The Plasma Proteins

STRUCTURE, FUNCTION, AND GENETIC CONTROL

Second Edition / Volume V

Edited by

FRANK W. PUTNAM

Department of Biology
Indiana University
Bloomington, Indiana

1987

ACADEMIC PRESS, INC.

Harcourt Brace Jovanovich, Publishers
Orlando San Diego New York Austin
Boston London Sydney Tokyo Toronto

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
24–28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data
(Revised for vol. 5)

Putnam, Frank W., ed.
The plasma proteins.

Includes bibliographies and indexes.
1. Blood proteins—Collected works. [DNLM: 1. Blood proteins. QY 455 P715] I. Title.
QP99.3.P7P87 1975 612'.12 75-3970
ISBN 0–12–568405–3 (v. 5 : alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

87 88 89 90 9 8 7 6 5 4 3 2 1

To Dorothy

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

Winona C. Barker (365), National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C. 20007

Barbara H. Bowman (1), Department of Cellular and Structural Biology, The University of Texas Health Science Center at San Antonio, San Antonio, Texas 78284

Frank W. Putnam (49), Department of Biology, Indiana University, Bloomington, Indiana 47405

Angelo M. Scanu (141), Departments of Medicine, Biochemistry, and Molecular Biology and the Lipoprotein Study Unit, University of Chicago, Pritzker School of Medicine, Chicago, Illinois 60637

Gerhard Schreiber (293), The Russell Grimwade School of Biochemistry, University of Melbourne, Parkville, Victoria 3052, Australia

Lars Sottrup-Jensen (191), Department of Molecular Biology, University of Aarhus, DK-8000 Aarhus C, Denmark

Funmei Yang (1), Department of Cellular and Structural Biology, The University of Texas Health Science Center at San Antonio, San Antonio, Texas 78284

Preface

In the short time since Volume IV appeared, the major focus on plasma proteins has turned from protein structure to gene structure and genomic organization. The approach of molecular biology and gene cloning, which was virtually unknown a decade ago, has come to dominate the thinking and experimentation of workers in the plasma protein field. The variety, versatility, and potential therapeutic value of plasma proteins have lured to this area many molecular biologists who hitherto had ignored it. The early results of their research are reflected in this volume. The advances already made portend an auspicious future.

In this volume the first chapter, by Bowman and Yang on the DNA sequences and chromosomal location of plasma protein genes, gives the current perspective and sets the background for the other chapters. The gene sequence of about 50 plasma proteins is already known as is also the chromosomal location of nearly as many. Chromosomal mapping and DNA sequencing are revealing surprising evolutionary relationships and are giving hints of unsuspected mechanisms of genetic control. DNA polymorphisms, both benign and malignant, are being detected by hybridization with plasma protein DNA probes. These are already proving useful in the detection of defective genes causing inherited diseases such as hemophilia. This theme is reiterated in subsequent chapters.

Certain families of plasma proteins have been at the forefront of research for more than a decade, in particular, the immunoglobulins and the lipoproteins. As reviewed in Chapter 2, the cloning and sequencing of immunoglobulin genes have elucidated the intricate relationship of immunoglobulin structure and antibody specificity to genomic organization. Yet, although the antibody dilemma has largely been solved, much remains to be clarified regarding the mechanism of biological effector functions. Similar progress has been made in the complex field of lipoproteins, and for the first time the structure of their tissue receptors is being elucidated. In Chapter 3 Scanu discusses these recent advances in relation to newly discovered genetic variants present either as normal phenotypes or in dyslipoproteinemic states associated with cardiovascular disease. The recent application of the techniques of molecular biology has already enabled elucidation of the gene structure and of the translational and maturation steps of every major plasma apolipoprotein.

In concert with the advances achieved through the new techniques of molecular biology, the classical approaches of protein chemistry continue to be productive even as they are being applied to larger and much more complex plasma proteins such as α_2-macroglobulin and the related thiol ester proteins of the complement family. In the first comprehensive review integrating this set of plasma proteins, Sottrup-Jensen describes the rapid advances since the discovery of the internal thiol ester bond. He emphasizes the unique features of protein structure and the conformational changes that characterize the assembly of macromolecular complexes in plasma and on cell surfaces. Such advances in the understanding of the structure and molecular genetics of plasma proteins should illuminate rather than overshadow their vital importance in physiological and pathological processes. In Chapter 5 Schreiber integrates the molecular biology, physiology, and pathology of plasma proteins and their response in inflammation. This is the first comprehensive review of its kind, and has special import for clinicians. As the other authors, Schreiber stresses the rapidity with which protein and DNA sequence information is accumulating. The Appendix updates the listing of plasma protein sequences and includes valuable data on the proteins listed.

As editor, I owe thanks to many for advice, encouragement, and help: to the contributors, who cooperated fully in every way; to my colleague Nobuhiro Takahashi, who read and advised on the manuscripts; to my efficient and understanding secretary, Ann Scales, who organized the literature database and acted as my editorial assistant; and, most of all, to my wife, Dorothy, for many years of encouragement and patience.

Frank W. Putnam

Contents of Other Volumes

1 DNA Sequencing and Chromosomal Locations of Human Plasma Protein Genes

Barbara H. Bowman and Funmei Yang

Department of Cellular and Structural Biology
The University of Texas Health Science Center at San Antonio
San Antonio, Texas 78284

I. Introduction

Recombinant DNA technology has dramatically facilitated the characterization and chromosomal mapping of genomes controlling human plasma proteins and has contributed to a deeper understanding of their evolutionary and functional relationships. At this time, plasma protein genes* have been mapped on 17 of the 22 human autosomes, as well as on the X chromosome. The distribution of the plasma protein genes in the human genome is shown in Fig. 1. The genes for

*In this chapter genes, alleles, and loci have been set in italic type to avoid confusion with abbreviated protein forms.

The Plasma Proteins, Vol. V

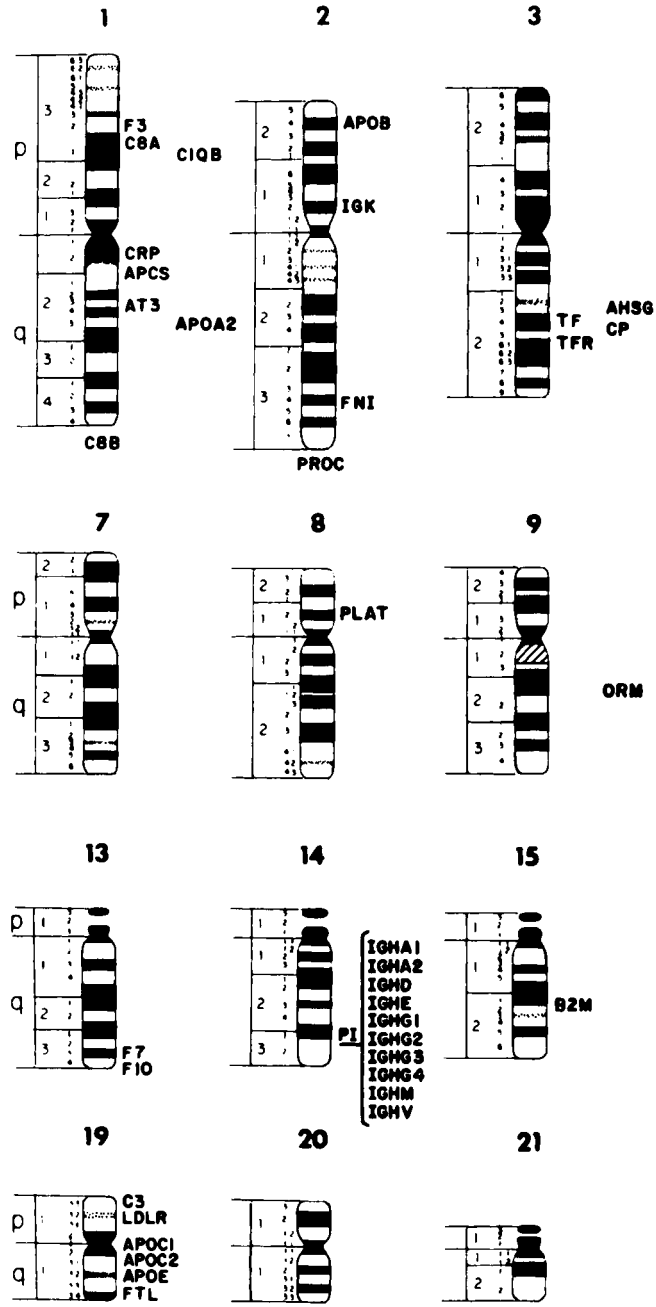
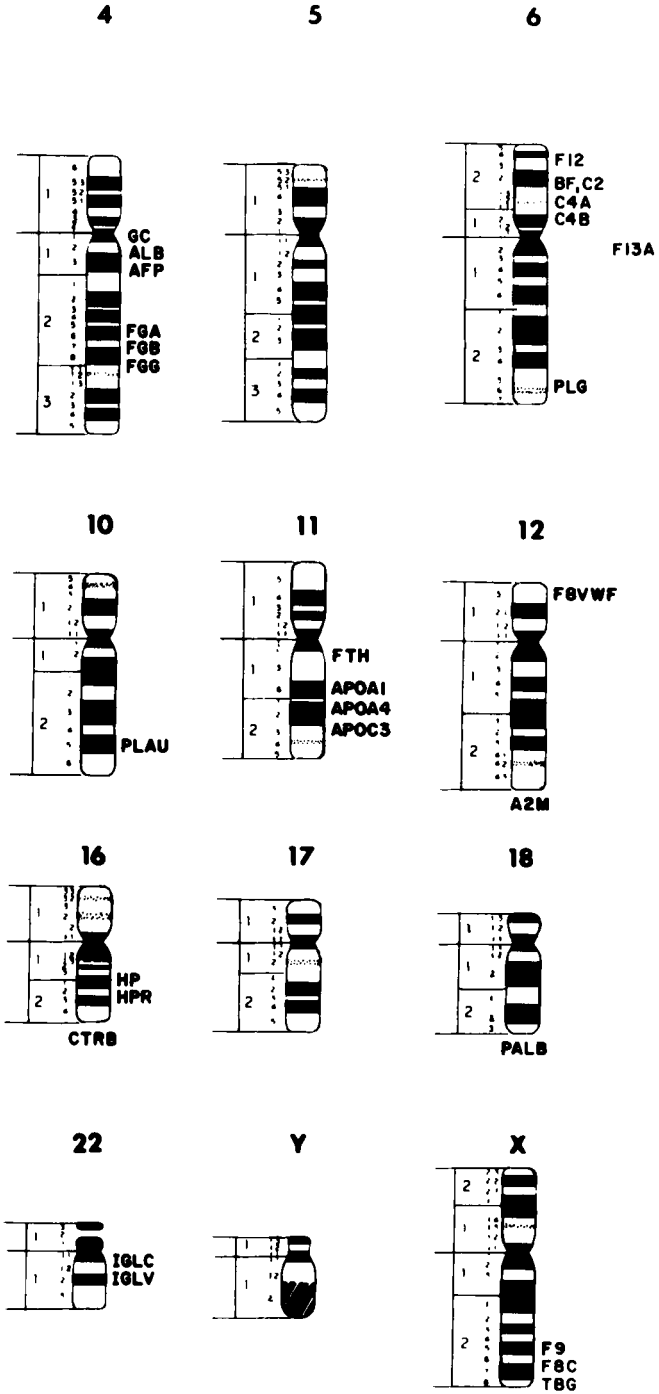


Fig. 1. Chromosomal assignments of human plasma protein genes. Symbols designating genes are located near the chromosomal region to which they have been mapped. Symbols of genes that have been localized to a chromosomal arm appear to the extreme right of the p or q arm, and symbols



of genes that have not been mapped to a specific region appear below the chromosome. Gene mapping information and explanation of symbols are given in Table I.

two transport proteins, the low-density lipoprotein (APOE) and transferrin (Tf), and their respective receptors (LDLR and TfR) are linked on chromosomes 19 (Francke *et al.*, 1984) and 3 (Yang *et al.*, 1984), respectively, raising the possibility of correlations between gene proximity and coordinated expression. Chromosomal mapping studies and DNA sequencing have been predictive of evolutionary relationships heretofore unsuspected. The gene encoding the vitamin D binding protein, also known as the group-specific component (Gc), has been characterized and found to have significant homologies to the major human plasma protein, albumin (Yang *et al.*, 1985a). The cloning, sequencing, and expression of the gene responsible for hemophilia A, caused by a deficiency of clotting factor VIII, has revealed a DNA sequence 187 kilobase pairs (kbp) long. The deduced amino acid sequence of factor VIII demonstrated an unexpected homology to ceruloplasmin, suggesting the presence of metal-binding sites in its structure (Vehar *et al.*, 1984). Sequence analysis of cDNAs encoding plasma proteins has furnished information on the posttranslational processing of haptoglobin (Yang *et al.*, 1983) and coagulation factor VIII. Sequences of heretofore unobserved signal peptides of plasma proteins are being deduced from the nucleotide sequences. Clones with DNA inserts encoding plasma proteins are also valuable probes that are being used to detect DNA polymorphisms within the human genome, thus providing genetic markers necessary for mapping genes that lead to inherited diseases. DNA polymorphisms that are closely linked to, and therefore cosegregate with, defective genes causing inherited diseases can be used for prenatal diagnosis of these diseases. This chapter is written about genes encoding three representative plasma proteins, haptoglobin, transferrin, and the group-specific component; it describes the genetic information gained by analysis of their DNA sequences and chromosomal locations.

