time, is blood banking. Blood or blood component therapy involves using different components from blood for a special purpose. Examples include red blood cells for oxygen transport, platelets for clotting problems, protein fraction as volume expanders, and clotting factors for coagulopathies. Modern medicine would be hard put to be as effective and efficient without the availability of blood components. These components are available because they can be collected from donors and stored for up to about one month (for cells) or more (for individual proteins).

The storage of blood and its components is not particularly sophisticated, despite the large amount of time and effort spent designing appropriate storage solutions, containers, etc. Whole blood is mixed with an anticoagulant solution, such as acid-citrate-dextrose or a similar solution containing metabolites such as adenine (for adenine nucleotide metabolism), and stored at about 4 °C for up to 21 days. The red blood cells do not need oxygen and can produce ATP through anaerobic glycolysis. The maintenance of soluble organic phosphates (DPG) during storage is also known to help preserve hemoglobin affinity for oxygen which could be better maintained in citrate-phosphate-dextrose solutions. With the increased fractionation of whole blood for the preparation of plasma proteins, methods to preserve red blood cells were sought. In the 1980s various solutions were developed for hypothermic storage. Most were saline-based solutions with adenine, glucose, and mannitol. Adenine and glucose were used for energy generation, and mannitol to prevent or retard hemolysis during storage due to osmotic events. The types of compounds that can be used to improve red blood cell or whole blood preservation are somewhat limited by the fact that the solutions are directly infused into the patient. Therefore, potentially useful agents that have dramatic effects on cell metabolism are difficult to get approved for use in products that are directly infused into the patient.

Isolated platelets, on the other hand, do not tolerate hypothermic storage, need to be kept at room temperature, and have a continual supply of oxygen for aerobic metabolism. A significant problem in the storage of platelets is the fall in pH due to glycolysis and the generation of acidic end-products of metabolism. Some improvements in preservation have been obtained by using good hydrogen ion buffers, such as phosphate or histidine in high concentrations. The increased acidity leads to structural changes and loss of viability.

Another widespread application of cold storage of tissue for transplantation is that of preservation of the cornea. Injuries leading to corneal scarring, various kinds of diseases, and the production of corneal opacity are leading causes of blindness. Cornea transplantation is a commonly used and highly successful procedure. It

Table 1. Some Clinical Applications of Cryobiology

Application	*Purpose*
Blood and blood products	
Whole blood	Blood cell and volume replacement for hemorrhagic shock, surgery
Concentrated red blood cells	Increase in red blood cell mass in accident victims, hemorrhagic shock, surgery
Platelet concentrates	Bleeding due to thrombocytopenia or thrombocytopathy, other platelet disorders
Fresh frozen plasma	Multiple coagulation disorders
Tissues	
Heart valves	Replacement of defective valves
Skin	Covering for burned victims
Cornea	Replacement for corneal opacity due to congenital diseases, injury, etc.
Pancreatic islets of Langerhans, neuronal tissues	Cell replacement for diabetes, Parkinson's disease, Alzheimer's
Bone, cartilage	Repair skeletal defects due to trauma and disease
Bone marrow	Cancers of various types
Organs	
Heart	
Liver	
Kidney	For organ replacement due to end-stage organ diseases
Lung	
Pancreas	
Surgical Procedures	
Cardioplegia	Arrest heart for open heart surgery repair of damaged vessels
Total body hypothermia	Repair of vessel disease (brain aneurysm)
Cryosurgery	Prostate cancer, liver cancer, hemorrhoidectomy

would not be so successful if a source of corneas were not available due to successful methods of preservation. It is essential in preservation of the cornea that the endothelial layer of cells remain viable. Storage of the enucleated eye at 0 °C to 4 °C in sealed jars containing gauze soaked in saline extend preservation from about 4 hr (at room temperature) to about 24 hours. When placed in preservation solution it is essential that corneal hydration remains normal. Increased hydration of the cornea leads to development of opacities and destruction of the endothelial cells. One of the earliest and best solutions was developed in 1974 (McCarey and Kaufman solution). This solution was made up of a tissue culture medium (TC 199) that also contained a colloid (dextran, 40,000 Da) and a mixture of antibiotics. Storage of corneas in this M-K solution at 4 °C provides viability of the transplant for about four days of storage. Corneas are also successfully stored by organ culture methods using tissue culture medium in the presence of a continuous supply of oxygen and carbon dioxide. In this case, the corneas are stored at near normothermic conditions (about 34 °C).

Skin has also been stored for use as a covering of wounds in burn victims. Preserved skin (allograft) serves as a mechanical and physiological barrier; decreases loss of water, protein, and heat; and provides strata for the regeneration of viable skin by the victim. The storage medium is usually a tissue culture solution and the skin is wrapped up in gauze that is soaked in this solution and kept for about one week.

There has been an increase in the use of cadaveric heart valves for patients with valvular defects. The valves are best stored by freezing but some success has been achieved by simple cold storage in an antibiotic medium made up of ingredients common to most tissue culture solutions. At a storage temperature of 4 °C there is a continual loss of viability of fibroblasts so that by three weeks there are practically no viable cells and the valves cannot be used.

The successes of organ transplantation over the past two decades are due to three factors: the development of surgical methods for implantation of organs, the discovery of effective immunosuppressive agents (cyclosporine, monoclonal antibodies), and methods to preserve the organs. Organ preservation was essential for increasing the number of organs available for patients with various end-stage organ diseases. For organ transplantation to be effective, cadaveric organs, obtained usually from accident victims, must be used. These organs need to be obtained in excellent condition and preserved for at least one to two days, so they can be shipped around the country to the appropriate recipient. Preservation of vascularized organs appears to be more complicated than preservation of single cells and tissues (heart valves, cornea, skin) due to the necessity to preserve different kinds of cells that make up the organ. Especially important is the preservation of the vascular system so that normal blood flow can be restored to the preserved and transplanted organ.

Based upon theoretical considerations of the mechanisms of hypothermic-induced cellular injury, we developed the University of Wisconsin organ preservation solution (UW solution) that has had a widespread and dramatic effect on organ preservation (Table 2). Prior to the development of this solution, the liver and pancreas could be preserved for only four to six hours. Thus, there was a large time constraint on liver and pancreas transplantation and many cadaveric organs were wasted. However, the UW solution increased preservation duration to 48 to 72 hours, and dramatically increased the quality and numbers of these organs transplanted. Furthermore, this solution appears effective for the preservation of the kidney for three days and the heart for at least 15 hours.

The development of this solution was based upon current knowledge of the mechanisms of cold and warm ischemic injury to complex tissues. In particular, the first requirement was to find a suitable impermeant that would prevent cell swelling due to the effects of hypothermia on the Na-pump. In the past, preservation solutions used saccharides such as glucose or mannitol, which were effective for the kidney but not for other organs. We found that oxidized milk sugar (lactobionic acid) was sufficiently large enough (358 Da), and had a negative charge, to remain

Table 2. Composition of University of Wisconsin (UW) Organ Preservation Solutions

Component	*Concentrations*	*Function*
Lactobionic acid	100 mM	Impermeant, suppresses cell swelling, Ca and Fe chelator
Raffinose	30 mM	Impermeant saccharide, suppresses cell swelling
Hydroxyethyl starch	5 g% (50 g/1)	Impermeant colloid, suppresses cell swelling
Phosphate	25 mM	Hydrogen ion buffer, stimulated phosphorylation and high-energy compound production
Adenosine	5 mM	Precursor for ATP synthesis
Glutathione	3 mM	Antioxidant, sulfhydryl reagent
Allopurinol	1 mM	Xanthine oxidase inhibitor, suppresses oxygen free radical production
Potassium	120 mM	Osmotic agent, high concentration to inhibit efflux from cell
Sodium	20 mM	Osmotic agent
Magnesium	5 mM	Membrane stabilizer, enzyme cofactor

Note: The pH of this solution is adjusted to 7.4 and has an osmolality of 320 mOsm/l. Prior to use, a glucocorticoid steroid (dexamethasone, 8 mg), insulin (200 units/ml), and an antibiotic (penicillin, 200,000 u/l) are added. The solution is used at 0 °C to 4 °C.

outside metabolically depressed cells and prevent cell swelling. Thus, lactobionate became an important ingredient for organ cold storage solutions. Furthermore, adenosine was added to stimulate ATP regeneration after transplantation; glutathione was also added to the solution as a free radical scavenger. Finally, a colloid, hydroxyethyl starch, was included to suppress hypothermic-induced edema, especially in vascular cells. This solution contains a high concentration of K^+ (120 mM) and low concentration of Na^+ (25 mM). This suppresses the leakage of K^+ out of metabolically depressed cells and conserves energy on reperfusion.

Although this preservation solution has had a positive impact on solid organ preservation, a number of organs still show delayed or no function following transplantation. For the kidney, about 20% to 30% of the recipients still require dialysis for up to two weeks until the transplant begins to regain full function. In liver transplantation, up to 9% of the preserved liver can show no function resulting in the need for a new transplant. Furthermore, up to 20% can show delayed function. This may be due to preservation injury which is manifested in injury developing upon reperfusion. Improving the methods of preservation could have a dramatic effect on these current problems in organ preservation and this will require further understanding of how hypothermia affects cells, tissues, and organs.

CRYOSURGERY

Cryosurgery is a method of therapy in which specialized instrumentation is used to achieve the *in situ* destruction of tumorous tissue by freezing. This surgical specialty has been accepted as standard procedures in such disciplines as derma-

Table 3. Scope of Damaging Actions Associated with Cryosurgery

Hypothermia—Indirect cryodestruction	
Metabolic uncoupling	Prefreeze sensitization
Energy deprivation	
Ionic imbalance	
Disruption of acid-base balance	
Waste accumulation	
Membrane phase transitions	
Cytoskeletal disassembly	
Frozen State—Direct cryodestruction	
Water solidification	
Hyperosmolality	
Cell-volume disruption	
Protein denaturation	
Tissue shearing	
Intracellular-ice propagation	
Membrane disruption	
Microvascular	
Thawed State	
Direct effects	
Continued intracellular freezing	
Recrystallization	
Hypoosmolality	
Indirect effects	
Tissue anoxia	
Thrombosis	
Edema	
Inflammation	
Cryoimmunologic action (?)	

tology, neurology, proctology, urology, gynecology, cardiology, otorhinolaryngology, and veterinary medicine. Compared with such conventional therapies as surgical excision, radiotherapy and immunotherapy, cryosurgery is generally considered a "less competitive" modality. The selection of cryosurgical instrumentation is critical because of the concern of the cryosurgeon to effectively control destruction of unwanted tissue.

An essential precursor to the development of cryosurgery was the development of liquefied gasses (oxygen, nitrogen, and hydrogen) which occurred in the period between 1870 and 1900. The availability of liquid air allowed investigators to treat various skin lesions including carcinomas. The technique relied on the application of liquid air to the lesion using a swab or by spraying. In the 1960s controlled cryosurgery became an acceptable technique in medicine. Closed probes capable of lowering tissue temperature to below −100 °C with the aid of liquid nitrogen found application in brain, liver, and rectum surgery.

There are two primary effects of cryosurgery: direct and indirect cryodestruction. The indirect cryodestruction is the result of hypothermia and the direct cryodestruction is the result of freezing of the tissue as indicated in Table 3. Hypothermia leads to disruption of metabolism and structural integrity of the tissue.

This can exacerbate the impact of actual freezing. During freezing, which begins in the extracellular compartments and microvascular network, ice forms as pure, solid water by excluding electrolytes and organics. Thus, cells become exposed to extremely hyperosmotic environments. The extracellular brine creates a driving force that draws water from the cell resulting in an increased tissue ice content plus shrunken cells, further damaging membranes and inducing protein denaturation. Microscopically visible ice within the cell is nearly always lethal. Also, as one cell freezes, communication channels between cells provide preferred routes of ice propagation to other cells. During the freezing progression, tissue will be trapped and experience a series of shearing forces which will abrasively disrupt structures. After completion of freezing, thawing is initiated. Thawing has direct and indirect damaging effects (Table 3). Upon completion of the cryosurgical procedure, the ablative effects will continue for weeks. Destruction of the microvascular network results in anoxia followed by tissue thrombosis. Local edema follows with activation of the inflammation cascade, macrophage invasion, and resorption of the necrotic lesion.

The progress of cryosurgery in the medical field has been limited primarily by two factors: (a) the lesser certainty of *in situ* monitoring and control of the freezing process, and (b) the inability of cryosurgical instruments to completely destroy tumors of large volume or irregular shape. The use of real-time ultrasound or magnetic resonance imaging techniques has greatly improved the accuracy of monitoring and controlling the cryosurgical process. Recently, a new probe system has been developed that has the desired freezing capacity for effective, controlled destruction of unwanted tissues. The AccuProb® system (Cryomedical Sciences, Inc., Rockville, MD) utilizes multiple probes to deliver subcooled liquid nitrogen with a flow capacity greatly increased over other cryosurgical probes and devices. With sophisticated electronic control the temperature of each probe can be independently controlled and maintained. This provides for more effective, rapid, and controlled destruction of unwanted tissue. This new device is finding application in prostate surgery, as well as surgery for hepatic metastases.

SUMMARY

Significant advances made in medicine are the result of interdisciplinary approaches to clinical problems. Thus, the science of cryobiology has found useful and important applications in medical advances. On the horizon are new methods to improve preservation of various tissues and organs. Longer-term and higher quality preservation will result from a greater understanding of how hypothermia and metabolic depression affect cell metabolism, as well as the ultrastructure of the cell and tissue. In addition, exciting possibilities exist for the use of cryopreservation (freezing) of tissues of clinical usefulness. By freezing, truly long-term storage of tissues can be obtained.

RECOMMENDED READINGS

Aggarwal, S.J., Baxter, C.R., & Diller, K.R. (1985). Cryopreservation of skin: An assessment of current clinical applicability. J. Burn. Res. 6, 469–476.

Belzer, F.O. & Southard, J.H. (1988). Principles of solid-organ preservation by cold storage. Transplant 45, 673–676.

Friedler, S., Giudice, L.C., & Lamb, E.J. (1988). Cryopreservation of embryos and ova. Fertil. Sterility 49, 743–764.

Fuller, B.J. (1991). The effects of cooling on mammalian cells. In: Clinical Applications of Cryobiology (Fuller, B.J. & Grout, B.W.W., eds.), 1st edn. Chapter 1, pp. 1–22, CRC Press, Boca Raton.

Gage, A.A. (1982). Current issues in cryosurgery. Cryobiology 19, 219–222.

Haneda, K., Sands, M.P., Thomas, R., Hessel, E.A., & Dillard, D.H. (1983). Prolongation of the safe interval of hypothermic circulatory arrest: 90 minutes. J Cardiovasc. Surg. 24, 15–21.

Meryman, H.T., Hornblower, L.S., & Syring, R.L. (1986). Prolonged storage of red cells at 4 °C. Transfusion 26, 500–505.

Moore G.I. (1987). Additive solutions for better blood preservation. Crit. Rev. Clin. Lab. Sci. 25, 2112–229.

Morris, G.J. & Clark, A. (1981). Effects of Low Temperatures on Biological Membranes. Academic Press, New York.

O'Connor, J.V., Wilding, T., Farmer, P., Sher, J., Ergin, M.A., & Griepp, R.B. (1986). The protective effect of profound hypothermia on the canine central nervous system during one hour of circulatory arrest. Ann. Thorac. Surg. 41, 255–259.

Scarborough, N.L. (1992). Current procedures for banking allograft human bone. Orthopedics 15, 1161–1167.

Smit Sibinga, C. Th., Das, P.C., & Meryman, N.T. (1990). Cryopreservation and Low Temperature Biology in Blood Transfusion. Kluwer Academic Publishers, Dordrecht, Netherlands.

Southard, J.H., Pienaar, H., McAnulty, J.F., D'Alessandro, A.M., Hoffmann, R.M., Pirsch, J.D., Kalayoglu, M., Sollinger, H.W., & Belzer, F.O. (1989). The University of Wisconsin solution for organ preservation. In: Transplantation Reviews (Morris, P.J. & Tilney, N.A., eds.), Chapter 3, pp. 103–130, Saunders, New York.

Taylor, M.J. (1986). Clinical cryobiology of tissues: Preservation of corneas. Cryobiology 23, 323–353.

Tisherman, S.A., Safar, P., Radovsky, A., Peirzman, A., Marrone, G., Kuboyama, K., & Weinrauch, V. (1991). Profound hypothermia (< 10 °C) compared to deep hypothermia (15 °C) improves neurologic outcome in dogs after two hours' circulatory arrest induced to enable resuscitative surgery. J. Trauma 31, 1051–1062.

This Page Intentionally Left Blank

Chapter 12

Malignant Hyperthermia

CALVIN JOHNSON and SUJATHA KOTAMRAJU

INTRODUCTION

Malignant hyperthermia (MH) is a genetic myopathy characterized by high temperatures and hypermetabolism and is triggered by volatile anesthetics and depolarizing muscle relaxants. With current understanding of MH, proper monitoring, and treatment, mortality has greatly decreased. Nevertheless, it continues to be an important syndrome which must be understood and respected in the practice of medicine.

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 399–409.

ISBN: 1-55938-808-0

HISTORY

Malignant hyperthermia was first described in Australia by Denborough of the Royal Melbourne hospital in 1960. He showed that an increase in temperature was an important feature of a family syndrome in a 21 year old MH patient. Dr. Denborough successfully treated the MH episode in this patient and later gave an uneventful spinal anesthesia to the same patient for other surgery. Discovery of a similar syndrome in the Landrace and Pietram pigs aided researchers in understanding the syndrome. Certain pigs, especially those which were inbred, would often experience "pork stress" syndrome when exposed to stress, such as separation, shipping, weaning, fighting, coitus, and slaughter. This syndrome was characterized by accelerated metabolism, muscle rigidity, high temperature, rapid deterioration of muscle, and death. This resulted in a pale, soft exudative pork that was unfit for making sausage. Investigation of animal models helped to pinpoint the major defect in skeletal muscle. It also led to early therapeutic trials with dantrolene and identification of triggering agents.

EPIDEMIOLOGY

Actual incidence of MH is difficult to estimate because clinical diagnosis is often questionable, diagnostic testing is often difficult to apply, and there is no central reporting agency for MH. As reported by Sessler in 1986, the incidence of MH is approximately 1 per 12,000 anesthetics in children and 1 per 40,000 anesthetics in adults. In 1989, Dubrow et al. put the incidence at 1 in 15,000 pediatric anesthetics and 1 in 100,000 adult anesthetics. Metz in 1986 suggested that the increased incidence in children may be the result of increased use of succinylcholine after induction in a pediatric population. Incidence of MH is greatly increased in anesthetics using depolarizing muscle relaxants. In an excellent study from Denmark by Ording in 1985, all anesthetics over a 6.5 year period were reviewed and the incidence of MH was calculated in relation to the type of anesthesia. All cases of MH occurred during general anesthesia, and more than 75% occurred during anesthesia that used a combination of potent volatile anesthetic agents and succinylcholine. The incidence of fulminant MH was found to be only 1:250,000 in total anesthetic procedures, 1:220,000 in general anesthetics and 1:62,000 in anesthetics where a combination of potent volatile anesthetic and succinylcholine was used. No cases were reported in regional or intravenous (IV) techniques.

Masseter muscle spasm, which sometimes is a warning sign of MH, was found in 1:12,000 anesthetics. However, masseter muscle spasm is not diagnostic of an impending MH crisis. Even though approximately 50% of patients with masseter muscle spasm have tested positive for MH by muscle biopsy, this does not reflect the true incidence of MH. Incidence of masseter spasm is reported to be as high as 1% in pediatric patients, yet MH occurs in only about one in 15,000 pediatric patients.

MH occurs more in men than in women (1.4:1) and is more common between the ages of three and 30. The incidence of MH is increased in people who reside in countries where the climate is prone to extreme temperature changes. In 1982, Lewandowski observed that the incidence of MH is also increased in patients undergoing strabismus surgery. Muscle problems in strabismus and MH probably are a related phenomena. Masseter spasm is also seen more often in strabismus surgery; occurrence is 2.8% compared to 1% with halothane induction in a general pediatric population. Increased incidence of MH was reported in head and neck surgery, and in otolaryngology procedures. This could be because these are commonly performed procedures in pediatrics.

Kaplan and Kelloner estimated that between 1972 and 1982, the number of deliveries with general anesthesia in the U.S., Canada, and Great Britain was 29,31,300. Only four cases of MH were reported in obstetric cases during this time. These statistics support Crawford's view that pregnancy may have a protective effect against MH (Crawford, 1972).

Not all MH susceptible patients experience a crisis with their first anesthetic. A susceptible patient can have one or more uneventful anesthetics and develop the MH crises during subsequent anesthetics. Mortality from MH was 90% in the early years, came down to 70% n 1975 and it was estimated to be 7% in 1980 and has remained relatively stable through the 1980s.

MOLECULAR GENETICS

The genetic trait for MH is not sex-linked and both men and women can inherit MH. Inheritance in humans appears to be autosomal dominant with variable penetrance. Studies of large families have documented an autosomal dominant pattern. McPherson and Taylor (1982) studied 93 families in whom MH occurred. Even though various patterns of inheritance did emerge in the study we should assume that 50% of children are at risk in MH susceptible families. Kalow and Britt (1992) suggested that, in some families, at least two different non-allelic genes are likely to be present, one of which is probably autosomal dominant but rare, and the other autosomal recessive but common.

Over the last 20 years abnormal calcium regulation has been proposed as the primary defect in MH. Much research has been carried out on a specific calcium release channel in sarcoplasmic reticulum. This channel has been labeled the Ryanodine Receptor (RyR) since it bonds to the plant alkaloid, ryanodine, with high specificity. The human gene for RyR is located in the short arm of chromosome 19, and a point mutation in this gene is correlated with inheritance of MH. Several observations support this view that a single RyR mutation is causal of MH in all breeds of pigs. Substitution of arginine 615 in the amino acid sequence is the only difference in susceptible compared to nonsusceptible swine in a study involving 450 pigs in 6 breeds. In some human susceptible families, substitution of thymine for cytosine at position 1840 of the RyR transcript results in a cysteine for arginine

substitution at position 614 of the amino acid sequence. These findings suggest that the cysteine for arginine mutation represents a shared calcium release channel pathogenesis between the human and the porcine MH syndrome. Linkage of the MH syndrome to RyR is, however, not observed in all human families with MH. Accordingly, other abnormal genes that may cause the condition are being sought. Scattered case reports suggest a possible association of MH with a variety of neuromuscular diseases. This association is very strong in the case of central core disease (CCD), where it is supported by clinical and laboratory evidence, including the proximity of the CCD gene to the RyR gene or chromosome.

PATHOPHYSIOLOGY

Pathogenesis of MH is not completely understood. Skeletal muscle, however, is the one tissue in MH with proven abnormalities, and it is further thought that the basic defect that causes the syndrome lies in the calcium regulation system found within the myoplasm. For example, calcium transport function appears to be decreased in the sarcoplasmic reticulum, mitochondria, and sarcolemma. Thus, the suggestion has been made that MH is characterized by a generalized membrane defect.

In patients susceptible to MH, a triggering agent causes a sustained rise in myoplasmic calcium. Excess calcium forces the cell into a state of hypermetabolism as the cell attempts to reverse this increase in calcium. The earliest change detected is an increase in lactate in venous blood leaving the muscle. This increase in lactate has a profound effect on systemic circulation, since muscle comprises 40% of body weight. The rise in lactate is followed by detectable signs of tissue hypoxia. Glycogen stores are depleted and ATP levels fall. Eventually, anaerobic glycolysis with accumulation of CO_2, lactic acid and heat is established. The cell's housekeeping functions stop. Ions and molecules cease to be actively transported and instead follow their natural concentration gradients across the cell membrane. Intracellular ions such as potassium, magnesium, phosphate, and later myoglobin and proteins diffuse outward into the extracellular space; sodium and calcium diffuse into the cell creating the muscle edema and tenderness evident after a crisis. With increased metabolism (both aerobic and anaerobic) heat is produced. Excess lactate formed during anaerobic metabolism is responsible for the acidosis and is transported to the liver, where it is converted to glucose and combusted to CO_2 and H_2O, producing even more heat and acidosis.

Signs and symptoms of sympathetic nervous system activity are invariably found in MH. Levels of catecholamines are markedly increased in MH. Whether activation of the sympathetic nervous system is a primary or a secondary response in the syndrome has not been fully elucidated. Gronert reported that stress-induced sympathetic hyperactivity can initiate a malignant hyperthermic episode in susceptible swine without a triggering agent. Stress-induced MH in humans has been inferred because susceptible families have been shown to have an increased incidence of sudden death. Gronert's reasons that activation of the sympathetic

nervous system is a secondary response to MH are due to the following observations. Increases in the levels of catecholamines follow rather than precede the changes in muscle metabolism and acid-base balance. In addition the catecholamine response is not necessary for inducing MH in swine. Total spinal blockade and the accompanying sympathetic denervation with stable levels of catecholamines have been shown not to affect onset, development, or characteristics of the MH reaction.

Acute episodes of MH most commonly happen in the operating room. However, it may also occur in the post-anesthetic recovery period. Acute episodes of MH depend on three variables. These are: a genetic predisposition, presence of a potent anesthetic or a nonanesthetic trigger, and absence of inhibiting factors. Anesthetic agents that trigger MH include halothane, enflurane, isoflurane, ether, methoxyflurane, cyclopropane, succinylcholine, and decamethonium. Nondepolarizing muscle relaxants block the effect of succinylcholine in triggering MH. They also attenuate the effect of volatile anesthetics. Based on studies, neither amide nor ester local anesthetics trigger MH. It seems that all the local anesthetics can safely be used in MH-susceptible patients and in MH crises (Wingard, 1974; Berkowitz and Rosenberg, 1989). But during crisis it is recommended that procainamide be used instead of lidocaine, due to inadequate data concerning the treatment of arrhythmia during MH crises with lidocaine.

Malignant hyperthermia may occur outside the operating room. Gronert has described a patient who had episodic fever and a positive muscle biopsy for MH. MH may occur in the postoperative recovery room or even on return to the ward.

Classic acute episodes of MH occur a few minutes after induction, where halothane or other potent inhalation agents and/or succinylcholine are used. Tachycardia and tachypnea result from underlying sympathetic nervous system stimulation. Muscle rigidity is noticed. The patient may break through neuromuscular blockade. Increased temperature, climbing at a rate of 1–2 °C every five minutes, follows. Other signs and symptoms of MH include mental derangement, muscle edema, tenderness, oliguria, and anuria. Cardiac arrest can occur in a treated or untreated patient. The myocardium is severely affected in MH.

Several musculoskeletal abnormalities have been linked to MH (see Table 1). In fact, many clinical conditions resemble MH. All these clinical states should be considered in the differential diagnosis of MH (see Table 2). The neuroleptic malignant syndrome (NMS), for example, is characterized by the slow onset of high fever, tachycardia, increased blood pressure, acidosis, muscle rigidity, and rabdomyolysis. The muscle biopsy is positive for the caffeine-halothane contracture test. NMS occurs after long-term exposure to antipsychotic medications, e.g., butyrophenomes, phenothiazine, monoamine oxidase inhibitors, and lithium. Addonizio and Susman (1977) pointed out that electroconvulsive therapy in these patients treated with succinylcholine does not trigger MH. Also, NMS does not seem to be inherited, and there are no case reports of NMS in MH-susceptible

Table 1. Associated Musculoskeletal Disorders in Patients Susceptible to Malignant Hyperthermia

Short stocky stature
Bulky muscles with rounded belly
Muscle hypertrophy
Atrophied muscle groups
Muscle cramps
Kyphoscoliosis
Strabismus
Joint hypermobility with spontaneous dislocation
Hernias
Club foot
Pectus carinatum
Hypoplastic mandible
Poor dental enamel
Central core disease
Osteogenesis imperfecta
Neuroleptic malignant syndrome (controversial)
Myotonia congenita
Archrogryposis
King-Denborough syndrome
Wolf-Hirschhorn syndrome

families. It is believed that the changes in NMS are a reflection of dopamine depletion in the central nervous system by psychoactive drugs. Dopamine agonists are often used in treating NMS. Even though there is no evidence, patients with a history of NMS are often treated as MH-susceptible by many physicians. In thyroid crisis, pyrexia is a prominent feature. But in thyroid crisis, arrhythmia is always atrial and acidosis is metabolic. Pheochromocytoma is differentiated from MH by the lack of a metabolic response.

Table 2. Conditions That May Initially Resemble Malignant Hyperthermia

Neuroleptic malignant syndrome
Sepsis
Thyrotoxicosis
Pheochromocytoma
Myotonic dystrophy
Trigeminal neuralgia
Succinylcholine-induced rhabdomyolysis
Hypoxic encephalopathy
Intrathecal administration of constant agents
Defective temperature-monitoring devices
Stimulation during light anesthesia

DIAGNOSTIC TESTING

Most MH-susceptible people lead a normal life and their disorder is only detected by specific diagnostic tests. Moreover, thorough family history going back at least two generations should give information about any reactions to anesthetic exposures. The plasma creatinine phosphokinase (CPK) levels in the resting, fasting state are elevated in 70% of affected people. Close relations of MH patients with elevated CPK levels may be considered MH prone and may be tested further. However, negative testing is of no predictive value.

Muscle contracture testing is the current standard diagnostic test for MH. All patients with a clinical history of MH or a clinical episode that indicates a possibility of MH, should have a muscle biopsy. A histologic examination of biopsied muscle must also reveal no significant evidence of muscle destruction for the test to be valid. For the surgical procedure of muscle biopsy, a nontriggering anesthetic and no prophylactic dantrolene are advised. Calcium blocking agents should be discontinued at the time of muscle biopsy.

The North American MH group has standardized the test. The procedures and the results obtained are as follows:

1. The halothane contraction test involves exposing viable muscle fibers to nonincremental 3% halothane, bubbled through the tissue bath for 10 min. A positive halothane contraction test is defined as a > 0.2–0.7 g contraction after exposure to 3% halothane for 10 min.
2. The caffeine contracture test involves exposing viable muscle fibers to incremental doses of caffeine from 0.5–32 mM, the concentration being increased every four min if no contraction develops. A positive caffeine contracture test is defined as the development of 0.2 g tension at 2 nM caffeine, or > 7% tension change from baseline with 2 mM caffeine.
3. In a joint halothane-caffeine contracture test, both caffeine and halothane are used. A positive test is defined as the development of 1 g contracture after exposure of viable muscle fibers to a concentration of 1 mM or less caffeine in the presence of 1% halothane.

False negative muscle contraction tests are very rare. To date, a negative muscle contraction test rules out MH. A false negative test can be explained by the presence of two types of muscle fibers in a MH susceptible patient, the response being dependent on the proportion of the two types of muscle fibers. The "K-type" designation is used to describe a patient who has a positive joint halothane-caffeine contracture, but a negative separate halothane or caffeine contracture. Whether K-type individuals are MH-susceptible or not is a controversial issue.

Newer tests that have not yet been introduced as diagnostic tests for MH are based on the enhanced release of calcium on exposure to halothane, e.g., in isolated monocytes from MH-susceptible patients, higher ratio of inorganic phosphate to

phosphocreatine as measured by nuclear magnetic resonance spectrometry, and delayed recovery of adenosine triphosphate after exercise as measured by nuclear magnetic resonance spectrometry.