II. Plasma Proteins Mapped on Human Chromosomes

A. Somatic Cell Hybrid Analysis

Before molecular techniques utilizing DNA hybridization were available, plasma protein genes were mapped on human chromosomes by detecting families in which karyotyping defined a chromosomal deletion that was accompanied by incompatibility of the inheritance of alleles of that locus. Gene mapping by somatic cell hybrid analysis has been particularly successful in utilizing radiolabeled cDNA probes encoding the protein to be mapped (Naylor and Sakaguchi, 1985). Hybridization of a plasma protein cDNA probe with panels of DNA from human–rodent somatic cell hybrids in which the human chromosomal

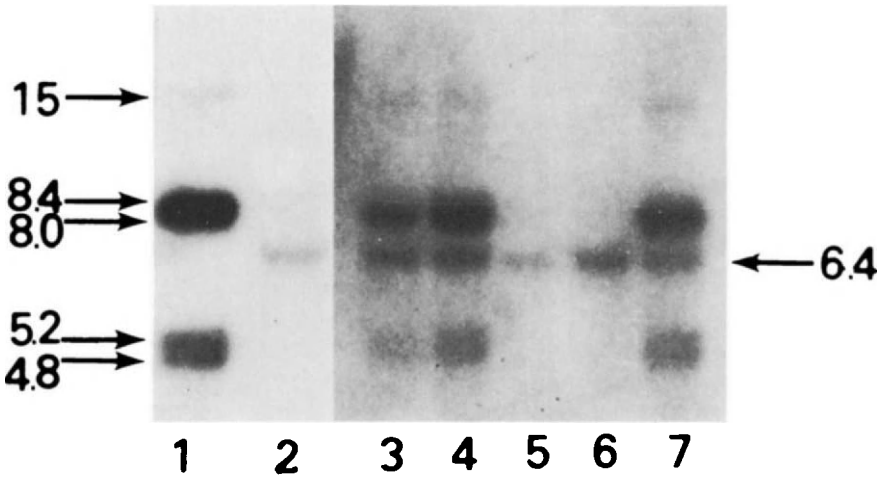


Fig. 2. Chromosomal mapping of the group-specific component (*Gc*) gene by somatic cell hybrid analysis (from Naylor and Sakaguchi, 1985). Hybridization of radiolabeled *Gc* cDNA and panels of endonuclease digests of human DNA (lane 1), mouse DNA (lane 2), and DNA of human–mouse hybrids (lanes 3–7). Human DNA contains five *Pvu*II fragments that hybridize with *Gc* cDNA. Lanes 3, 4, and 7 are hybrids that contain human chromosome 4. Mouse DNA contains one fragment, 6.4 kbp, that hybridizes with human *Gc* cDNA.

complements had previously been defined has facilitated mapping the gene encoding human group-specific component to chromosome 4. Figure 2 is an autoradiograph illustrating the hybridization of radiolabeled *Gc* cDNA with DNA from human–mouse somatic cell hybrids, three of which are known to carry human chromosome 4 (Naylor and Sakaguchi, 1985). The method has been particularly informative when using a human parental cell line having a chromosomal rearrangement affecting the area of interest. For example, a human cell line having a chromosome 3/17 rearrangement affecting only the long arm of chromosome 3 was used in mapping the transferrin gene (Yang *et al.*, 1984). Hybridization of *Tf* cDNA with genomic DNA from a human–rodent hybrid line containing the long arm (3q21-qter region) of human chromosome 3, but not its short arm, p, mapped the *Tf* locus to 3q.

B. *In Situ* Hybridization

The precise chromosomal region in which a gene is located can be determined by the *in situ* hybridization technique developed for single copy genes by Harper and Saunders (1981). This technique, when used with chromosome banding

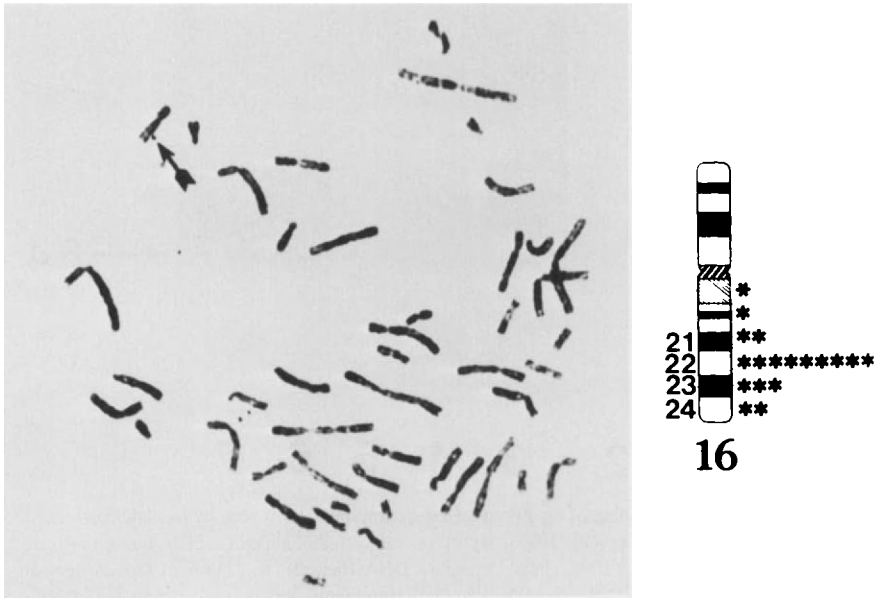


Fig. 3. Chromosomal mapping of human haptoglobin gene by *in situ* hybridization (from McGill *et al.*, 1984b). The arrow designates a silver grain, signaling hybridization of *Hp* cDNA and a chromosomal region on the long arm of chromosome 16. When 60 or more chromosomal spreads were analyzed, over 50% of the silver grains were found over 16q22.

techniques, has proved a most valuable tool for identifying the chromosomal band in which a gene is located. *In situ* hybridization of radiolabeled *Hp* cDNA with human chromosomal spreads is illustrated in Fig. 3. After hybridization and autoradiography, the *Hp* locus was visualized by the presence of silver grains over the chromosomal band containing the locus. In the past five years, *in situ* hybridization and somatic cell hybrid analysis utilized together have significantly increased the number of plasma protein genes mapped to human chromosomes; at the end of 1985 a total of 65 plasma proteins had been mapped to 19 human chromosomes.

Chromosomal mapping has revealed families of duplicated genes that have remained on the same chromosome following the evolutionary events leading to their duplication. The gene families detected thus far include the transferrin family consisting of transferrin and melanoma antigen p97, mapping on the long arm of chromosome 3. The genes encoding the albumin family, including α-fetoprotein (AFP), albumin (ALB), and group-specific component (Gc), are closely linked on chromosome 4. Chromosome 4 also contains the loci for the fibrinogen genes, FGA, FGB, and FGG, The complement components, C2,

C4a, and C4b, map on the short arm of chromosome 6. The clotting factors, F12 and F13A, also map to chromosome 6. Of the apolipoprotein family, APOA1, APOA4, and APOC3 map on chromosome 11; APOC1 and APOC2 map on chromosome 19. The genes encoding the immunoglobulin heavy chain family, IGHG, IGHA, IGHM, IGHE, and IGHD, map on chromosome 14.

Some gene family members have been separated by translocation to other chromosomes, some alone and some with another family member. The immunoglobulin kappa chain gene is found on chromosome 2 while the lambda chain gene is on chromosome 22. One gene of the complement component family, C8, is translocated to the long arm of chromosome 1, while C3 is on chromosome 19. A family of homologous proteins, α_2-macroglobulin and complement components C3 and C4, all contain a unique activatable β-cysteinyl-α-glutamyl thiol ester. These genes are scattered on chromosomes 12, 19, and 6, respectively (Sottrup-Jensen *et al.*, 1985). In the family of blood clotting factors, factor XIIIA is on chromosome 6, factors VII and X are on chromosome 13, and factors VIII and IX are on the X chromosome. The genes encoding haptoglobin and haptoglobin-related, members of the serine protease family, are on chromosome 16, as is the homologous gene chymotrypsinogen B, but they are separated from genes encoding other homologs, trypsin on chromosome 7q and elastase 1 on chromosome 12. A summary of the plasma proteins that have been mapped to human chromosomes is given in Table I. The position on the long or short arm, designated q and p, respectively, is given when known. The references provided are limited by space; therefore, an attempt has been made to provide the first published work or the report presenting the most precise chromosomal location for each gene. The symbols used for the genetic markers appearing in the catalogs of mapped genes follow the guidelines approved by the group responsible for this valuable compilation (Shows and McAlpine, 1982). Each gene symbol consists of capitalized letters and, occasionally, numbers and is presented in this manner in Fig. 1. The marker name for each plasma protein symbol is given in Table I. In the text of this chapter, however, the common abbreviations of the plasma proteins are used.

C. DNA Polymorphisms

Interregional mapping of genes can be carried out utilizing nucleotide polymorphisms. DNA polymorphisms have been valuable in genetic linkage analysis because they segregate in families in a Mendelian manner. It is estimated that in the sequence of genomic DNA, polymorphism occurs as often as 1 in 50 to 1 in 200 nucleotide base pairs. Many of the DNA polymorphisms mask or create recognition sites for restriction endonucleases and are therefore responsible for variation in the number and size of fragments produced by digestion of DNA with endonucleases. The utilization of benign nucleotide polymorphisms for

TABLE I

Chromosomal Mapping of Human Plasma Protein and Related Genes

Protein	Symbol	Chromosome	Arm	Region	Reference
Albumin	ALB	4	q	q11-q13	Kurnit *et al.* (1982)
α_2-Fetoprotein	AFP	4	q	q11-q13	Harper and Dugaiczyk (1983)
α-HS-glycoprotein	AHSG	3	q		Cox and Francke (1985)
α_2-Macroglobulin	A2M	12			Kan *et al.* (1985)
Amyloid P component, serum	APCS	1	q	q12-q23	Mantzouranis *et al.* (1985)
Apolipoprotein A-I	APOA1	11	q	q13-qter	Cheung *et al.* (1984)
Apolipoprotein A-II	APOA2	1		p21-qter	Lackrer *et al.* (1984)
Apolipoprotein A-IV	APOA4	11	q	q13-qter	Schamaun *et al.* (1984)
Apolipoprotein B	APOB	2	p	pter-p23	Law *et al.* (1985)
Apolipoprotein C-I	APOC1	19		cen-q13.2	Tata *et al.* (1985)
Apolipoprotein C-II	APOC2	19		cen-q13.2	Humphries *et al.* (1984a)
Apolipoprotein C-III	APOC3	11		q13-qter	Karathanasis *et al.* (1983)
Apolipoprotein E	APOE	19		cen-q13.2	Olaisen *et al.* (1982)
Antithrombin III	AT3	1	q	q23-q25	Kao *et al.* (1984)
β_2-Microglobulin	B2M	15	q	q22	Goodfellow *et al.* (1975)
Properdin factor B	BF	6	p	p21.3	Allen (1974); Olaisen *et al.* (1983)
Ceruloplasmin	CP	3	q		Weitkamp (1983); Naylor *et al.* (1985)
Chymotrypsinogen B	CTRB	16			Honey *et al.* (1984)
Complement component 1q,β	C1QB	1	p		Solomon *et al.* (1985)
Complement component 2	C2	6	p	p21.3	Alper (1981); Carroll *et al.* (1984)
Complement component 3	C3	19	p	p13.3-p13.2	Lachmann (1982)
Complement component 4A	C4A	6	p	p21.3	Alper (1981); Carroll *et al.* (1984)
Complement component 4B	C4B	6	p	p21.3	Alper (1981); Carroll *et al.* (1984)
Complement component 8A	C8A	1	p	p36.2-p22.1	Mevag *et al.* (1983); Alper *et al.* (1983)
Complement component 8B	C8B	1			Rogde *et al.* (1985)
Coagulation factor III	F3	1	p	pter-p21	Carson *et al.* (1985)
Coagulation factor VII	F7	13	q	q34	Pfeiffer *et al.* (1982)

Coagulation factor VIII C	F8C	X	q	q28	Gitshier *et al.* (1985)
Coagulation factor VIII VWF	F8VWF	12	p	pter-p12	Ginsburg *et al.* (1985)
Coagulation factor IX	F9	X	q	q26-q27.3	Camerino *et al.* (1984)
Coagulation factor X	F10	13	q	q34	Stoll and Roth (1980); Pfeiffer *et al.* (1982)
Coagulation factor XII	F12	6	p	p23	Pearson *et al.* (1982)
Coagulation factor XIII A	F13A	6		p23-qter	Olaisen *et al.* (1985)
C-reactive protein	CRP	1	q	q12-q23	Whitehead *et al.* (1983)
Ferritin H	FTH	11	q	q13	Boyd *et al.* (1984); McGill *et al.* (1984a)
Ferritin L	FTL	19	q	q13.3-q13.4	Caskey *et al.* (1983); McGill *et al.* (1984a)
Fibrinogen A chain	FGA	4	q	q26-q28	Kant *et al.* (1985); Henry *et al.* (1984)
Fibrinogen B chain	FGB	4	q	q26-q28	Kant *et al.* (1985); Henry *et al.* (1984)
Fibrinogen γ chain	FGG	4	q	q26-q28	Kant *et al.* (1985)
Fibronectin	FN1	2	q	q32.3-qter	Koch *et al.* (1982)
Haptoglobin	HP	16	q	q22.1	McGill *et al.* (1984b)
Haptoglobin-related	HPR	16	q	q22.1	Maeda *et al.* (1984)
Group-specific component	GC	4	q	q12-q13	Weitkamp *et al.* (1970); Yang *et al.* (1985a); McCombs *et al.* (1985)
Immunoglobulin polypeptides					
α_1	IGHA1	14	q	q32.3	Croce *et al.* (1979); McBride *et al.* (1982)
α_2	IGHA2	14	q	q32.3	Croce *et al.* (1979); McBride *et al.* (1982)
δ	IGHD	14	q	q32.3	Croce *et al.* (1979); McBride *et al.* (1982)
ϵ	IGHE	14	q	q32.3	McBride *et al.* (1982)
γ 1,2,3,4	IGHG1,2,3,4	14	q	q32.3	Hobart *et al.* (1981)
μ	IGHM	14	q	q32.3	Croce *et al.* (1979); Hobart *et al.* (1981)
Heavy	IGHV	14	q	q32.3	Hobart *et al.* (1981); McBride *et al.* (1982)
κ	IGK	2	p	p12	Malcolm *et al.* (1982)
λ constant	IGLC	22	q	q11.1-q11.2	Erikson *et al.* (1981)
λ variable	IGLV	22	q	q11.1-q11.2	Anderson *et al.* (1984)
Low-density lipoprotein receptor	LDLR	19	p	p13.2-p13.1	Francke *et al.* (1984)
Orosomucoid	ORM	9	q		Cox and Francke (1985)
Prealbumin	PALB	18			Wallace *et al.* (1985)

(*continued*)

TABLE I (*Continued*)

Protein	Symbol	Chromosome	Arm	Region	Reference
Protease inhibitor (α_1-antitrypsin)	PI	14	q	q32.1	Darlington *et al.* (1982)
Plasminogen	PLG	6	q	q25-qter	Murray *et al.* (1985)
Plasminogen activator					
Tissue	PLAT	8	p	p12	Rajput *et al.* (1985)
Urokinase	PLAU	10	q	q24-qter	Riccio *et al.* (1985)
Protein C	PROC	2			Rocci *et al.* (1985)
Thyroxine-binding globulin	TBG	X			Daiger *et al.* (1982)
Transferrin	TF	3	q	q21-q24	Yang *et al.* (1984)
Transferrin receptor	TFRC	3	q	q26.2-qter	Goodfellow *et al.* (1982); Rabin *et al.* (1985)

prenatal diagnosis has had dramatic impact on human genetic disorders in the past four years (Caskey and White, 1983). DNA polymorphisms have been identified that are closely or loosely linked to the genes responsible for sickle-cell anemia, the multiple mutants of β-thalassemia, citrullinemia, Lesch–Nyhan syndrome, growth hormone deficiency, Duchenne-type muscular dystrophy, phenylketonuria, and Huntington's disease. Although there is a wide diversity of restriction endonucleases and each has a highly specific recognition site on the nucleotide sequence of DNA, it is sometimes impossible to find one capable of detecting a DNA polymorphism involving the lesion of the mutant gene. This can be overcome by synthesizing an oligonucleotide that includes the altered nucleotide, or the normal nucleotide sequence. Woo *et al.* (1983) discovered that in the mutant Z allele of the α_1-antitrypsin locus the gene lesion involved a G to A transition. This point mutation does not create or destroy a restriction enzyme recognition site; therefore, an oligonucleotide specific for the normal genotype was synthesized and used under defined conditions as a hybridization probe. Fetuses lacking the normal gene can be identified in prenatal diagnosis procedures using this methodology. The Japanese form of hereditary amyloidosis can be detected by a change in restriction site due to a mutation affecting the gene-encoding prealbumin (Nakazato *et al.*, 1984). DNA polymorphisms are also useful when they are located near a defective gene causing an inherited disease. A DNA polymorphism of a restriction site in the lipoprotein ApoC-II cDNA can be used to follow ApoC-II deficiency in families (Humphries *et al.*, 1984b). The power of using linkage of DNA polymorphisms to approach a genetic disease for which other avenues of investigation have proved unsuccessful is illustrated by the recent success of mapping Huntington's disease to chromosome 4 (Gusella *et al.*, 1983).

The radiolabeled cDNAs encoding plasma proteins are extremely valuable for detecting DNA polymorphisms on panels of restriction digests of human lymphocyte DNA. DNA polymorphisms, often called RFLPs (restriction fragment length polymorphisms) (Botstein *et al.*, 1980), are detected by hybridization with radiolabeled cDNA on filter blots containing endonuclease digests of DNA obtained from human lymphocytes. Since the genes encoding the plasma proteins are scattered throughout the human genome, they furnish an excellent resource in gene mapping studies. Linkage and the distance between a gene and a DNA polymorphism are established by statistical analysis of the rate of crossover that occurs between the two loci.

III. Gene Organization Revealed by DNA Sequences Encoding Human Plasma Proteins

The plasma protein genes encoding albumin, α-fetoprotein, the group-specific component, haptoglobin, and transferrin have evolved from simple, smaller

genes in ancient, ancestral species to larger genes in vertebrates. The evolutionary events leading to the structure of vertebrate plasma proteins include gene duplication, gene conversion, point mutation, translocation, and perhaps reverse transcription.

In the evolution of primitive to complex organisms, gene duplications contribute heavily to the generation of new biochemical mechanisms (Doolittle, 1985). In addition to generating multiple genes with the same function, duplication of genetic material can also produce intragenic amplifications which increase the function of one protein product by increasing the number of active sites. Albumin, α-fetoprotein, the group-specific component (Gc), ceruloplasmin, haptoglobin, hemopexin, and transferrin are among a large number of proteins in modern vertebrates that contain homologous internal domains. Intragenic duplications have contributed to the evolutionary refinement of these proteins. For example, in the divergence of vertebrates from prochordates, the transferrin gene duplicated into two homologous domains, doubling its iron binding capacity. Haptoglobin, transferrin, and Gc are valuable protein models to study because their intragenic duplications and triplications allow us to observe the results of divergence after a defined period in evolution. This is particularly useful in repeated stretches of genes that are still located in tandem on the same chromosome and presumably under the same control mechanisms.

François Jacob (1983) has reviewed protein systems in which evolution has borrowed from what already exists, either transforming a system to give it a new function or combining diverse systems to construct a more complex one. Jacob referred to this process as molecular "tinkering." One of the best examples of molecular tinkering among the human plasma proteins is haptoglobin (Bowman and Kurosky, 1982).

A. Haptoglobin

Haptoglobin is an α_2-glycoprotein that binds free hemoglobin in plasma of humans and other vertebrates. The mature haptoglobin molecule has a tetrachain structure $(\alpha\beta)_2$ with disulfide bonds linking the polypeptide chains. It can usually be visualized after electrophoresis and chromatography because of the pink color associated with its capacity to complex with hemoglobin. Its concentration in serum, normally 30–190 mg/100 ml, is increased during inflammation and decreased after hemolysis. One functional role for haptoglobin is known: protecting the kidneys from tissue destruction by binding free hemoglobin in the event of hemolysis. Recent molecular studies have reaffirmed that haptoglobin is an excellent model for studying the evolutionary impact of genetic events. Effects of point mutations, site replacements, duplication, gene conversion, crossing-over, reverse transcription, and retroviral insertion are displayed in the human haptoglobin locus.

1. Point Mutations

Smithies and Walker (1955) found three haptoglobin types in human populations corresponding to phenotypes Hp1, Hp2, and Hp2-1. Inherited variations in the smaller α subunit pair of the circulating Hp tetrachain structure are responsible for the common haptoglobin phenotypes. Two electrophoretically distinct haptoglobin polypeptides, hp1Fα and hp1Sα, are products of the genes Hp^{1F} and Hp^{1S}, respectively. The molecular weight of the hp1Fα and Hp1Sα chains is 8900 (M_r). Protein analysis indicated the two chains differed by an amino acid substitution, which was lysine at residue 53 in hp1Fα and glutamic acid at the same position in hp1Sα, accounting for the differences in electrophoretic migration at acid pH (Smithies *et al.*, 1962a). Analyses of cDNAs encoding the *Hp* genes recently revealed that another amino acid variation occurred at residue 52 of the hpα polypeptides where Asp and Asn were deduced in the HP^{1F} and Hp^{1S} genes, respectively (van der Straten *et al.*, 1984; Maeda *et al.*, 1984; Brune *et al.*, 1984).

2. Site Replacements

Two additional nucleotide site replacements distinguish the Hp^{1F} and Hp^{1S} cDNAs. Residue 47, Val, is found in both hpα chains; however, it is encoded by GTA in Hp^{1F} and GTG in Hp^{1S}. The second site replacement occurs in the codons specifying Asn in residue 51, AAT in Hp^{1F}, and AAC in Hp^{1S} (Brune *et al.*, 1984; van der Straten *et al.*, 1984). The differences found between the Hp^{1F} and Hp^{1S} cDNAs are summarized in Fig. 4. The four differences characteristic of the Hp^{1F} and Hp^{1S} cDNAs detected thus far have permitted us to distinguish their contributions in the crossover event that gave rise to the Hp^2 gene, the third allele of the haptoglobin system (Yang *et al.*, 1983).