TREATMENT OF MALIGNANT HYPERTHERMIA

Dantrolene is the mainstay of MH treatment. It has long been available for the treatment of muscle spasm in cerebral palsy and similar diseases. It is a hydantoin derivative that was first synthesized in 1967, and reported to be effective in the treatment of porcine MH in 1975. Also in 1975, dantrolene was shown to be more effective than procainamide in the treatment of human MH, which until that time was the drug of choice. However, the intravenous preparation was not made available until November 1979. It significantly lowered mortality. The half-life of dantrolene is estimated to be 6–8 hr. Dantrolene's primary mode of action is the reduction in calcium release by the sarcoplasmic reticulum. Dantrolene also exerts a primary antiarrhythmic effect by increasing atrial and ventricular refractory periods. Side effects of dentrolene include hepatotoxicity, muscle weakness, ataxia, blurred vision, slurred speech, nausea, and vomiting. Dantrolene is not contraindicated in pregnancy, but it does cross into breast milk and its effect on the neonate is unknown.

When MH is diagnosed, the following steps should be taken immediately. All anesthetic agents should be discontinued. Hyperventilation with 100% O_2 should be instituted at high flow rates. Help should be called for at this stage. Dantrolene (10 mg/kg) should be given IV and cooling methods by whatever means available should be started. Iced IV fluids, cooling of body cavities with sterile cold fluids, cooling blankets, and heat exchange with a pump oxygenator can all be used. Research, however, favors using convection evaporation rather than ice-water baths as the most effective method. Cooling should be stopped when temperature reaches 38–39 °C to avoid inadvertent hypothermia. Changing rubber and plastic anesthesia tubing will speed anesthetic washout, as these tubes are anesthetic reservoirs. The CO_2 reservoir should be changed as early as possible during treatment. Arterial blood gases should be obtained and acidosis should be corrected with bicarbonate. Urinary output should be monitored and maintained at 1–2 ml/Kg/hr. Cardiac arrhythmias usually respond with the treatment of acidosis. If the arrhythmia persists, then procaine amide should be given. Cardiac glycosides may worsen a crisis since they accelerate calcium release from the sarcoplasmic reticulum in skeletal muscle and hence increase the efflux of calcium from muscles, e.g., cardiac muscle. Hyperkalemia should be treated slowly. Hyperkalemia responds to effective treatment with dantrolene. More recently, a study of several cases by Harioka et al. (1990) reveals the safe and effective use of calcium to counter the effects of hyperkalemia in fulminant MH in humans. But calcium use is risky in this situation. It should be reserved for treating poor cardiac function or intractable arrhythmia.

Dantrolene should be repeated after 5–8 hr. Bicarbonate, procaine amide, and other drugs should be repeated as needed. Treatment of disseminated intravascular coagulation is symptomatic. Early diagnosis and treatment of MH is essential. After effective treatment, the patient should be watched closely in the intensive care unit for recurrence of MH, myoglobinuric renal failure, disseminated intravascular coagulation, muscle weakness, and electrolyte imbalance.

ANESTHESIA FOR SUSCEPTIBLE PATIENTS

The pretreatment of MH-susceptible patients with oral or intravenous dantrolene prior to surgery in order to avoid a crisis is controversial. Most physicians do not recommend prophylactic pretreatment except in patients who have had a previously documented episode. However, if pretreatment is desired, it is recommended that therapy be begun with intravenous dantrolene in a dose of 2 mg/Kg just prior to induction of anesthesia. This prevents the uncertainty of predictive blood values associated with the use of the oral route. The adverse effects of intravenous dantrolene prophylaxis include phlebitis and tissue necrosis. Patients who receive prophylactic treatment with oral dantrolene often complain of incapacitation, gastrointestinal irritation, prolonged drowsiness, and clinically significant respiratory muscle weakness.

Generally, the patient should also be as comfortable and stress-free as possible. Premedication is obtained with agents such as midazolam, while avoiding phenothiazines and large doses of the cholinergics. Anesthesia should consist of nitrous oxide, barbiturates, opiates, tranquilizers, and nondepolarizing muscle relaxants (see Table 3). All inhaled anesthesia should be delivered through an anesthesia machine free of potential triggering agents. Despite these precautions, some patients may develop MH. Therefore, all susceptible patients should be monitored closely. Monitoring should include temperature, end-tidal carbon dioxide, arterial blood gases, and electrocardiography. Intravenous and local anesthesia are considered safe for susceptible patients and should be considered as alternatives whenever possible.

Table 3.

Drugs That Can Be Safely Used in MH	*Drugs That Should Be Avoided in MH*	*Controversial Drugs*
Barbiturates	All potent inhalation agents (except N_2O)	Calcium
Narcotics		Ketamine
Local anesthetics	Succinylcholine	Curare, Metocurine
Antibiotics	Decamethonium	Phenothiazine
Antiinflammatory drugs	Potassium	Digoxin
Pancuronum		
Atracrurium		
Vecuronium		
N_2O		
Droperidol		

SUMMARY

Malignant hyperthermia is a genetic myopathy inherited by autosomal dominant pattern with variable penetrance. Masseter muscle spasm and many neurological conditions are associated with this condition. A susceptible person, when exposed to a triggering agent, may manifest MH syndrome. The triggering agent usually is either a potent inhalation anesthetic or succinylcholine. The classic acute episode of MH occurs in the operating room a few minutes after induction. Signs and symptoms include increased temperature, tachycardia, hypercarbia, tachypnea, and muscle rigidity. Coagulopathy, myocardial failure, and renal failure follow these symptoms. Diagnosis of susceptibility to this syndrome is by family history, and by testing biopsied muscle for halothane-caffeine contracture. Dantrolene is the drug of choice for prophylaxis, as well as for the treatment of the acute crisis. Monitoring and supportive treatment should be aggressive during the crisis and should continue post-operatively. Mortality with the MH crisis has greatly decreased in the last two decades with proper understanding of this syndrome.

REFERENCES

Britt, B.A. (1985). Malignant hyperthermia. Can. Anaesth. Soc. J. 32, 666–677.

Crawford, J.S. (1972). Hyperpyrexia during pregnancy. Lancet 1, 1244.

Denboroug, M.A., & Phil, D. (1985). Current concepts of the etiology and treatment of malignant hyperthermia susceptible patients (letter). Anesthesiology 62, 99.

Denborough, M.A., Galloway, G.J., & Hopkinson, K.C. (1982). Malignant hyperpyrexia and sudden infant death. Lancet 2, 1068–1069.

Dubrow, T.J., Wackym, P.A., Abdul Rasool, I.H., & Moore, T.C. (1989). Malignant hyperthermia: Experience in the prospective management of either children. J. Ped. Surg. 24, 163–166.

Gronert, G.A. (1980), Malignant hyperthermia. Anesthesiology 53, 395–413.

Gronert, G.A. (1986). Malignant hyperthermia. In: Anesthesia (Miller, R.D., ed.), Vol. 3, pp. 1971–1994. Churchill Livingstone, New York.

Harioka, T., Sone, T., & Toda, H. (1990). Ca^{++} release channel (ryanodine receptor) of skeletal muscle sarcoplasmic reticulum. J. Biol. Chem. 265, 2244–2256.

Huckell, F.V. (1986). Cardiac manifestations of malignant hyperthermia susceptibility. Circulation 58, 916–925.

Johnson, C., & Edleman, K. (1993). Malignant hyperthermia: A review. J. Perinatology 12, 61–71.

Kalow, W., Britt, B.A., & Chan, F.Y. (1992). Epidemiology and inheritance of malignant hyperthermia. Intl. Anesth. Clinic 17, 119–140.

Kaplan, R.F., & Kellner, K.R. (1985). More on malignant hyperthermia during delivery. Am J. Obst. Gynec. 152, 5608.

Larew, R.E.B. (1989). Malignant hyperthermia, quick recognition to avoid death. Post Grad Med. 85, 117–118.

MacLennan, D.H., Duff, C., Zorgato, F., et al. (1990). Ryanodine receptor gene is a candidate for predisposition to malignant hyperthermia. Nature 343, 559–561.

Merz, B. (1986). Malignant hyperthermia: Nightmare for anesthesiologists and patients. Medical News 255, 709–715.

Moore, D.C. (1986). Ester or amide local anesthetics in malignant hyperthermia—Who knows? Anesthesiology 54, 294–296.

Ording, H. (1985). Incidence of malignant hyperthermia in Denmark. Anesth. Analg. 64, 700–704.

Rosenberg, H., & Seitman, D. (1989). Pharmacogenetics. In: Clinical Anesthesia (Barash, Cullen, & Stoelting, eds.), pp. 459–480.

Schwartz, L., Rockoff, M.A., & Koka, B.V. (1984). Masseter spasm with anesthesia: Incidence and implications. Anesthesiology 61, 772–775.

Tomarken, J.L., & Britt, B.A. (1987). Malignant hyperthermia. Ann. Emerg. Med. 16, 1253–1265.

Wingard, D.W. (1974). Malignant hyperthermia: A human stress syndrome? Lancet 4, 1450.

Zorgato, F., Fugii, J., Otsu, K., et al. (1990). Molecular cloning of encoding cDNA human and rabbit forms of Ca^{++} release channel (ryanodine receptor) of skeletal muscle sarcoplasmic reticulum. J. Biol. Chem. 265, 2244–2256.

RECOMMENDED READINGS

Fill, M., Stefani, E., & Nelson, T.E (1991). Abnormal sarcoplasmic reticulum Ca release channels in malignant hyperthermia skeletal muscle. Biophys. J. 59, 1085–1090.

Harrison, G.G. (1979). Porcine malignant hyperthermia. Intl. Anesthesiol. Clin. 17, 25–62.

Nelson, T. E. (1988). SR function in malignant hyperthermia. Cell 9, 257–265.

Mackenzie, A.E., Allen, G., Lahey, D., Crossen, M.L., Nolan, K., Mettler, G., Worton, R.G., MacLennan, D.H., & Korneluk, R. (1991). A comparison of the caffeine halothane muscle contracture test with the molecular genetic diagnosis of malignant hyperthermia. Anesthesiology 75, 4–8.

Meissner, G. (1994). Ryanodine receptor/Ca^{2+} release channels and their regulation by endogenous effectors. Ann. Rev. Physiol. 56, 485–508.

Otsu, K., Phillips, M.S., Khanna, V.K., de Leon, S., & MacLennan, D.H. (1992). Refinement of diagnostic assays for a probable causal mutation for porcine and human malignant hyperthermia. Genomics 13, 8325–837.

This Page Intentionally Left Blank

Chapter 13

The Human Heat Shock Response

DONALD A. JURIVICH

INTRODUCTION

Physiologic stress evokes a molecular response essential to protecting the human body from cellular injury and death. Over the last two decades, molecular determinants of cellular responses to stress have become better understood, as well as their relationship to human health, development, aging, and disease. Molecular re-

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 411–462.

ISBN: 1-55938-808-0

sponses to physiologic stress were first called the heat shock (HS) response, a term persistently used even though many stimuli other than heat can lead to an identical response (Hightower, 1991; Morimoto, 1993). The HS response was first observed by Ritossa who noticed a distinctive pattern of chromosomal banding after *Drosophila* cells were exposed to thermal stress (Ritossa, 1962). Suggesting that HS depleted energy storage molecules such as ATP, Ritossa tested oxidative uncouplers such as dinitrophenol and sodium salicylate (Ritossa, 1963). These also evoked chromosomal puffing, but did not directly prove the signal which triggers the HS response. In fact, one of the key questions about the HS response is how cells sense change in their external environment and transduce a signal or signals that result in altered gene expression. Once cells detect unfavorable conditions, they rapidly shift their transcriptional machinery towards the increased expression of HS genes. Although the observation was first made in *Drosophila*, all organisms appear to have the capacity for heat-inducible expression of a limited class of genes. This response generally occurs at 3–5 °C above the usual, homeostatic temperature of the organism (Lindquist and Craig, 1988). As such, the HS response has provided an excellent paradigm for analyzing inducible gene expression. Furthermore, it provides a good model for understanding how the expression of other genes is regulated during periods of cellular stress (Schmidt and Abdulla, 1988; Atkinson et al., 1990). Ultimately, through a better understanding of the molecular switching that occurs during cellular stress, we may learn new approaches to enhance cell survival and minimize cellular injury in several clinical conditions including hypoxia, inflammation, infections, and neurodegenerative disorders.

Efforts to elucidate the mechanism of HS gene expression have been advancing rapidly, and the first part of this chapter will focus on many of the seminal discoveries regarding the regulation of the HS response in cultured cells. Although much is known about the sequence of events leading to the increased expression of the HS genes, many important questions remain unanswered (Morimoto, 1993). For instance, the transcription factor mediating elevated expression of the HS genes has been extensively characterized, yet it is not known exactly how this factor or other regulatory molecules monitor the external environment in response to unfavorable conditions. Similar questions arise concerning specific functions of the HS proteins. Many of these proteins are present during usual, homeostatic conditions while preferentially accumulating during stressful conditions. It is not clear whether these proteins assume additional functions during stress or whether they perform in the same capacity during nonstressful conditions. These and other unanswered questions about basic mechanisms of the HS response attest to the complexity of molecular events initiated by physiologic stress. As we learn more about the basic mechanisms of the human HS response, many biomedical considerations arise. With this in mind, the second part of the chapter will address the biomedical relevance of the HS response, linking *in vitro* observations with aspects of human development, aging, and disease. In this section, questions common to investiga-

tions of molecular mechanisms of human health and disease are apparent. For instance, is the HS response transient or can certain conditions lead to a constitutive expression? Is there a defective HS response which is inherited or acquired, perhaps through senescence? How are cellular functions such as growth and differentiation affected by the HS response? How is the HS response linked to inflammation and autoimmunity? Is pharmacologic manipulation of the HS response possible and relevant to certain clinical goals? These questions will lend themselves to an overview of how our understanding of the HS response could have important biomedical applications.

THE HUMAN HEAT SHOCK CYCLE

Transcriptional Regulation of the Heat Shock Response

The human HS cycle can be considered broadly as a period which leads to the dramatic shift in activities of the transcriptional and translational machinery followed by eventual recovery and resumption of original activities preceding stress. Figure 1 depicts many of the key events in the HS cycle for a typical human cell line such as cervical carcinoma-derived HeLa cells. Most cells respond in an identical fashion, but some cell types that have distinctive HS responses. These differences are manifested by shifts in the relative concentrations of accumulated HS proteins and possibly in the pattern of posttranslational modifications. In all cases, however, the cellular stress response is heralded by induction of a specific transcription factor whose DNA binding activity facilitates increased expression of one or more of the stress-inducible genes.

Thermal stress has been most commonly employed for analyzing inducible responses to stress, but cellular stress can occur in many ways. Depending upon the type of stress, distinctive patterns of gene expression may be seen. For instance, glucose depletion will result in the accumulation of glucose regulated proteins (grp), whereas exposure of cells to heat leads to the accumulation of HS proteins (hsp). Thermal stress will also mildly stimulate the expression of some grp genes; thus, there is overlap in the transcriptional responses to different types of physiologic stress. HS is the best understood response to cellular stress, and thus serves as the primary model for understanding cytoprotective molecular events. Cultured cells respond almost immediately to elevated temperatures, with 42 °C eliciting the briskest reaction in most human cell lines grown at 37 °C. The temperature set-point for activation of the HS response is not absolute either between or within cell lines. For example, some cell lines may require 43 °C rather than 42 °C for maximal HS gene expression. Cells that maximally respond to 42 °C may demonstrate a similar response at 41.5 °C if grown at temperatures slightly below 37 °C (Abravaya et al., 1991a). In this case, the difference between the basal growth and the HS temperature appears to be the determining factor. The relationship of basal temperature to

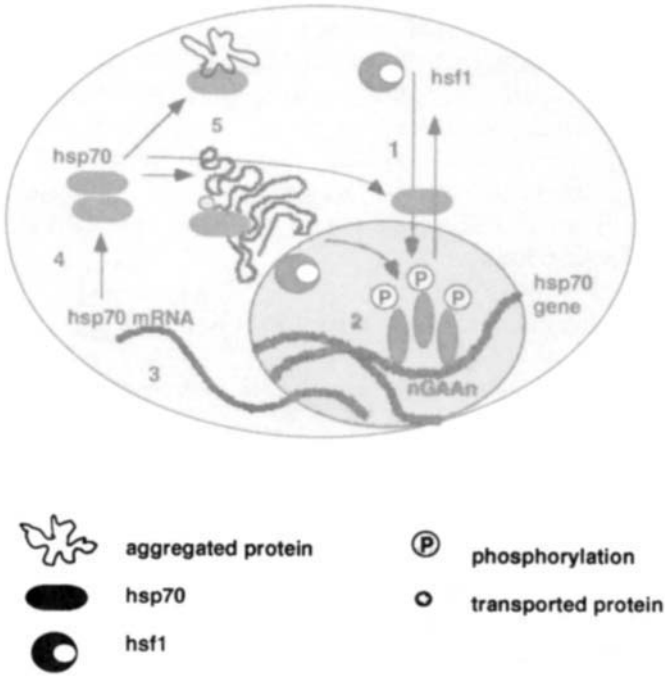


Figure 1. (1) The heat shock factor, HSF1, is folded upon itself during nonstressful conditions and is located in the cytoplasm and nucleus. Upon heat shock the cytoplasmic form enters the nucleus and associates with other monomers to form a trimer capable of binding DNA at the heat shock elements contained in the promoter region of the heat shock genes (nGAAn). (2) HSF1 is phosphorylated and increased expression of the heat shock genes occurs. (3) Intronless mRNA for the heat shock genes (e.g., hsp70) accumulates in the cytoplasm. (4) hsp70 is synthesized and accumulates as a dimer or it interacts as a monomer with other proteins. (5) hsp70 helps dissociate aggregated protein, acts as a chaperone, and helps in the folding of new and previously existing proteins. It may assist HSF1 in returning to its non-DNA binding form during recovery from heat shock.

temperatures that induce the HS response is not fully understood (Clos et al., 1993). Somehow the cell is able to adjust its sensing mechanism and triggers the HS response according to the ambient temperature. For example, cells that have previously been heat shocked require at least a full degree higher than the preceding HS to evoke the same level of response. This *in vitro* phenomenon may be important to certain clinical conditions such as thyroid disease where basal body temperatures may shift above and below 37 °C. Hypothyroidism leading to lower basal body temperatures could cause cells to be hypersensitive to activation of the HS response

by fever or other inducers of HS gene expression. On the other hand, hyperthyroidism and elevated body temperatures could desensitize cells to a full HS response, perhaps increasing the possibility of cellular or tissue injury. Despite a plethora of knowledge about the human HS response from cultured cells, little is known about how this response is regulated in the human body. The possibility that basal body temperatures play a role in the inducibility of the human HS response is intriguing. Factors other than core body temperature may intersect with the cellular stress response. For instance, Holbrook has found endocrine determinants of the HS response (Blake et al., 1991). Furthermore, different tissues elaborate different levels of HS gene expression despite the same level of stress (Blake et al., 1990). The basis for discordant expression of HS genes among tissue is yet undetermined, but it does point to local tissue factors in regulating the level of HS gene expression. This observation emphasizes the complexity of the HS response. It also points out the vigilance needed when interpreting data from cell culture studies and extrapolating its potential relevance to regulation of the HS response *in vivo*.

One of the most difficult aspects of relating data from *in vitro* studies to molecular responses in the human body is the severity of temperature required to induce HS gene expression in cultured cells. Forty-two degrees Celsius is an extraordinarily high temperature, and rarely, if ever, is this level of heat realized. In most instances, temperatures at or below 40 °C are encountered in clinical conditions, and this level of heat is generally insufficient to drive the HS response in cultured cells. Nevertheless, HS proteins are found to accumulate in a variety of clinical conditions, thus suggesting that factors other than heat play a role in the human HS response. Animal studies have shown that exercise-generated heat is sufficient to induce the HS response (Salo et al., 1991). Other examples of how the HS response is generated in the human body at slightly elevated temperatures are more indirect. For instance, inflammation may influence the temperature at which the HS response is activated since increased levels of HS proteins are found in degenerative arthritis (Kubo et al., 1985). Recent studies with cultured cells favor this possibility. If cells are exposed to levels of arachidonate found during inflammation, they exhibit an increased sensitivity to the heat-inducible expression of the HS genes (Jurivich et al., 1993). Thus, instead of the usual 42 °C necessary to induce HS gene expression in cultured HeLa cells, exposure of these cells to arachidonate leads to activation of the HS transcription factor at temperatures of 39 °C to 40 °C. The effect of arachidonate on activation of the HS transcription factor raises the possibility that free fatty acids and, perhaps, other regulatory molecules intersect with the HS response and influence the temperature at which the HS response occurs. Thus, extremely high temperatures used with cultured cells may not be required by the human body to initiate the HS response.

As more information is obtained about the relationship of temperature levels and the HS response, further analysis is needed to determine the relationship between inducing the HS response in one part of the body and other organ systems. Insight

into this question may be derived from plant studies. When plants are confronted with a viral infection within a leaf, a local and a systemic protective response similar to the HS response is elicited (Malamy et al., 1990). Part of this response includes the local and systemic accumulation of salicylate, a compound which signals cytoprotective gene expression in plants. Curiously, salicylate also affects the human HS response (Jurivich et al., 1992). Based upon this observation, we can postulate that endogenous factors produced by the human body may resemble the signal plants produce when under stress of a viral infection. Acute phase reactants circulating in the body in response to infection and inflammation may provide a comparable effect. For instance, γ-interferon and the antiviral response has been associated with the increased accumulation of HS proteins (Dubois et al., 1988). Thus, temperature alone may not dictate the human HS response. Rather, induction of the HS response is apt to occur from multiple stimuli which converge upon a common pathway leading to HS gene expression.

Induction of the HS response can occur by a variety of agents without elevating the ambient temperature. Other inducers of the HS response at 37 °C are listed in Table 1. Even though the HS response can be induced by a variety of stimuli, the kinetics of the response are not always identical. For instance, heat induces the HS response within minutes, whereas heavy metals induce the response only after several hours (Mosser et al., 1988). Some of the clinically relevant inducers of the HS response are hypoxia, oxidant injury, and certain eicosanoids (Christman et al., 1985; Currie, 1987; Amici et al., 1992). All of these inducers are thought to cause protein damage and protein unfolding which in turn activate the HS response through specific induction of protein-DNA binding in the promoter region of the HS genes (Kelley and Schlesinger, 1978; Hightower, 1980; Ananthan et al., 1986). Microinjection of denatured proteins into intact cells will induce the HS response in the absence of heat, but it is not known how the accumulation of damaged protein specifically activates a heat-inducible transcription factor (HSF) into a state which allows it to bind to specific DNA sequences known as the HS elements. The

Table 1. Inducers of the Heat Shock Response

Stress
Heat–fever
Heavy metals
Oxidant injury
Oxidative uncouplers
Ischemia
Infection (viral, bacterial, and parasitic)
Inflammation (prostaglandins a, d, j, and arachidonate)
Drugs (nonsteroidal anti-inflammatory, antineoplastic agents, alcohol)
Denatured proteins (amino acid analogs)
Nonstressful Conditions
Growth stimuli
Differentiation
Mitosis

induction of HSF-DNA binding appears to be complex (Wu et al., 1990). At very high temperatures HSF-DNA interactions occur independently of protein synthesis, whereas at lower HS temperatures, HSF-DNA interactions are dependent upon protein synthesis. The effect of cycloheximide at low and high HS temperatures has suggested that disruption of nascent polypeptides is the primary signal for the HS response at low temperatures, while damage of pre-existing proteins is the signal at high temperatures.

Once activated, the response can be dissected into three phases: transcriptional, translational, and recovery (DiDomenico et al., 1982). The transcriptional phase is characterized by activation of protein-DNA interactions in the promoter region of the HS genes. Specifically, the HS transcription factor (HSF) is transformed into its DNA binding state (Kingston et al., 1987; Larson et al., 1988; Goldenberg et al., 1988). The heat-inducible HSF has been designated HSF1 in contrast to other members of this transcription factor family (HSF2 and HSF3) (Sarge et al., 1991; Baler et al., 1993; Nakai and Morimoto, 1993). HSF1 is the primary transcription factor activated by heat (Clos et al., 1990). The role of the other HS factors during stress is not clear, and they may serve other functions under nonstressful conditions. For example, HSF2 is induced when K562 erythroleukemic cells are differentiated by hemin (Sistonen et al., 1993). In many cells examined thus far HSF1 pre-exists in the cytoplasm and nucleus (Sarge et al., 1992). Heat shock does not induce the synthesis of this transcription factor but does alter its conformation and increases its affinity for DNA. HSF1 is a leucine zipper DNA binding protein that contains three leucine zippers near the amino terminus and one leucine zipper near the carboxyl terminus (Figure 2). The arrangement of these leucine zippers is thought to contribute to the regulation of HSF1 during nonstressful conditions, as well as provide the appropriate motif for specific interactions with the HS gene promoter region (Rabindran, 1993). Lower organisms lack the hydrophobic heptad repeats in the COOH terminal region and, unlike human cells, their HSF is constitutively bound to DNA (Sorger et al., 1987). This observation has suggested that the COOH-terminal region suppresses aggregation of human HSF1, perhaps through intramolecular interactions with the other hydrophobic regions near the amino terminus (Rabindran et al., 1993). Deletion mutations also substantiate the putative role of the COOH-terminal region in autoregulating HSF1 multimerization. Removal of either the heptad repeat or a 12-residue element in the COOH terminus results in constitutive HSF1-DNA binding. This demonstrates at least two areas of

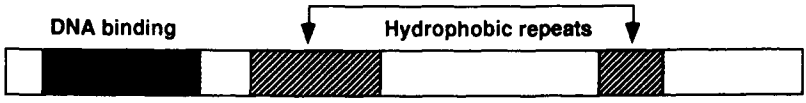


Figure 2. Schematic of HSF1 map depicting the DNA binding region at the amino terminus and the regulatory regions for protein-protein interactions in the COOH-terminus.

HSF1 necessary for regulating its non-DNA binding state. The non-DNA binding form of HSF1 is suggested to be a globular protein resembling a horseshoe conformation.

How hydrophobic interactions within HSF1 are disrupted leading to its unfolding and oligomerization is not known. If recombinant HSF1 is overexpressed in *Escherichia coli*, it spontaneously forms higher order complexes with DNA binding potential at 37 °C (Sarge et al., 1993). Thus, HSF1 does not appear to have the ability to spontaneously suppress its oligomerization when high concentrations of the protein accumulate in bacteria. This suggests that some negative regulator of HSF1 is absent in bacteria. It has been proposed that this negative regulator is one or more of the HS proteins (Morimoto et al., 1992; Mosser et al., 1993). These end-products of the heat shock response are thought to participate in the refolding of HSF1, returning it back to a non-DNA binding state. Proof of the HS proteins causing the refolding of HSF1 is indirect. An interaction between HSF1 and hsp70 is apparent on electromobility shift assays (EMSA) when antibody to hsp70 causes retardation of the HSF1-DNA complex. However, other assays do not capture the putative interaction of HSF1 with hsp70, thus suggesting that the interaction, if it exists *in vivo*, is transient (Westwood and Wu, 1993). Whether proteins other than hsp70 act as a negative regulatory molecule for HSF1 *in vivo* is not certain.

Upon HS, the monomeric form of HSF1 is thought to unfold from a horseshoelike structure after interactions between the amino and carboxy terminal leucine zippers are disrupted. Posttranslational modification of HSF1 is unlikely to contribute to oligomerization since anti-HSF1 antibodies can activate HSF1 (Zimarino et al., 1990b). How self-dissociation of HSF1 is achieved is not known, but the elongated HSF1 has the ability to interact with other HSF1 molecules via hydrophobic-heptad repeats near the DNA binding pocket and to form multimeric structures which acquire DNA binding potential. Hydrophobic residues at the COOH terminus may also interact with C-terminal residues from other HSF1 molecules. Unlike most transcription factors which form either homo- or heterodimeric units, HSF1 appears to become a homotrimer (Sorger and Nelson, 1989). Higher orders of oligomerization may also exist for HSF1, but the trimers are sufficient for binding to specific regions of DNA known as the HS elements (HSEs) (Zimarino and Wu, 1990). HSEs are nGAAn arrays located upstream of the transcriptional start site for the HS genes. Some members of the HS gene family have HSEs in more than one location (Abravaya et al., 1991a). Analysis of the hsp70 gene has revealed at least two pentameric arrays of HSEs. The proximal HSE is sufficient for the induction of hsp70 gene expression, and it remains to be seen whether the distal HSE has a role in the regulation of this particular HS gene.

During 42 °C HS, HSF-DNA binding can be detected within minutes. Maximal activity peaks around 30–40 minutes. One model for HSF1 activation suggests that hsp70 or hsc70 acts as a brake for HSF-DNA binding activity, which is released when denatured or damaged protein accumulates during stress and competes with

HSF for hsp70 binding (Baler et al., 1992; Morimoto, 1993). This model has been reinforced by the observation that microinjection or electroporation of denatured proteins into cells induces HSF1-DNA binding. However, extracts made from cells grown at 37 °C do not manifest HSF-DNA binding activity if the extract is combined with denatured proteins. The lack of inducibility by denatured protein of HSF oligomerization *in vitro* can be interpreted in a couple of ways. First, HSF oligomerization may occur *in vivo* after a signal arises when denatured proteins accumulate. This may be an ionic signal or a change in the redox potential of HSF. Some type of posttranslational modification of HSF is unlikely to account for the oligomerization process. So far, phosphorylation is the only known posttranslational modification of HSF and it is known that oligomerization can occur in the absence of HSF phosphorylation (Sorger and Pelham, 1988; Jakobsen and Pelham, 1988; Sorger, 1990). Glycosylation does not appear to regulate HSF oligomerization, although fatty acylation/sulfonation have not been ruled out. When recombinant HSF protein is overexpressed in bacteria, it assumes both monomeric and multimeric forms, even though there is no apparent stress which would account for the induction of HSF oligomerization. One might argue that high levels of any protein expressed in bacteria are a type of stress when artificially induced by IPTG. However, higher levels of bacterial HS proteins are not observed, thus suggesting that HSF can oligomerize in the absence of stress. It is intriguing that HSF oligomerization can occur at 37 °C and that this process is not strictly a thermal-dependent event.

Perhaps the strongest model of HSF activation is that the transcription factor itself is the sensor of stress. Rather than evoking a model in which HSF1 dissociates from a regulatory molecule, HSF1 activation could occur simply through the disassociation of intramolecular bonds. The globular form of HSF then becomes elongated, which in turn exposes hydrophobic residues for interactions with other unfolded HSF molecules. Because there has not been convincing evidence of HSF-protein interactions in the nonstressed state of the cell, it is unlikely that damaged or denatured proteins act as a direct signal for HSF oligomerization. This observation raises the possibility of secondary signals that induce HSF oligomerization. To address this problem, experiments have been undertaken which attempt to induce oligomerization of HSF *in vitro*. HSF1 isolated from unstressed cells can be exposed to a variety of agents which affect HSF1 conformation (Mosser et al., 1990). Curiously, heat itself is a poor inducer of HSF oligomerization *in vitro*. Denatured proteins, unlike the *in vivo* situation, do not induce HSF oligomerization. The most potent inducer of HSF1-DNA binding is a nonionic detergent, NP-40. The ability of NP-40 to activate HSF1 *in vitro* may be through its effect on molecular interactions within HSF1. Disruption of intramolecular hydrophobic interactions could give way to intermolecular ones which allow HSF multimerization. Calcium and acidic pH also appear to convert the monomeric into the trimeric form of HSF *in vitro*. The fact that calcium and alterations in pH stimulate

HSF-DNA binding in whole cell extracts suggest that denatured proteins may trigger ionic fluxes that regulate HSF-DNA binding. Although calcium would appear to be a good second messenger candidate, calcium ionophores induce glucose-regulated proteins and not HS proteins. Thus, fluxes in intracellular calcium appear to be insufficient to induce HSF-DNA binding. Experiments interfering with the proton pump or changing the extracellular pH have not led to the induction of HSF-DNA binding. In some instances, however, an acidic extracellular environment will enhance the induction of HSF-DNA binding by compounds such as sodium salicylate (Jurivich et al., 1992). Although low extracellular pH enhances intracellular levels of salicylate, this compound has the property of reducing intracellular pH. Thus, intracellular ionic fluxes may have a potentially important role in regulating the conversion of HSF from its monomeric to a trimeric form. There is little evidence to support other known signaling molecules in HSF oligomerization. Induction of diacylglycerol and cAMP have not led to the induction of HSF-DNA binding, and it remains to be seen if any second messengers are involved in triggering HSF-DNA binding.