3. Intragenic Duplication

Intragenic duplication appears to be a common mechanism in the evolution of plasma proteins. Haptoglobin has been conspicuous among the plasma proteins for sustaining intragenic duplications. The partial gene duplication which produced the Hp^2 allele was the product of an unequal crossover. The crossover apparently occurred in a pair of human chromosome 16's in the germ cells of a heterozygote, with the genotype Hp^{1F}/Hp^{1S} (Smithies *et al.*, 1962b; Yang *et al.*, 1983; Maeda *et al.*, 1984). This event occurred in evolution after the divergence of man, approximately four to five million years ago. Haptoglobin, therefore, contains one of the newest intragenic duplications known among the plasma proteins. The hp2α polypeptide encoded by the Hp^2 gene was shown to have a molecular weight of 16,000 (M_r), nearly double that of the hp1Fα or hp1Sα

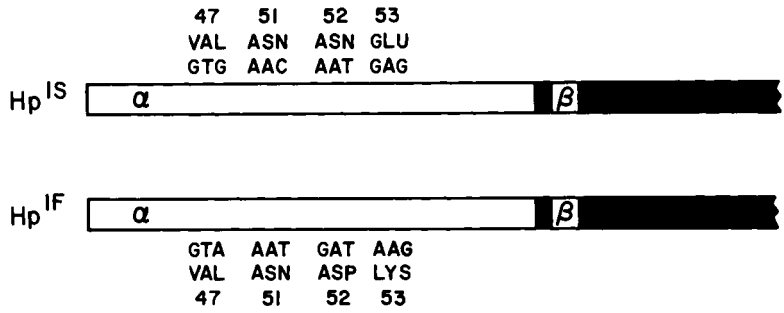


Fig. 4. Differences in the Hpα domains encoded by Hp^{1S} and Hp^{1F}. Two amino acid differences at residues 52 and 53 result from G to A transitions. The site replacements at residues 47 and 51 are also caused by transitions that change the codons without affecting the amino acids.

chains. This and the presence of sequences unique to hp1Fα or hp1Sα in the hp2α chain led Smithies *et al.* (1962b) to propose that Hp^2 was a partially duplicated gene formed by a rare nonhomologous crossover event which fused the Hp^{1F} and Hp^{1S} genes. This was recently substantiated by sequence analysis of Hp^2 cDNA (Yang *et al.*, 1983) and analysis of all three haptoglobin genes (Maeda *et al.*, 1984) (see Appendix).

Until the sequence of Hp^2 cDNA was determined, it had been impossible to align the contributions of the Hp^{1S} and Hp^{1F} genes. Three Hp^2 cDNA sequences and one gene sequence characterized thus far (van der Straten *et al.*, 1983; Yang *et al.*, 1983; Raugei *et al.*, 1983; Maeda *et al.*, 1984) have established the same alignment, Hp^{2FS}. The amino acid substitutions and nucleotide site replacements noted above in the Hp^{1F} and Hp^{1S} alleles were observed in tandem in Hp^2, providing evidence that the mutations responsible for the four nucleotide changes (Fig. 4) occurred before the evolutionary event that produced the Hp^2 gene. In fact, there has been no evidence of a mutation occurring in the coding region of the Hp^{2FS} gene since its origin.

Among other plasma proteins that are products of intragenic duplications, the homologous domains of the Hp^2 cDNA display the least evolutionary divergence, having nucleotide sequences 98% identical (Yang *et al.*, 1983), and provide a dramatic contrast to the two homologous domains of transferrin, which are 50% identical (Yang *et al.*, 1984); the three homologous domains of albumin, which are 40–50% identical (Sargent *et al.*, 1981); and the three homologous domains of Gc, which are 36% identical (Yang *et al.*, 1985a).

There have been complex differences observed among the introns of the Hp^{1F} and Hp^{1S} genes that may have stemmed from an ancient gene conversion between the primordial Hp^1 gene and *Hpr* (Nobuyo Maeda, personal communication). *Hpr* is another gene in the *Hp* complex that is discussed below. When the

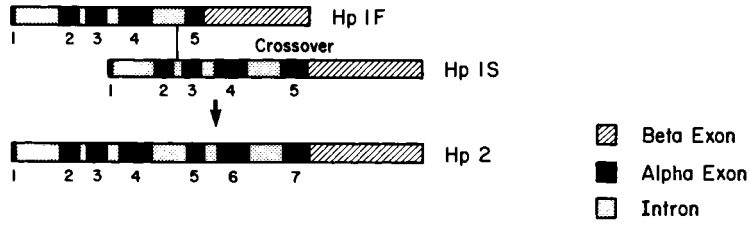


Fig. 5. Chromosomal sites of unequal crossover event that produced Hp^2. The crossover probably took place in the germ cells of a heterozygote with the genotype of Hp^{1S}/Hp^{1F}. The recombination was between the fourth intron of Hp^{1F} and the second intron of Hp^{1S}. The introns and exons are not drawn to scale.

genomic DNA of the Hp^2 gene was sequenced, the hpα1F and hpα1S domains were shown to differ at 37 positions, including 27 transitions, 8 transversions, and 2 differences in length. Four of the differences were in the coding regions noted above and the others were in introns (Maeda *et al.*, 1984).

The site of the crossover event leading to the partial duplication produced in the Hp^2 gene was characterized by Maeda *et al.* (1984) and is demonstrated in Fig. 5. The nucleotide sequences of the *Hp* genes demonstrated that the site of the unequal crossover was between the fourth intron of Hp^{1F} and the second intron of Hp^{1S}. Probably the most remarkable feature of the DNA sequences in which the crossover took place is the lack of similarity between the relevant regions of the Hp^{1F} and Hp^{1S} derived sequences, limited to a region of identity in only a few base pairs (Fig. 6).

Additional *Hp* duplications have been noted in population studies, but do not occur as often as Hp^{2FS}. Nance and Smithies (1963) found two haptoglobin variants, Hp^{2FF} and Hp^{2SS}, in a Brazilian population. Constans and Viau (1977a) described Hp^{2FF} and Hp^{2SS} in Basques and Baronnies. Contiguous gene duplication generates subsequent unequal crossing-over between homologous chromosomes. Therefore, in regions where genetic duplications are found there is an increased probability that subsequent duplications will occur because of the increased chromosomal homology. Giblett (1964) described a haptoglobin type,

Fig. 6. Details of the crossing-over that formed the Hp^2 allele (from Maeda *et al.*, 1984). Bases identical to Hp^2 are tied by vertical lines. The box shows the region within which the crossing-over occurred. [Reprinted by permission from *Nature* (*London*). Copyright © 1984 Macmillan Journals Limited.]

called haptoglobin Johnson, that appeared to be a triplicated hpα chain, Hp^3, probably the product of unequal crossing-over between chromosomes in Hp^2 homozygotes (Dixon, 1966).

Oliviero *et al.* (1985b) have reported molecular evidence of triplication within the haptoglobin Johnson gene. Direct gene analysis by hybridization of genomic DNA endonuclease digests and radiolabeled Hp^2 cDNA indicated a threefold tandem repeat of a 1.7-kbp DNA segment. A nine-exon model of the triplicated gene was proposed in which exons 3 and 4 were repeated three times.

4. Gene Duplication: Hpr, the Haptoglobin-Related Gene

Raugei *et al.* (1983) noted a second haptoglobinlike sequence in the haploid genome of man. Hybridization of radiolabeled *Hp* cDNA with *Eco*RI digests of human DNA revealed polymorphic fragments independent of those contributed by the common alleles, Hp^{1F}, Hp^{1S}, and Hp^2. Maeda *et al.* (1984) cloned the haptoglobin-related gene and named it *Hpr*. Maeda (1985) demonstrated that *Hpr* was tightly linked to, and in the same orientation of, the *Hp* gene, only 2.2 kbp away from the 3′ noncoding region of human Hp. These results were in agreement with those of Bensi *et al.* (1985). Maeda (1985) confirmed that *Hpr* was formed by 6.5 kbp of DNA duplicated from Hp^1. The length of the coding region of *Hpr* is only one codon (in the leader sequence) longer than that of Hp^1. The schematic representation of *Hp* and *Hpr* on chromosome 16 is given in Fig. 7.

Sequence analysis of DNA demonstrated that the difference between Hp^1 and *Hpr* constitutes 6.4%. There are 27 amino acid changes between the amino acid sequence deduced from *Hpr* and the precursor to Hp^{1F} (Bensi *et al.*, 1985; Maeda, 1985). One of the amino acid differences is a substitution of Cys-15 by Phe in *Hpr*. This Cys residue would be expected to be important in stabilizing the Hp tetramer by cross-linking its αβ monomers. There is, however, an additional Cys residue at position 157 of the β domain in *Hpr*, not present in Hp^1, which might serve the same purpose and cross-link the Hp monomers. Even if this does not occur, results from characterizing dog haptoglobin indicated that haptoglobin monomers can form a tetramer without Cys-15. The substitution of Phe for Cys-15 in dog haptoglobin does not contribute to a reduction of molecular weight nor electrophoretic charge. Only when dog haptoglobin is denatured with urea does it dissociate and decrease in molecular weight to that of an αβ subunit (Kurosky *et al.*, 1979).

Other differences found between *Hpr* and Hp^1 were two amino acid alterations in the signal peptides, one extra amino acid codon in *Hpr*, 9 amino acid differences in the α domain, and 16 differences in the β domain (Maeda, 1985). Based on differences in the first introns of the two genes, *Hpr* was estimated to be 30 million years old (Maeda, 1985).

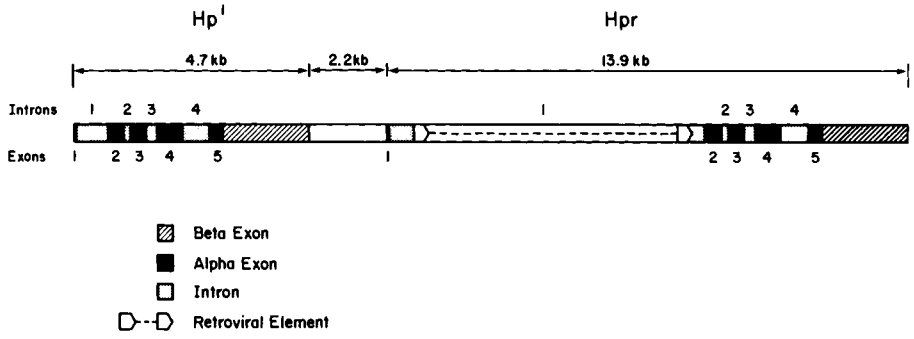


Fig. 7. Tandem arrangement of the haptoglobin (*Hp¹*) and haptoglobin-related (*Hpr*) genes on human chromosome 16 (drawn from information in Bensi *et al.*, 1985, and Maeda, 1985). The two genes are separated by 2.2 kbp and have similar number and positions of exons and introns. Unlike *Hp¹*, the first intron of the *Hpr* gene is increased to over 9 kbp in length because of the presence of a retroviruslike element. The introns and exons are not drawn to scale.

The first intron of *Hpr* holds one of the surprises typical of haptoglobin history; the presence of a retroviruslike element has increased the first intron to over 10 kbp in length (Fig. 7). It is the first example of a retroviruslike element being found within a defined gene in the human genome. The retroviral sequence contained in *Hpr* intron 1 is different from any retrovirus sequence found before in human DNA and has a potential primer-binding site homologous to a mouse isoleucine tRNA (Maeda, 1985).

It is not yet understood if the *Hpr* gene is expressed. There are no apparent reasons to suspect it of being a pseudogene since sequence comparisons of the first exon and the 5′-flanking region indicate putative TATA and CAAT boxes. The number of introns and the position of the intron–exon junctions are identical to those in the *Hp¹* genes (Maeda, 1985; Bensi *et al.*, 1985). However, no *Hpr* mRNA has been detected in fetal or adult liver. If expressed, *Hpr* is estimated to be at the lower limit of detection (Bensi *et al.*, 1985). Maeda (1985) has suggested that the retroviruslike element may influence the expression of *Hpr*, *Hp*, and other acute-phase reactants. Therefore, characterization of its function may yield new information related to gene expression. Her prediction is consistent with the number of important discoveries made first while studying haptoglobins.

5. Homology of Haptoglobin and the Serine Proteases

Determination of the amino acid sequence of haptoglobin revealed a significant homology of the β domain with members of the serine protease family, a group of proteolytic enzymes that includes trypsin, chymotrypsin, thrombin, elastase, and some members of the serum complement components (Barnett *et*

al., 1970; Kurosky *et al.*, 1980; Bowman and Kurosky, 1982). The genes encoding these proteins are scattered widely throughout the human genome; however, the genes encoding haptoglobin, haptoglobin-related, and chymotrypsin B have remained linked on chromosome 16.

Analysis of the amino acid sequence of haptoglobin and the serine proteases revealed that the hpβ domain was 29–33% identical to bovine trypsin, chymotrypsin, porcine elastase, human thrombin, and plasmin (Kurosky *et al.*, 1980). As Doolittle (1981) pointed out, haptoglobin assumed a specificity more stringent than its proteolytic ancestors; it no longer has enzymatic activity but binds free hemoglobin irreversibly. Haptoglobin has lost proteolytic activity because the active sites of the serine proteases, His-57 and Ser-195, have been replaced by Lys and Ala, respectively (Kurosky *et al.*, 1980). Kurosky *et al.* (1980) had demonstrated a significant homology in the α domain of haptoglobin with both the activation peptides of the serine proteases and the kringle domain (Magnusson *et al.*, 1975) found in thrombin, tissue plasminogen activator, and plasmin, the greatest similarity being demonstrated with the fifth kringle structure of plasminogen. After the *Hp* cDNA was characterized, homologous sequences in amino acids and nucleotide bases were detected near the α–β junction of human haptoglobin and the tissue-type plasminogen activator (Yang *et al.*, 1983). The homology of haptoglobin and the serine proteases has recently been emphasized by characterization of the posttranslational events contributing to the tetrameric structure of mature haptoglobin.

6. Posttranslational Processing of Haptoglobin

After *Hp* cDNA had been identified and sequenced, an interesting posttranslational event by which the haptoglobin tetramer is formed was characterized (Yang *et al.*, 1983). The series of steps involving limited proteolysis that led to the configuration of circulating haptoglobin is shown in Fig. 8. This type of posttranslational processing has also been observed in ceruloplasmin, factor VIII, and several complement components. The amino acid sequence deduced from *Hp* cDNA sequence predicted an unexpected COOH-terminal arginine in the 143rd position of the α chain that had never been found in the human haptoglobin sequence (Kurosky *et al.*, 1980). This information demonstrated that the α and β chains of haptoglobin were products of posttranslational proteolysis and that the DNAs specifying Hp α and β polypeptides are not only linked but are continuous in the cDNA sequence. It seemed likely that limited proteolysis of the haptoglobin precursor results in cleavage behind the 143rd residue arginine in the α chain. Hanley and Heath (1985) have characterized an enzyme in rat plasma that cleaves prohaptoglobin in this precise manner. Two polypeptides, the α and β, are thus formed and are held together by disulfide bonds. The carboxyl-terminal arginine is removed from the α chains by circulat-

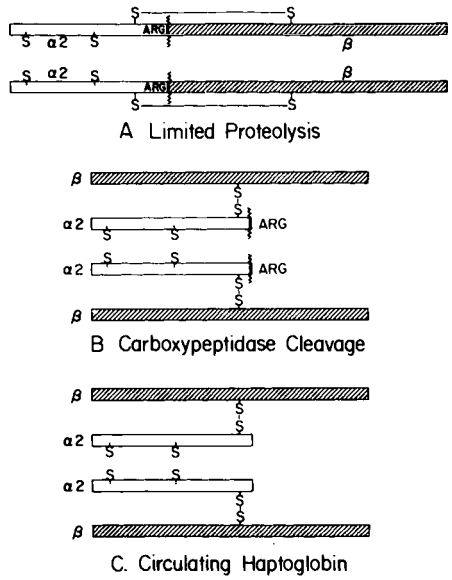


Fig. 8. Model for maturation of human haptoglobin (from Yang *et al.*, 1983). Contiguous cDNAs encode the Hpα2 and Hpβ chains, separated by one codon corresponding to an arginine residue in the COOH-terminal portion of the Hpα2 chain. This is the postulated site of limited proteolysis (A). The COOH-terminal arginine is removed by circulating carboxypeptidase (B). The mature haptoglobin tetramer ($\alpha\beta_2$) is held by disulfide bonds, the positions of which have not been specified (C).

ing carboxypeptidase in an early posttranslational step. This model has been supported by observations of *in vitro* synthesis of single polypeptide chains of haptoglobin by cells from rabbit (Chow *et al.*, 1983) and rat (Haugen *et al.*, 1981). Transcription of a single *Hp* mRNA also clarified earlier results obtained from characterizing the Hp polymers in plasma (Fuller *et al.*, 1973). Hp2 and Hp2-1 form polymers of increasing molecular weight by addition of Hp polypeptides through formation of disulfide bonds. The amino acid composition of the major polymers of Hp indicated that each differed from the next smaller polymer by the addition of an αβ unit. It now seems clear that the Hp polymer series are formed by increments of single Hp (αβ) polypeptide chains.

A comparison of the posttranslational cleavage sites of haptoglobin (Yang *et al.*, 1983) and tissue-type plasminogen activator (Pennica *et al.*, 1983) demonstrates a high degree of identity in the amino acids and nucleotides (Fig. 9). When the homologous regions in the two proteins are compared, 12 out of 31 amino acids are identical and 4 additional residues are chemically similar. In this region 44 out of 93 nucleotides are identical and all the nucleotides encoding the Pro-Trp-Gln-Ala sequence are identical in the two genes.

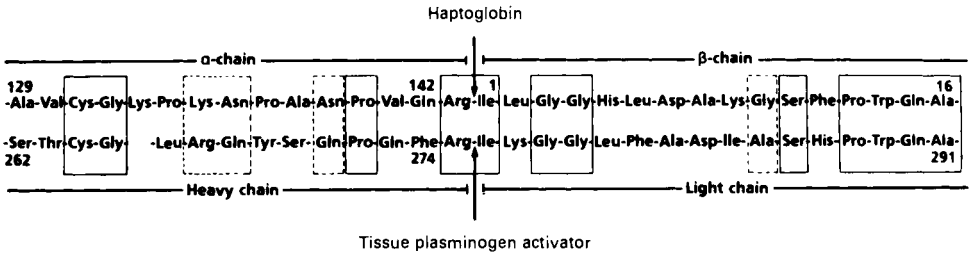


Fig. 9. Homologous regions of human haptoglobin and tissue-type plasminogen activator (from Yang *et al.*, 1983). Arrows indicate potential cleavage sites between arginine and isoleucine that occur at the α–β junction of haptoglobin and at the heavy–light chain junction of plasminogen activator. The amino acid residues are numbered above and below the sequences of haptoglobin and plasminogen activator, respectively. One gap was introduced in the plasminogen activator sequence to maximize homology. Identical residues are enclosed by solid lines and chemically similar residues by broken lines.

7. Reverse Transcription

Comparison of the haptoglobin gene to the genomic organization of its close relative, chymotrypsin, revealed unexpected results. Craik *et al.* (1984) pointed out that serine protease genes, including rat trypsin I and II, mouse kallikrein, rat chymotrypsin B, rat elastase I and II, and the serine protease domain of human complement factor B, demonstrated similarity in their intron numbers and junctions as well as their sequence and structural homologies. Chymotrypsin's gene contained eight introns that interspersed seven exons (Bell *et al.*, 1984). Although the β domain of haptoglobin is 37% identical to chymotrypsin in amino acid sequence and has significant homology to other members of the serine protease family, there were no introns found upon sequencing the β domain of the *Hp* gene and comparing it to the sequence of the *Hp* cDNA (Maeda *et al.*, 1984; Yang *et al.*, 1983). This suggests that the β domain of the *Hp* gene arose from the insertion of a processed gene, i.e., a reverse transcript, obtained from mRNA of a chymotrypsin ancestral gene. A proposed scheme is shown in Fig. 10. An alternative, but much less likely, mechanism would be the removal of the eight introns and consolidation of the seven exons characteristic of the chymotrypsinogen gene by repeated unequal crossing-over accompanied by removal of intervening sequences. If true, haptoglobin would be the first example in which a reverse transcript is present in the coding sequence of a functional gene (Bowman *et al.*, 1985). Reverse transcripts are processed genes that have been enzymatically formed from vertebrate mRNAs by the enzyme reverse transcriptase. Retroviruses synthesize reverse transcriptase and can reinsert processed genes into the host DNA (Bishop, 1983). The reverse transcript gained by

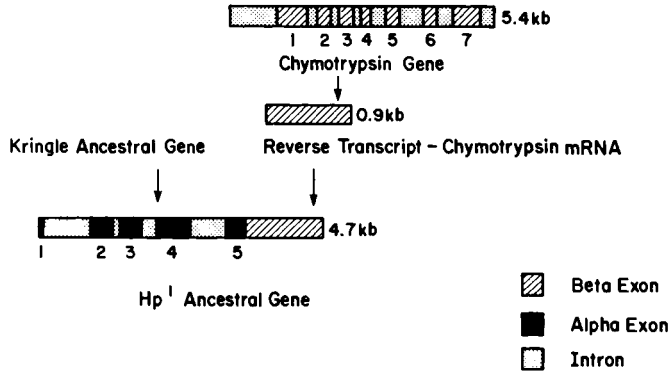


Fig. 10. Contribution of a reverse transcript to *Hp* gene formation (from Bowman *et al.*, 1985). The scheme proposes the origin of the β domain from mRNA of a chymotrypsin gene ancestor. Although there is strong homology of the Hp β chain and chymotrypsin, genomic DNA encoding the β chain contains no introns compared to the eight introns of the chymotrypsin gene. The α domain is derived from a kringle ancestor (Kurosky *et al.*, 1980). The ancestor of the kringle gene is shown with four introns to correspond with the Hpα domain, although the intron number of kringle domains characterized thus far varies from 0 to 1. The introns and exons are not drawn to scale.

the host usually contains a 3′ poly(A) tail, a sequence missing in the DNA encoding the β domain of haptoglobin. Sequencing *Hp* cDNA did reveal an A-G-T-G-G-A repeat that occurred three times in the 3′ noncoding region; its relationship, if any, to reverse transcription remains unknown (Yang *et al.*, 1983). The fusing of the reverse transcript that encoded the β domain with a genomic sequence of the kringle region, the evolutionary ancestor of the α domain, is likely to have produced the ancestral gene of *Hp*, *Hpr*, and perhaps other genes belonging to the serine protease and complement families. Reverse transcription joins other interesting genetic events such as gene conversion, gene duplication, viral insertion, and point mutation in contributing to the evolution of haptoglobin.

8. Gene Mapping and DNA Polymorphisms Detected by Hp cDNA

The genes encoding the family of serine proteases are separated throughout the human genome, although the genes encoding haptoglobin, haptoglobin-related, and chymotrypsin have remained linked on human chromosome 16. The identification of the *Hp* cDNA allowed us to confirm the chromosomal mapping, which had been established earlier for the Hpα chain by Magenis *et al.* (1970) after studying a large kindred demonstrating the segregation of *Hpα* alleles with a recurrent chromosome 16 break. The chromosomal break point was identified as

TABLE II

Human Haptoglobin DNA Polymorphisms[a]

	Restriction fragment (kbp)		
Hp Genotype	*BCl*I[b]	*Eco*RI[c]	*Pst*I[d]
2	11.5	11.8 (10.1)	5.9
1	9.6 (4.0)	10.1	4.2
r	6.5	8.3	4.4 (2.9 + 1.5)

[a]Polymorphic fragments are shown within parentheses.
[b]Hill *et al.* (1985).
[c]Maeda *et al.* (1984).
[d]Oliviero *et al.* (1985a).

16qh (heterochromatin) by Hecht *et al.* (1971), and independently by Ferguson-Smith and Aitken (1978) as the distal region of 16cen-q22. In our studies, radiolabeled *Hp* cDNA hybridized to a site on chromosome 16, precisely on the 22nd band of the long arm (McGill *et al.*, 1984b). The *in situ* hybridization of *Hp* cDNA and band 16q22 on human chromosomal spreads is shown in Fig. 3.

Probing human lymphocyte DNA with radiolabeled *Hp* cDNA has also furnished information about the subregional mapping of the haptoglobin locus. Intron differences among the Hp^2, Hp^{1F}, and Hp^{1S} alleles and the *Hpr* gene have facilitated the detection of these genes by filter hybridization according to the method of Southern. The DNA polymorphisms reported to date within the *Hp* locus are summarized in Table II.