Signaling cascades during cellular stress are likely to be complex. Not only is there a signal for HSF oligomerization, but there is a signal for HSF phosphorylation which may not be the same as the oligomerization signal. Phosphorylation of HSF has been observed in a number of species and its role in regulating the HS response is under investigation. In yeast, HSF is constitutively bound to HS elements in the promoter region of the HS genes, and transcriptional activation of the HS genes accompanies HSF phosphorylation (Sorger et al., 1987). In human and other mammalian cells, HSF-DNA binding is induced by stress, and at some point, phosphorylation of HSF is increased either before, during, or after HSF binds to DNA. What is known thus far is that increased phosphorylation is not required for oligomerization of HSF. This has been observed when mouse HSF is overexpressed in *E. coli* (Sarge et al., 1993) and when HSF-DNA binding is activated by salicylate in human cells (Jurivich et al., 1993). Because there is little or no transcription of the HS genes with sodium salicylate-inducible HSF-DNA binding, it is likely that HSF phosphorylation contributes to the transcriptional proficiency of RNA polymerase II (Jurivich et al., 1992). Other investigators have also found that HSF-DNA binding in murine cells is insufficient to drive HS gene transcription (Hensold et al., 1990). Collectively, these data suggest that HS gene expression requires more than the mere occupation of HSF1 on the HSE promoter.

Components other than HSF contributing to the overall regulation of the HS genes undoubtedly include alterations in or around the RNA polymerase complex (Lis and Wu, 1993). RNA polymerase II is poised near the transcriptional start site of HS genes (Gilmour and Lis, 1986) and its processivity apparently is dependent upon DNA-protein and protein-protein interactions in the promoter region of the HS genes (Rougvie and Lis, 1988). HSF phosphorylation may contribute to

protein-protein interactions or possibly DNA bending which is favorable to RNA polymerase activation and/or recruitment.

The intracellular location of HSF phosphorylation is not certain because in most cells examined thus far, there are two pools of HSF, one in the nucleus and the other in the cytoplasm. During HS the cytoplasmic pool joins the nuclear pool. This process may require molecular chaperones: e.g., hsp70/hsc70 are observed to cotranslocate into the nucleus at the same time that HSF1 moves from the cytoplasm into the nucleus. The factors that lead to compartmentalization of HSF1 are not known. Western blots of HSF1 from cells at 37 °C show at least two molecular weight species of HSF1, and it may be that the two forms represent different cellular fractions of HSF1. Cell fractionation studies are wrought with problems of intranuclear contents leaking into the cytoplasmic fraction; thus, it may be difficult to determine which of the HSF1 forms are represented in the nucleus and cytoplasm. One possibility is that intranuclear phosphorylation of HSF contributes to a charge effect which retains the transcription factor in the nucleus. Charge effects are known to contribute to compartmentalization and in some instances exclusion from intracellular compartments. Whether HSF becomes phosphorylated as a multimer or as a monomer is not known. Because the monomeric form of HSF1 is partially phosphorylated it may be that both the monomer and dimer are substrates for protein kinase activity.

Phosphorylation of HSF1 is rapid, and one or more protein kinases are likely to be activated upon HS. An alternative explanation is that heat inactivates a phosphatase which is more active than the HSF kinase at 37 °C. Inactivation of the phosphatase by heat would allow the presumptive heat stable HSF kinase activity to predominate, thus increasing the phosphorylation of HSF1.

HSF1 phosphorylation must be sensitive to nonheat inducers of HSF-DNA binding activity because HSF1 phosphorylation can be achieved at 37 °C by other inducers of the HS response. HSF1 contains polypeptide sequences that could serve as substrates for well characterized protein kinases, but few of these are known to be heat inducible. One family of protein kinases, the S6 protein kinases, have already been shown to exhibit heat inducible activity; however, their peak level of activity during HS occurs well after the maximal induction of HSF phosphorylation (Jurivich et al., 1991). Thus, other protein kinases are likely to be directly linked to the phosphorylation of HSF. Some of the putative protein phosphorylation sites on HSF include motifs for protein kinase C, casein kinase, and enterokinase. There are tyrosine sequences that match substrates for known tyrosine kinases, but whether these residues are accessible to phosphorylation is not established.

When considering the role of phosphorylation in the regulation of the HS response, it is indeed curious that oxidative stress and heat induce a protein tyrosine phosphatase at the transcriptional level (Keyse and Emslie, 1992). Whether this phosphatase has any role in the regulation of HSF phosphorylation is not known, but it does indicate that both transcriptional and translational regulation of signaling

events occur during cellular stress. Induction of a phosphatase by heat at the transcriptional level offers a plausible explanation of how HSF may be down-regulated during recovery of cells from stress. The rapidity with which HSF is phosphorylated makes it unlikely that stress induces synthesis of a HSF kinase. Rather, posttranslational modifications of preexisting HSF kinase are likely to contribute to HSF phosphorylation.

Phosphorylation of HSF substantially enhances the transcriptional activity of HS gene expression which may be up to 100-fold of basal levels after HSF1 binds to the promoter element. Heat shock will increase the C-terminal-domain-kinase activity in cell extracts, and this action may enhance the activity of RNA polymerase II that is bound to HS genes (Legagneux et al., 1990). Whether this kinase activity also affects HSF1 phosphorylation is not known, but increased HS gene expression appears to occur as long as HSF1 is bound to the promoter region. The CTD kinase complex contains multiple proteins, and it is quite possible that one or more of these proteins is also regulated by stress.

In contrast to the heat inducible expression of the HS genes, constitutively low expression of the HS genes does occur when HSF1 is not bound to the promoter region (Figure 1). This low level of HS gene expression is due to other cis-acting elements contained in the promoter region of the HS genes (Wu et al., 1986, 1987; Schiller et al., 1988). One of the unique features of hsp70 gene regulation is that it has a bipartite regulatory region that is responsive to growth signals as well as to stress signals (Wu and Morimoto, 1985; Phillips and Morimoto, 1991). The proximal promoter region which includes SP1, CCAAT, and ATF binding sites allows for hsp70 gene expression during growth and differentiation. The distal promoter region contains the HSEs which regulate the stress-inducible expression of the HS genes. Deletion mutations have demonstrated that the proximal promoter is not essential for heat-inducible expression of the hsp70 gene, but it does contribute to the maximal expression of this gene during stress (Williams and Morimoto, 1990). More recently, protein binding to GAGA sequence elements facilitates HSF-DNA binding (Tsukiyama et al., 1994). GAGA factor is proposed to displace histones or to disrupt nucleosomes which in turn potentiate HSF binding to the HS promoter (Lis and Wu, 1993). Thus, a series of events involving protein-DNA interactions, as well as protein-protein interactions, are essential for full transcriptional activation of the HS genes.

While activation of HSF1 and HSF2 results in the increased expression of HS genes, recent studies indicate that there are distinct differences between these transcription factors. As noted previously, HSF1 is the heat-inducible form of the HS factors, whereas HSF2 is activated by hemin (Sistonen et al., 1992). This observation has indicated a possible role of HSF2 in cellular differentiation. It also indicates that the activation domains of the two transcription factors are quite different. Presumably the DNA binding properties of the two transcription factors are similar, but recent experiments indicate otherwise (Kroger and Morimoto,

1993). Gel shift analysis has shown that three binding elements are sufficient for DNA-protein binding *in vitro*, whereas DNA footprint analysis suggests that at least four and preferably five sites are required for HSF1 binding to DNA. Interestingly, HSF2 prefers two or three nGAAn binding sites. This observation raises the possibility that functional diversification of the HS factors is based upon variations in their DNA binding properties as well as differences in their activation domains. Differences in the activation domains of the HS factors are just beginning to be realized. First, HSF2 does not become phosphorylated like HSF1. Second, the control form of HSF2 appears to reside as a dimer rather than as a monomer. Although there are differences between HSF1 and HSF2, their formation of oligomers with at least three DNA binding domains poses a curious paradox in which an odd number of DNA binding domains are able to bind at DNA sites with two-fold symmetry (Lis and Wu, 1993). Another unresolved question concerning activation of the HS factors is whether they can form heteromeric complexes like other types of transcription factors. For instance, the transcription factor AP-1 contains heteromeric complexes of fos and jun, as well as homodimer jun-jun interactions (Angel and Karin, 1991). Studies have shown to date that HSF exists in homomeric states, but this does not exclude the possibility that they interact with other leucine zipper transcription factors. Finally, it is known that HSF binds to multiple sites on eukaryotic chromosomes (Westwood et al., 1991). In many cases, this represents transactivation of HS genes, but given the limited number of heat inducible genes, it raises the possibility that repression of other genes occurs through HSF-DNA interactions.

By convention, the HS genes are designated by the molecular weight of their corresponding protein. There are roughly three broad classes of heat shock genes: high molecular weight (90–120 kDa), average molecular weight (60–70 kDa), and

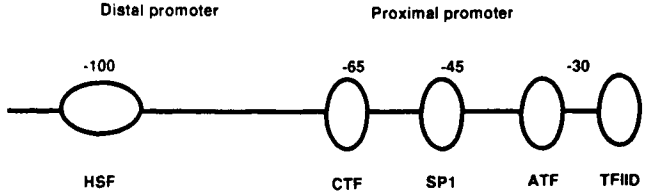


Figure 3. Schematic representation of the human hsp70 gene promoter. HSF = binding to heat shock elements (nGAAn), CTF = binding to CCAAT, SP1 = GGCGGG binding elements, ATF = TGACGTCA binding elements (activator protein-I), TFIID = TATA binding elements. The proximal promoter is sensitive to induction by serum, IL-2, forskolin, TPA, E1A protein from adenovirus, and herpes simplex infection. The distal promoter is responsive to many forms of stress, including heat, ethanol, heavy metals, prostaglandins, arachidonate, and hypoxia.

Table 2. The hsp70 Gene Family

Gene	*Location*
hsp70	cytoplasmic, nuclear
hcs70	cytoplasmic, nuclear
hsp72	nuclear
hsp75	mitochondrial
grp78	endoplasmic reticulum

low molecular weight (10–50 kDa) species. Generally, the 70 and 90 kDa HS genes are the most highly inducible, but this may vary among cell lines and tissue. Differences between cells in their heat inducible expression of certain HS genes may reflect special needs of cells for certain HS proteins. Likewise, the basal expression of HS genes varies from cell to cell. Mosser found that overexpression of hsp70 can lead to an abbreviated period of HSF-DNA binding, and it is possible that pre-existing levels of hsp70 help regulate the magnitude and duration of the HS response (Mosser et al., 1993).

HS genes are represented on several chromosomes, and some, such as the hsp70 gene, are found on multiple chromosomes (Harrison et al., 1987). The hsp70 gene family is among the most well characterized of the HS genes. There are five and perhaps more members of this gene family (Table 2) which correspond to approximately 10 comparable gene sequences. Southern blot analysis has revealed two genes each for p72 and grp78. Hsp70 is represented by five genes. The hsp70 genes are encoded on chromosomes 6, 14, and 21. Hsp70 maps on chromosome 6 within the HLA class III region, which also contains the genes for complement and tumor necrosis factor (Goate et al., 1987; Sargent et al., 1989). The location of hsp70 on chromosome 14 is particularly interesting because familial Alzheimer's disease maps to this locus, suggesting a possible involvement of the HS genes in neurodegenerative diseases (Schellenberg et al., 1992). Other HS genes show a similar diversification and distribution throughout the genome. The purpose of this redundancy is not apparent, nor is it known whether all loci for the HS genes are transcriptionally active. Presumably, chromosomal location of a particular HS gene dictates its level of induction.

As more is learned about the chromosomal effects on HS gene expression, it is important to point out that these genes are actually a subset of inducible responses to cellular stress. Not all of these inducible responses involve HSF, and this indicates that cells have diversified transcriptional responses to cope with different types of stress. This diversification is manifested by glucose regulated genes (grp), as well as the metallothionein and oxidant-injury genes (Watswich, 1988; Storz et al., 1990; Devary et al., 1992; Skroch et al., 1993; Xu, 1993).

Translational Control of the Heat Shock Response

While transcriptional regulation of the HS response has been the center of many studies, posttranscriptional regulation of this response is equally important (Yost and Lindquist, 1986; Yost et al., 1990). Human hsp70 mRNA is from an intronless gene, thus mRNA splicing is not required during HS. In fact, mRNA splicing is inhibited by heat and the degree of inhibition is proportional to the intensity of HS. Many mRNAs that have intervening sequences and require splicing are not duly processed (Yost and Lindquist, 1986). This allows for the preferential accumulation of hsp70 mRNA and conserving the translational apparatus for the nearly exclusive expression of the HS proteins. The block in mRNA splicing persists even after recovery from HS. Interestingly, cells that have fully recovered exhibit the ability to splice mRNA after another HS stimulus. The implication here is that the accumulation of HS proteins is protective of the splicing machinery, a property that may contribute, in part, to the phenomenon of thermotolerance. This observation represents a more generalized phenomenon in which usual cellular activities are shutdown by an initial HS but through a process of adaptation they become functional if stress reoccurs. The duration of this adaptation to stress either *in vitro* or in the human body is not known.

HS also affects ribonucleoproteins. Overall, there is a decrease in the protein/RNA ratio. Antibodies against certain ribonucleoproteins have been used as a probe in heat shocked cells and these demonstrate a loss of some RNA particles normally seen at 37 °C.

Transcripts synthesized during HS exhibit longer lengths than those synthesized at 37 °C. This is due to an extensive poly A tail which further demonstrates the effect of HS on RNA processing. Some hsp70 mRNA pre-exists in a polysome fraction at 37 °C which is a likely mechanism to rapidly generate hsp70 even before transcriptional regulation of the gene is in full force. Most polysomes at 37 °C are disassociated and replaced with polysomes containing mRNA from HS genes. The disappearance of pre-existing polysomes is proportional to the level of HS. Interestingly, HS does not appear to affect the efficiency of translation (Lindquist, 1980). Ribosomes accumulate on hsp70 mRNA at the same rate as other mRNAs at 37 °C.

The HS-induced block in translation of pre-existing mRNA is multifactorial (Spradling et al., 1977). There does not appear to be an overt modification of the mRNA nor does there appear to be a heat-induced melting of the pre-existing mRNA. Heat shock blocks translation of pre-existing mRNA at the level of elongation while in other instances at the level of initiation. The 5′ end of the hsp70 gene evidently is responsible for ribosomal recognition during HS and may account for the preferential accumulation of this message by the translational machinery during HS. Deletion of the 5′ untranslated leader results in the absence of translation

of hsp70 mRNA at HS temperatures. At physiologic temperatures, the mutated gene will be translated.

Heat-inducible changes in the translational machinery may also account for the dramatic switch to hsp synthesis (McGarry and Lindquist, 1985). How stress modifies translation is not certain. Originally it was thought that dephosphorylation of a ribosomal protein during HS led to the preferential expression of hsps, but phosphorylation of this protein was later found to be induced by HS in mammalian cells (Jurivich et al., 1991). Other ribosomal proteins are modified by HS, but none of these modifications have been conclusively linked to altered ability of the ribosomes to translate hsp mRNA. One set of experiments has indicated that while most mRNAs require cap binding proteins, hsp mRNA may be excluded from this requirement. The high adenine content of hsp mRNA has been cited as a possible means to minimize secondary structure, and thus circumvent RNA unwinding activity which is potentially altered by HS. Finally, a specific protein may preferentially recognize the 5′ end of the hsp mRNA and not other mRNA, thus shepherding it to the translational machinery which has somehow been modified to recognize this RNA-protein complex (McMullin and Hallberg, 1986).

Heat shock stabilizes the levels of mRNA for the heat shock genes by increasing the half-life of the messenger RNA. Hsp mRNA degradation upon recovery is largely dependent upon the degree of HS. The higher the level, the slower the degradation of hsp mRNA during recovery. Presumably, this is a means to sustain hsp synthesis proportional to the degree of injury. That is, high concentrations of hsps are needed for repair when cell injury is extensive. Thus, levels of hsp synthesis is a tightly regulated process at both the transcriptional and translational level. Finally, it is important to note that some non-HS genes such as c-fos are also regulated by heat through a posttranscriptional mechanism. The significance of elevated c-fos expression during heat shock is not known, and the effects of heat on c-fos mRNA contrasts with most other mRNAs which exhibit a shorter half-life during HS than during growth at normal temperatures (Gubits and Fairhurst, 1988).

Heat Shock Protein Functions

Hsp70 is the first hsp to be synthesized during HS and it is the most prevalent of all HS proteins. This suggests that it has a central role in protecting cells from injury during stress. Recent investigations have focused on the function of this and other HS proteins (Pelham, 1990). Clearly, induction of hsps is associated with thermotolerance (McAlister and Finkelstein, 1980). Cells exposed to a nonlethal HS accumulate hsps and then become resistant to an otherwise lethal stress. The overexpression of the hsp70 gene by stable transfection has increased cellular resistance to injury by heat, tumor necrosis factor, and oxidant injury (Li et al., 1991, 1992). Other hsps may also provide this level of protection (Sanchez and Lindquist, 1990; Bansal et al., 1991). Extending observations on the cytoprotective

action of HS in cells, investigators have found that organs can be protected from damage by a prior HS. This observation was made with retina which were protected from light-induced damage after hyperthermia (Barbe et al., 1988). Hsps are thought to contribute towards thermotolerance by assisting in the refolding of proteins and protecting nascent polypeptides from denaturation and aggregation. However, experiments suggest that HS proteins may not have a role in the induction of thermotolerance (Andersen et al., 1986). Cycloheximide treatment during HS blocks hsp synthesis and yet does not prevent cells from becoming thermotolerant. Acquisition of thermotolerance under these circumstances is explained by enhanced function of some HS proteins after they become phosphorylated, e.g., hsp27 (Landry et al., 1992).

Understanding the function of hsps has been the focus of numerous investigations (Pelham, 1986). Each HS protein family appears to have specific functions, yet studies from bacteria indicate that interactions between the hsp families occur in a coordinated fashion (Goloubinoff et al., 1989; Silver and Way, 1993). This assemblage of hsp complexes appears to be essential for some hsp functions (Langer et al., 1992). How a HS protein interacts with another hsp is not clear, but presumably hsp-hsp interactions provide some level of protein conformational changes essential to their function (Perdew and Whitelaw, 1991). Posttranslational modification of some HS proteins may also be a means to alter their function. For instance, heat-inducible phosphorylation of the bacterial HS protein groEL changes its protein binding property and enhances the release of unfolded proteins by ATP hydrolysis (Yu et al., 1992).

Some of the observed functions of hsps include protein folding (Hartl and Neupert, 1990; Gething and Sambrook, 1992) and dissociating protein aggregates (Braell et al., 1984; Ostermann et al., 1989; Beckmann et al., 1990). When polypeptides are synthesized on the ribosomes, HS proteins immediately interact with the nascent polypeptide chains (Bochkareva et al., 1988; Flynn et al., 1989). This interaction may protect the new polypeptide from degradation by proteases prior to folding into a competent protein. The hsp-polypeptide interaction may also be necessary for proper folding of the protein. Older axioms that the primary sequence of a polypeptide dictates its tertiary structure are now being modified with observations that certain classes of proteins require hsp-assistance for final conformation (Agard, 1993). In addition to folding newly synthesized proteins, HS proteins also recognize denatured proteins. Depending upon the degree of protein damage, the hsp may assist in the refolding of the protein or it may shuttle it into the protein degradation pathways (Hershko, 1988). In fact, a small molecular weight class of HS proteins called ubiquitin are utilized as tags for targeting certain proteins for degradation, and other HS proteins are utilized in the processing of proteins targeted for lysosomal degradation (Chiang et al., 1989).

Perhaps one of the most celebrated functions of hsp is molecular chaperoning (Chirico et al., 1988; Ellis, 1987; Craig, 1990, 1993; Lis and Wu, 1993). Protein

translocation through intracellular membranes is the result of hsp-protein interactions (Deshaies et al., 1988). Hsp70 and hsc70 are capable of transporting proteins into the nucleus (Shi and Thomas, 1992), but one of the major questions in the field of chaperones is how intracellular traffic is directed. Compartmentalization of some proteins is based upon leader sequences specific for a certain organelle, but other proteins lack these identifying sequences, and it is not clear how they are retained in a specific organelle. Hsp-assisted translocation could assure protein retention in the organelle, and in some instances, posttranslational modification of the protein within the organelle is responsible for its localization. How the hsp initially matches a protein with its targeted organelle is of great interest. Several clones of the hsp70 gene family have been found in mitochondria and the endoplasmic reticulum. These organelle-specific hsps may act as platforms for accepting proteins delivered by another hsp from the cytoplasm.

In analyzing hsp-protein interactions, the association of hsp70 with the tumor suppressor protein p53 has been of great interest (Clarke et al., 1988; Findlay et al., 1988). Hsp70 has been found to alter p53 conformation *in vitro* and the interaction is dependent upon the carboxy-terminus of p53. The interaction of hsp70 with the carboxy terminus of p53 is important because this is the region of p53 responsible for oligomerization and nuclear localization. The role of hsp70 in regulating p53 assembly or disassembly has been suggested but unproved. The potential role of hsp70 in regulating p53 conformation has been cited as one explanation of how this factor can regulate cell growth. Mutant p53 is often observed in a variety of cancers, and it has been proposed that mutant p53-hsp70 interactions jam an otherwise fluid regulatory system for cell growth. This interaction suggests a regulatory role for hsp70 at the transcriptional level. Hsp-regulatory protein interactions may either block or assist protein-protein interactions necessary for activating RNA polymerase II. As HS proteins appear to preserve intracellular protein integrity, they may also enhance certain cellular functions during stress. For example, hsp90 appears to increase the activity of casein kinase when this enzyme is exposed *in vitro* to elevated temperatures (Miyata and Yahara, 1992). The hsp-enzyme interaction also prevents self-aggregation of the enzyme. The implication of this observation is that enhancement of certain protein kinase activity may be necessary for cells to preserve intracellular signaling which might otherwise go awry during stress. Other protein-hsp interactions have been observed, such as hsp-pp60src complexes (Hutchison et al., 1992; Ikawa and Weinberg, 1992) and hsp-immunophilin-glucocorticoid receptor complexes (Tai et al., 1992).

Hsps serve other functions during nonstressful conditions. Hsp90 acts as a molecular brake for the steroid receptor family (Bresnick et al., 1988; Bagchi et al., 1990, 1991). The complex of hsp90 and the glucocorticoid receptor prevents DNA binding of the receptor until the ligand for the receptor is present (Cadepond et al., 1991). Hsp90, like hsp70, has been found to effect protein folding (Wiech et al., 1992). Other hsps are involved in such regulation and it is becoming apparent that

complexes involving several of the hsps contribute to the regulation and function of other proteins.

Studies on hsp70 have indicated that it has a protein binding domain and an ATP binding domain. The ability of hsp70 to bind ATP has been helpful in the purification of this HS protein (Welch and Ferimisco, 1985). The protein can exist as a dimer, or the monomer can bind to a substrate. The dissociation of the dimer or the monomer-substrate complex is thought to occur upon ATP hydrolysis because non-hydrolyzable ATP analogues prevent the release of protein from hsp70. The hsp70 monomer is thought to interact with hydrophobic residues exposed in unfolded polypeptides. This prevents protein aggregation. ATP hydrolysis dissociates hsp70 from the polypeptide, thus allowing it to fold without the hindrance of hydrophobic residues from other proteins. It is also possible that ATP hydrolysis provides a level of energy necessary for hsp70 to act as a catalyst of protein folding.

ATP hydrolysis contributes to the hsp70 function of disaggregation. Hsp70 has been studied for its property of removing clathrin from coated vesicles. Breakdown of these protein aggregates occurs in the presence of ATP although ATP is not needed for clathrin-hsp70 association. It is likely that under conditions of ATP depletion during stress, hsp70 is associated with protein aggregates in an ATP-independent manner and that during recovery ATP triggers the disaggregation of proteins accumulated during cell injury.

More recently, investigations have been undertaken to understand the function of the carboxyl domain of the hsp70 proteins. Binding assays have determined that one of the members of the hsp70 family, grp78, has a substrate binding site that interacts with a seven-residue peptide sequence which contains aliphatic side chains. These proteins are located in the rough endoplasmic reticulum (ER) and are induced by glucose deprivation, glycosylation inhibitors, calcium ionophores, and heat. Certain mutations block import to the ER while others block export from the ER. Thus, grp78 serves a chaperoning function for the ER (Flynn et al., 1991). Perhaps one of the most important aspects of its function relates to the processing of immunoglobulins and the humoral immune system (Haas and Meo, 1988). As the carboxyl domain of the HS proteins contains a peptide binding site, it also appears that this region also contains an oligomerization domain near the substrate binding site (Meyers and Morimoto, unpublished observations). The relationship between HS protein oligomerization and function during stress and nonstressful conditions is currently being investigated.

ROLE OF THE HEAT SHOCK RESPONSE IN MEDICINE

Understanding the regulation of HS proteins is critical to knowledge about molecular responses in health and disease. In the past, most studies of physiologic stress focused on the "fight or flight" response, and analysis was limited to quantitation of circulating hormone levels. As physiologic stress is intimately tied to the

neuroendocrine responses, one area of interest is how the HS response is linked to hormonal activity. In rodent models, simple restraint of the animal induces the HS response at the transcriptional level (Blake et al., 1991). Manipulation of the adrenal-cortical axis appears to interfere with this response, indicating a potential endocrine basis for some aspects of the HS response (Blake et al., 1991).

While recent research has delineated the role of heat in regulating gene expression, there has been a strong historical interest in the role of heat in the pathogenesis of disease, as well as its therapeutic value. In fact, there has always been a medical query as to whether fever is beneficial (Kluger et al., 1975; Haahr and Mogensen, 1977). The febrile response is thought to occur as a result of circulating cytokines such as interleukin 1, TNF, and interferon which affect the central nervous system (Dinarello et al., 1984, Morimoto et al., 1987; Kettelhut and Goldberg, 1988). Body temperature is tightly regulated and central control of this homeostatic mechanism can be altered as part of the febrile response during infection. Furthermore, exposure to extreme heat, and even pharmacologic agents, can cause core body temperature to increase (MacLennan and Phillips, 1992). To date, very little is known about molecular controls of thermal regulation in the CNS, but if transcriptional responses are to be evoked, the HS factors are likely candidates for mediating these responses.

In contrast to our understanding of temperatures associated with the HS response in cultured cells, our observations about the HS response *in vivo* are more limited. Curiously, internal body temperatures can be increased by up to 2 °C by simply restraining rodents (Briese and DeQuijada, 1970). Based upon cell culture studies, this slight increase in temperature would be insufficient to induce HSF-DNA binding activity. However, some studies suggest *in vivo* HS gene expression can occur at temperatures below those used with cultured cells (Blake et al., 1990). This observation has raised the possibility that local tissue or systemic factors intersect with heat and induce HSF-DNA binding. Identification of endogenous factors which may influence the sensitivity of HSF to temperature-induced oligomerization is being actively investigated, and there are several potential candidates. For instance, prostaglandin levels increase during fever, and some of these have been found to induce HSF-DNA binding activity (Amici et al., 1992; Holbrook et al., 1992). The implication of this observation is that independent activators of the HS response such as heat and prostaglandins might act synergistically in leading to the HS response when temperatures are elevated by only 2–3 °C. Experiments with cultured cells support this hypothesis (Jurivich et al., 1993), but additional studies are needed to show whether HSF-DNA binding does in fact occur in tissue during febrile episodes. Animals exposed to increased external temperatures will manifest a HS response (Blake et al., 1990). This observation suggests that external heat and perhaps heat stroke are inducers of the HS response *in vivo*. This observation also raises the possibility that local body heat induces the HS response. Local body heat is often employed clinically to relieve pain and reduce inflammation. Thus, one of

the unexplored effects of heat therapy could very well be enhanced cytoprotection against inflammation through induction of the HS response.

In addition to considering the role of the HS response in febrile episodes and environmental heat exposure, there are other conditions where abnormal regulation of heat could intersect with the HS response. For instance, malignant hyperthermia is a condition in which there is a genetic predisposition towards drug-induced fever. It can be lethal (Manzerra and Brown, 1990). The anesthetic halothane is typically implicated as an inducer of malignant hyperthermia, although other drugs are known to be associated with this response. Curiously, halothane induces a 70 kDa-class of HS proteins in rodents (Van et al., 1992). Sequelae from malignant hyperthermia include neuronal, hepatic, and renal damage which indirectly suggest that these tissues are less capable of mounting a cytoprotective HS response relative to other tissues. Genetic linkage studies have not yet identified the gene for familial malignant hyperthermia, but animal studies have found stress-induced death accompanying fever in swine to be linked to abnormalities in calcium channels. These observations suggest that a HS response in an animal model of malignant hyperthermia may be defective, perhaps through abnormal regulation of intracellular calcium levels. Further investigations in this area could link abnormalities of the HS response to genes that may not be heat inducible themselves, but are necessary for function of the heat inducible genes.

While the increased expression of HS proteins is thought to be generally beneficial, there are medical considerations where this may not be good. For example, increased levels of HS proteins in tumors might render therapy less effective. Resistance to thermal killing of cells as well as to chemotherapy and other modalities of cancer treatment would be the consequence of an up-regulated heat shock response. On the other hand, increased expression of the hsps in noncancerous cells would be valuable in cancer patients so as to protect normal tissue from the toxic effects of therapy. Uncoupling of the HS response in cancer cells would clearly be desirable, but selective uncoupling of the response, while preserving it in normal tissue would be a formidable challenge. Other problems might arise if the clinical goal were to be an uncoupling of the HS response. For example, one could argue that hsp70 is essential to cell growth and cytoprotection. Based upon this knowledge, gene therapy directed at selectively blocking hsp70 gene expression in tumor cells to enhance cytotoxicity and slow growth would be valuable. However, such an attempt would likely be compensated by the increased expression of other HS genes. Rather than disrupting expression of a particular HS gene by introducing an antisense oligonucleotide, a more global blockade of the HS response may be obtainable by interfering with HSF-DNA binding activity. Bioflavinoids are an example of compounds that have the potential to disrupt HSF activation, thus providing a pharmacologic means to modulate the HS response (Kioka et al., 1992). The key goal here would be to selectively block HSF activation by concentrating pharmacologic uncouplers of the HS response in tumors.