Hp^{1F} and Hp^{1S} can be distinguished after digesting human DNA with the endonuclease *Xba*I (Maeda *et al.*, 1984; Oliviero *et al.*, 1985a; Hill *et al.*, 1985). The DNA polymorphism first detected in the *Hp* locus was with the endonuclease *Eco*RI (Raugei *et al.*, 1983) and was mapped by Maeda *et al.* (1984) to approximately 4 kbp 3′ to the Hp^2 gene. Hill *et al.* (1985) discovered a DNA polymorphism with *Bcl*I that mapped upstream of the Hp^1 gene.

The presence of an Hp-like sequence was detected in human DNA digests by utilizing the endonuclease *Eco*RI (Raugei *et al.*, 1983). The *Hpr* gene can also be observed in all human genomes after digestion with *Hin*dIII, *Bam*HI, *Bgl*II, *SST*I, and *Xba*I. A DNA polymorphism in the *Hpr* gene was found by Oliviero *et al.* (1985a) after digestion of human lymphocyte DNA with *Pst*I. The segregation of the variant was followed in a family where the heterozygous father transmitted the gene to two of his three children.

The triplicated gene, Hp^3 (Johnson), was detected with *Bgl*II, *Bam*H1, and *Hin*dIII (Hill *et al.*, 1985) and with *Hin*dIII, *Eco*RI, and *Pst*I (Oliviero *et al.*, 1985b).

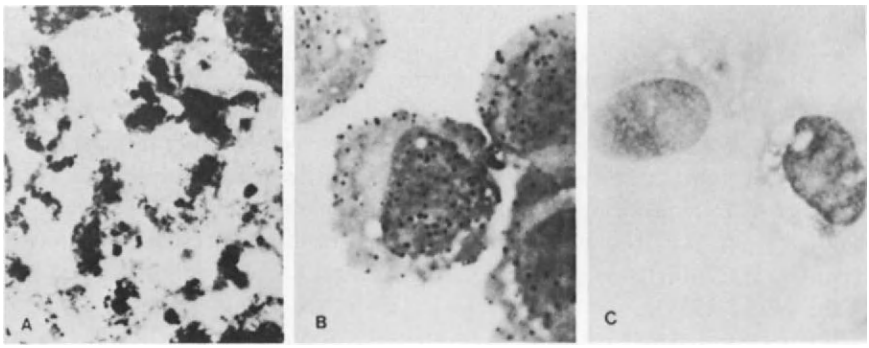


Fig. 11. Transcription of human haptoglobin mRNA in liver and in hepatocytes (from Lum *et al.*, 1985a). *In situ* hybridization with radiolabeled human *Hp* cDNA with tissue sections of liver (A) and a hepatoma cell line (B) demonstrates *Hp* mRNA by the appearance of silver grains. No transcription is seen in fibroblasts (C).

9. Expression of the Hp Gene

Haptoglobin is synthesized in the liver. Transcription of *Hp* mRNA has not been detected in any other tissue by *in situ* histohybridization utilizing radiolabeled human *Hp* cDNA as a probe (Lum *et al.*, 1985). Transcription of *Hp* mRNA can be seen in Fig. 11. van der Straten *et al.* (1985) reported *in vitro* expression of human *Hp2* after inserting *Hp2* and α_1-antitrypsin (*PI*) cDNA in expression vectors for yeast and *Escherichia coli*. In yeast, *PI* was expressed in concentrations 100–1000 times higher than *Hp* although the relative concentrations of the two proteins in the blood are approximately the same, and the levels of mRNA encoding the two proteins were the same in yeast. An interesting question is whether or not the extensive posttranslational processing necessary for the formation of mature haptoglobin can be carried out in yeast.

B. Transferrin

Transferrin is a plasma protein of biological interest not only because of its role as a growth factor for normal and malignant proliferating cells but also because of its interesting evolutionary history. The structure and function of transferrin have recently been reviewed by Putnam (1984). Transferrin transports iron from the intestine, reticuloendothelial system, and lower parenchymal cells to all proliferating cells in the body. Each transferrin molecule can bind two atoms of ferric iron (Aisen and Listowsky, 1980). Iron binding imparts a pinkish color to the glycoprotein and is carried out by the side chains of specific tyrosine,

histidine, and arginine residues that have been conserved in evolution (Williams, 1982; MacGillivray *et al.*, 1983).

The concentration of transferrin in plasma, approximately 250 mg/100 ml, can increase in pregnancy and chronic iron deficiency, although transferrin is not considered an acute-phase reactant. Decreased concentrations can accompany chronic infection, cirrhosis, and starvation. Transferrin is a polypeptide composed of 679 amino acids in addition to carbyohydrate chains. The sequence deduced from *Tf* cDNA added one amino acid to the 678 residues previously reported (MacGillivray *et al.*, 1983), an additional Leu-562 following Leu-561 (Yang *et al.*, 1984).

The entry of transferrin into cells involves receptor-mediated endocytosis (Newman *et al.*, 1982; Dautry-Varsat *et al.*, 1983; Iacopetta *et al.*, 1983). A cell can have as many as 400,000 transferrin receptors. The receptor and transferrin are internalized and the complex passes through an acidic compartment; the iron is delivered and the receptor plus apotransferrin return to the membrane, where apotransferrin is released into the extracellular environment.

In 1984 the *Tf* cDNA was cloned, characterized, and chromosomally mapped (Yang *et al.*, 1984). The nucleotide sequence revealed that there was a 19-residue leader sequence, homologous to the chicken ovotransferrin leader sequence (Thibodeau *et al.*, 1978; Jeltsch and Chambon, 1982), followed by nucleotides encoding the 679 amino acids, a stop signal, and an additional 171 nucleotides in the 3′ noncoding region. A comparison of the nucleotide sequences in the two homologous domains of the *Tf* cDNA provided information related to the intragenic duplication of the *Tf* gene.

1. Transferrin Evolution

A. INTRAGENIC DUPLICATION. The vertebrate *Tf* gene appears to be the descendant of a primordial gene that existed in prochordates approximately 500 million years ago (Williams, 1982). The ancient gene was approximately one-half the size of the vertebrate gene. Figure 12 illustrates the evolution of the transferrin family of genes. In Fig. 12 the gene duplications responsible for human transferrin, lactotransferrin, and p97 are depicted as occurring after the divergence of birds and mammals; however, very recent characterization of the DNA sequence (Rose *et al.*, 1986) provides evidence that p97 probably diverged from the other members of the transferrin family before humans and chickens diverged from one another.

Martin *et al.* (1984) described an iron binding protein one-half the size of vertebrate transferrin, approximately 41,000 Da, in *Pyura stolonifera*. *Pyura* is a prochordate having a notochord in its larval stage. Hybridization studies of endonuclease digests of the DNA from a related prochordate, *Pyura haustor*, demonstrated strong hybridization patterns with human *Tf* cDNA. The precise

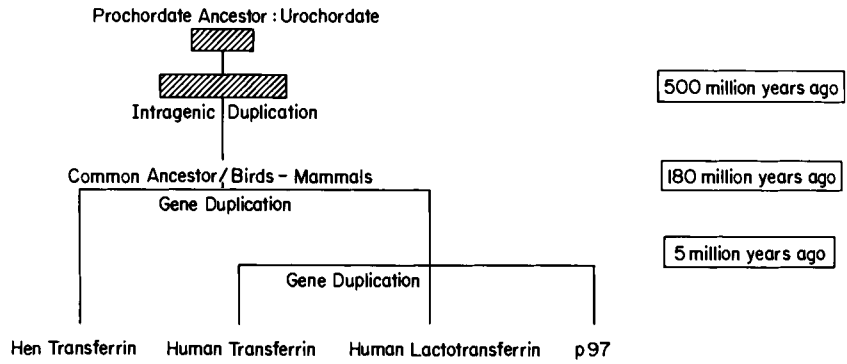


Fig. 12. Evolution of genes encoding the transferrin family (based on Yang *et al.*, 1985b). The vertebrate *Tf* gene appears to be the descendant of a primordial gene that existed in prochordates approximately 500 million years ago (Williams, 1982). Proteins in the transferrin family in vertebrates are controlled by genes that have evolved by duplication of DNA sequences 18 kbp long or longer. These include hen transferrin, human transferrin, human lactotransferrin, and p97.

size of the *Pyura* gene has not been determined, but preliminary studies indicate that it is significantly smaller than the vertebrate gene (Yang *et al.*, 1985b).

The intragenic duplication of the *Tf* ancestral gene carried a distinct evolutionary advantage: transferrin could bind twice as many iron ions. Studies with erythroid precursors have shown that iron is more rapidly taken up from diferric than from monoferric transferrin (Huebers *et al.*, 1981).

The amino acid sequence of human transferrin (MacGillivray *et al.*, 1983) and chicken ovotransferrin (Williams *et al.*, 1982) reflected the *Tf* intragenic duplication. The two homologous domains in human transferrin have approximately 37% identity when the amino-terminal domain (residues 1–336) is aligned with the carboxyl sequence (residues 337–678).

The determination of the nucleotide sequence of *Tf* cDNA indicated that, during evolution, selection acted more strongly on some regions of the *Tf* exons than on others. Of the amino acids identical in the amino and carboxyl domains of transferrin, only 50% have identical nucleotide codons. Yet, when the nucleotide base sequences were compared after aligning the two domains of the *Tf* cDNA, there were three regions, designated homology blocks A, B, and C, that revealed 73%, 64%, and 62% identity of nucleotide sequence, respectively (Fig. 13). Although the basis for evolutionary constraints on the three homology blocks is unknown, blocks A, B, and C do contain codons for tyrosine and histidine residues predicted to be functional sites important in iron binding. For example, of the eight pairs of tyrosine residues conserved in the amino and carboxyl domains of both human (MacGillivray *et al.*, 1983) and hen transferrin (Williams *et al.*, 1982), six pairs are encoded by nucleotides in the homology

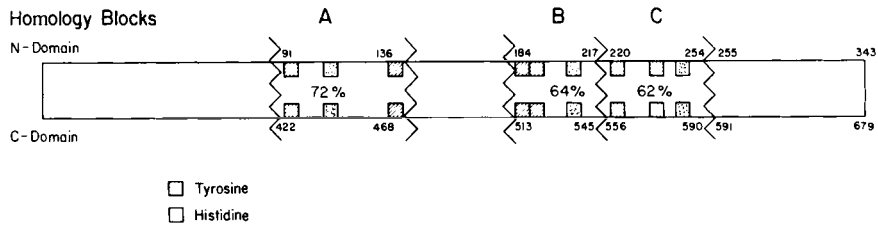


Fig. 13. Homology blocks obtained by aligning the nucleotide sequences of the amino and carboxyl domains of the *Tf* cDNA sequence (based on Yang *et al.*, 1984). The percentages within the three homology blocks signify the nucleotide identity. The tyrosine and histidine residues, identical in the two domains, have been predicted to be functional sites important in iron binding.

blocks. During evolution, the intragenic gene duplication producing a longer transferrin polypeptide with two active sites was followed by repeated gene duplications producing additional new iron binding proteins discussed below.

B. GENE DUPLICATION. Proteins in the transferrin family are controlled by genes that have evolved by duplication of DNA sequences 18 kbp long or longer. Mammalian descendants include serum transferrin; p97, a melanoma cell antigen present in very low amounts in normal cells; and lactoferrin, a secreted protein present in milk and other body fluids. In the past two years, extensive information has become available in the field of iron binding proteins, with recent determinations of the organization and sequences of *Tf* cDNA (Yang *et al.*, 1984), most of the *Tf* gene (Park *et al.*, 1985), the transferrin receptor (*TfR*) gene (McClelland *et al.*, 1984; Schneider *et al.*, 1984), the *p97* gene (Brown *et al.*, 1985), and the lactoferrin amino acid sequence (Metz-Boutigue *et al.*, 1984). The genes encoding *Tf*, the *TfR*, and *p97* have been mapped to the long arm of chromosome 3 by hybridization of their cDNAs with human chromosomal spreads (Yang *et al.*, 1984; Rabin *et al.*, 1985; Le Beau *et al.*, 1985). The chromosomal location of the gene encoding lactoferrin is unknown.