If uncoupling of the HS response is a desirable goal in clinical oncology, then enhancement of the response would be a goal of clinicians interested in minimizing the untoward effects of ischemia, inflammation, and other disorders associated with cell death. For instance, enhancement of the HS response in neuronal cells would be beneficial in stroke and neurodegenerative disorders. An example might be progressive neuronal cell death induced by the toxic accumulation of β-amyloid in Alzheimer's disease. The effects of accumulated β-amyloid might be reduced by enhancing hsp expression in the brain. Based upon what we know of hsp function, increased levels of hsp should improve the degradation of damaged proteins, enhance disassociation of aggregated proteins, and generally improve protection of cells in a neurodegenerating environment. Unresolved questions are whether increased expression of hsp can be achieved over the long term and whether sustained expression of hsp alters cellular function in some manner.

A salient issue concerning the HS response and human disease is whether a constitutive stress response occurs in certain diseases. *In vitro* studies indicate that the HS response is transient despite the continued stress of heat (Abravaya et al., 1991a). Once the stress response has been activated in cells, they become refractory to inducing the response again unless the stress is harsher than the first one. These observations may have some relevance to our understanding of certain chronic diseases that portend chronic stress. For instance, Alzheimer's disease might be viewed as a chronic stress state leading to neuronal injury. Support of this view is based upon the increased levels of HSF-regulated genes found in brain tissue. These include some HS genes, β-amyloid precursor protein, and heme oxygenase (Perez et al., 1991; Abe et al., 1991). If *in vitro* studies, which show an eventual attenuation of the HS response despite the continued stimulus of heat, are applicable to Alzheimer's disease, one might expect attenuation of the stress response in chronically stressed neurons. This process would exhaust their capability to mount a stress response, thus leading to cell death. On the other hand, the human HS response may not attenuate as it does in cultured cells, and this may account for the abnormally high levels of stress-inducible gene products observed in brain tissue from Alzheimer's patients. Clearly, more information is needed to determine the long term molecular consequences of chronic physiologic stress *in vivo*.

Another area which is not fully understood is the relationship of HS protein levels to overall cellular functions, especially if chronic stress occurs in certain human diseases. Increased levels of HS proteins have been observed in degenerative arthritis which on the one hand could be cytoprotective, and on the other hand, could have long-term detrimental effects. Levels of HS proteins appear to be tightly regulated and conditions which perturb hsp levels could disrupt routine cellular activities. For example, the HS response usually leads to down-regulation of cellular growth (Rowley et al., 1993). One of the potential consequences of a chronic stress response during inflammation might be that cells become refractory to mitogenic stimuli. Altered cell growth would delay tissue repair and, in this

respect, the HS response potentially would be detrimental. Thus, one of the goals in analyzing stress responses during human disease is to determine the short and long term consequences of this response.

Despite many of the elegant studies that have been completed on the HS response, we still do not know whether the human HS response can go awry and subsequently contribute to disease. There is redundancy in many HS gene families, and one could imagine the possibility that mutations in one or more of the HS genes could exist in the human population. This consideration can be likened to p53, a potent antitumor regulatory molecule that manifests numerous mutations in transformed cells. The mutations lead to a dysfunctional protein which is no longer capable of inhibiting DNA synthesis. In an analogous fashion, mutations in one of the hsp70 genes could give rise to a dominant-recessive event in which a certain pool of hsp70 is functionally defective. This could have far-reaching consequences, such as slowing the processing of damaged proteins, disrupting intracellular protein traffic, and reducing the cell's ability to withstand stress. The predictable consequence would be a cell prone to further damage and degeneration. The possibility that acquired or inheritable defects in the HS response lend themselves to functional problems at the cellular level remains to be explored as a possible mechanism of disease.

Most analyses of the HS response in human disease have focused on the accumulation of hsp either by Western blots or *in situ* hybridization. Because the accumulation of hsp can be either through translational or transcriptional mechanisms, it is not immediately clear which mechanism contributes to the increased levels of hsp observed in certain diseases. Thus, the actual role of the HS factors in human disease can only be inferred from *in vitro* experiments. In many of the cases under consideration, the key question is whether HSF is induced into its DNA binding state. A corollary question is whether conditions arise in which HSF binding ought to be triggered but is not. HSF is known as an inducible transcription factor, yet certain cell lines exhibit constitutively oligomerized HSF. This condition does not necessarily lead to a constitutive HS response, and it demonstrates that the HS response is pleiotrophic. Equally important, there may be defects in the triggering of HSF into its DNA-bound state or problems with the translocation of HSF into the nucleus from the cytoplasm. Even if HSF is properly activated and found in the nucleus, there are chromosomal effects which may prevent its binding to DNA. This would allow differential expression of HS genes in tissue where higher levels of one class of HS genes may be more desirable than another class. As we learn more about the normal regulation of HSF we can turn to questions about its potential involvement in human disease or, conversely, questions about how to upregulate this factor under conditions where activation of the HS response would be clinically advantageous.

Heat Shock Response in Immunology and Infectious Diseases

The role of HS proteins in infections and inflammation has stimulated considerable research interest. The elevated expression of HS proteins is thought to be associated with the febrile response, but equally important, the HS response may also be a mechanism for survival of the pathogen (Christman et al., 1985). Levels of HS proteins may dictate the pathogenicity of a microbe (Caruso et al., 1987) or its resistance to antibiotics (Hallet et al., 1990). Epidemiologic studies have revealed that HS proteins from infecting organisms are one of the most potent antigens for the human immune system (Young et al., 1988; Adeleye et al., 1991). For example, patients with malaria exhibit antibodies to hsp70 and 90 from *P. falciparum*. Similar examples can be found for other parasitic infestations, including schistosomiasis and filariasis. Certain bacterial infections result in a humoral response to the groEL protein, the bacterial equivalent to the human hsp60. Circulating antibodies to groEL can be found in patients with tuberculosis, syphilis, Lyme disease, and many bacterial diseases (Shanafelt et al., 1991). Despite the ubiquitous nature of certain HS proteins being an antigen in numerous infections, it is not clear whether a humoral response to HS proteins from an invading organism assists in the host defense system, especially when one considers that these proteins are intracellular. Many of the infections that induce antibodies to HS proteins are confronted with T cell immunity and intracellular killing by the host. The humoral response is not the primary host defense mechanism in these instances. It may be that antibodies to parasitic and bacterial HS proteins are useful in host clearance of microorganisms successfully killed by the immune system. However, antibodies to hsp90 are reportedly protective against systemic infections by yeast (Matthews et al., 1991).

Diagnostic and therapeutic considerations have evolved from our understanding of the HS response in infectious diseases. From a diagnostic perspective, it may be possible to test sera of patients for immunoreactivity to certain HS proteins from various pathogens as a means to identify the type of infection, especially when traditional methods of microbial isolation and culture are slower than what is necessary for judgments about clinical therapy. Because gene knock-out studies have indicated the essential nature of hsp to bacterial cell growth and survival, pharmacologic interference with one or more of the bacterial HS proteins could provide a novel approach towards treating infections. Some of the hsp utilize ATP for their function, and drugs targeted towards interference with the ATP binding site would potentially block their activity; however, the conservation of ATP binding sites in proteins between species would make this pharmacologic approach less plausible. A more likely approach in interfering with the bacterial HS cascade might be undertaken by blocking the substrate-binding site. Again, specificity of a certain bacterial HS protein for a prokaryotic substrate would be an ideal target. Identification of a unique prokaryotic substrate would also have to take into account

the reversible nature of HS protein-substrate interactions, and ideally one would want a therapeutic agent that specifically and irreversibly binds to the prokaryotic HS protein, thus rendering it inactive. An alternative approach to interfering with HS protein substrate binding sites would be to introduce polypeptides corresponding to interactive sites between one or more of the HS proteins. Heat shock proteins can assemble with each other. This assemblage is much like the aggregation of complement factors in their formation of the membrane attack complex, and there may be similarities between the two protein systems given that HS proteins map to the same chromosomal region as some members of the complement family. Blocking the cascade of HS protein-hsp assembly would presumably uncouple HS protein functions which require specific interactions between different members of the hsp family. The key to such a quest would be specificity of interfering with bacterial hsp-hsp interactions and not human ones. Although there is substantial homology between hsp proteins among species, interference with bacterial HS proteins may be achieved if there is divergence between bacterial and human interactive sites of the HS proteins.

The HS response is tightly linked to the wide range of conditions encountered by microbial pathogens. For instance, parasites have life cycles that range from poikilothermic hosts (arthropods) to homeothermic hosts (mammals). The sudden changes in ambient temperature appear to elicit a typical HS response where there is an increase in HS gene expression and a decrease in the synthesis of other proteins. Some investigators have suggested that certain parasites generate a HS response as a normal part of their development. Studies with *Leishmania* have indicated an increase in pathogenicity when hsp synthesis is elevated. Upon infecting a human host, many parasites induce fever which could act as a further stimulus of the pathogen's stress response. Certain forms of trypanosomes associated with the febrile state demonstrate higher levels of hsps than other forms (Davis et al., 1987).

Parasitic elaboration of HS gene mRNA is slightly different than in humans. Trypanosomes produce a polycistronic HS mRNA that requires trans-splicing. Like mammalian cells, trans-splicing is generally inhibited by stress in trypanosomes, with the exception of the HS messenger RNA (Muhich and Boothroyd, 1989). Thus, this parasite has evolved a means to preferentially express HS proteins that involve a posttranscriptional mechanism. The exact means by which trypanosomes sustain mRNA splicing for their HS mRNA is an area of active investigation.

Bacterial responses to stress have both transcriptional and translational points of regulation. *E. coli* demonstrate both cold- and heat-inducible gene expression. The heat-inducible response is mediated by σ^{32} which is transiently increased and interacts with the RNA holoenzyme to specifically increase heat shock gene expression (Young et al., 1990). Studies regarding the bacterial HS response have revealed a level of feedback inhibition accounting for the transient nature of the HS response. The activity of the 32 kDa sigma factor is negatively regulated by the

products of HS gene expression, namely, the DnaK, DnaJ, and GrpE HS proteins (Gamer et al., 1992). These observations have served as a model for understanding the regulation and attenuation of the human HS response, although the actual role of HS proteins in attenuating HSF activity *in vivo* remains unproved. Intracellular accumulation of bacterial HS proteins during intracellular killing of certain bacterial pathogens by white blood cells could potentially affect the HS response of the host. Whether such interactions between components of the human and bacterial HS response occur is open to investigation.

Undoubtedly there is a pitched battle between the efficacy of the human HS response relative to the invading organism. For instance, production of bacterial HS proteins may be a means to defend against host-killing mechanisms. White blood cells produce hydrogen peroxide and other reactive oxygen intermediates as a means to kill bacteria. H_2O_2 is a potent inducer of the stress response in bacteria (Walkup and Kogoma, 1989), and bacteria that accumulate hsps are more resistant to macrophage killing than bacteria which do not evoke a stress response (Morgan et al., 1986). Bacteria may encounter several inducers of the stress response, including heat, superoxide damage, sequestration of metals, and hypoglycemia. In addition to normal processing of host proteins, certain infectious agents may use human-derived HS proteins to their advantage. Interactions with some members of the hsp70 gene family (grp78) and viral protein products have been observed, suggesting that intracellular pathogens hook into the host's chaperone system as a means to advancing their life cycle (Macejak and Luftig, 1991). As grp78 interacts with unfolded protein from the influenza virus, some human HS proteins may act as inadvertent facilitators of intracellular viral assembly (Watowich et al., 1991). In fact, certain viruses actually induce hsp synthesis in infected cells (Phillips et al., 1991). Adenovirus is known to induce elevated HS gene expression via E1A protein and its interaction with the proximal promoter region of the hsp70 gene. Herpes simplex virus 1 (HSV1) induces hsp70 also, but this is not a property of all DNA viruses. Hsp90 is also overexpressed during HSV1 infection (La and Latchman, 1988). The synthesis occurs regardless of lysis, so the induction is not necessarily due to cytolytic effects of the virus per se. Other viruses specifically induce the expression of the glucose regulated stress proteins, presumably via viral products which interact with the promoter region of the grp genes. Exposure of mice to influenza A virus causes pulmonary macrophages to increase their synthesis of HS proteins. Thus, there are several examples where viral pathogens specifically up-regulate the HS response, potentially linking viral viability to host cell viability.

While pathogens may interact with the host's HS response, the expression of HS proteins in immunocytes is thought to be an integral component of the immune system at several levels. Certain hsps such as hsc70 exist in cells under normal conditions, and levels of hsp70 will increase when normally quiescent lymphocytes are stimulated and become lymphoblasts (Ferris et al., 1988). Regulation of growth-inducible hsp70 is thought to occur via protein-DNA interactions in the

proximal promoter of the hsp70 gene. Curiously, heat and fever appear to augment the primary humoral response (Downing et al., 1987). Increased IgM expression is observed *in vitro* when cultured B cells are exposed to slightly elevated temperatures (Jampel et al., 1983). Polymorphonuclear leukocytes also seem to be regulated by heat, thus suggesting that the HS response is an integral component of a proficient immune system (Johansen et al., 1983; Eid et al., 1987). Phagocytosis induces HS protein synthesis (Cleget and Polla, 1990). Interferon treatment of immunocytes will augment the accumulation of HS proteins (Chang et al., 1991), thus suggesting that cytokines work in synergy with HS gene expression. The cytokine transforming growth factor beta-1 upregulates the HS response by a posttranscriptional mechanism (Takenaka, and Hightower, 1993). Finally, a member of the hsp70 HS family appears to be involved in antigen presentation (Vanbuskirk et al., 1989). All of these observations attest to the interrelationship of the HS response and immunologic function.

One of the questions confronting investigators in the HS field is whether fever or other acute phase reactants can induce HS gene expression. *In vitro* studies utilize extraordinary temperatures of 42 °C and higher. Core body temperatures may approach 40 °C as a result of fever. In most *in vitro* systems, this temperature does not lead to the HS response. However, there are reports that fever induces the increased synthesis of hsps in peripheral blood lymphocytes (Ciavarra, 1990). This response was observed in mononuclear cells exposed to febrile temperatures and in cells isolated from a medical intern who developed fever.

The purpose of increased hsp expression in white blood cells is not fully understood, but one likely function is cytoprotection against toxic mediators of inflammation such as tumor necrosis factor. In fact, overexpression of hsp70 by transfecting cells with a plasmid containing the hsp70 gene confers protection against the otherwise lethal effects of TNF (Jaattela et al., 1992). Hsps also protect cells from acidic environments and the protein denaturing effects of heat that accompany the inflammatory response. In addition to protecting cells from unfavorable extracellular environments, hsps undoubtedly afford protection from intracellular conditions that are potentially damaging. Activation of hsp synthesis in response to phagocytosis is a good example where the intracellular release of H_2O_2 and other mediators of oxidative damage could harm white blood cells if not for the HS proteins. Other functions of HS proteins have been found which contribute to a successful immunologic response. In particular, antigen processing cells such as B-lymphocytes and macrophages utilize a HS protein to facilitate their communication with T cells (Vanbuskirk et al., 1989). This protein acts as a chaperone for processing antigens and delivering them to the cell membrane surface in conjunction with the MHC complex. Similar functions can be ascribed to other HS proteins such as grp78 (BiP) which interacts with some immunoglobulin heavy chains and other secretory proteins. These and other examples indicate a central role of the HS proteins in mediating immune function.

As HS proteins are intimately tied to immune function, mechanisms regulating the HS response in the immune system need to be elucidated. Conditions associated with inflammation appear to be potent modulators of the HS response (Winrow et al., 1990). HSF-DNA binding is activated by some prostaglandins and arachidonate, thus suggesting that the arachidonate cascade works in conjunction with elevated temperatures in regulating the HS response (Amici et al., 1992; Jurivich et al., 1993). Heat induces phospholipase A_2 activity and leads to the accumulation of arachidonate (Calderwood, 1990). Other inflammatory conditions such as tissue hypoxia and altered pH may also contribute to the overall induction of the stress response. The convergence of multiple inducers of the HS response during inflammation is a plausible explanation of how elevated HS gene expression can occur in the human body at temperatures that are typically lower than those used in the laboratory. In fact, current evidence indicates that inflammatory compounds such as arachidonate lower the temperature threshold for the activation of HSF into its DNA-bound state (Jurivich et al., 1993). Inflammatory conditions not only alter the sensitivity of HSF to heat induction, but they also appear to enhance the duration of HSF-DNA binding activity. Whether HSF-DNA binding occurs transiently or constitutively during inflammation is not known.

One area of controversy regarding the heat shock response and the immune system involves regulation of γδ T cells. γδ cells are a subclass of T-lymphocytes identified because of their cell surface receptors which differ from most T cells that express heterodimer receptors αβ. These cells are relatively rare in the immune system, located in lymphoid and epithelial tissue. The function of these cells is not entirely known, although they may influence the development of αβ cells and promote cytolytic activity. As such, these cells may be the front line of responses to infections. One of the key observations about γδ cells is that they are reactive to HS proteins. Parasitic and bacterial heat shock proteins act as ligands for γδ cells, possibly in conjunction with the major histocompatibility complex, and they cause these cells to proliferate. The observations that γδ cells exhibit a very limited repertoire of V_γ and V_δ gene segments has raised the possibility that these cells are quite restricted in their recognition of antigens and that ubiquitous HS proteins represent a class of proteins which the γδ cells might recognize regardless of the invading organism. γδ cells encountering mycobacterium-expressed HS proteins leads to the proliferation of these cells, increased expression of the IL-2 receptor and autocrine expression of IL-2 (Haregewoin et al., 1989). The mycobacterium HS protein, hsp65, appears to be the most antigenic for the γδ cells (Ab et al., 1990; Born et al., 1990). In fact, γδ cells are most reactive to an epitope contained within amino acid sequences 180–196 of hsp65 from a variety of bacteria when splenic-derived antigen presenting cells are employed.

Given the reactivity of γδ cells to mycobacterial hsp65, a popular hypothesis regarding autoimmune disease has been that γδ cells cross-react with human hsp60

or other human proteins which have sequence homology to the mycobacterial hsp65 (Agues et al., 1990). Other T cell populations are also reactive to mycobacterial antigens (Pope et al., 1991, 1992). In favor of the hypothesis that bacterial infections evoke an autoimmune response revolving around HS proteins has been the observation that bacterial and human hsp65 equivalents share approximately one-half of their amino acid sequences. This has led to several experiments demonstrating the immunoreactivity of γδ cells to self-HS proteins. In a mouse model, PPD-reactive γδ cells were found to weakly react to a peptide sequence from a murine hsp60 class of proteins. Synovial-derived T cells react to an epitope of hsp65 (Gaston et al., 1990). Curiously, HS of cultured lymphocytes from mice exposed to extracts of mycobacteria resulted in a disproportionate representation of γδ cells. Furthermore, murine macrophages stimulated *in vivo* with influenza virus were found to develop surface membrane epitopes similar to mycobacterium hsp65. This observation has led to the possibility that γδ cells are cytotoxic for macrophages exhibiting hsp-epitopes on their membrane surface and that this serves as a means of immunologic clearance and resolution of inflammation (Koga et al., 1989). Thus, γδ cells may act as hsp scavengers and serve a dual function of responding (i) to pathogens when their HS proteins are part of an antigen presentation and (ii) to immunocytes when they express surface membrane HS proteins as a signal for self-clearance of stressed cells (Ottenhoff et al., 1988).

While some data are suggestive that self-reactivity to HS proteins plays a normal role in autoregulation of the immune system, others have considered the role HS proteins may have in patients with autoimmune disease, especially rheumatoid arthritis. Synovial fluid from a rheumatoid patient was found to have mycobacterium-reactive γδ cells. Antibodies reactive to mycobacterial hsp65 are also reactive to human chondrocytes (Kimura et al., 1991). Rodent models of autoimmunity have been used to demonstrate the increased number of γδ cells reactive to hsp65 and to self-hsp65 homologues. Despite this relationship, subjects without autoimmune disease can be found with γδ cells self-reactive to HS proteins (Munk et al., 1989; Fischer et al., 1992). This observation has diminished the likelihood that autoreactivity of T cells to HS proteins is the primary mechanism underlying autoimmune disease. Nevertheless, autoimmune diseases such as lupus and rheumatoid arthritis appear to be associated with circulating antibodies to one or more of the HS proteins (Muller et al., 1988; Tsoulfa et al., 1989). Healthy human subjects also manifest circulating antibodies to HS proteins (Wakui et al., 1991), thus raising concern about the specificity and role of autoantibodies to heat shock proteins in human health and disease (Grandia et al., 1991). In light of these concerns, investigators continue to pursue the possibility that biological mimicry involving the bacterial and human HS proteins may account for some inflammatory diseases (Smiley and Hoffman, 1991).

Heat Shock Response in Oncology

In most instances encountered clinically, the cytoprotective effects of the HS proteins are beneficial, but this effect is not necessarily beneficial when treating malignant tumors. Heat therapy for tumor regression has been used with mixed results (Abe and Hiraoka, 1990). The rationale for thermal treatment of tumors is that they are more sensitive than normal tissue to heat-induced killing and it does not lead to systemic toxicity. Limitations of thermal therapy include the delivery of heat locally to the tumor, especially if it is not superficial, and the development of thermotolerance. The induction of thermotolerance in tumors is especially disadvantageous because the multidrug resistance (MDR) genes are regulated, in part, by the HS response (Chin et al., 1990). In fact, even certain chemotherapeutic agents induce HSF-DNA binding activity which enhances the tolerance of tumors to all types of stress (Kroes et al., 1991). The induction of the HS response may occur as a natural course of tumor growth when one considers low blood flow and hypoxia as contributory to the induction of the HS response.

Analysis of HS proteins in cells derived from tumors has indicated a wide range of hsp70 levels. Other hsps appear to be associated with certain tumors. For instance, small molecular weight hsps have been found in certain breast tumors. Expression of these hsps may be a marker of disease (Tetu et al., 1992). Special attention has been given to the heat and estrogen inducible regulation of the 24 kDa protein (Fuqua et al., 1989). Monitoring levels of this and other HS proteins (Jameel et al., 1992) may be a sensitive measure of disease recurrence. Manipulation of the HS response in mammary epithelial cell lines also appears to promote tumorigenicity (Lebeaus et al., 1991). On the other hand, a recent and novel experiment was conducted in which the HS response was employed to manipulate immunosurveillance of tumors (Lukacs et al., 1993). The bacterial HS protein hsp65 was introduced in a macrophage tumorigenic cell line and the expression of hsp65 in these cells rendered them nontumorigenic. The mechanism for this response presumably is through immunologic recognition of a stress protein, perhaps by cytotoxic $\gamma\delta$ T cells.

Because HS proteins contribute to thermotolerance and reduced sensitivity of tumors to remitive therapy, attention has been given to the possibility that the HS response can be pharmacologically blocked. The flavanoid family of drugs appear to have such an effect and is discussed below. However, the actual role of the HS proteins in thermotolerance is controversial. Many studies have indicated that overexpression of a HS protein results in the cell's increased tolerance of stress as measured by cell viability. Overexpression of hsp70 or hspl04 confers thermotolerance (Solomon et al., 1991; Sanchez et al., 1992). Because none of these studies tightly regulated levels of hsp70 or other hsps, it is unknown how much hsp should accumulate before the cell is protected. The usual heat induction of hsp70 at the translational level is only 2–3-fold above levels seen at 37 °C. Many cell lines are

resistant to stable transfection and integration of a heat shock gene, suggesting that overexpression of hsps over the long term is deleterious and that hsp levels are closely monitored within the cell. What is curious about the studies which have successfully overexpressed a HS protein is that even though there are several classes of HS proteins, only one apparently needs to be overexpressed to achieve thermotolerance. The stable expression of hsp70, for instance, does not seem to elevate the expression of other HS proteins. The ability of only one hsp to confer thermotolerance does raise the question whether HS proteins share common functions. Although they may be versatile, most studies indicate separate and coordinated functions for each class of HS proteins. Thus, it is not clear how increased levels of one HS protein accomplishes all of the tasks which a whole set of HS proteins presumably perform in conferring thermotolerance.

Other studies have challenged the concept that HS proteins are necessary for thermotolerance. For instance, cycloheximide treatment of cells, which prevents the induction of hsp synthesis, does not prevent the induction of thermotolerance. Phosphorylation and not the accumulation of hsps has been cited as the reason cells become thermotolerant, even though they have been prevented from increasing the expression of their hsps (Ciocca et al., 1992). In prokaryotic cells, phosphorylation of groEL leads to the enhanced function of this HS protein (Yu et al., 1992). Thus, examples do exist where the posttranslational modification of pre-existing hsps could potentially lead to enhanced function necessary for thermotolerance. Of course, factors other than the HS proteins could contribute to the development of stress-tolerance. Even in light of experiments suggesting that increased expression of one hsp is beneficial, it is likely that multiple factors are involved in the regulation and acquisition of thermotolerance.

Heat Shock Response in Cardiovascular Disease

The role of the HS response in heart and blood vessel diseases has become increasingly evident over the past few years. Hypoxia is a potent inducer of the HS response and provides cardioprotection (Aldashev et al., 1991; Williams and Benjamin, 1991). To date, myocardial cells have been found fully capable of inducing the HS response (Mehta et al., 1988; Benjamin et al., 1992). In addition to increased expression of the HS genes, hyperthermia appears to enhance the accumulation of β-TGF mRNA (Flanders et al., 1993). This observation illustrates how genes other than the HS genes are regulated by stress. Whether increased levels of β-TGF contribute to the cytoprotection of cardiac cells is not known, but enhancement of the HS response in the myocardium offers the clinical goal of maximizing cytoprotection and limiting tissue necrosis during ischemia.

Neuroendocrine determinants of stress have been found in rodent studies linking endothelial cell responses to hormonal regulation of the HS response (Blake et al., 1991). When rats are physically restrained some of their tissues elaborate HS genes,

especially in the adrenal and pituitary glands. Curiously, hypophysectomy will prevent elevated HS gene expression in adrenal tissue, suggesting that a circulating factor such as ACTH can evoke HSF-DNA binding activity. Injection of ACTH into hypophysectomized animals restores hsp70 mRNA induction in restrained rodents, but it does not appear to activate HSF-DNA binding directly as cultured cells are insensitive to the effects of ACTH. Vascular tissue also produces hsp70 mRNA in response to restraint stress, and this response is blocked by α-adrenergic antagonists (Udelsman et al., 1993). Thus, there are at least two examples of how endocrine responses are associated with the HS response. These observations have important implications for cardiovascular disease. For instance, blockage of the α-adrenergic component of the HS response may be one means to prevent cardiac hypertrophy which is associated with the increased expression of HS proteins (Snoeckx et al., 1992).

Studies on hypertension have been linked to the HS response (Hamet et al., 1990a,b). Ambient temperatures affect blood pressure in hypertensive patients. Animals genetically predisposed to systolic hypertension have an increased sensitivity to environmental stress. They increase their core body temperature to elevations in ambient temperature much more rapidly and at higher levels than normotensive animals. This has suggested that genes contributing to high blood pressure are associated with genes involved in thermal responses. Hypertensive rats and humans demonstrate higher transcription of the hsp70 gene than normotensive subjects. By nuclear run-on analysis of kidney tissue, hsp70 gene expression was at least two-fold higher in hypertensive rats than in their normotensive counterparts. This led to the increased accumulation of hsp70 mRNA. Why hypertensive rats exhibit higher levels of transcription of the hsp70 gene is not clear, although increased levels of HSF-DNA binding for a given level of HS is one possible explanation (Hashimoto et al., 1991).

In addition to altered hsp70 gene expression, an hsp70 gene restriction length polymorphism has been found in hypertensive rats but not in normotensive rats (Hamet et al., 1992). Whether this results in a functionally abnormal stress response during hypertension is not certain, but it raises the possibility that the environmental and genetic determinants of hypertension are due to abnormal regulation of the HS genes. Despite the increased expression of hsp70 in cells from hypertensive rats, cardiomyocytes from these animals exhibit an increased sensitivity to heat-induced cell death when compared to cells from normotensive animals. The restriction fragment length polymorphism of the hsp70 gene in hypertensive animals indirectly suggests that a defective hsp70 gene is somehow contributory to the increased sensitivity to cell death in the hypertensive animals.

Cardiac hypertrophy appears to be mediated by HS proteins (Izumo et al., 1988). Cardiac myocytes exposed to a hemodynamic stress have been found to increase their levels of heat shock proteins (Delcayre et al., 1988). Although experiments involving interference with HS protein synthesis were not done in these studies,

clinical conditions such as hypertension which contribute to cardiac hypertrophy might be altered by strategies which block HS protein expression.

Perhaps some of the most intriguing studies on the clinical relevance of the HS response come from experiments linking HS proteins and atherosclerosis. Immunization of rabbits with hsp65 leads to atherosclerotic plaque formation (Xu et al., 1992). The HS proteins are also associated with the plaques, perhaps demonstrating an extracellular role of HS proteins. The mechanism of plaque formation through immunologic reactions to stress proteins remains to be elucidated, but these observations suggest a novel means by which coronary disease can evolve. Atherosclerosis also appears to alter the distribution of hsp70 (Johnson et al., 1993). While overall levels of hsp70 do not change during atherogenesis, there is a marked redistribution in aortic tissue suggesting cellular differences in responding to cytotoxic environments that accompany the necrotic cores of atherosclerotic plaques. Cells not elaborating hsp70 are thought to be susceptible to cell death and contribute to the progression of necrosis within the plaque.

Heat Shock Response in Neurons and Neurodegenerative Disorders

The HS response is vital to neuronal cytoprotection during ischemic, febrile, and epileptic episodes (Vass et al., 1988, 1989; Wong et al., 1992; Abe and Kogure, 1993). Heat shock protein mRNA levels increase after brain trauma and transient ischemia (Nowak et al., 1990; Gozales et al., 1989). Neuronal damage can also occur from neurotransmitter toxicity and HS protects cultured cells from this effect (Rordorf et al., 1991). Not all neuronal cell lines manifest the same type of HS protein production. Observations on neuronal cell lines have indicated that there may be a selective production of HS proteins. A retinoblastoma cell line minimally induces hsp70 gene expression after thermal stress, whereas a glial cell line fully induces hsp70 mRNA (Mathur and Morimoto, unpublished observations). By comparison, hsp90 gene expression is induced equally well in both cells. The discordant expression of HS genes in a single neuronal cell line has raised the possibility that different chromosomes containing the various HS genes are not equally responsive to binding HSF. *In vivo* footprint analysis has confirmed the lack of HSF-DNA binding at the hsp70 promoter region in the aforementioned retinoblastoma cell line, but whether this selective repression of the hsp70 gene occurs in normal brain tissue is not known. It does raise the possibility that selective expression of the HS genes is somehow linked to specialized cellular function. Furthermore, this observation and others have raised the question whether the HS response is different for various cells comprising the central nervous system, and perhaps accounting for survival of some cells and loss of others during neurodegenerative disorders. A similar question is being asked about potential regional differences within the brain for elaborating HS proteins (Kawagoe et al., 1993).

One of the common themes emerging from studies on neuronal degenerative disorders is that they are associated with the HS response in some manner. In certain respects, past clinical observations on patients with multiple sclerosis (MS) have indirectly linked the HS response with this disease. In MS patients there is an increased sensitivity to elevations in core body temperatures. This clinical observation has heightened interest in the role of the HS response in mediating this and possibly other neurodegenerative diseases (Garofalo et al., 1991). While little has been done to characterize the increased sensitivity to temperature, a few studies have focused on the potential for hsps to act as autoantigens in MS. The T cell proliferative response in patients with MS is enhanced by the presence of *M. tuberculosis* hsp70 and not to the *M. bovis* hsp65. γδ T cells are found to aggregate in MS plaques (Selmaj et al., 1991). CNS 90 and 60 hsps are thought to contribute as autoantigens for the clonal expansion of T cells because these hsps are overexpressed in MS plaques when compared to normal neuronal tissue (Wucherpfenning et al., 1992). Immature oligodendrocytes were found to have strong immunoreactivity for hsp65 whereas mature cells did not (Selmaj et al., 1992). This finding plus the increased association of γδ cells with immature oligodendrocytes suggest that autoimmune destruction of immature oligodendrocytes contributes to the demyelination associated with MS. Human oligodendrocytes are in fact lysed by γδ cells *in vitro*, suggesting that hsp-immunoreactive T cells are cytolytic for brain tissue (Freedman et al., 1991). Unresolved is how γδ cells target neuronal tissue and whether their clonal expansion is truly due to a HS protein stimulus.