Park *et al.* (1985) characterized the major part of the *Tf* gene, 24 kbp, that coded for 70% of the protein. The 17 exons of human *Tf* agree closely in size with the exons found in the hen ovotransferrin gene (Jeltsch and Chambon, 1982). The human transferrin gene, however, is twice as long as the hen ovotransferrin gene because of the increased length of its introns. Furthermore, comparisons of the length and positions of the introns within the amino and carboxyl domains of the *Tf* gene were similar, indicating that the splicing patterns existed in the primordial gene before intragenic duplication occurred. The homology of the splicing pattern in the corresponding exon pairs, 3/10, 4/12, 5/13, and 6/14, led Park *et al.* (1985) to propose an unequal crossover event, shown in Fig. 14, to explain the intragenic duplication in the ancestral gene of human transferrin.

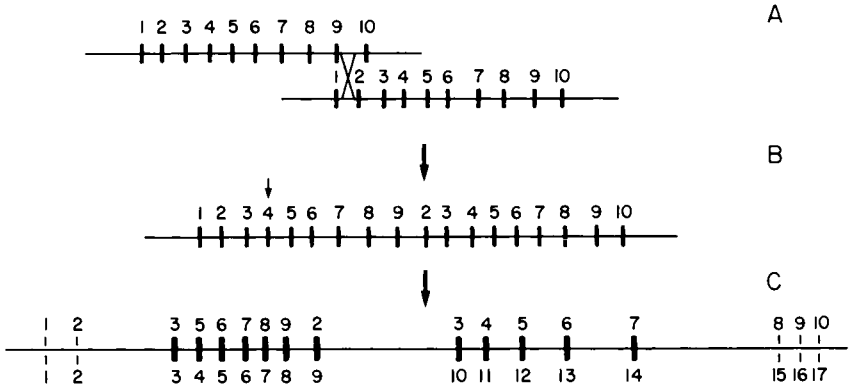


Fig. 14. Proposed scheme of the evolution of ovotransferrin and human transferrin genes (from Park *et al.*, 1985). The common ancestor of transferrin duplicated by intragenic crossing-over (A) leading to a duplicated ancestor that had lost one of its leader peptide coding exons and one of the terminal 3′ exons (B). During evolution, exon 4 (short arrow) was deleted on the 5′ side of the gene. In the duplicated ancestor, independent evolution led to different intron sizes for the ovotransferrin (B) and the human transferrin gene (C).

The homology of the N-terminal regions of p97, lactoferrin, and transferrin is shown in Fig. 15. The genomic sequences of human *Tf* and *p97* have emphasized the extensive homology in the two genes. Similar intron/exon patterns in the genomic DNAs of *Tf* and *p97* reflected striking evolutionary homology, which was confirmed when the 16 exons of *p97* were compared to the exons characterized in *Tf*. The *p97* gene, however, was missing exon 10 found in *Tf* but contained a new exon encoding a C-terminal region of 24 hydrophobic and uncharged amino acids that could serve to anchor p97 to the plasma membrane. The amino acid sequences deduced from the *p97* and *Tf* cDNAs reflect slightly less than 40 to 45% identity. Brown *et al.* (1985) observed strong internal homology in the genomic DNA encoding the amino and carboxyl domains of *p97*.

2. Gene Mapping

A. SOMATIC CELL HYBRID ANALYSIS. The transferrin gene was mapped to human chromosome 3 by hybridization analysis of radiolabeled *Tf* cDNA and

p97				Gly	Met	Glu	Val	Arg	Trp	Cys	Ala	Thr	Ser	Asp	?	Glu
Transferrin			Val	Pro	Asp	Lys	Thr	Val	Arg	Trp	Cys	Ala	Val	Ser	Glu	His	Glu
Lactotransferrin	Gly	Arg	Arg	Arg	Arg	Ser	Val	Gln	Trp	Cys	Ala	Val	Ser	Gly	Pro	Glu

Fig. 15. Homology of the amino-terminal sequences of p97, transferrin, and lactotransferrin (reprinted by permission from Brown *et al.*, 1982. Copyright © 1982, Macmillan Journals Limited.)

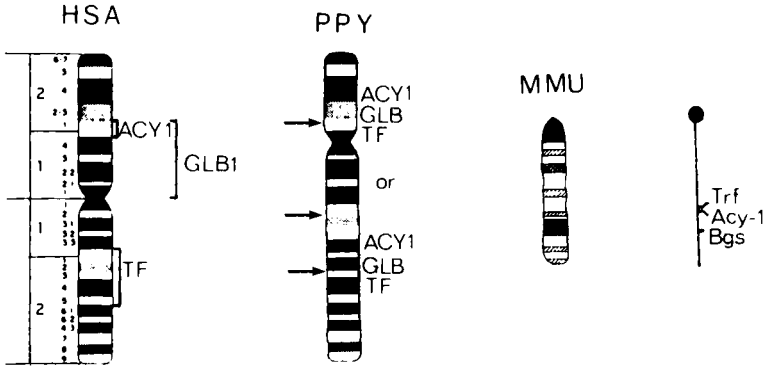


Fig. 16. Rearrangement of chromosome 3 in primates and mice (from Naylor and Sakaguchi, 1985). A rearrangement of chromosome 3 involving a pericentric inversion separated the *Tf* gene from other markers of 3q during evolution. HSA (*Homo sapiens*, PPY (*Pongo pygmaeus*, orangutan), MMU (*Mus musculus*, mouse). The last structure shows the linkage of genes on the mouse homolog of human chromosome 3. β-Galactosidase is GLB in primates, Bgs in mice. Aminoacylase-1 is ACY1 in primates, Acy-1 in mice. Transferrin is TF in primates, Trf in mice.

DNA of mouse–human hybrid cells containing defined complements of human chromosomes (Yang *et al.*, 1984). Utilization of human parental lines having specific rearrangements of chromosome 3 demonstrated that the *Tf* gene was on the long arm of 3, i.e., 3q. This was unexpected since mapping studies in mice had established linkage of mouse transferrin, aminoacylase (ACY1), and β-galactosidase (GLB) to chromosome 9, a homolog of human chromosome 3p (Naylor *et al.*, 1982); comparative studies have established that many linkage groups are conserved in widely divergent species. However, after evaluation of chromosome patterns of primates (deGrouchy *et al.*, 1978; Yunis and Prakash, 1982) it became clear that a rearrangement of chromosome 3 involving a pericentric inversion had separated the transferrin gene during primate evolution from the other markers of 3p. Humans, gorillas, and chimpanzees have the rearranged chromosome 3, while orangutans do not. The linkage group (*Tf*, *ACY1*, and *GLB*) affected by chromosome 3 inversion is shown in Fig. 16.

B. *IN SITU* HYBRIDIZATION. Hybridization of radiolabeled *Tf* cDNA with metaphase spreads of human chromosomes revealed that the *Tf* gene was located in the region of 3q21-25 (Yang *et al.*, 1984). Genes coding for the transferrin receptor, *TfR*, and the melanoma antigen, *p97*, are also located on the long arm of chromosome 3. The transferrin receptor had previously been mapped at 3q22-qter and 3q26.2, respectively (Miller *et al.*, 1983; Rabin *et al.*, 1985), while *p97* had been located on chromosome 3 by somatic cell hybrid analysis (Plowman *et al.*, 1983). Comparison of *in situ* hybridization patterns of chromosomes from a normal and a malignant cell line containing a rearranged chromosome 3 identi-

fied the chromosomal locations of these related proteins to be *Tf* on 3q21, *TfR* on 3q26, and *p97* within 3q28-q29 (Le Beau *et al.*, 1985). Comparison of *in situ* hybridization patterns of chromosomes from a normal and a malignant line associated with thrombocythemia and abnormal megakaryocytopoiesis indicates a break at 3q21 that splits the *Tf* gene, implicating a role of *Tf* in the pathogenesis of some human tumors.

C. OTHER MEMBERS OF THE *Tf* LINKAGE GROUP. The mapping of human *Tf* placed two other plasma proteins of the *Tf* linkage group, ceruloplasmin (Cp) and α_2-HS-glycoprotein, on chromosome 3 (Yang *et al.*, 1984). Linkage of *Tf* and *Cp* was found earlier by family studies that indicated there was a distance of 10 centimorgans between the *Tf* and *Cp* genes (Weitkamp, 1983). We have recently cloned human cDNA encoding *Cp* and confirmed its location on chromosome 3 (Bowman *et al.*, 1985; Naylor *et al.*, 1985). The cDNA encoding ceruloplasmin (*Cp*) was identified by screening a human liver library with two oligonucleotide probes based on amino acid sequences (538–543 and 666–672) reported by Takahashi *et al.* (1984). The *Cp* cDNA sequence is almost complete (Yang *et al.*, 1986) and confirms the similarity to factor VIII demonstrated by Vehar and coworkers (1984). A comparison of the genomic structures of these two genes will be valuable in analyzing their evolutionary relationship.

Therefore, the loci of several genes related in function, *Tf*, *TfR*, *p97*, and *Cp*, are linked on chromosome 3. Ceruloplasmin is said to serve as an oxidase that converts ferrous to ferric iron. Only the ferric form of iron can specifically be bound by transferrin. Functional and chromosomal relationships again raise a question of gene proximity and coordinated expression.

D. DNA POLYMORPHISMS DETECTED WITH *Tf* cDNA. Two inherited DNA polymorphisms have been detected by hybridization of *Tf* cDNA with endonuclease digests of human DNA (Yang *et al.*, 1985b). One, a 6.1-kbp fragment, was detected with *Eco*RI. This polymorphism is shown in Fig. 17. In a panel of DNAs digested with *Pvu*II, a second polymorphism, a 1.9-kbp fragment, was detected in 65% of the Caucasians and in 25% of the Hispanics examined. Schaeffer *et al.* (1985) examined the DNA from individuals having genetic variations of transferrin. Lymphocyte DNA from individuals having the TfD1 phenotype displayed DNA polymorphisms detected with *Eco*RI. The hybridizing probe was an insert of *Tf* cDNA that encoded amino acids 98 to the C-terminal end.

3. *Tf* Expression

A. *IN SITU* HYBRIDIZATION STUDIES DETECT *IN VIVO* EXPRESSION. The availability of *Tf* cDNA has facilitated microscopic identification of sites of transcription of *Tf* mRNA in human tissue by *in situ* hybridization (Lum, 1986; Lum *et*

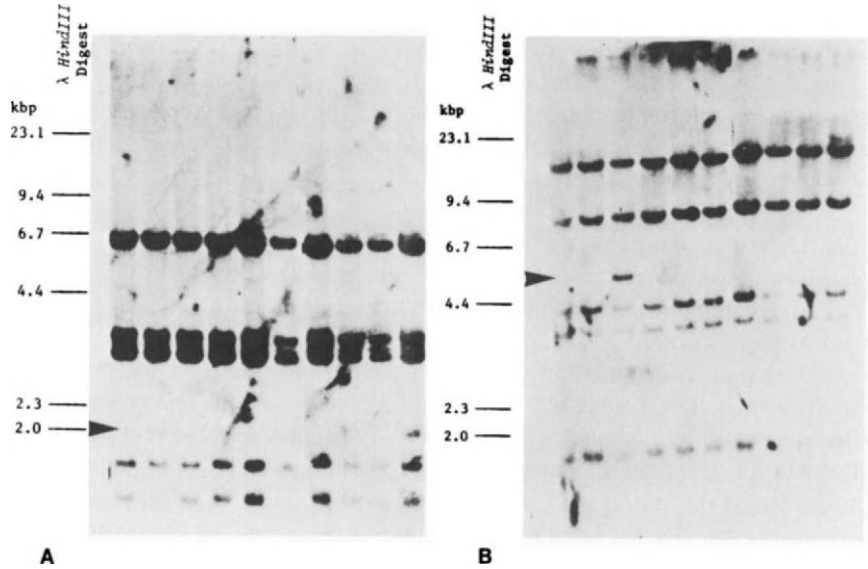


Fig. 17. Transferrin DNA polymorphisms detected in human lymphocyte DNAs digested with *Pvu*II (A) and *Eco*RI (B) (from Yang *et al.*, 1985b). Arrows indicate the polymorphic fragments.

al., 1985). Although the liver is the major site of transferrin synthesis, there had been reports (Soltys and Brody, 1970; Nishiya *et al.*, 1980; Broxmeyer *et al.*, 1983) suggesting that human peripheral blood mononuclear cells can also synthesize transferrin. Lum *et al.* (1985, 1986) localized *Tf* mRNA transcription in human lymphocytes, specifically in the $T4^+$ helper/inducer subset (Lum *et al.*, 1986). Application of the *in situ* hybridization technique in peripheral blood mononuclear cells is shown in Fig. 18. In peripheral blood mononuclear cells, approximately 2–5% were seen having silver grains denoting *Tf* mRNA hybridization (Fig. 18A). After T cells were isolated by rosette formation with sheep red blood cells, 10–20% of the T cells transcribed *Tf* mRNA (Fig. 18B). No *Tf* mRNA has been detected in B cells (Fig. 18C). The capacity to synthesize transferrin in a localized area such as a lymph node would impart a significant advantage to the vertebrate host during inflammation when Tf-bound iron is decreased in the circulation.