The expression of HS proteins in neurodegenerative disorders may be in response to chronic stress. Neurons are sensitive to cell death when exposed to certain neuroexcitatory peptides, and HS proteins may protect against this effect (Lowenstein et al., 1991). There have been strong indications that the HS response is associated with Alzheimer's dementia (Pappolla et al., 1992). Brain tissue from patients with Alzheimer's exhibit hsp expression (Perez et al., 1991). Neuropathological findings in heat-stressed rodents resembled changes in brain tissue from Alzheimer patients (Papasozomenos and Su, 1991). In this model, both the heat-stressed rat brain and Alzheimer brain tissue manifested increased levels of phosphorylated tau protein, thus suggesting a possible link between the stress response and the etiology of Alzheimer's disease. Recently, a novel gene was isolated using an antibody raised against protein extracts from Alzheimer's brain tissue, and the sequence of this gene corresponds to the bacterial hsp, DnaJ (Raabe and Manley, 1991). The fact that DnaJ and DnaK (the bacterial equivalent to hsp70) interact in bacteria to cause protein folding and disaggregation suggests that this type of protein-protein interaction is important to neuronal function. mRNA for the DnaJ homologue in human brain tissue is most highly expressed in the neuronal layers. In addition to finding elevated expression of HS genes in Alzheimer's disease, there is evidence that HSF may be regulating β-amyloid gene expression. Studies have now indicated that the upstream promoter region of the β-amyloid

gene contains a HS element which is potentially capable of binding HSF (Salbaum et al., 1989; Izumi et al., 1992). Furthermore, β-amyloid mRNA is increased by heat shock (Abe et al., 1991). Although it is not clear that HSF is constitutively active in Alzheimer's disease, the fact that both HS genes and other genes that contain HSEs are found to have elevated expression strongly suggests a role of and the stress response in neuronal damage.

Pharmacologic Effects on the Heat Shock Response

Full and partial induction of the HS response has been associated with several classes of drugs. The nonsteroidal antiinflammatory drugs (NSAIDs) have unique effects on the human heat shock response. Sodium salicylate, one of the first agents used in inflammation, is a potent inducer of HSF-DNA binding activity (Jurivich et al., 1992). Salicylate has the biochemical property *in vitro* of uncoupling HSF-DNA binding activity from transcription of the HS genes. Salicylate-treated cells exhibit full HSF-DNA binding by both gel shift and *in vivo* footprint analysis. Despite this protein-DNA interaction, there is little expression of HS genes. This observation indicates that salicylate affects the first step of the HS response, i.e., oligomerization of HSF into its DNA bound state, but does not induce further steps in the activation of the HS response. The ability of HS to induce full transcriptional activity in salicylate-treated cells indicates that salicylate is not simply poisoning or inhibiting other components of the HS response. Evidently, salicylate fails to induce some additional signal pertinent to the elevated expression of the HS genes. Because salicylate does not lead to increased phosphorylation of HSF1 like HS, the inability of salicylate to fully induce the HS response appears to be due to its inability to activate a HSF1 kinase.

The ability of heat to fully trigger the HS response in salicylate treated cells may have clinical implications. For instance, the untoward effects of salicylism may be reduced by measures to increase the core body temperature of the patient in addition to other commonly accepted measures to counteract toxic levels of salicylate. Perhaps more important to the widespread use of NSAIDs, is whether they provide cytoprotective effects because of their potential impact on HS gene expression. *In vitro* experiments have indicated that clinically relevant levels of salicylate influence the HS response in two ways (Jurivich, 1992). First, there is a synergistic effect with heat. The combination of salicylate and heat increases the time that HSF-DNA binding occurs relative to HS alone. Second, prolongation of HSF-DNA binding activity results in the increased expression of the HS genes. While verification of this effect is pending in the patient population, the implication is that NSAIDs have a cytoprotective effect by enhancing HS gene expression. Because salicylate does not appear to increase HSF kinase activity, it is likely that the synergistic effect of salicylate and heat is at the level of HSF oligomerization and, possibly, inhibition of factors which lead to the attenuation of HSF-DNA binding.

Not all NSAIDs appear to mimic the effects of salicylate, although indomethacin does (Lee, 1995). Indomethacin has been found to lower the temperature threshold at which HSF is activated into its DNA bound state. Thus, cells treated with indomethacin will respond to 40–41 °C with HSF oligomerization, whereas untreated cells do not respond until 41.5–42 °C is reached. This observation has raised the possibility that effectors of the HS response can act synergistically, converging upon the pathway(s) that are responsible for HSF oligomerization. The actual signals that trigger HSF oligomerization are elusive, and the known properties of NSAIDs do not resolve this question. NSAIDs can lower intracellular pH and alter intracellular calcium levels. Both of these changes can induce HSF oligomerization in extracts from cells grown at 37 °C, but whether one or both of these changes serve as a primary signal for HSF unfolding and oligomerization *in vivo* is unknown. The mechanism by which salicylate induces HSF-DNA binding appears to require intact cellular processes because addition of salicylate to whole cell extracts does not directly induce HSF oligomerization as in the case of heat, calcium, or low pH. Furthermore, the effects of salicylate on the cell are due to intracellular rather than extracellular changes because nonpermeable salicylates do not induce HSF-DNA binding. Furthermore, HSF-DNA binding attenuates almost immediately after cells are washed of salicylate.

Agents other than the NSAIDs have been found to affect the HS response. In particular, the bioflavanoid, quercetin, has been found to selectively inhibit the induction and accumulation of the HS proteins (Hosokawa et al., 1990). This compound inhibits the synthesis of hsp27, hsp70, and hsp90 while not significantly altering the production of most other proteins. One notable exception is the inhibition of the MDR1 gene by quercetin in cultured cells (Kioka et al., 1989). Hepatocarcinoma cells increase MDR1 gene expression upon exposure to arsenate and this is inhibited by treatment of the cells with quercetin. The mechanism for the drug's action appears to be at the level of transcription. Quercetin has the property of inhibiting HSF-DNA binding both *in vitro* and *in vivo*, thus accounting for the block of HS gene expression. Curiously, the drug does not inhibit other DNA-protein interactions such as the octamer binding proteins. How quercetin prevents the oligomerization of HSF is not known.

Blockage of hsc70 transcription and translation was noticed in HL-60 cells treated with N-methylformamide, an agent which induces differentiation (Beere et. al., 1993). This drug-effect is curious because it does not appear to alter hsp70 or hsp90 transcription. The means by which N-methylformamide selectively inhibits hsc70 expression and not other HS genes has not been elucidated, but interference with constitutively expressed HS proteins is likely to be through disruption of protein binding at basal promoter elements and not HSF-HSE binding activity. Selective disruption of transcription factor binding to its respective DNA binding element is an area of great interest in the development of pharmacologic agents which alter gene expression. In the case of N-methylformamide, the main concern

is whether disruption of hsc70 transcription is due to the drug itself or the differentiation process.

Other drugs such as the neuroleptic, haloperidol, inhibit the induction of hsp70 mRNA in rodent neurons (Sharp et al., 1992). Although this observation needs to be confirmed in the human population, it raises the possibility that an age-dependent defect in the production of HS proteins is exacerbated by a drug which is commonly used in demented elderly patients. The potential for certain pharmacologic agents to inhibit the HS response could increase the risk for untoward effects of atherosclerosis and hypoxia. A similar concern may be raised with certain calcium channel blockers which also have been found to reduce the synthesis of HS proteins in cardiac myocytes (Low-Friedrich and Schoeppe, 1991).

In contrast to quercetin, N-methylformamide, and some calcium channel blockers, other compounds may enhance the accumulation of HS proteins. For example, the anticonvulsant drug, valproic acid has been reported to increase one type of HS protein in glial cells (Martin and Regan, 1988). Cisplatin has been found to induce small HS proteins and thermotolerance in cultured cells (Oesterreich et al. 1991). The antitumor nitrosourea drugs also evoke HS gene transcription (Kroes et al., 1991) via activation of the HS factor. In other instances, enhancement of the HS response may be through posttranscriptional mechanisms. Nutrients such as glutamine enhance the accumulation of HS proteins (Sanders and Kon, 1991). These observations raise the possibility that the HS response can be manipulated, and that clinical interventions can be considered which might enhance the HS response, especially in those patients at risk for ischemic heart or brain disease.

In addition to affecting HSF-DNA interactions, drug-interactions with HS proteins have also been noted. Immunophilins, a class of proteins that bind immunosuppressant drugs, have been found with the glucocorticoid receptor complex which includes hsp90 and hsp70 (Tai et al., 1992). This observation has suggested that as peptidyl-prolyl binding proteins immunophilins participate in the assembly of protein complexes. Whether immunosuppressants such as FK506 or rapamycin interfere with immunophilin function and glucocorticoid receptor regulation is not known, but there is the possibility that drugs can affect peptide binding sites, thus altering the function of chaperones and related classes of protein molecules.

Heat Shock Response in Environmental Toxicology, Human Development, and Aging

There is increasing interest in the relationship of the HS response to environmental toxicology. As induction of the HS response portends resistance to otherwise lethal stresses, many environmental hazards have the potential for inducing the HS response (Finnell et al., 1992). Activity of the dioxin receptor is affected by HS proteins (Nemoto et al., 1990; Pongratz et al., 1992) and it is likely that the HS

proteins are involved in processing damaged proteins caused by a variety of environmental hazards, including ultraviolet radiation, industrial dyes, and heavy metals. Given that some pharmacologic agents uncouple the HS response, it is conceivable that certain environmental toxins could do the same. In this case, organisms exposed to environmental toxins would be at increased risk of cellular injury and death. Interference with the HS response could occur at the transcriptional level or it may involve direct interactions with the chaperones. For instance, chaperones which have an affinity for a particular chemical toxin could alter the balance of chaperone functions and lead to altered cellular processes. Immunologic activities which depend upon the cytoprotective action of the HS proteins would predictably be hampered. Cell growth is known to be altered by physiologic stress and the M and S-phases of cellular division are particularly susceptible to the effects of cellular stress (Westra and Dewey, 1971). Cells in the M or S phase of cell division which are exposed to stress have an increased susceptibility to cell death and induction of polyploidy. These *in vitro* observations may account for the increased incidence of birth defects that are attributed to areas where levels of toxic waste are high (Finnell et al., 1991). In fact, there are several reports that birth defects can be observed in animals exposed to elevated temperatures (Pleet et al., 1981; Webster et al., 1985; Gericke et al., 1989).

Separate from the damaging effects of the stress response on development, it appears that expression of certain HS proteins (by non-stress-inducing stimuli) is linked to specific phases of embryogenesis (Hahnel et al., 1986; Gruppi et al., 1991). For instance, maternal mRNA for hsp90 is translated in developing embryos (Bedard and Brandhorst, 1986). Other studies have suggested a role for the transcription factor HSF2 in the development of the gonadal system (Sarge et al., 1993). Because this factor itself may be sensitive to the effects of physiologic stress, as evidenced by its loss of DNA binding potential *in vitro* (Sarge et al., 1991), toxic-inducers of the stress response could potentially interfere with HSF2 transcriptional regulation and provide a possible explanation for stress-induced birth defects.

Even though environmental toxicology could have profound effects on the HS response, there are other considerations where manipulation of the HS response could be beneficial. For example, constitutive induction of the HS response in bacteria could be a potential means to develop super-microbes which could withstand harsh environments while engaging in their genetically-designed function such as scouring oil spills and enzymatically converting toxins into harmless substances.

Enhancement of the human HS response may eventually be a clinical goal in dealing with the aging population which may be more susceptible to environmental stress. Several investigators have now shown an age-dependent defect in the regulation of the HS response at the transcriptional level (Fargnoli et al., 1990). This defect has been observed in human peripheral blood lymphocytes and in

fibroblast cell lines (Deguchi et al., 1988; Liu et al., 1989; Fassen et al., 1989). Curiously, as one serially passages human fibroblasts in culture, the cells lose their ability to induce HSF-DNA binding as they lose their proliferative potential (Liu et al., 1989). The mechanism leading to this loss of HSF induction is not known. This observation has suggested that posttranslational modifications of HSF or interactions with a negative regulator account for age-dependent changes in the triggering mechanism of HSF. However, cells that have entered the G_0 phase of the cell cycle may also be refractory to induction of HSF-DNA binding activity by certain inducers of the HS response such as the prostaglandins (Holbrook et al., 1992). Thus, it may not be senescence alone that dictates changes in the cellular response to stress, and there may be a fundamental difference in the HS response between post-mitotic cells and cells that have retained their mitotic potential.

SUMMARY

The HS response entails a shift in transcription and protein synthesis when cells are under siege by heat, hypoxia, inflammation, and a variety of other conditions unfavorable to the cell. In addition to altering patterns of transcription and protein synthesis, physiologic stress causes structural changes and alters cell growth. This reaction allows cells to husband their resources until better times arrive. Transient activation of the HS response assists in the repair of cellular injury and it enhances cellular resistance to further damage if unfavorable conditions re-emerge. Heat shock of cultured cells has proven to be an excellent model for dissecting the regulation of inducible gene expression, translational controls, and intracellular protein traffic. While it has provided a model for studying basic mechanisms of cellular function, it is clear that the HS response has important implications for human health and disease.

Alterations in the cytoprotective stress response are being sought as possible explanations for the etiology of several diseases including Alzheimer's disease and some types of autoimmune disease. Modulation of the HS response may be a valuable intervention for patients predisposed to myocardial or brain ischemia. Exogenously added HS proteins appear to prevent retinal damage, and delivery of HS proteins to sites of organ injury may facilitate the healing process and protect against further damage (Margulis et al., 1991). As induction of the HS response has potential clinical value, its uncoupling may also be a promising therapeutic modality. For instance, uncoupling the HS response in tumor cells might increase the efficacy of chemotherapy or even heat therapy. Uncoupling of HS protein function in bacteria may be a novel means to combat infectious agents in lieu of or in addition to antibiotic therapy. Decidedly, there is a wide spectrum of possibilities in citing clinical applications of the HS response.

While much has been learned about the HS response, our understanding is still incomplete. Why, for instance, are there multiple HS transcription factors? Are some of these factors utilized during development and cellular differentiation?

Another paradox is why the HS factor forms trimers when most other factors form dimers. Additional information is needed on how HS factors are reversibly activated and how HSF1 is phosphorylated.

In addition to questions about transcriptional regulation of the HS response, many are directed at the function of HS proteins as chaperones. How they accomplish this function and how they guide intracellular proteins to their appropriate compartment are areas of active research and discussion.

Fundamental knowledge about regulation of the HS response holds great promise for clinical applications. The HS paradigm is an excellent example of how basic science questions can eventually merge with clinical questions, leading to novel approaches in managing human disease.

REFERENCES

Ab, B.K., Kiessling, R., Van, E.J.D., Thole, J.E., Kumararatne, D.S., Pisa, P., Wondimu, A., & Ottenhoff, T.H. (1990). Induction of antigen-specific CD4+ HLA-DR restricted cytotoxic lymphocytes as well as nonspecific nonrestricted killer cells by the recombinant mycobacterial 65 kDa heat shock protein. Eur. J. Immunol. 20, 369–377.

Abe, K., St. George-Hyslop, P.H., Tanzi, R.E., & Kogure, K. (1991). Induction of amyloid precursor protein mRNA after heat shock in cultured human lymphoblastoid cells. Neurosci. Lett. 125, 169–171.

Abe, K. & Kogure, K. (1993). Selective gene expression after brain ischemia. Progr. Brain Res. 96, 221–236.

Abe, M. & Hiraoka, M. (1990). Hyperthermia in combination with radiation in the treatment of cancers. In: Stress Proteins in Biology and Medicine (Morimoto, R., Tissieres, A., Georgopoulos, C., eds.), pp. 117–130. CSHL Press, Cold Spring Harbor.

Abravaya, K., Phillips, B., & Morimoto, R.I. (1991a). Attenuation of the heat shock response in HeLa cells is mediated by the release of bound heat shock transcription factor and is modulated by changes in growth and in heat shock temperatures. Genes and Dev. 5, 2117–2127.

Abravaya, K., Phillips, B., & Morimoto, R.I. (1991b). Heat shock-induced interactions of heat shock transcription factor and the human hsp70 promoter examined by in vivo footprinting. Mol. Cell. Biol. 11, 586–592.

Adeleye, T.A., Colston, M.J., Butler, R., & Jenner, P.J. (1991). The antibody repertoire to proteins of Mycobacterium leprae. J. Immunol. 147, 1947–1953.

Agard, D.A. (1993). To fold or not to fold. Science 260, 1903–1904.

Aguas, A., Esaguy, N., Sunkel, C.E., & Silva, M.T. (1990). Cross-reactivity and sequence homology between the 65 kilodalton mycobacterial heat shock protein and human lactoferrin, transferrin, and DR beta subsets of major histocompatibility complex class II molecules. Infect. Immun. 58, 1461–1470.

Aldashev, A.A., Agibetov, K.A., Yugai, A.A., & Shamshiev, A.T. (1991). Specific proteins synthesized in human lymphocytes during hypoxia. Am. J. Physiol. 261, S92–96.

Amici, C., Sistonen, L., Santoro, M.G., & Morimoto, R.I. (1992). Antiproliferative prostaglandins activate heat shock transcription factor. Proc. Natl. Acad. Sci. USA 89, 6227–6231.

Ananthan, J., Goldberg, A.L., & Voellmy, R. (1986). Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes. Science 232, 522–524.

Anderson, R.L., Tao, T.W., Betten, D.A., & Hahn, G.M. (1986). Heat shock protein levels are not elevated in heat resistant B16 melanoma cells. Radiat Res. 105, 240–246.

Angel, P. & Karin, M. (1991). The role of jun, fos, and the AP-1 complex in cell proliferation and transformation. Biochim. Biophys. Acta 1072, 129–157.

Ashburner, M. (1970). Pattern of puffing activity in the salivary gland chromosomes of *Drosophila*. V. Response to environmental treatments. Chromosoma 31, 356–376.

Atkinson, B.G., Blaker, T.W., Tomlinson, J., & Dean, R.L. (1990). Ferritin is a translationally regulated heat shock protein of avian reticulocytes. J. Biol. Chem. 265, 14156–14162.

Aujame, L. & Firko, H. (1988). The major inducible heat shock protein hsp68 is not required for acquisition of thermal resistance in mouse plasmacytoma cell lines. Mol. Cell. Biol. 8, 5486–5494.

Bagchi, M.K., Tsai, S.Y., Tsai, M.J., & O'Malley, B.W. (1990). Identification of a functional intermediate in receptor activation in progesterone-dependent cell free transcription. Nature 345, 547–550.

Bagchi, M.K., Tsai, S.Y., Tsai, M.J., & O'Malley, B.W. (1991). Progesterone enhances target gene transcription by receptor free of heat shock proteins hsp90, hsp56, and hsp70. Mol. Cell. Biol. 11, 4998–5004.

Baler, R., Welch, W.J., & Voellmy, R. (1992). Heat shock gene regulation by nascent polypeptides and denatured proteins: hsp70 as a potential autoregulatory factor. J. Cell. Biol. 117, 1151–1159.

Baler, R., Dahl, G., & Voellmy, R. (1993). Activation of human heat shock genes is accompanied by oligomerization, modification, and rapid translocation of heat shock transcription factor HSF1. Mol. Cell. Biol. 13, 2486–2496.

Bansal, G.S., Norton, P.M., & Latchman, D.S. (1991). The 90-kDa heat shock protein protects mammalian cells from thermal stress but not from viral infection. Exp. Cell Res. 195, 303–306.

Barbe, M.F., Tytell, M., Gower, D.J., & Welch, W.J. (1988). Hyperthermia protects against light damage in the rat retina. Science 241, 1817–1820.

Beckmann, R.P., Mizzen, L.A., & Welch, W.J. (1990). Interaction of hsp70 with newly synthesized proteins: Implications for protein folding and assembly. Science 248, 2241–2249.

Bedard, P.A. & Brandhorst, B.P. (1986). Translational activation of maternal mRNA encoding the shock protein hsp90 during sea urchin embryogenesis. Dev. Biol. 117, 286–293.

Beere, H.M., Morimoto, R.I., & Hickman, J.A. (1993). Investigations of mechanisms of drug-induced changes in gene expression: N methylformamide-induced changes in synthesis of the M(r) 72,000 constitutive heat shock protein during commitment of HL-60 cells to granulocyte differentiation. Cancer Res. 53, 3034–3039.

Benjamin, I.J., Horie, S., Greenberg, M.L., Alpern, R.J., & Williams, R.S. (1992). Induction of stress proteins in cultured myogenic cells. J. Clin. Invest. 89, 1685–1689.

Blake, M.J., Gershon, D., Fargnoli, J., & Holbrook, N.J. (1990). Discordant expression of heat shock protein mRNAs in tissues of heat stressed rats. J. Biol. Chem. 265, 15275–15279.

Blake, M.J., Udelsman, R., Feulner, G.J., Norton, D.D., & Holbrook, N.J. (1991). Stress induced heat shock protein 70 expression in adrenal cortex: An ACTH-sensitive, age-dependent response. Proc. Natl. Acad. Sci. USA 88, 9873–9877.

Bochkareva, E.S., Lisson, N.M., & Girshovich, A.S. (1988). Transient association of newly synthesized unfolded proteins with the heat shock GroEL protein. Nature 336, 254–257.

Born, W., Hall, L., Dallas, A., Boymel, J., Shinnick, T., Young, D., Brennan, P., & O'Brien, R. (1990). Recognition of a peptide antigen by heat shock-reactive gamma delta T lymphocytes. Science 249, 67–69.

Braell, W.A., Schlossman, D.M., Schmid, S.L., & Rothman, J. (1984). Dissociation of clathrin coats coupled to hydrolysis of ATP: Role of an uncoating ATPase. J. Cell Biol. 99, 734–741.

Bresnick, E.H., Sanchez, E.R., & Pratt, W.B. (1988). Relationship between glucocorticoid receptor steroid binding capacity and association of the Mr 90,000 heat shock protein with the unliganded receptor. J. Steroid Biochem. 30, 267–269.

Briese, E. & DeQuijada, M.G. (1970). Colonic temperature of rats during handling. Acta Physiol. Latinoam. 20, 97–102.

Cadepond, F., Schweizeer, G.G., Segard, M.I., Jibard, N., Hollenberg, S.M., Giguere, V., Evans, R.M., & Baulieu, E. (1991). Heat shock protein 90 as a critical factor in maintaining glucocorticosteroid receptor in a nonfunctional state. J. Biol. Chem. 266, 5834–5841.

Calderwood, S.K., Bornstein, B., Farnum, E.K., & Stevenson, M.A. (1989). Heat shock stimulates the release of arachidonic acid and the synthesis of prostaglandins and leukotriene B4 in mammalian cells. J. Cell. Physiol. 141, 325–333.

Caruso, M., Sacco, M., Medoff, G., & Maresca, B. (1987). Heat shock 70 gene is differently expressed in Histoplasma capsulatum strains with different levels of thermotolerance and pathogenicity. Mol. Microbiol. 1, 151–158.

Chang, C.C., Konno, S., & Wu, J.M. (1991). Enhanced expression of heat shock protein and mRNA synthesis by type I interferon in human HL-60 leukemic cells. Biochem Intl. 24, 369–377.

Chiang, H.L., Terlecky, S.R., Plant, C.P., & Dice, J.F. (1989). A role for a 70 kDa heat shock protein in lysomsomal degradation of intracellular proteins. Science 246, 382–385.

Chin, K.V., Tanaka, S., Darlington, G., Pastan, I., & Gottesman, M.M. (1990). Heat shock and arsenite increase expression of the multi-drug resistance (MDR1) gene in human renal carcinoma cells. J. Biol. Chem. 265, 221–226.

Chirico, W.J., Waters, M.G., & Blobel, G. (1988). 70 kDa heat shock related proteins stimulate protein translocation into microsomes. Nature 332, 805–810.

Christman, M.F., Morgan, R.W., Jacobson, F.S., & Ames, B. (1985). Positive control of a regulon for defenses against oxidative stress and some heat shock proteins in Salmonella typhimurium. Cell 41, 753–762.

Ciavarra, R.P. & Simeone, A. (1990). T lymphocyte stress response. Cell Immunol. 129, 363–376.

Ciocca, D.R., Fuqua, S.A., Lock-Lim, S., Toft, D.O., Welch, W.J. & McGuire, W.L. (1992). Response of human breast cancer cells to heat shock and chemotherapeutic drugs. Cancer Res. 52, 3648–3654.

Clarke, C.F., Cheng, K., Frey, A.B., Stein, R., Hinds, P.W., & Levine, A.J. (1988). Purification of complexes of nuclear oncogene p53 with rat and E. coli heat shock protein: In vitro dissociation of hsc70 and dnaK from murine p53 by ATP. Mol. Cell. Biol. 8, 1206–1215.

Cleget, M. & Polla, B.S. (1990). Erythrophagocytosis induces heat shock protein synthesis by human monocytes-macrophages. Proc. Natl. Acad. Sci. USA 87, 1081–1085.

Clos, J., Westwood, J.T., Becker, P.B., Wilson, S., Lambert, K., & Wu, C. (1990). Molecular cloning and expression of a hexameric Drosophila heat shock factor subject to negative regulation. Cell 63, 1085–1097.

Clos, J., Rabindran, S., Wisniewski, J., & Wu, C. (1993). Induction temperature of human heat shock factor is reprogrammed in Drosophila cell environment. Nature 363, 252–254.

Craig, E.A. (1990). Role of hsp70 in translocation of proteins across membranes. In: Stress Proteins in Biology and Medicine (Morimoto, R., Tissieres, A., Georgopoulos, C., eds.), pp. 279–286, CSHL Press, Cold Spring Harbor.

Craig, E.A. (1993).Chaperones—helpers along the pathways to protein folding. Science 260, 1902–1903.

Currie, R.W. (1987). Effects of ischemia and perfusion temperature on the synthesis of stress-induced (heat shock) proteins in isolated and perfused rat hearts. J. Mol. Cell. Cardiol. 19, 795–808.

Davis, C.E., Guiney, D.G., & Clomerauer, M.E.M. (1987). Molecular strategies of parasite invasion. UCLA Symp. Mol. Cell. Biol. 42, 169–179.

Delcayre, C., Samuel, J.L., Marotte, F., Best, B.M., Mercadier, J.J., & Rappaport, L. (1988). Synthesis of stress proteins in rat cardiac myocytes 2–4 days after imposition of hemodynamic overload. J. Clin. Invest. 82, 460–468.

Deguchi, Y., Negoro, S., & Kishimoto, S. (1988). Age-related changes of heat shock protein gene transcription in human peripheral blood mononuclear cells. Biochem. Biophys. Res. Comm. 157, 580–584.

Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A., & Schekman, R. (1988). kA subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. Nature 332, 800–805.

Devary, Y., Gottlieb, R.A., Smeal, T., & Karin, M. (1992). The mammalian ultraviolet response is triggered by activation of Src tyrosine kinases. Cell 71, 1081–1091.

DiDomenico, B.J., Bugaisky, G.E., & Lindquist, S. (1982). The heat shock response is self-regulated at both the transcriptional and post-transcriptional level. Cell 31, 593–603.

Dinarello, C.A., Berheim, H.A., Duff, G.W., Le, H.V., Nagabhushan, T.L., Hamilton, N.C., & Coceani, F. (1984). Mechanisms of fever induced by recombinant human interferon. J. Clin. Invest. 74, 906–913.

Downing, J.F., Taylor, M.W., Wie, K.-M., & Elizondo, R.S. (1987). In vivo hyperthermia enhances plasma antiviral activity and stimulates peripheral lymphocytes for increased synthesis of interferon gamma. J. Interferon Res. 7, 185–193.

Dubois, M.F., Mezger, V., Morange, M., Ferrieux, C., Lebon, P., & Bensaude, O. (1988). Regulation of the heat-shock response by interferon in mouse L cells. J. Cell Physiol. 137, 102–109.

Eid, N.S., Kravath, R.E., & Lanks, K.W. (1987). Heat shock protein synthesis by human polymorphonuclear cells. J. Exp. Med. 165, 1448–1452.

Ellis, J. (1987). Proteins as molecular chaperones. Nature 328, 378–379.

Fargnoli, J., Kunisada, T., Fornace, A.J., Schneider, E.L., Holbrook, N.J. (1990). Decreased expression of heat shock protein 70 mRNA and protein after heat treatment in cells of aged rats. Proc. Natl. Acad. Sci. USA 87, 846–850.

Fassen, A.E., O'Leary, J.J., Rodysill, K.J., Bergh, N., Hallgren, H.M. (1989). Diminished heat shock protein synthesis following mitogen stimulation of lymphocytes from aged donors. Exp. Cell. Res. 183, 326–334.

Ferris, D.K., Harel, B.A., Morimoto, R.I., Welch, W.J., & Farrar, W.L. (1988). Mitogen and lymphokine stimulation of heat shock proteins in T lymphocytes. Proc. Natl. Acad. Sci. USA 85, 3850–3854.

Findlay, C.A., Hinds, P.W., Tan, T.H., Eliyahu, D., Oren, M., & Levine, A.J. (1988). Activating mutations for transformation by p53 produce a gene product that forms a hsc70–p53 complex with an altered half-life. Mol. Cell. Biol. 8, 531–539.

Finnell, R.H., Mohl, V.K., & Englen, M.D. (1991). In vitro analysis of the murine heat shock response: Implications for reproductive toxicology. Toxicol. Lett. 58, 297–308.

Finnell, R.H., Ager, P.L., Englen, M.D., & Bennett, G.D. (1992). The heat shock response: A potential to screen teratogens. Toxicol. Lett. 60, 39–52.

Fischer, H.P., Sharrock, C.E., & Panayi, G.S. (1992). High frequency of cord blood lymphocytes against mycobacterial 65 kD heat shock protein. Eur. J. Immunol. 22, 1667–1669.

Flanders, K.C., Winokur, T.S., Holder, M.G., & Sporn, M.B. (1993). Hyperthermia induces expression of transforming growth factor beta in rat cardiac cells in vitro and in vivo. J. Clin. Invest. 92, 404–410.

Flynn, G.C., Chappell, T.G., & Rothman, J.E. (1989). Peptide binding and release by proteins implicated as catalysts of protein assembly. Science 245, 385–390.

Flynn, G.C., Pohl, J., Flocco, M.T., & Rothman, J.E. (1991). Peptide-binding specificity of the molecular chaperone BiP. Nature 353, 726–730.

Freedman, M.S., Ruijs, T.C., Selin, L.K., & Antel, J.P. (1991). Peripheral blood gamma delta T cells lyse fresh human brain derived oligodendrocytes. Ann. Neurol. 30, 794–800.

Fuqua, S.A., Blum, S.M., & McGuire, W.L. (1989). Induction of the estrogen-regulated 24 kD protein by heat shock. Cancer Res. 49, 4126–4129.

Gamer, J., Bujard, H., & Bukua, B. (1992). Physical interaction between heat shock proteins DnaK, DnaJ, & GrpE & the bacterial heat shock transcription factor sigma 32. Cell 69, 833–842.

Garofalo, O., Kennedy, P.G., Swash, M., Martin, J.E., Luthert, P., Anderton, B.H., & Leigh, P.N. (1991). Ubiquitin and heat shock protein expression in ALS. Neuropath. Appl. Neurobiol. 17, 39–45.

Gaston, J.S., Life, P.F., Jenner, P.J., Colston, M.J., & Bacon, P.A. (1990). Recognition of a mycobacteria-specific epitope in the 65 kD hsp by synovial fluid derived T cell clones. J. Exp Med. 171, 831–841.

Gericke, G.S., Hofmeyer, G.J., Laburrn, H., & Isaacs, H. (1989). Does heat damage fetuses? Med. Hypothesis 29, 275–278.

Gething, M. & Sambrook, J. (1992). Protein folding in the cell. Nature 355, 33–45.

Gilmour, D.S. & Lis, J.T. (1986). RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene *Drosophila melanogaster* cells. Mol. Gen. Genet. 6, 3984–3989.

Goate, A.M., Cooper, D.N., Hall, C., Leung, T.K., Solomon, E., & Li, L. (1987). Localization of a human heat shock hsp70 gene sequence to chromosome 6 and detection of two other loci by somatic cell hybrid and RFLP analysis. Hum. Gen. 75, 123–128.

Goldenberg, C.J., Luo, Y., Fenna, M., Baler, R., Weinmann, R., & Voellmy, R. (1988). Purified human factor activates heat shock promoter in a HeLa cell free transcription system. J. Biol. Chem. 263, 19734–19739.

Goloubinoff, P., Christeller, J.T., Gatenby, A.A., & Lorimer, G.H. (1989). Reconstitution of active dimeric ribulosebiphosphate carboxylase from an unfolded state depends on two chaperonin proteins and magnesium ATP. Nature 342, 844–889.

Gozales, M.F., Shiraishi, K., Hisanaga, K., Sagar, S.M., Mandabach, M., & Sharp, F.R. (1989). Hsp as markers of neural injury. Brain Res. Mol. Brain Res. 6, 93–100.

Grandia, A.A., Devisser, H., Vanembden, J.D.A., Vanderzee, R., Vanderberg, W.B., & Hazenberg, M.P. (1991). Natural antibodies to 65 kD mycobacterial heat shock protein in rats do not correlate with susceptibility for mycobacterium-tuberculosis induced adjuvant arthritis. Immunobiology 182, 127–134.

Gruppi, C.M., Zakeri, Z.F., & Wolgemuth, D.J. (1991). Stage and lineage regulated expression of 2 hsp90 transcripts during mouse germ cell differentiation and embryogenesis. Mol. Reprod. Dev. 28, 209.

Gubits, R.M. & Fairhurst, J.L. (1988). c-fos mRNA levels are increased by the cellular stressors, heat shock and sodium arsenite. Oncogene 3, 163–168.

Guidon, P.T.J. & Hightower, L.E. (1986). The 73 kD heat shock cognate protein purified from rat brain contains nonesterified palmitic and stearic acids. J. Cell. Physiol. 128, 239–245.

Haahr, S. & Mogensen, S. (1977). Function of fever. Lancet II, 613.

Haas, I.G. & Meo, T. (1988). cDNA cloning of the immunoglobulin heavy chain binding protein. Proc. Natl. Acad. Sci. USA 85, 2250–2254.

Hahnel, A.C., Gifford, D.J., Heikkila, J.J., & Schultz, G.A. (1986). Expression of the major heat shock protein hsp70 family during early mouse embryo development Teratogen. Carcinog. Mutagen. 6, 493–510.

Hallett, P., Mehlert, A., & Maxwell, A. (1990). *E. coli* cells resistant to the DNA gyrase inhibitor, ciprofloxacin, overproduce a 60 kD protein homologous to GroEL. Mol. Microbiol. 4, 345–353.

Hamet, P., Malo, D., Hasimoto, T., & Tremblay, J. (1990a). Increased transcription of a major stress gene in spontaneously hypertensive mice. Hyperten. 15, 904–908.

Hamet, P., Tremblay, J., Malo, D., Kunes, J., & Hashimoto, T. (1990b). Genetic hypertension is characterized by the abnormal expression of a gene localized in major histocompatibility complex hsp70. Transplant. Proc. 22, 2566–2567.

Hamet, P., Kong, D., Pravenec, M., Kunes, J., Kren, V., Klir, P., Sun, Y.L., & Tremblay, J. (1992). Restriction fragment length polymorphism of hsp70 gene, localized in the RT1 complex, is associated with hypertension in spontaneously hypertensive rats. Hypertension 19, 611–614.

Haregewoin, A., Soman, G., Hom, A.C., & Finberg, R.W. (1989). Human γδ T cells respond to mycobacterial heat shock protein. Nature 340, 309–310.

Harrison, G.S., Drabkin, H.A., Kao, F.-T., Hartz, I.M., Chu, E.H.Y., Wu, B.J., & Morimoto, R.I. (1987). Chromosomal location of human genes encoding major heat shock protein hsp70. Somatic Cell Mol. Genet. 13, 119–130.

Hartl, F.-U. & Neupert, W. (1990). Protein sorting to mitochondria: Evolutionary conservations of folding and assembly. Science 247. 930–938.

Hashimoto, T., Mosser, R.D., Tremblay, J., & Hamet, P. (1991). Increased accumulation of hsp70 mRNA due to enhanced activation of heat shock transcription factor in spontaneously hypertensive rats. J. Hyperten. Suppl. 9, S170–171.

Hensold, J.O., Hunt, C.R., Calderwood, S.K., Housman, D.E., & Kingston, R.E. (1990). DNA binding of heat shock factor to the heat shock element is insufficient for transcriptional activation in murine erythroleukemia cells. Mol. Cell Biol. 10, 1600–1608.

Hershko, A. (1988). Ubiquitin-mediated protein degradation. J. Biol. Chem. 263, 15237–15240.

Hightower, L.E. (1980). Cultured animal cells exposed to amino acid analogues or puromucin rapidly synthesize several polypeptides. J. Cell. Physiol. 102, 407–427.

Hightower, L.E. (1991). Heat shock, stress proteins, chaperones, and proteotoxicity. Cell 66, 191–197.

Holbrook, N.J., Carlson, S.G., Choi, A.M.K., & Fargnoli, J. (1992). Induction of hsp70 gene expression by the antiproliferative prostaglandin PGA2: A growth dependent response mediated by activation of heat shock transcription factor. Mol. Cell Biol. 12, 1528–1534.

Hosokawa, N., et al. (1990). Flavinoids inhibit the expression of heat shock proteins. Cell Struct. Funct. 15, 393–401.

Hutchison, K.A., Brott, B.K., De, L.J.H., Perdew, G.H., Jove, R., & Pratt, W.B. (1992). Reconstitution of the multiprotein complex of pp60src, hsp90, and p50 in a cell free system. J. Biol. Chem. 267, 2902–2903.

Ikawa, S. & Weinberg, R.A. (1992). An interaction between p21 ras and heat shock protein hsp60, a chaperonin. Proc. Natl. Acad. Sci. USA 89, 2012–2016.

Izumi, R., Yamada, T., Yoshikai, S., Sasaki, H., Hattori, M., & Sakaki, Y. (1992). Positive and negative regulatory elements for the expression of the Alzheimer's disease amyloid precursor-encoding gene in mouse. Gene 112, 189–195.

Izumo, S., Nadal, G.B., & Mahdavi, V. (1988). Protooncogene induction and reprogramming of cardiac gene expression produced by pressure overload. Proc. Natl. Acad. Sci. USA 85, 339–343.

Jaattela, M., Wissing, D., Bauer, P.A., & Li, G.C. (1992). Major heat shock protein hsp70 protects tumor cells from tumor necrosis factor cytotoxicity. EMBO J. 11, 3507–3512.

Jakobsen, B.K. & Pelham, H.R.B. (1988). Constitutive binding of yeast heat shock factor to DNA in vivo. Mol. Cell. Biol. 8, 5040–5042.

Jameel, A., Skilton, R.A., Campbell, T.A., Chander, S.K., Coombes, R.C., & Luzmani, Y.A. (1992). Clinical and biological significance of hsp89 alpha in human breast cancer. Intl. J. Cancer 50, 409–415.

Jampel, H.D., Duff, G.W., Gershon, R.K., Atkins, E.A., & Durum, S.K. (1983). Fever immunoregulation: III. Hyperthermia augments the primary *in vitro* humoral immune response. J. Exp. Med. 157, 1229–1238.

Johansen, K.S., Berger, E.M., & Repine, J.E. (1983). Effect of temperature on polymorphonuclear leukocyte function. Acta Pathol. Microbiol. Immunol. Scand. Sect C, Immunol. 91, 355–359.

Johnson, A.D., Berberian, P.A., & Bond, M.G. (1993). Atherosclerosis alters the localization of hsp70 in human and macaque aortas. Exp. Mol. Path. 58, 155–168.

Jurivich, D.A., Chung, J., & Blenis, J. (1991). Heat shock induces two distinct S6 protein kinase activities in quiescent mammalian fibroblasts. J. Cell. Physiol. 148, 252–259.

Jurivich, D.A., Sistonen, L., Kroes, R.A., Morimoto, R.I. (1992). Effect of sodium salicylate on the human heat shock response. Science 255, 1243–1245.

Jurivich, D.A., Sistonen, L., Sarge, K.D., & Morimoto, R. (1994). Arachidonate is a potent modulator of heat shock gene expression. Proc. Natl. Acad. Sci. USA 91, 2280–2284.

Jurivich, D.A., Pachetti, C., Qui, L., & Welk, J.F. (1995). Salicylate triggers heat shock factor differently than heat. J. Biol. Chem. 270, 24489–24495.

Kawagoe, J., Abe, K., & Kogure, K. (1993). Regional difference of hsp70 and hsc70 heat shock mRNA inductions in rat hippocampus after transient global ischemia. Neurosci. Lett. 153, 165–168.

Kelley, P.M. & Schlesinger, M.J. (1978). The effect of amino acid analogues and heat shock gene expression in chicken embryo fibroblasts. Cell 15, 1277–1286.

Kettelhut, I.C. & Goldberg, A.L (1988). Tumor necrosis factor can induce fever in rats without activating protein breakdown in muscle or lipolysis in adipose tissue. J. Clin. Invest. 81, 1384–1389.

Keyse, S.M. & Emslie, E.A. (1992). Oxidative stress and heat shock induce a human gene encoding a protein-tyrosine phosphatase. Nature 359, 644–647.

Kimura, L., Plymer, M., McLean, L., Yamaga, K., & Lance, E. (1991). Reaction of antibody to myocbacterial 65 kDa heat shock protein with human chondrocytes. J. Autoimmun. 4, 881–892.

Kingston, R.E., Schuetz, T.J., & Larin, Z. (1987). Heat-inducible human factor that binds to a human hsp70 promoter. Mol. Cell. Biol. 7, 1530–1534.

Kioka, N., Hosokawa, N., Komano, T., Hirayoshi, H., Nagat, K., & Ueda, K. (1992). Quercetin, a bioflavonoid, inhibits the increase of human multi-drug resistance gene (MDR1) expression caused by arsenite. FEBS Lett. 301, 307–309.

Kluger, M.J., Ringler, D.H., & Anver, M.R. (1975). Fever and survival. Science 188, 166–168.

Koga, T., Wand, W.A., DeBruyn, J., Munk, M.E., Schoel, B., & Kaufmann, S.H. (1989). T cells against a bacterial heat shock protein recognize stressed macrophages. Science 245, 1112–1115.

Kroes, R.A., Abravaya, K., Seidenfeld, J., & Morimoto, R.I. (1991). Selective activation of human heat shock gene transcription by nitrosurea antitumor drugs mediated by isocyanate-induced damage and activation of heat shock transcription factor. Proc. Natl. Acad. Sci. USA 88, 4825–4829.

Kroeger, P.E., & Morimoto, R.I. (1994). Selection of new HSF1 and HSF2 DNA-binding sites reveals difference in trimer cooperativity. Mol. Cell. Biol. 14, 7592–7603.

Kubo, T., Towle, C.A., Mankin, H.J., & Treadwell, B.V. (1985). Stress-induced proteins in chondrocytes from patients with osteoarthritis. Arthritis Rheum. 28, 1140–1145.

La, T.N.B. & Latchman, D.S. (1988). A cellular protein related to heat shock protein 90 accumulates during herpes simplex virus infection and is overexpressed in transformed cells. Exp. Cell Res. 178, 169–179.

Landry, J., Chretien, P., Laszlo, A., & Lambert, H. (1991). Phosphorylation of hsp27 during development and decay of thermotolerance in Chinese hamster cells. J. Cell Physiol. 147, 93–101.

Landry, J., Lambert, H., Zhou, M., Lavoie, J.N., Hickey, E., Weber, L.A., & Anderson, C.W. (1992). Human hsp27 is phosphorylated at serines 78 and 82 by heat shock and mitogen activated kinases that recognize the same amino acid motif as S6 kinase II. J. Biol. 267, 794–803.

Langer, T., Lu., C, Echols, H., Flanagan, J., Hayer, M.K., & Hartl, F.-U. (1992). Successive action of Dnak, DnaJ, & GroEL along the pathway of chaperone mediated protein folding. Nature 356, 683–689.

Larson, J.S., Schuetz, T.J., & Kingston, R.E. (1988). Activation in vitro of sequence specific DNA binding by a human regulatory factor. Nature 335, 372–375.

Lebeaus, J., Le, C.C., Prosperi, M.T., & Goubin, G. (1991). Constitutive overexpression of a 89 kD heat shock protein gene in the HBL100 human mammary cell line converted to a tumorigenic phenotype by the EJ/T24 Harvey ras oncogene. Oncogene 6, 1125–1132.

Lee, B.S., Chen, J., Angelidis, C., Jurivich, D.A., & Morimoto, R.I. (1995). Pharmacological modulation of heat shock factor 1 by antiinflammatory drug results in protection against stress-induced cellular damage. Proc. Natl. Acad. Sci. USA 92, 7207–7211.

Lee, H., Kraus, K.W., Wolfner, M.F., & Lis, J. (1992). DNA sequence requirements for generating paused polymerase at the start of hsp70. Genes Dev. 6, 284–295.

Legagneux, V., Morange, M., & Bensaude, O. (1990). Heat-shock and related stress enhance RNA polymerase II C-terminal-domain kinase activity in HeLa cell extracts. Eur. J. Biochemistry 193, 121–126.

Li, G. & Laszlo, A. (1985). Thermotolerance in mammalian cells: A possible role for the heat shock proteins. In: Changes in Eukaryotic Gene Expression in Response to Environmental Stress (Atkinson, B.G. & Walden, D.B. eds.), pp. 349–371, Academic Press, Orlando.

Li, G.C., Li, L.G., Liu, Y.K., Mak, J.Y., Chen, L.L., & Lee, W.M. (1991). Thermal response of rat fibroblasts stably transfected with the human 70 kDa heat shock protein-encoding gene. Proc. Natl. Acad. Sci. USA 88, 1681–1685.

Li, G.C., Li, L.G., Liu, R.Y., Rehman, M., & Lee, W.M. (1992). Heat shock protein hsp70 protects cells from thermal stress even after deletion of its ATP binding domain. Proc. Natl. Acad. Sci. USA 89, 2036–2040.

Lindquist, S. (1980). Translational efficiency of heat induced messages in *Drosophila melanogaster* cells. J. Mol. Biol. 137, 151–158.

Lindquist, S. & Craig, E.A. (1988). The heat-shock proteins. Ann. Rev. Genet. 22, 631–677.

Lis, J. & Wu, C. (1993). Protein traffic on the heat shock promoter-parking, stalling, and trucking along. Cell 74, 1–4.

Liu, A.Y., Lin, Z., Choi, H.S., Sorhage, F., & Li, B. (1989). Attenuated induction of heat shock gene expression in aging diploid fibroblasts. J. Biol. Chem. 264, 12037–12045.

Low-Friedrich, I. & Schoeppe, W. (1991). Effects of calcium channel blockers on stress protein synthesis in cardiac myocytes. J. Cardiovasc. Pharmacol. 17, 800–806.

Lowenstein, D.H., Chan, P.H., & Miles, M.F. (1991). The stress protein response in cultured neurons: Characterization and evidence for a protective role in excitotoxicity. Neuron 7, 1053–1060.

Lukacs, K.V., Lowrie, D.B., Stokes, R.W., & Colston, M.J. (1993). Tumor cells transfected with a bacterial heat shock gene lose tumorigenicity and induce protection against tumors. J. Exp. Med. 178, 343–348.

Macejak, D.G. & Luftig, R.B. (1991). Association of Hsp70 with the adenovirus type 5 fiber protein in infected Hep-2 cells. Virology 180, 120–125.

MacLennan, D.H. & Phillips, M.S. (1992). Malignant hyperthermia. Science 256, 789–793.

Malamy, J., Carr, J.P., Klessig, D.F., & Raskin (1990). Salicylic acid: A likely endogenous signal in the resistance response of tobacco to viral infection. Science 250, 1002–1004.

Manzerra, P. & Brown, I.R. (1990). Time course of induction of a heat shock gene (hsp70) in the rabbit cerebellum after LSD *in vivo*: Involvement of drug-induced hyperthermia. Neurochen. Res. 15, 53–59.

Margulis, B.A., Sandler, S., Eizirik, D.L., Welsh, N., & Welsh, M. (1991). Liposomal delivery of purified heat shock protein hsp70 in rat pancreatic islets as protection against interleukin 1 beta-induced impaired beta cell function. Diabetes 40, 1418–1422.

Martin, M.L. & Regan, C.M. (1988). The anticonvulsant sodium valproate specifically induces the expression of a rat glial heat shock protein which is identified as the collagen type IV receptor. Brain Res. 459, 131–137.

Matthews, R.C., Burnie, J.P., Howat, D., Rowland, T., & Walton, F. (1991). Autoantibody to heat shock protein 90 can mediate protection against systemic candidosis. Immunology 74, 20–24.

McAlister, L. & Finkelstein, D.B. (1980). Heat shock proteins and thermal resistance in yeast. Biochem. Biophys. Res. Commun. 93, 819–824.

McGarry, T.J. & Lindquist, S. (1985). The preferential translation of *Drosophila* hsp70 mRNA requires sequences in the untranslated leader. Cell 43, 903–911.

McMullin, T.W. & Hallberg, R.L. (1986). Effect of heat shock on ribosome structure: Appearance of a new ribosome associated protein. Mol. Cell. Biol. 6, 2527–2535.

Mehta, H.B., Popovich, B.K., & Dillman, W.H. (1988). Ischemia induces changes in the level of mRNAs coding for stress protein 71 and creatin kinase M. Circ. Res. 63, 512–517.

Miyata, Y. & Yahara, I. (1992). The 90-kDa heat shock protein, hsp90, binds and protects casein kinase II from self aggregation and enhances its kinase activity. J. Biol. Chem. 267, 7042–7047.

Morgan, R.W., Christman, M.F., Jacobson, F.S., Storz, G., & Ames, B. (1986). Hydrogen peroxide-inducible proteins in Salmonella typhimurium overlap with heat shock and other stress proteins. Proc. Natl. Acad. Sci. USA 83, 8059–8063.

Morimoto, A., Murakami, M., Takada, M., Teshirogti, S., & Watanabe, T. (1987). Fever and the acute phase response induced in rabbits by human recombinant interferon. J. Physiol. 391, 209–218.

Morimoto, R.I., Sarge, K.D., & Abravaya, K. (1992). Transcriptional regulation of heat shock genes. J. Biol. Chem. 267, 21987–21990.

Morimoto, R.I. (1993). Cells in stress: Transcriptional activation of heat shock genes. Science 259, 1409–1410.

Mosser, D.D., Theodorakis, N.G., & Morimoto, R.I. (1988). Coordinate changes in heat shock element binding activity and hsp70 gene transcription rates in human cells. Mol. Cell. Biol. 8, 4736–4744.

Mosser, D.D., Kotzbauer, P.T., Sarge, K.D., & Morimoto, R.I. (1990). *In vitro* activation of heat shock transcription factor DNA binding by calcium and biochemical conditions that affect protein conformation. Proc. Natl. Acad. Sci. USA 87, 3748–3752.

Mosser, D.D., Duchaine, J., & Massie, B. (1993). The DNA binding activity of the human heat shock transcription factor is regulated in vivo by hsp70. Mol. Cell. Biol. 13, 5427–5438.

Muhich, M.L. & Boothroyd, J.C. (1989). Synthesis of trypanosome hsp70 mRNA is resistant to disruption of trans-splicing by heat shock Mol. Cell. Biol. 8, 3837–3846.

Muller, S., Briand, J.P., & Van, R.M.H. (1988). Presence of antibodies to ubiquitin during the autoimmune response associated with systemic lupus erythematosis. Proc. Natl. Acad. Sci. USA 85, 8176–8180.

Munk, M.E., Schoel, B., Modrow, S., Karr, R.W., Young, R.A., & Kaufmann, S.H. (1989). T lymphocytes from healthy individuals with specificity to self-epitopes shared by the mycobacterial and human 65 kD heat shock protein. J. Immunol. 143, 2844–2849.

Nakai, A., & Morimoto, R.I. (1993). Characterization of a novel chicken heat shock transcription factor, heat shock factor 3, suggests a new regulatory pathway. Molecular & Cellular Biology 13, 1983–1997.

Nemoto, T., Mason, G.G., Wilhelmsson, A., Cuthill, S., Hapgood, J., Gustafsson, J.A., & Poellinger, L. (1990). Activation of the dioxin and glucocorticoid receptors to a DNA binding state under cell free conditions J. Biol. Chem. 265, 2269–2277.

Nowak, T.S., Bond, U., & Schlesinger, M.J. (1990). Heat shock levels in brain and other tissues after hyperthermia and transient ischemia. J. Neurochem. 54, 451–458.

Otenhoff, T.H., Kale, A.B., VanEmbden, J.D.A., Thole, J.E.R., & Kiessling, R. (1988). The recombinant 65-kD heat shock protein of Mycobacterium bovis bacillus calmette guerin/M. tuberculosis is a target molecule for CD4 + cytotoxic T lymphocytes that lyse human monocytes. J. Exp. Med. 168, 1947–1952.

Ostermann, J., Horwich, A.L., Neupert, W., & Hartl, F.-U. (1989). Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. Nature 341, 125–130.

Osterreich, S., Schunck, H., Benndorf, R., & Bielka, H. (1991). Cisplatin induces the small heat shock protein hsp25 and thermotolerance in Ehrlich ascites tumor cells. Biochem. Biophys. Res. Comm. 180, 243–248.

Papasozomenos, S.C. & Su, Y. (1991). Altered phosphorylation of tau protein in heat shocked rats and patients with Alzheimer's disease. Proc. Natl. Acad. Sci. USA 88, 4543–4547.

Pappolla, M.A., Omar, R.A., Kim, K.S., & Robakis, N.K. (1992). Immunochemical evidence of antioxidant stress in Alzheimer's disease. Am. J. Pathol. 140, 621–628.

Pelham, H.R.B. (1986). Speculations on the functions of the mayor heat shock and glucose regulated proteins. Cell 46, 959–961.

Pelham, H.R.B. (1990). Functions of the hsp70 protein family: An overview. In: Stress Proteins in Biology Medicine (Morimoto, R., Tissieres, A., Georgopoulos, C., eds.), pp. 287–300, CSHL Press, Cold Spring Harbor.

Perdew, G.H. & Whitelaw, M.L. (1991). Evidence that the 90 kD heat shock protein exists in cytosol in heteromeric complexes containing hsp70 and three other proteins with Mr of 63,000, 56,000, 50,000. J. Biol. Chem. 266, 6708–6713.

Perez, N., Sugar, J., Charya, S., Johnson, G., Merril, C., Bierer, L., Perl, D., Haroutunian, V., & Wallace, W. (1991). Increased synthesis and accumulation of heat shock proteins in Alzheimer's disease Brain Res. Mol. Brain Res. 11, 249–254.

Phillips, B. & Morimoto, R.I. (1991). Transcriptional regulation of human hsp70 genes: Relationship between cell growth, differentiation, virus infection, and the stress response. In: Results &

Problems in Cell Differentiation, 17, Heat shock and Development (Hightower, L., Nover, D.. eds.), pp. 167–187, Springer-Verlag, Berlin.

Phillips, B., Abravaya, K., & Morimoto, R.I. (1991). Analysis of the specificity and mechanism of the transcriptional activation of the human hsp70 gene during infection by DNA viruses. J. Virol. 11, 5680–5692.

Pleet, H., Grahan, J.M., & Smith, D.W. (1981). Central nervous system and facial defects associated with maternal hyperthermia at four to 14 weeks gestation. Pediatrics 67, 785–789.

Pongratz, I., Mason, G.G., & Poellinger, L. (1992). Dual roles of the 90 kD heat shock protein hsp90 in modulating functional activities of the dioxin receptor. J. Biol. Chem. 267, 13728–13734.

Pope, R.M., Wallis, R.S., Sailer, D., Buchanan, T.M., & Pahlavani, M.A. (1991). T cell activation by mycobacterial antigens in inflammatory synovitis. Cell Immunol. 133, 95–108.

Pope, R.M., Lovis, R.M., & Gupta, R.S. (1992). Activation of synovial fluid T lymphocytes by 60 kD heat shock proteins in patients with inflammatory synovitis. Arthritis Rheum. 35, 43–48.

Raabe, T. & Manley, J.L. (1991). A human homologue of the *E. coli* DnaJ heat shock protein. Nuc. Acids Res. 19, 6645.

Rabindran, S.K., Haroun R.I., Clos, J., Wisniewski, J., & Wu, C. (1993). Regulation of heat shock factor trimer formation: Role of a conserved leucine zipper. Science 259, 230–234.

Ritossa, F.M. (1962). A new puffing pattern induced by a temperature shock and DNP in *Drosophila*. Experientia 18, 571–573.

Ritossa, F.M. (1963). New puffs induced by temperature shock DNP and salicylate in salivary chromosomes of *Drosophila melanogaster*. Drosophila Inf. Service 37, 122–123.

Rordorf, B., Koroshetz, W.J., and Boventre, J.V. (1991). Heat shock protects cultured neurons from glutamate toxicity. Neuron 7, 1043–1051.

Rougvie, A.E & Lis, J.T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene in *D. melanogaster* is transriptionally engaged. Cell 54, 795–804.

Rowley, A., Johnston, G.C., Butler, B., Werner-Washburne, M., & Singer, R.A. (1993). Heat shock-mediated cell cycle blockage and G1 cyclin expression in the yeast saccomyces cerevisiae. Mol. Cell. Biol. 13, 1034–1041.

Salbaum, J.M., Weidemann, A., Masters, C.L., & Beyreuther, K. (1989). The promoter of Alzheimer's disease amyloid A4 precursor gene. Prog. Clin. Biol. Res. 317, 277–283.

Salo, D.C., Donovan, C.M., & Davies, K.J. (1991). Hsp70 and other possible heat shock or oxidative stress proteins are induced in skeletal muscle, heart, and liver during exercise. Free Radic. Biol. Med 11, 239–246.

Sanchez, Y. & Lindquist, S.L. (1990). Hsp104 is required for induced thermotolerance. Science 248, 1112–1115.

Sanders, M.M. & Kon, C. (1991). Glutamine is a powerful effector of heat shock protein expression in *Drosophila* Kc cells. J. Cell Physiol. 146, 180–190.

Sarge, K.D., Zimarino, V., Holm, K., Wu, C., Morimoto, R.I. (1991). Cloning and characterization of two mouse heat shock factors with distinct inducible and constitutive DNA binding ability. Genes Dev. 5, 1902–1911.

Sarge, K.D., Murphy, S.P., & Morimoto, R.I. (1993). Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA binding activity, and nuclear localization and can occur in the absence of stress. Mol. Cell. Biol. 13, 1392–1407.

Sargent, C.A., Dunhan, I., Trowsdale, J., & Campbell, R.D. (1989). Human major histocompatibility complex contains genes for the major heat shock protein hsp70. Proc. Natl. Acad. Sci. USA 86, 1968–1972.

Schellenberg, G.D., Payami, H., Wysman, E.M., Orr, H.T., Goddard, K.A., Anderson, L., Nemens, E., White, J.A., Alonso, M.E., & Ball, M.J. (1992). Genetic linkage evidence for a familial Alzheimer's Disease locus on chromosome 14. Science 258, 668–671.

Schiller, P., Amin, J., Ananthan, J., Brown, M.E., Scott, W.A., & Voellmy, R. (1988). Cis acting elements involved in the regulated expression of human hsp70 gene. J. Mol. Biol. 203, 97–105.

Schmidt, J.A. & Abdulla, E. (1988). Down regulation of IL-1β biosynthesis by inducers of the heat shock response. J. Immunol. 141, 2027–2034.

Selmaj, K., Brosnan, C.F., & Raine, C.S. (1991). Colocalization of lymphocytes bearing gamma delta T-cell receptor and heat shock protein hsp65 oligodendrocytes in multiple sclerosis. Proc. Natl. Acad. Sci. USA 88, 6452–6456.

Selmaj, K., Brosnan, C.F., & Raine, C.S. (1992). Expression of heat shock protein 65 by oligodendrocytes *in vivo* and *in vitro*: Implications for multiple sclerosis. Neurology 42, 795–800.

Shanafelt, M.C., Hindersson, P., Soderberg, C., Mensi, N., Turck, C.W., Webb, D., Yssel, H., & Pertz, G. (1991). T cell and antibody reactivity with the Borrelia burgdorferi 60 kD heat shock protein in Lyme arthritis. J. Immunol. 146, 3985–3992.

Sharp, F.R., Butman, M., Wang, S., Koistinaho, J., Graham, S.H., Sagar, S.M., Noble, L., Berger, P., & Longo, F.M. (1992). Haloperidoe prevents induction of the hsp70 heat shock gene in neurons injured by phencyclidine, MK801, and ketamine. J. Neurosci Res. 33, 605–616.

Shi, Y. & Thomas, J.O. (1992). The transport of proteins into the nucleus requires the 70 kD heat shock protein or its cytosolic cognate. Mol. Cell. Biol. 12, 2186–2192.

Silver, P.A. & Way, J.C. (1993). Eukaryotic DnaJ homologs and the specificity of hsp70 activity. Cell 74, 5–6.

Sistonen, L., Sarge, K.D., Phillips, B., Abravaya, K., & Morimoto, R.I. (1992). Activation of heat shock factor 2 during hemin-induced differentiation of human erythroleukemia cells. Mol. Cell. Biol. 12, 4104–4111.

Skroch, P., Buchman, C., & Karin, M. (1993). Regulation of human and yeast metallothionein gene transcription by heavy metal ions. Prog. Clin. Biol. Res. 380, 113–128.

Smiley, J.D. & Hoffman, W.L. (1991). The role of infections in the rheumatic diseases: molecular mimicry between bacterial and human stress proteins? Am. J. Med. Sci. 301, 138–149.

Snoeckx, L.H., Cornelussen, R., Vandervusse, G.J., & Reneman, R.S. (1992). The expression of heat shock protein 70 kD in the hypertrophied and nonhypertrophied heart of heat shocked rats. FASEB J. 6(4), 1259–1264.

Solomon, J.M., Rossi, J.M., Golic, K., McGarry, T., & Lindquist, S. (1991). Changes in hsp70 alter thermotolerance and heat shock regulation in *Drosophila*. New Biologist 3, 1106–1120.

Sorger, P.K. & Pelham, H.B.R. (1988). Yeast heat shock factor is an essential DNA binding protein that exhibits temperature-dependent phosphorylation. Cell 54, 855–864.

Sorger, P.K. & Nelson, H.C.M. (1989). Trimerization of a yeast transcriptional activator via a coiled-coil motif. Cell 59, 807–813.

Sorger, P.K., Lewis, M.J., & Pelham, H.B.R. (1987). Heat shock factor is regulated differently in yeast and HeLa cells. Nature 329, 81–84.

Sorger, P.K. (1990). Yeast heat shock factor contains separable transient and sustained response transcriptional activators. Cell 62, 793–805.

Spradling, A., Pardue, M.L., & Penman, S. (1977). Messenger RNA in heat shocked *Drosophila* cells. J. Mol. Biol. 109, 559–587.

Storz, G., Tartaglia, L.A., & Ames, B.N. (1990). Transcriptional regulator of oxidative stress-inducible genes: Direct activation by oxidation. Science 248, 189–194.

Tai, P.K., Albers, M.W., Chang, H., Faber, L.E., & Schreiber, S.L. (1992). Association of a 59 kilodalton immunophilin with the glucocorticoid receptor complex. Science 256, 1315–1318.

Takenaka, I.M. & Hightower, L.E. (1993). Regulation of chicken hsp70 and hsp90 family gene expression by transforming growth factor beta 1. J. Cell Physiol. 155, 54–62.

Tetu, B., Lacasse, B., Bouchard, H.L., Lagace, R., Hout, J., & Landry, J. (1992). Prognostic influence of hsp27 expression in malignant fibrous histiocytoma: A clinicopathological and immunohistochemical study. Cancer Res. 52, 2325–2328.

Theodorakis, N.G., Zand, D.J., Kotzbauer, P.T., Williams, G.T., & Morimoto, R.I. (1989). Hemin induced transcriptional activation of the hsp70 gene during erythroid maturation in K562 cells is due to a heat shock factor mediated stress response. Mol. Cell. Biol. 9, 3166–3173.

Tsoulfa, G., Rook, G.A., Bahr, G.M., Sattar, M.A., Behkehani, K., Young, D.B., Mehlert, A., Van Emblen, J.D., Hay, F.C., & Isenberg, D.A. (1989). Elevated IgG antibody levels to the mycobacterial 65 kD heat shock protein are characteristic of patients with rheumatoid arthritis. Scand. J. Immunol. 30, 519–527.

Tsukiyama, T., Becker, P.B., & Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 367, 525–532.

Udelsman, R., Blake, M.J., Stagg, C.A., Li, D.-G., Putney, J., & Holbrook, N.J. (1993). Vascular heat shock protein expression in response to stress. J. Clin. Invest. 91, 465–473.

Van, D.R.A., Mostafapour, S., Marsh, H.M., Li, Y., & Chopp, M. (1992). Immunocytochemical detection of the 72 kDa heat shock protein in halothane-induced hepatotoxicity in rats. Life Sci. 50, 5–6.

Vanbuskirk, A., Crump, B.L., Margoliash, E., & Pierce, S.K. (1989). A peptide binding protein having a role in antigen presentation is a member of the hsp70 heat shock family. J. Exp. Med. 170, 1799–1809.

Vass, K., Welch, W.J., & Nowak, T.S.J. (1988). Localization of 70 kD stress protein induction in gerbil brain after ischemia. Acta Neuropath. (Berl.) 77, 128–135.

Vass, K., Berger, M.L., Nowak, T.S.J., & Welch, W.J. (1989). Induction of stress protein hsp70 in nerve cells after status epilepticus in the rat. Neurosci. Lett. 100, 259–264.

Wakui, H., Itoh, H., Tashima, Y., Kobayasyi, R., Nakamoto, Y., & Miura, A.B. (1991). Specific antibodies against the stress inducible 72 kD protein, a member of the heat shock protein hsp70, in healthy human subjects. Intl. J. Biochem. 23, 975–978.

Walkup, L.K. & Kogoma, T. (1989). *E. coli* proteins inducible by oxidative stress mediated by the superoxide radical. J. Bacteriol. 171, 1476–1484.

Watowich, S. & Morimoto, R.I. (1988). Complex regulation of heat shock and glucose responsive genes in human cells. Mol Cell. Biol. 8, 393–405.

Watowich, S., Morimoto, R.I., & Lamb, R.A. (1991). Flux of the paramyxovirus hemagglutinin-neuraminidase glycoprotein through the endoplasmic reticulum activates transcription of the grp78-Bip gene. J. Virol. 65, 3690–3697.

Webster, W.S., Gernain, M.A., & Edwards, M.J. (1985). The induction of microphthalmia, encephalocele, and other heat defects following hypothermia during the gastrulation process in the rat Teratology 31, 73–82.

Welch, W.J. & Feramisco, J.R. (1985). Rapid purification of mammalian 70,000 Dalton stress proteins: Affinity of the proteins for nucleotides. Mol. Cell. Biol. 5, 1229–1237.

Welch, W.J. (1991). The role of heat shock proteins as molecular chaperones Curr. Opin. Cell Biol. 3, 1033–1038.

Westra, A. & Dewey, W.C. (1971). Variation in the sensitivity to heat shock during the cell cycle of Chinese hamster cells *in vitro*. Intl. J. Radiat. Biol. 19, 467–477.

Westwood, J.T., Clos, J., & Wu, C. (1991). Stress-induced oligomerization and chromosomal relocalization of heat shock factor. Nature 353, 822–827.

Westwood, J.T. & Wu, C. (1993). Activation of *Drosophila* heat shock factor: Conformational change associated with monomer to trimer transition. Mol. Cell. Biol. 13, 3481–3486.

Wiech, H., Buchner, J., Zimmermann, R., & Jakob, U. (1992). Hsp90 chaperones protein folding *in vitro*. Nature 358, 169–170.

Williams, G.T. & Morimoto, R.I. (1990). Maximal stress-induced transcription from the human hsp70 promoter requires interactions with the basal promoter elements independent of rotational alignment. Mol. Cell. Biol. 10, 3125–3136.

Williams, R.S. & Benjamin, I.J. (1991). Stress proteins and cardiovascular disease. Mol. Biol. Med. 8, 197–206.

Winrow, V.R., McLean, L., Morris, C.J., & Blake, D.R. (1990). The heat shock response and its role in inflammatory disease. Ann. Rheum. Dis. 49, 128–132.

Wong, M.L., Weiss, S.R., Gold, P.W., Doi, S.Q., Banerjee, S., Lucinio, J., Lad, R., Pst, R.M., & Smith, M.A. (1992). Induction of constitutive heat shock protein 73 mRNA in the dentate gyrus by seizures. Brain Res. Mol. Brain Res. 13, 19–25.

Wu, B. & Morimoto, R.I. (1985). Transcription of the human hsp70 gene is induced by serum stimulation. Proc. Natl. Acad. Sci. USA 82, 6070–6074.

Wu, B., Kingston, R.E., & Morimoto, R.I. (1986). Human hsp70 promoter contains at least two distinct regulatory domains. Proc. Natl. Acad. Sci. USA 83, 629–633.

Wu, B., Williams, G.T., & Morimoto, R.I. (1987). Deletion of three protein binding sites in the serum-regulated promoter of the human gene encoding the 70 kDa heat shock protein. Proc Natl. Acad. Sci. USA 84, 2203–2207.

Wucherpfenning, K.W., Newcombe, J., Li, H., Keddy, C., Cuzner, M.L., & Hafler, D.A. (1992). γδ T-cell receptor repertoire in acute multiple sclerosis lesions. Proc. Natl. Acad. Sci. USA 89, 4588–4592.

Xu, C. (1993). cDNA cloning of a mouse factor that activates transcription from a metal response element of the mouse metallothionein-I gene in yeast. DNA Cell Biol. 12, 517–525.

Xu, Q. et al. (1992). Induction of arteriosclerosis in normocholesterolemic rabbits by immunization with heat shock protein 65. Arterioscler. Thromb. 12, 789–799.

Yost, H.J. & Lindquist, S. (1986). RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell 45, 185–193.

Yost, H.J., Petersen, R.B., & Lindquist, S. (1990). Post-transcriptional regulation of heat shock protein synthesis in *Drosophila* In: Stress Proteins in Biology and Medicine (Morimoto, R., Tissieres, A. Georgopoulos, C., eds.), pp. 379–409, CSH Laboratory Press, Cold Spring Harbor.

Young, D., Lathigra, R., Hendrix, R., Sweetser, D., & Young, R.A. (1988). Stress proteins are immune targets in leprosy and tuberculosis. Proc. Natl. Acad. Sci. USA 85, 4267–4270.

Young, D.B., Mehlert, A., & Smith, D.F. (1990). Stress proteins and infectious diseases In: Stress Proteins in Biology and Medicine (Morimoto, R., Tissieres, C., & Georgopoulos, C., eds.), pp. 131–166, CSHL Press, Cold Spring Harbor.

Yu, M., Goldberg, S., & Goldberg, A.L. (1992). Heat shock in *Escherichia coli* alters the protein-binding properties of the chaperonin groEL by inducing its phosphorylation. Nature 357, 167–169.

Zimarino, V. & Wu, C. (1987). Induction of sequence-specific binding of *Drosophila* heat shock activator without protein synthesis. Nature 327, 727–730.

Zimarino, V., Tsai, C., & Wu, C. (1990a). Complex modes of heat shock factor activation. Mol. Cell. Biol. 10, 752–759.

Zimarino, V, Wilson, S., & Wu, C. (1990b). Antibody mediated activation of *Drosophila* heat shock factor *in vitro*. Science 249, 546–549.

Chapter 14

Cell Culture

MARY L. TAUB

INTRODUCTION

The growth and functional properties of animal cells in the body are regulated by the hormonal and nutritional status of the whole animal (Sato, 1975). In response to changes in the external environment, different sets of endocrine hormones are secreted into the blood, affecting a number of different tissues. In addition, factors produced in the local tissue environment can also affect animal cells, including autocrine factors (produced by the cells themselves), and paracrine factors (produced by adjacent cells). Included among these local tissue factors are extracellular matrix proteins and growth factors (Figure 1). As animal cells are in a complex dynamic environment in the whole animal (*in vivo*), their regulatory processes are

Principles of Medical Biology, Volume 4
Cell Chemistry and Physiology: Part IV, pages 463–483.

ISBN: 1-55938-808-0

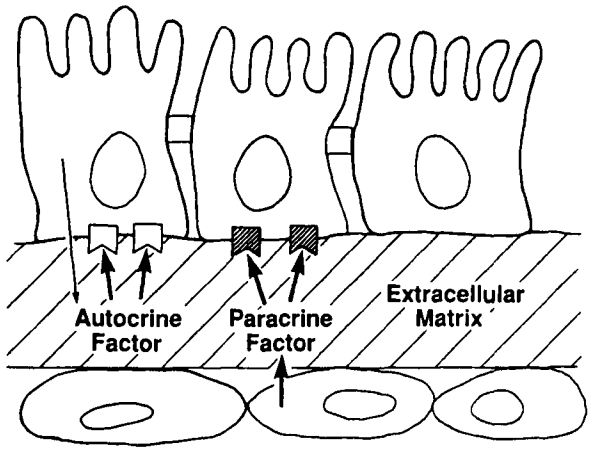


Figure 1. Autocrine and paracrine factors. Many animal cells secrete regulatory factors called autocrine factors, which then interact with specific receptors on the surface of the same cells, so as to modulate cell function. Animal cells may also secrete regulatory factors called paracrine factors, which interact with specific receptors on different cells in the same locale, modulating their functional properties.

more readily examined in a more simplified environment external to the animal (*in vitro*).

When animal cells are removed from the body, the cells can be maintained (or cultured) for prolonged periods, provided the cells are in an appropriate alternative environment. Environmental factors of importance for animal cell culture include the culture medium, substratum (or surface of the culture vessel), and temperature. The substratum (or surface) is a significant factor for those animal cells which grow attached to the surface of the culture container. A number of types of animal cells (such as lymphoid cells) grow in suspension.

Animal cell cultures that are initiated from cells removed directly from the animal are called primary cultures (Figure 2). Primary cultures include both explant cultures (i.e., cultures initiated from small pieces of intact tissue), as well as cultures initiated from preparations of individual or dispersed cells (obtained from intact tissue by mechanical or proteolytic disruption). Nerve fiber explant cultures in blood plasma were among the earliest types of tissue cultures (Harrison, 1907). Cells grow out from such tissue explants and form a single layer of cells completely filling the tissue culture vessel surface. Such cell cultures are called confluent monolayers. Confluent monolayers can then be treated with trypsin, so as to remove the individual cells from the culture vessel surface. The resulting cell suspension is then transferred into other culture containers, so that more viable monolayer

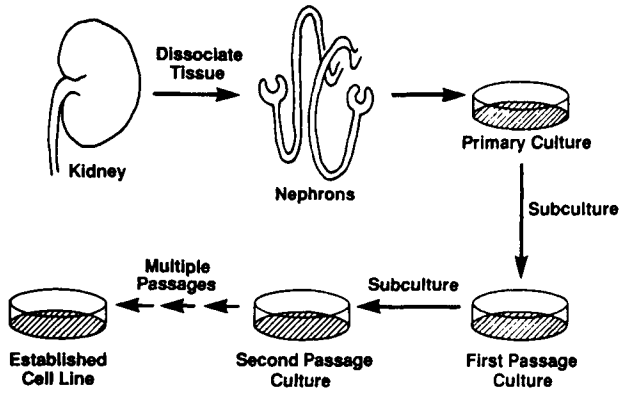


Figure 2. Initiation of primary cultures and their establishment as cell lines. Primary cultures are initiated from tissue removed from animals. Following multiple subculturings, cell lines in some cases are established.

cultures are obtained. This procedure, trypsinization, developed by Rous and Jones in 1916, is routinely utilized to this day.

Following the initial subculturing of primary cultures, first passage cultures are obtained. Second passage cultures are obtained following the second subculturing of the primary cells. Most normal animal cells that are removed from the animal and are put into culture are only "short-term cultures," that is, the cultures eventually degenerate (i.e., undergo a crisis of senescence), and die following a number of population doublings (Figure 3).

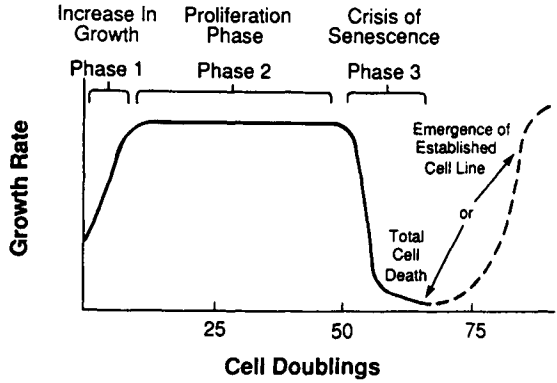


Figure 3. Growth phases of short-term cultures. Upon initiation of short-term cultures, growth rapidly increases (Phase 1). A stable growth rate is maintained for a defined number of passages (Phase 2). Finally, most cells undergo a crisis of senescence and ultimately die (Phase 3). In some cases immortal cell lines are established.

The events which occur prior to the death of short-term cultures have been referred to as the process of cellular aging, or apoptosis. Whether *in vitro* cellular aging is due to an inherited genetic program, or whether *in vitro* aging is simply a consequence of an imperfect cell culture environment is a matter of controversy. The possibility that the cellular aging which occurs *in vitro* resembles cellular aging *in vivo* is unresolved.

Occasionally, some of the animal cells in short-term cultures do not die, but instead survive indefinitely. These types of animal cell cultures, which can divide indefinitely, are called established cell lines. Established animal cell lines have been obtained from both normal and tumorigenic cells. Immortalized animal cell lines have also been successfully obtained from short-term cultures following their transformation with appropriate oncogenes.

DIFFERENTIATED ANIMAL CELLS IN CULTURE

The animal cell lines which were initially established in culture were primarily derived from malignant, or carcinogen-treated tissues. Included among these cell lines are mouse L cells (derived from muscle tissue treated with carcinogen), and HeLa (derived from a human cervical carcinoma) (Gey et al., 1952). Most of the initial established animal cell lines were fibroblasts (including mouse L cells), and did not express differentiated functions present in the tissue of origin. The lack of success in developing differentiated cell lines was attributed by many to a process of "dedifferentiation" which was presumably occurring in culture. The end result of this process of dedifferentiation was thought to be the fibroblast phenotype.

This hypothesis was essentially disproven by Dr. Gordon Sato, whose laboratory established a number of differentiated cell lines in culture from transplantable tumors. The tumor cells gradually adapted to the culture situation, as the cells underwent serial passages from the *in vitro* cell culture environment, to the animal, and then back again to the *in vitro* cell culture environment. Included among the cell lines established by this method are the B104 rat neuroblastoma, C6 rat glioma, and Y1 mouse adrenocortical cell lines. B104 rat neuroblastoma cells retain a number of differentiated neural functions in culture, as indicated by the very striking outgrowth of neurites from monolayer cultures (Figure 4). The C6 cell line retains differentiated functions typical of glial cells, including the S-100 protein. GH_3 rat pituitary cells synthesize growth hormone *in vitro*, as is observed with pituitary cells *in vivo*. Similarly, Y1 mouse adrenocortical cells exhibit the capacity for steroidogenesis, a process which is stimulated by ACTH and cyclic AMP (Schimmer, 1980). The morphology of Y1 cells is also affected by ACTH and cyclic AMP (Figure 5).

Embryonal carcinoma cell lines have also been established in culture from transplantable teratocarcinomas. The embryonal carcinoma (EC) stem cells present in the tumors actually differentiated, and give rise to a variety of different types of embryonic and adult cells. When these EC cells are cultured, they still possess the

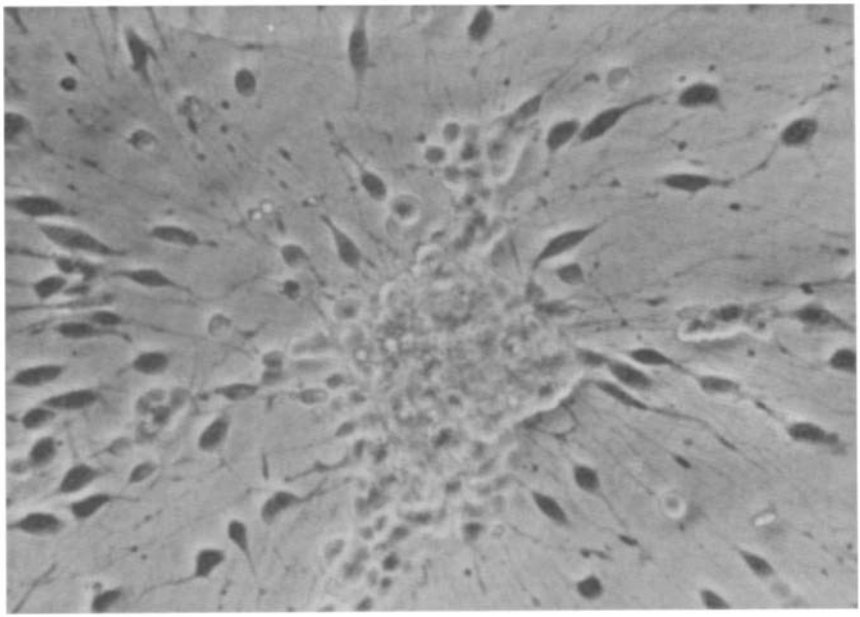


Figure 4. Neurite outgrowth by LA-N-1 human neuroblastoma cells in culture. LA-N-1 human PNS neuroblastoma cells were grown for five days in N2 medium (as described by Bottenstein and Sato, 1979) on a polylysine and fibronectin-modified surface. The cells were plated in clumps, rather than as a single cell suspension, which enhances neurite extension. Very long processes result, and exhibit varicosities along their length. Most of the cells have migrated from the central clump. (Photo courtesy of Dr. Jane Bottenstein.)

capacity to differentiate in response to particular inducers. F9 is a mouse EC cell line that differentiates into either parietal endoderm (following treatment with retinoic acid (vitamin A) and dibutyryl cyclic AMP), or into visceral endoderm (during aggregation in the presence of retinoic acid alone). Another EC cell line, 1003, remains undifferentiated when grown in serum-containing medium, but differentiates primarily into neurons when maintained in the absence of serum.

Included among other differentiating cell lines which have been established in culture, is the human promyelocytic cell line HL-60, which differentiates into more mature myeloid cells upon treatment with retinoic acid and prostaglandin E_1 (PGE_1). Friend erythroleukemia cells differentiate into hemoglobin-producing cells when treated with either dimethyl sulfoxide, or hexamethylene bisacetamide.

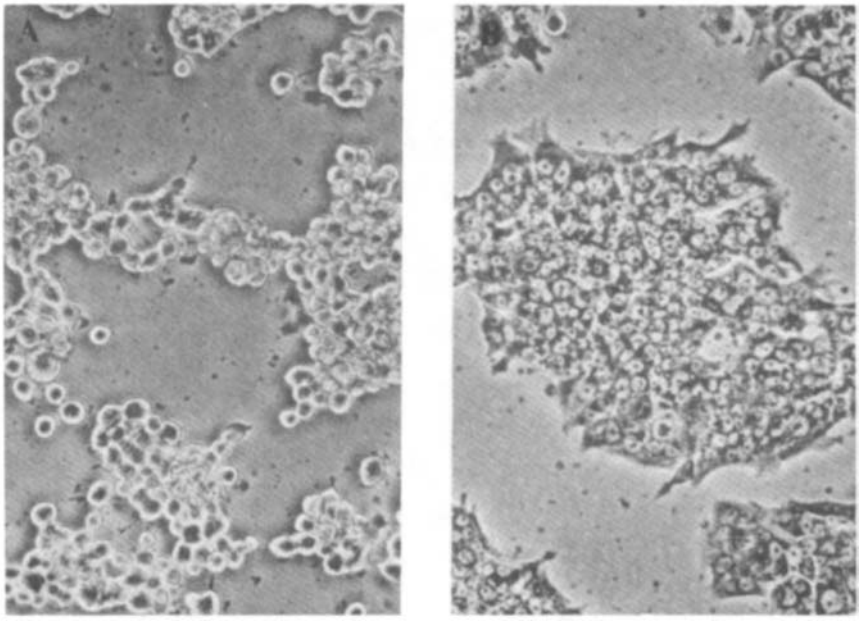


Figure 5. Y1 adrenocortical cells in culture. Y1 mouse adrenocortical tumor cells (**a**) in the absence and (**b**) in the presence of 1 mM 8-bromocyclic AMP overnight. (Photos courtesy of Margaret Wong and Dr. Bernard Schimmer.)

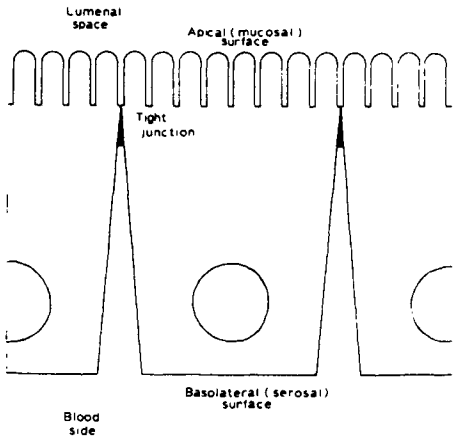


Figure 6. Polarized epithelial cells in culture. Epithelial cells in culture possess an apical surface with microvilli that faces the tissue culture medium (equivalent to the lumenal side of the cells *in vivo*), and a basolateral surface that faces the tissue culture dish (equivalent to the blood side of the cells *in vivo*).

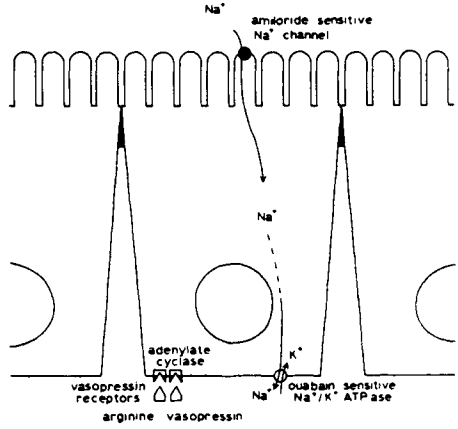


Figure 7. Transport of salt and water across kidney tubule epithelial cells.

In the 1950s only a few established cell lines were developed from normal differentiated cells. One such cell line, Madin Darby Canine Kidney (MDCK), was established from the kidney of a normal cocker spaniel in 1958. MDCK cells retain the structural and functional polarity typical of transporting epithelial cells. Their apical surface faces the culture medium, and their basolateral surface faces the tissue culture dish. The cells are interconnected by tight junctions (Figure 6). In confluent monolayers, ion transport systems such as the Na^+,K^+-ATPase assume a polarized orientation in the plasma membrane. Sodium and water are transported across the monolayer from the apical to the basolateral surface (Figure 7), and become entrapped between the monolayer and the dish. The resultant hydrostatic pressure causes groups of adjacent cells to become physically elevated from the dish (Figure 8). These groups of cells, called domes or hemicysts, can be observed microscopically (Figure 9).

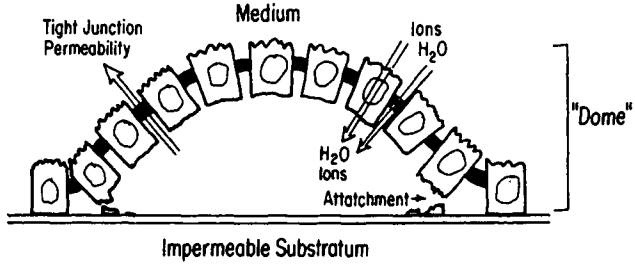


Figure 8. Diagrammatic schema for dome formation.

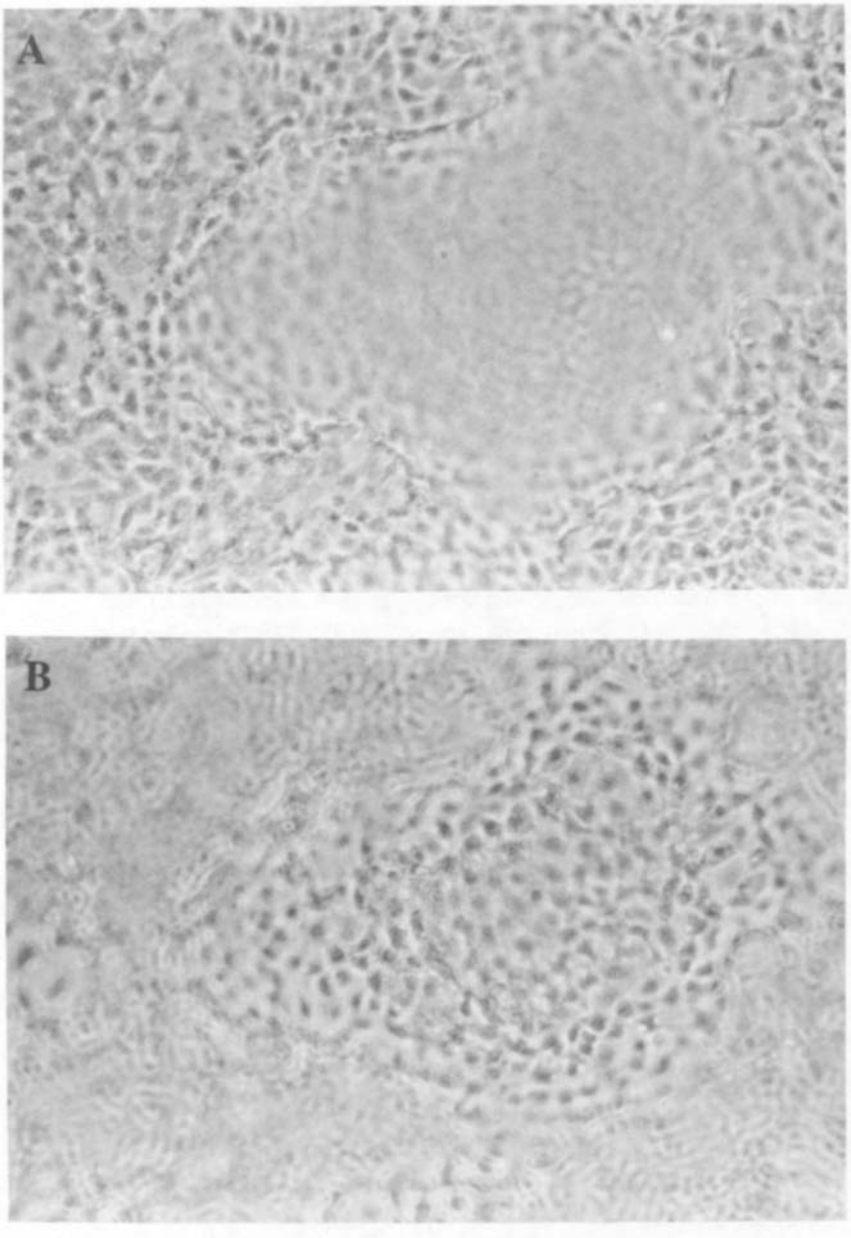


Figure 9. Domes in MDCK monolayers. Domes are the groups of cells that are slightly out of focus under the microscope when focusing on the monolayer (**a**). The cells in the monolayer are out of focus when focusing on the cells in the domes (**b**). 100 × magnification.

TISSUE CULTURE MEDIUM

Historically, the development of animal cell culture systems has been dependent upon the development of new types of tissue culture media. Mouse L cells and HeLa cells were developed using a balanced salt solution supplemented with blood plasma, an embryonic tissue extract, and/or serum. In 1955 Eagle developed a nutritionally defined medium, containing all of the essential amino acids, vitamins, cofactors, carbohydrates, salts, and small amounts of dialyzed serum (Table 1). He demonstrated that this minimal essential medium (MEM) supported the long-term growth of mouse L and HeLa cells. Eagle's MEM was so well defined that the omission of a single essential nutrient eventually resulted in the death of these animal cells in culture.

Eagle's MEM with serum rapidly became a standard growth medium for culturing animal cells *in vitro*. A number of variations of this medium were developed, including Dulbecco and Vogt's modified Eagle's essential medium (DMEM) (Table 2). DMEM contains nonessential as well as essential amino acids. The essential amino acids and vitamins are at concentrations which are significantly elevated as compared to MEM.

Eagle's MEM was constructed assuming that serum is an essential requirement for the growth of animal cells *in vitro*. We now know that the serum requirement for the growth of animal cells *in vitro* can be replaced by a number of specific defined components, which are present in the serum (Sato, 1975). Included among the types of components which are utilized in lieu of serum are hormones, growth factors, cell attachment factors, nutrients, and transport binding proteins (Table 3). Growth factors utilized include epidermal growth factor (EGF), and hormones utilized include steroid hormones (such as hydrocortisone), and peptide hormones

Table 1. Minimum Essential Medium

Compound	*Concentration, mM*	*Compound*	*Concentration, mM*
L-Amino acids		Salts	
Arginine	0.6	NaCl	116.0
Cystine	0.1	KCl	5.4
Glutamine	2.0	$CaCl_2$	1.8 (0)
Histidine	0.2	$MgCl_2 \cdot 6H_2O$	1.0
Isoleucine	0.4	$NaH_2PO_4 \cdot 2H_2O$	1.1 (11)
Leucine	0.4	$NaHCO_3$	23.8
Lysine	0.4	Vitamins	
Methionine	0.1	Choline	0.006
Phenylalanine	0.2	Folic acid	0.0025
Threonine	0.4	Inositol	0.011
Tryptophan	0.05	Nicotinamide	0.00825
Tyrosine	0.2	Pantothenate	0.002
Valine	0.4	Pyridoxal	0.005
Carbohydrate		Riboflavin	0.00025
Glucose	5.5	Thiamine	0.003

Table 2. Dulbecco's Modified Eagle Medium

Compound	*Concentration, mM*	*Compound*	*Concentration, mM*
Chemicals		L-Isoleucine	0.8
Sodium chloride	109.4	L-Leucine	0.8
Potassium chloride	5.4	L-Lysine·HCl	0.8
Calcium chloride	1.8	L-Methionine	0.2
Magnesium sulfate	0.8	L-Phenylalanine	0.4
Sodium phosphate monobasic	0.9	L-Serine	0.4
Dextrose	5.5	L-Threonine	0.8
Inositol	0.04	L-Tryptophan	0.08
Ferric nitrate	0.00025	L-Tyrosine	0.4
Sodium bicarbonate	44.0	L-Valine	0.8
Sodium pyruvate	1.0	Vitamins	
Phenol red	0.04	Choline chloride	0.024
Amino acids		Folic acid	0.009
L-Arginine·HCl	0.4	Nicotinamide	0.033
L-Cystine	0.2	Calcium pantothenate	0.008
L-Glutamine	4.0	Pyrodoxal·HCl	0.02
Glycine	0.4	Riboflavin	0.001
L-Histidine·HCl·H_2	0.2	Thiamine·HCl	0.012

(such as parathyroid hormone). Most of these factors are added at physiologic concentrations. A common exception is insulin, which is often added at superphysiologic concentrations. Presumably, insulin is required at higher dosages because the insulin is acting through receptors for insulin-like growth factors (IGFs) rather than for insulin. Insulin has a considerably lower affinity for IGF receptors than for insulin receptors. A large number of animal cells in culture require insulin as well as transferrin in order to grow serum free. Transferrin is an iron-binding protein in the blood which facilitates the entry of iron into cells.

Differentiated cells derived from different tissues generally require different groups of these growth supplements in order to grow completely serum free. GH_3 mouse pituitary cells require insulin, somadomedin-C, transferrin, thyroid releasing hormone (TRH), parathyroid hormone, selenium, and cadmium for long-term growth in Ham's F12 medium lacking serum (Table 4) (Hayashi and Sato, 1976). In contrast, HeLa cells require insulin, transferrin, EGF, fibroblast growth factor

Table 3. Serum Free Medium Supplements

Hormones
Growth factors
Cell attachment factors
Transport binding proteins
Nutrients

Table 4. Growth Supplements for Two Cell Lines in Serum-Free Medium

Human cervical carcinoma cells	
Insulin	5 μg/ml
Transferrin	5 μg/ml
Hydrocortisone	10^{-8}M
Fibroblast growth factor (FGF)	100 ng/ml
Epidermal growth factor (EGF)	5 ng/ml
GH_3 rat pituitary cells	
Insulin	5 μg/ml
Transferrin	5 μg/ml
Triiodothyronine (T_3)	3×10^{-11}M
Thyrotropin releasing hormone (TRH)	1 ng/ml
Parathyroid hormone (PTH)	0.5 ng/ml
Fibroblast growth factor (FGF)	1 ng/ml
Somatomedin C	1 ng/ml

(FGF), and hydrocortisone to grow serum free (see Table 4) (Hutchings and Sato, 1978). Both of these cell types can grow serum free with appropriate supplements at the same rate as that obtained in serum-supplemented medium. Both cell types can also survive over extended culture periods in hormonally defined serum-free medium.

F9 embryonal carcinoma cells have a simple set of growth supplements which are required for growth in serum-free medium: insulin, transferrin, and fibronectin (Rizzino and Sato, 1978). Fibronectin is a component of the extracellular matrix and facilitates the attachment of the cells to the culture dish. In addition, high density lipoprotein (HDL) has been observed to promote the growth of F9 cells serum-free.

Vitamins and lipids are often required for animal cells to grow in serum-free medium. Phosphoethanolamine and ethanolamine are key additives that facilitate the growth of the mammary tumor cell line 64024 (Kano-Sueoka and Errick, 1981). In addition, ethanolamine promotes the growth of human lymphocytes and mouse hybridoma cells. Short-term cultures of human diploid lung and foreskin fibroblasts grow in medium that includes among its supplements soybean lecithin, cholesterol, sphingomyelin, and vitamin E.

The mechanisms by which the growth supplements in serum-free medium act are still not understood. In order to achieve an understanding of the biochemical basis for the hormonal and growth factor requirements of animal cells, the basic mechanism of action of hormones and growth factors must be determined. The biochemical basis for the nutritional requirements of animal cells can only be determined when we have an understanding of the metabolism of the different types of animal cells.

The physiological significance of the growth requirements for established animal cell lines in serum-free medium is still an unresolved matter. Cultures of

differentiated cells derived from a number of tissues (e.g., neural, renal, and ovarian) and established animal cell lines derived from the same tissue often possess the same growth requirements. For example, the B104 rat neuroblastoma cell line grows in serum-free medium supplemented with insulin, transferrin, progesterone, putrescine, selenium, and fibronectin. Progesterone and putrescine are the unique growth supplements for this cell type. Remarkably, these medium supplements also support the growth of primary cultures of neurons, as well as established mouse and human neuroblastoma cell lines serum free.

A major problem in the culturing of normal primary cells in culture is fibroblast overgrowth. Although primary cell cultures may initially consist of differentiated cells almost exclusively, a minor subpopulation of fibroblasts may eventually become the predominant cell type in the cultures. Such fibroblast overgrowth occurs when the cell culture conditions favor the growth of fibroblasts, rather than the differentiated cells. Included among the culture conditions that affect fibroblast overgrowth are the serum, as well as the basal culture medium (e.g., Eagle's MEM).

A good example of the problems encountered with the use of serum with primary cultures is illustrated by the case with cultured kidney cells. The kidney epithelial cell line MDCK grows in serum-free medium supplemented with five supplements:

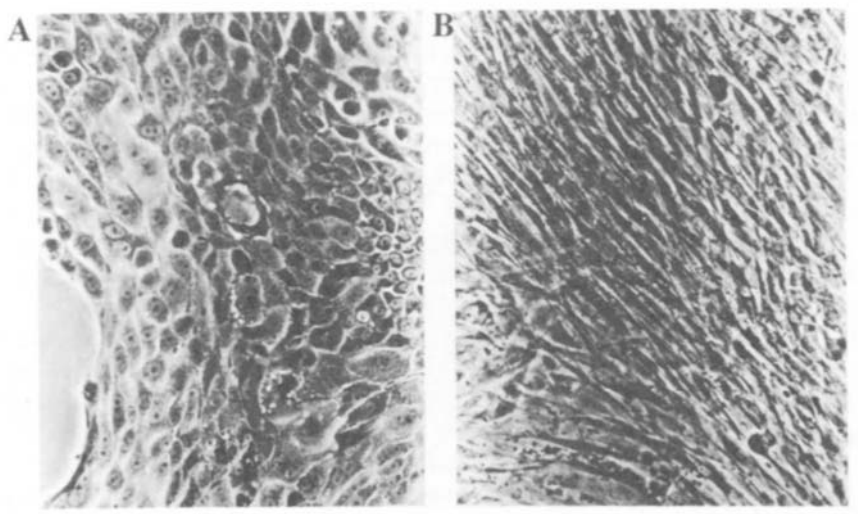


Figure 10. Primary cultures of mouse kidney cells. Primary cultures of kidney epithelial cells derived from 10-day-old mice were grown either in hormonally defined medium with five supplements (5 μg/ml insulin, 5 μg/ml transferrin, 25 ng/ml PGE_1, 5×10^{-8}M hydrocortisone, and 5×10^{-12} M T_3), or in medium supplemented with 10% fetal calf serum. After 10 days, primary cultures still were epithelial in morphology serum free (**a**) but were overgrown with fibroblasts with serum (**b**). (Taub et al., 1979; with permission.)

Table 5. Nutrient Mixture F-12

Compound	*Concentration, mM*	*Compound*	*Concentration, mM*
L-Arginine·HCl	1.0	Phenol red	3.3×10^{-3}
Choline chloride	0.1	Glucose	10.0
L-Histidine·HCl·H_2O	0.1	L-Glutamine	1.0
L-Isoleucine	0.03	Riboflavin	1.0×10^{-4}
L-Leucine	0.1	Sodium pyruvate	1.0
L-Lysine·HCl	0.2	L-Cysteine·HCl	0.2
L-Methionine	0.03	L-Asparagine	0.1
L-Phenylalanine	0.03	L-Proline	0.3
L-Serine	0.1	Putrescine·2HCl	1.0×10^{-3}
L-Threonine	0.1	Vitamin B_{12}	1.0×10^{-3}
L-Tryptophan	0.01	L-Alanine	0.1
L-Tyrosine	0.03	L-Aspartic acid	0.1
L-Valine	0.10	L-Glutamic acid	0.1
Biotin	3.0×10^{-5}	Glycine	0.1
Calcium pantothenate	1.0×10^{-3}	Hypoxanthine	0.03
Niacinamide	3.0×10^{-4}	myo-Inositol	0.10
Pyridoxine·HCl	3.0×10^{-4}	Lipoic acid	1.0×10^{-3}
Thiamin·HCl	1.0×10^{-3}	Thymidine	3.0×10^{-3}
KCl	3.0	$CuSO_4$	1.0×10^{-5}
Folic acid	3.0×10^{-3}	$ZnSO_4$	3.0×10^{-3}
Na_2HPO_4	1.0	NaCl	130
$FeSo_4$	3.0×10^{-3}	$NaHCO_3$	14
$MgCl_2$	0.6	Linoleic acid	3.0×10^{-4}
$CaCl_2$	0.3		

insulin, transferrin, prostaglandin E_1 triiodothyronine, and hydrocortisone. This same set of supplements permits primary cultures of baby mouse kidney epithelial cells to grow to confluence without fibroblast overgrowth. In contrast fibroblast overgrowth is obtained when serum supplemented medium is used (Figure 10).

The importance of the basal medium in culturing normal cells was demonstrated by Dr. Richard Ham (Ham and McKeehan, 1979). Dr. Ham's laboratory developed a number of different types of culture media, in which the concentrations of the individual nutritional components have been optimized to support the growth of specific types of normal, differentiated cells. The most widely used is nutrient mixture F12, which contains nonessential, as well as essential, amino acids, a number of lipids, and trace elements (Table 5).

GROWTH PROPERTIES OF ANIMAL CELLS IN CULTURE

Most normal animal cells in culture grow at a logarithmic rate until they achieve a particular density (i.e., their saturation density) (Figure 11). A typical growth rate for animal cells in culture is one doubling per day, and a typical saturation density

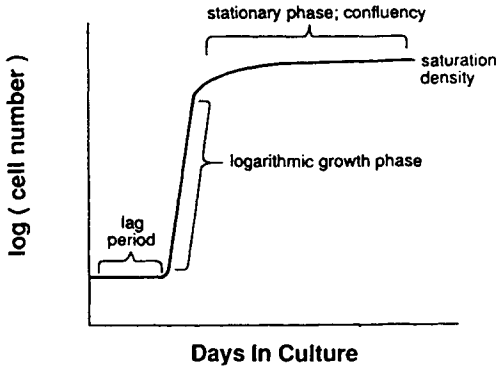


Figure 11. Growth characteristics of animal cells in culture.

is 5×10^4 cells/cm^2. The cells growth rate and saturation density depend upon the inherent properties of the cells themselves, as well as upon their environment. When most normal animal cells which attach to the bottom of culture dishes reach their saturation density, they form a confluent monolayer of cells which completely covers the culture dish. However the saturation density of the cells can be altered by the culture conditions, as well as by changes in the genotype of the cells themselves.

When animal cells are plated into sterile plastic tissue culture dishes at a low density (e.g., 10 cells/cm^2), clones of cells, or colonies, are formed. A clone is a group of adjacent cells that are tightly packed together. All of the cells in such a clone generally originate from a single cell which attached to the culture dish in that locale during the plating procedure, and subsequently divided. Only a fraction of the animal cells that are plated into culture dishes at a low density attach and divide to form such clones. This fraction (or percentage) is called the efficiency of plating. Animal cells have characteristic plating efficiencies, depending upon the inherent properties of the cells themselves, as well as upon their environment.

The ability of animal cells to form clones at a low plating density is an important property. The clones of cells can be isolated using cloning cylinders, and homogeneous populations of cells can be grown for subsequent studies. Under appropriate environmental conditions, particular types of cells (either normal, mutant, or genetically transformed) can be selectively grown out from the rest of the cell population. Both positive selection procedures (which stimulate the growth of the cell type of interest), and negative selection procedures (which selectively kill all cells but the cell type of interest) are utilized. An example of a positive selection procedure is the use of glucose-free medium for the selective growth of gluconeogenic cells. Cells which cannot make their own glucose eventually die in the absence of glucose, or another alternative source of sugar. An example of a negative selection procedure is the use of bromodeoxyuridine, for the selective growth of

those cells which lack thymidine kinase. Bromodeoxyuridine is an analog of thymidine, which is toxic following metabolism by thymidine kinase. Thymidine kinase-negative mutants survive killing by the drug.

The use of such selection procedures, followed by cloning, permit genetic studies to be feasible with established animal cell lines. Not only is the isolation of genetic mutants possible, but in the isolation of cells which are transformed with plasmids containing a selectable marker, as well as another gene of interest, is also possible. For example, mammalian cells transformed with a plasmid containing a G418 resistance gene and the c-myc oncogene (for example), will survive treatment with the drug G418. G418 is an aminoglycoside related to gentamycin, which is toxic to bacteria, yeast, and mammalian cells. G418 resistance genes encode a protein product (aminoglycoside phosphotransferase), which metabolizes G418, rendering the drug nontoxic.

TRANSFORMATION OF ANIMAL CELLS

Some established cell lines were derived from malignant tissue. Many of these cell lines can form tumors when injected into susceptible animals. Other established cell lines are not tumorigenic. However, exposure to carcinogens, and irradiation can cause these cells to form tumors in susceptible animals. In addition, transformation can be caused by spontaneous mutations, by growth factors, and by viral (or oncogenic) transformation (Table 6). Malignant transformation is defined as consisting of the series of events that cause normal cells to develop the capacity to form tumors.

A number of changes are observed in the properties of animal cells in culture following transformation, including an increased growth rate, the ability to form multilayers, and to grow to a higher saturation density (Table 7). The adhesion properties of animal cells are also often altered concomitantly with malignant transformation. Normal adherent animal cells must attach to an appropriate substratum if they are to grow. When normal adherent cells are placed in a semisolid medium such as soft agar, they will remain viable, but will not grow. However

Table 6. Some Tumor Viruses Which Cause Cellular Transformation

Virus	*Tumor*
Retroviruses	
Human T-cell leukemia virus (HTLV)	Human T-lymphocytic tumors
Harvey murine sarcoma (Ha-ras)	Bladder, mammary, skin carcinoma
Kirsten murine sarcoma (Ki-ras)	Lung, colon carcinoma
Hepatitis B virus	Human liver cancer
Papilloma virus	Human cervical carcinoma
Simian virus 40 (SV40)	Fibrosarcomas (in hamsters)
Epstein-Barr virus	Burkitt's lymphoma
Adenovirus	Sarcoma

Table 7. Cellular Alterations Following Transformation

Altered cell morphology
Increased growth rate
Increased saturation density
Formation of multilayers
Reduced adhesion to substratum
Colony formation in soft agar
Reduced serum requirement
Altered growth factor requirement
Tumor formation in athymic nude mice

many transformed cells have altered requirements for adherence, and when suspended as single cells in soft agar they grow, forming colonies.

The serum requirement of many transformed animal cells is also reduced. The reduction in the serum requirement can, in many cases, be explained by a loss of some of the cells' hormone, growth factor or adhesion requirements. The altered hormone, growth factor, or attachment requirements can be studied by growing the cells in serum-free medium.

Some, but not all, animal cells in culture become malignant following their transformation in culture. Malignancy is defined by the cells' ability to form tumors in appropriate animal hosts. A good general assay of the malignancy of many different types of animal cells is their ability to form tumors in nude mice. Nude mice lack a thymus gland, and thus lack the ability to reject foreign implants. In addition these animals lack hair (i.e., are nude). Many (but not all) malignantly transformed cells *in vitro* also exhibit other cellular changes that are associated with transformation in general, such as anchorage-independent growth, or the loss of a growth factor requirement. However, the observation of these other cellular alterations is not necessarily associated with malignancy.

If a loss of a growth factor requirement is observed in malignantly transformed cells, the loss of the growth factor requirement can be attributed to a number of different molecular changes in the cultured cells. In some cases, the loss of a growth factor requirement *in vitro* can be explained by the production of the growth factor by the transformed cells. Transformed cells can also lose a growth factor requirement as the consequence of an alteration in the signal transduction pathway activated by the growth factor.

The growth requirement for EGF is a good example in this regard. EGF stimulates the growth of many different types of animal cells in culture. In order to initiate the growth response, EGF interacts with specific EGF receptors localized in the plasma membrane, activating a tyrosine-specific protein kinase, which is an intrinsic part of the receptor (Figure 12). As a consequence, specific proteins are phosphorylated at tyrosine residues, and some of these proteins (which are also

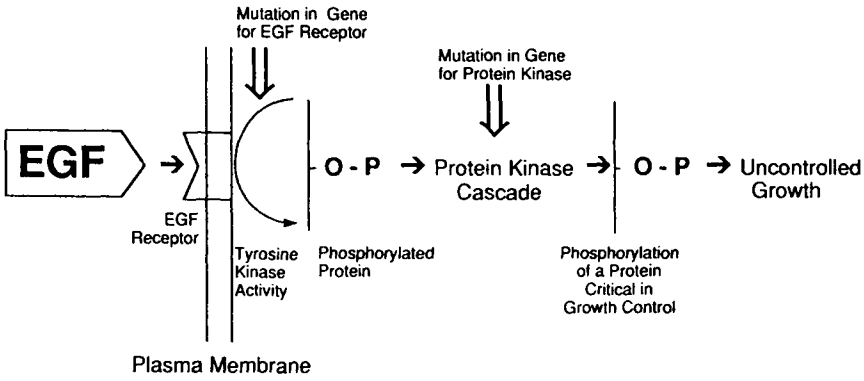


Figure 12. An alteration in a signal transduction pathway may be associated with malignant transformation.

protein kinases) in turn phosphorylate additional proteins. The ultimate result of this protein kinase cascade is that a protein which is critical in growth control is phosphorylated.

The loss of the EGF requirement for growth may result from a number of different types of genetic changes. First, a specific point mutation in the gene for the EGF receptor may cause the receptor to lose affinity for EGF. Such cells would be EGF nonresponsive, and would only achieve an equivalent response if an alternative signal transduction pathway is operative. Second, another point mutation in the gene for the EGF receptor may instead render the tyrosine kinase activity of the receptor permanently active, even in the absence of exogenous EGF (i.e., extracellular EGF). Such cells would also be EGF nonresponsive, but would grow as though EGF was nevertheless present. Finally, a deletion mutation may occur which has a similar effect: a tyrosine kinase which is permanently active. Indeed the v-erb B oncogene encodes a truncated EGF receptor that lacks the EGF binding site on the EGF receptor (Figure 13), but possesses a permanently activated tyrosine-specific protein kinase.

An oncogene encodes for an oncoprotein that can cause the transformation of normal cells. Included among the different types of oncoproteins are growth factors, growth factor receptors, protein kinases, guanine nucleotide binding proteins (G proteins), and nuclear proteins. V-erb B (described above) is an example of an oncogene which encodes for a protein kinase. The ras oncogene is an example of an oncogene which encodes for G proteins with GTPase activity. Ras G proteins are activated by the binding of GTP and inactivated by the binding of GDP (the breakdown product of GTP by the GTPase). These G proteins, when present in the plasma membrane, presumably affect normal signal transduction pathways. Included among the oncogenes which encode for nuclear proteins are myb, which is

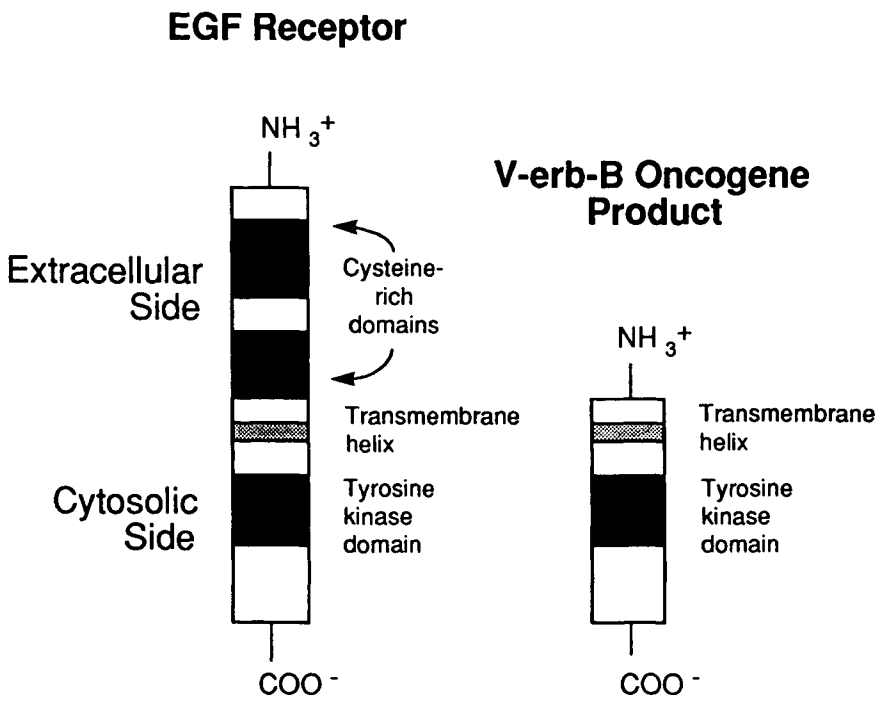


Figure 13. Structure of the EGF Receptor and the v-erb-B oncogene product.

expressed in leukemia, and myc, which is expressed in neuroblastoma. In addition to these previously defined classes of oncogene products are many as of yet undefined types of oncogenes and their protein products.

An oncogene product (or oncoprotein) can theoretically be any protein that is important in growth control. Whether only alterations in particular oncoproteins involved in growth regulation result in malignant transformation, or whether alterations in any protein involved in growth regulation can result in malignant transformation is still unknown.

However, most normal, untransformed cells do contain cellular proto-oncogenes, which encode information for protein products which can cause malignant transformation if activated by specific mutations. In addition, most normal cells are susceptible to transformation by viruses which contain proto-oncogenes within their genome. The insertion of the viral protooncogene into the genome may result in the activation of the oncogene, causing the cell to assume a transformed phenotype. In addition, the proto-oncogene may become active as the consequence of cellular (or viral) mutations which affect either the structure of the protein product itself, or the level of expression of the protein. The result of the activation of the oncogene in the above cases is cellular transformation.

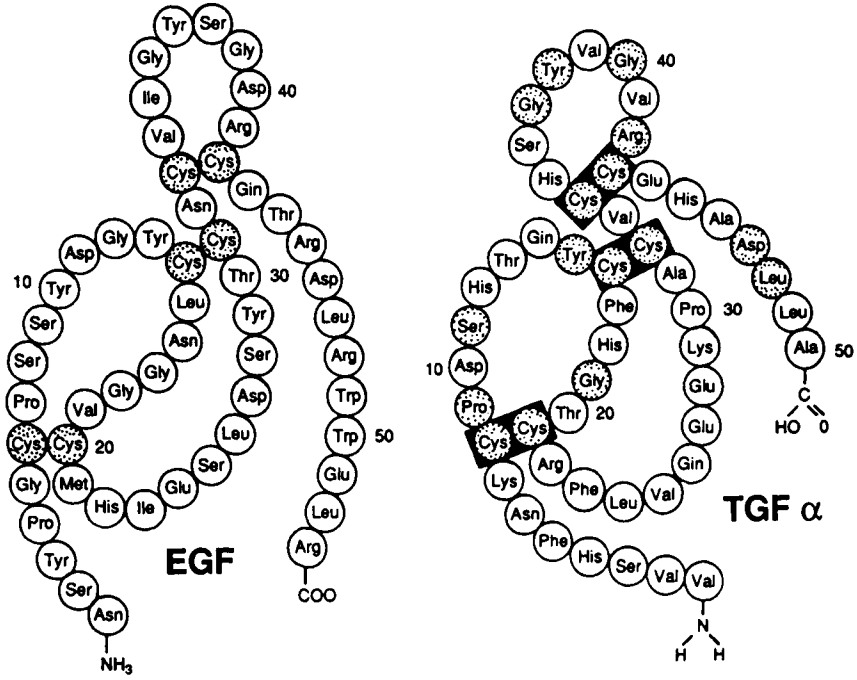


Figure 14. Amino acid sequence of EGF and TGF alpha. Shaded residues of TGF structure are the residues homologous to those in EGF.

As described above, particular growth factors produced by cells can be classified as oncogene products, which cause cellular transformation. For example, many animal cells produce transforming growth factors (TGFs). Two major types of transforming growth factors, TGF alpha and TGF beta, have been found. TGF alpha is structurally very similar to EGF (Figure 14). TGF alpha binds to EGF receptors so as to initiate its biological effects. TGF beta binds to a separate class of receptors on the plasma membrane.

TGFs are secreted by animal cells following their biosynthesis. Subsequently, these TGFs may either stimulate the growth of the very cells that have produced them (in this case, the TGFs act as autocrine factors), or may stimulate the growth of other adjacent cell types (in this case, the TGFs act as paracrine factors).

The normal rat kidney (NRK) cell line is an untransformed cell line established from rat kidney cells. However, when NRK cells are treated with both TGF alpha and TGF beta, the cells assume a transformed phenotype. In addition to exhibiting an altered morphology in monolayer culture, NRK cells possess the capacity for anchorage-independent growth while being treated with both TGF alpha and TGF

beta. Anchorage-independent growth is assayed by the ability of NRK cells to form colonies in soft agar. When TGF beta and TGF alpha are removed from the culture medium, NRK cells again assume the characteristics of normal cells. Thus these effects of TGF alpha and TGF beta are not heritable.

SUMMARY

Animal cells in the intact animal are in a complex, dynamic environment. Studies concerning the regulation of the growth and the expression of the differentiated functions of animal cells are best conducted using *in vitro* primary cell culture systems (initiated directly from the animal), as well as established animal cell lines (which can grow indefinitely). A number of differentiated animal cell lines derived from different tissues are available for study, including neuroblastoma, glioma, adrenocortical, kidney tubule epithelial, and embryonal carcinoma. The development of tissue culture medium has been a limiting factor in the development of new cell culture systems as well as in studies of the control of cell growth and function. Hormonally defined serum-free media have alleviated these limitations considerably. Both normal and malignantly transformed animal cells can be studied *in vitro*. Transformed cells exhibit such altered *in vitro* properties as elevated growth rate, increased saturation density, loss of growth factor requirements, and the ability to form colonies in soft agar. The molecular mechanisms responsible for cellular transformation are poorly understood. However, during malignant transformation, alterations very likely occur in signal transduction pathways such as those initiated by EGF. Such alterations may originate from the activation of cellular or viral oncogenes, mutations, or externally applied factors, such as TGF alpha and TGF beta. *In vitro* cells in culture have been observed to have close similarities to animal cells *in vivo*. Thus *in vitro* studies concerning such growth factors may have physiological significance. When using animal cell cultures, it is critical to constantly refer back to the *in vivo* situation. Similarly, the results of *in vivo* studies should be further examined *in vitro*.

ACKNOWLEDGMENTS

The author thanks Drs. Jane Bottenstein and Bernard Schimmer for photographs of differentiated cells, and The Educational Technology Department of The State University of New York at Buffalo for illustrations. This work was supported in part by The National Institutes of Health Grant 9 R01 DK4028607.

REFERENCES

Bottenstein, J.E. & Sato, G.H. (1979). Growth of a rat neuroblastoma cell line in serum-free supplemented medium. Proc. Natl. Acad. Sci. USA 76, 514–517.

Eagle, H. (1955). The specific amino acid requirements of a mammalian cell (strain L) in tissue culture. J. Bio. Chem. 214, 839–885.

Gey, G.O., Coffman, W.D., & Kubiek, M.T. (1952). Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. Cancer Res. 12, 264–265.

Ham, R.G. & McKeehan, W.L. (1979). Media and growth requirements. Methods in Enzymol. 58, 44–93.

Harrison, R.G. (1907). Observations on the living developing nerve fiber. Proc. Soc. Exptl. Biol. Med. 4, 140–143.

Hayashi, I. & Sato, G. (1976). Replacement of serum by hormones permits growth of cells in a defined medium. Nature 259, 132–134.

Hutchings, S.E. & Sato, G.H. (1978). Growth and maintenance of HeLa cells in serum-free medium supplemented with hormones. Proc. Natl. Acad. Sci. USA 75, 901–904.

Kano-Sueoka, T. & Errick, J.E. (1981). Effects of phosphoethanolamine and ethanolamine on growth of mammary carcinoma cells in culture. Exp. Cell Res. 136, 137–145.

Martin, G.R. & Timpl, R. (1987). Laminin and other basement membrane components. Ann. Rev. Cell. Biol. 3, 57–85.

Rizzino, A.A. & Sato, G. (1978). Growth of embryonal carcinoma cells in serum free medium. Proc. Natl. Acad. Sci. USA 75, 1844–1848.

Rous, R. & Jones, F.S. (1916). A method for obtaining suspensions of living cells from the fixed tissues and for the plating out of individual cells. J. Exp. Med. 23, 549–555.

Sato, G.H. (1975). The role of serum in cell culture. In: Biochemical Actions of Hormones, Vol III. (Litwack, G., ed.), pp. 391–396, Academic Press, New York.

Schimmer, B. (1980). Cyclic nucleotides in adrenocortical function. Adv. Cyclic Nucleotide Res. 13, 181–214.

Taub, M. & Sato, G. (1980). Growth of functional primary cultures of kidney epithelial cells in defined medium. J. Cell. Physiol. 105, 369–378.

Taub, M., Chuman, L., Saier, M.H., & Sato, G. (1979). Growth of a kidney epithelial cell line (MDCK) in hormonally defined serum-free medium. Proc. Natl. Acad. Sci. USA 76, 3338–3342.

RECOMMENDED READINGS

Barbicid, M. (1987). Ras genes. Ann. Rev. Biochem. 56, 779–827.

Carpenter, G. (1987). Receptors for epidermal growth factor and other polypeptide mitogens. Ann. Rev. Biochem. 56, 881–914.

Derynck, R. (1988). Transforming growth factor alpha. Cell 54, 593–595.

Green, M.R. (1989). When the products of oncogenes and anti-oncogenes meet. Cell 56, 1–3.

Pollack, R. (1981). Readings in Mammalian Cell Culture. Cold Spring Harbor Laboratories. Cold Spring Harbor, New York.

Sato, G.H. (1980). Differentiated Animal Cell Lines. Alan R. Liss, New York.

Sporn, M.B., Roberts, A.B., Wakefield, L.M., & de Crombrugghe, B. (1987). Some recent advances in the chemistry and biology of transforming growth factor beta. J. Cell Biol. 105, 1039–1045.

Tooze, J. (1973). The Molecular Biology of Tumor Viruses. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

This Page Intentionally Left Blank

INDEX