B. Expression of *Tf* by transgenic mice. Transgenic mice were utilized by McKnight *et al.* (1983) to follow the transferrin gene during development and differentiation. After the intact chicken transferrin (Ch *Tf*) gene was transplanted into fertilized mouse eggs, approximately 15–30% of the offspring were found to carry the Ch *Tf* DNA integrated into the mouse genome either as a single or as

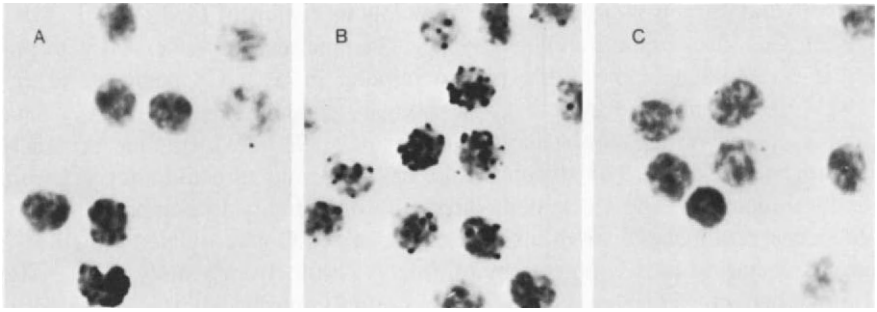


Fig. 18. Transcription of *Tf* mRNA in peripheral blood mononuclear cells (A) and T lymphocytes (B). B lymphocytes have not been observed to transcribe *Tf* mRNA (C) (from Lum *et al.*, 1985b).

multiple copies. Six of seven transgenic mice expressed the Ch *Tf* gene in several tissues; however, in five out of six there was a 5- to 10-fold preferential expression of Ch *Tf* mRNA in liver compared to other tissues. More than 80% of Ch *Tf* mRNA transcribed in mouse liver was found to be actively engaged in protein synthesis; Ch *Tf* was also detected circulating in mouse serum. The most exciting implication of this study is that the tissue specificity may be encoded in the *Tf* gene sequence itself.

C. Human Group-Specific Component

The group-specific component (Gc), an α_2-globulin of 51,000 molecular weight (Cleve and Bearn, 1962), is the major vitamin D-binding protein in serum (Daiger *et al.*, 1975). Gc appears in human populations as three common genetic phenotypes: Gc1, Gc2, and Gc2-1. With the application of isoelectric focusing (Constans and Viau, 1977b) many additional genetic subtypes have been reported (Cleve and Patutschnick, 1977).

1. Genetic and Posttranslational Variations

Gc1 migrates electrophoretically as two bands because of a posttranslational event involving sialic acid (Svasti and Bowman, 1978; Svasti *et al.*, 1979). Binding of vitamin D_3 by Gc is accompanied by a decrease in isoelectric point and an increase in electrophoretic mobility at alkaline pH, indicating a conformational change (Svasti and Bowman, 1978). Gc has also been reported to bind G-actin (Van Baelen *et al.*, 1980) and to be spatially associated with IgG on lymphocyte membranes (Petrini *et al.*, 1985).

Amino acid differences in the sequences of Gc1 and Gc2 include a Thr residue in Gc1 that is glycosylated and carries a sialic acid group; the threonine-linked

carbohydrate group is absent in the homologous region of Gc2 (Svasti *et al.*, 1979). Gc1 also contains an *N*-acetylgalactosamine residue associated with this same carbohydrate moiety that is also missing in Gc2 (Coppenhaver *et al.*, 1983). In the inherited disease cystic fibrosis, patients characteristically have mucins of abnormal consistency (Allen *et al.*, 1980). Mucin, an extremely heterogeneous mixture of glycoproteins, has numerous threonine-linked carbohydrate moieties. The Gc system offers the opportunity to compare the concentration and metabolism of a single characterized O-glycosylated protein in a group of cystic fibrosis subjects and normals. Results from a study of 90 cystic fibrosis patients, 57 heterozygotes, and 46 normal controls indicated very significant lower mean concentrations of Gc1 and Gc2-1 in the sera of cystic fibrosis patients compared with both heterozygotes and controls (Coppenhaver *et al.*, 1981). In contrast, other plasma proteins appeared to be unchanged or elevated in concentration. The diminished concentrations of Gc1 and Gc2-1 in cystic fibrosis patients may be due to an impaired step in posttranslational processing of O-glycosylated proteins.

When the common Gc genetic types Gc1 and Gc2 were examined by fingerprinting, tryptic peptide maps of the Gc1 and Gc2 differed by two peptides related to structural differences involving at least three amino acid residues (Svasti *et al.*, 1979). Analysis of cDNAs have accurately provided the number of genetic differences in the two common *Gc* genes (Yang *et al.*, 1985b; Cooke and David, 1985).

The identity of *Gc2* cDNA in a human liver library followed by characterization of its sequence has recently been completed (Yang *et al.*, 1985b). One clone, Gc2p, contained an 1805-bp nucleotide insert including the entire Gc2 coding sequence. It consisted of a 5′ untranslated region of 153 bp, a 48-bp sequence encoding a putative 16-amino acid leader sequence, a 1374-bp sequence encoding the 458 amino acids of the mature Gc protein, and 230 bp in a 3′ untranslated region. Insert Gc2p encodes 458 amino acids in the mature protein providing a minimum molecular weight of 51,240. Cooke and David (1985) have characterized human *Gc1* cDNA; a comparison of this sequence with *Gc2* cDNA (Yang *et al.*, 1985a) reveals four amino acid differences in amino acid residues 152 (Glu in Gc1, Gly in Gc2), 311 (Arg in Gc1, Glu in Gc2), 416 (Glu in Gc1, Asp in Gc2), and 420 (Thr in Gc1, Lys in Gc2). The last two substitutions confirmed previous results of Svasti *et al.* (1979), who sequenced tryptic peptides characteristic of Gc1 and Gc2. The deduced amino acid analysis demonstrated homology with albumin and α-fetoprotein and an intragenic triplication within the *Gc* gene.

2. *Intragenic Triplication*

In 1976 Brown analyzed the complete albumin amino acid sequence and demonstrated that this protein evolved by intragenic triplication of a small pri-

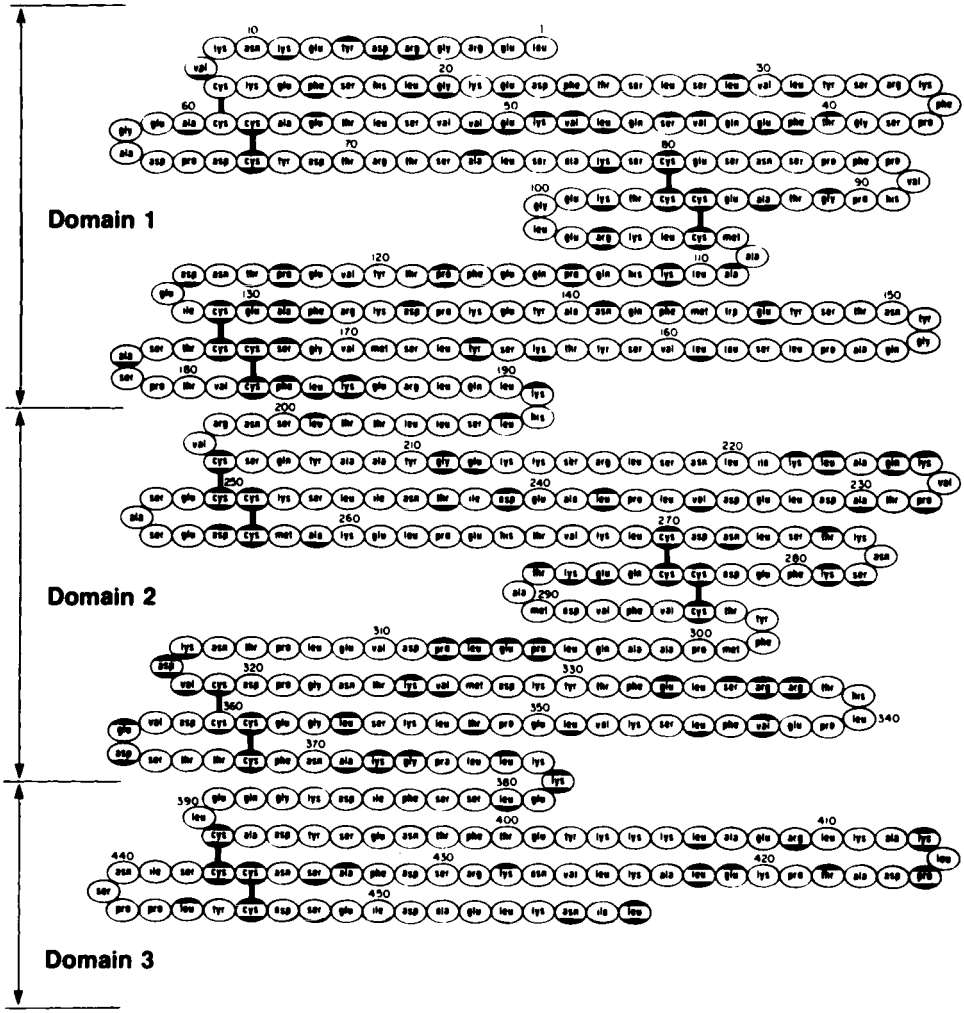


Fig. 19. Triple-domain structure of mature human Gc arranged according to the albumin model proposed by Brown (1976) (from Yang *et al.*, 1985a). Amino acids homologous to those of mature human α-fetoprotein and albumin are indicated by shading at the top and bottom, respectively, of the ovals representing the residues.

mordial gene. The albumin sequence analysis demonstrated that the double disulfide bonds generated a pattern of loops that was repeated threefold and defined three structural domains. The original gene coded for a protein corresponding to one domain of the vertebrate albumin gene. The Gc sequence is arranged in Fig. 19 in the same configuration as the albumin model proposed by Brown (1976).

The pattern of half-cystine bridges that contribute to the double loops forming

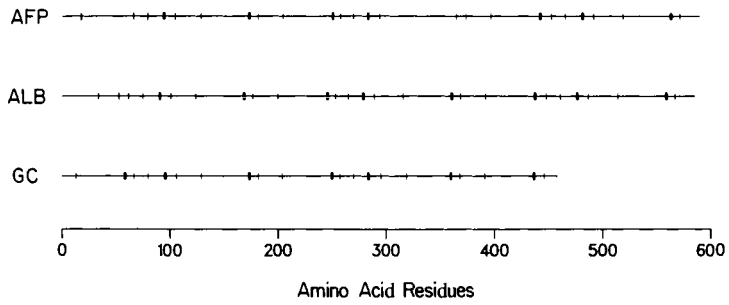


Fig. 20. Conserved locations of the disulfide and double disulfide bonds in sequences of Gc, α-fetoprotein, and albumin (from Yang *et al.*, 1985a). Thin vertical bars represent Cys residues and thick vertical bars represent Cys-Cys sequences.

the three domains of Gc, albumin, and α-fetoprotein is highly conserved (Fig. 20). Three homologous domains in each of the three proteins are defined by the double disulfide bonds. When the structure deduced for Gc was compared to the triple domain structures of albumin and α-fetoprotein, it was observed that Gc had lost two-thirds of the coding sequence of domain 3. Therefore, in Gc's third domain, only one of the three large double loops that usually comprise a typical domain was retained (Fig. 21). The last, incomplete domain is also the site for at least part of the structural changes that account for the common Gc alleles, Gc1 and Gc2 (Yang *et al.*, 1985a).

A comparison of the amino acid sequences of the three domains of Gc confirms that Gc, like albumin and α-fetoprotein, is the evolutionary product of intragenic triplication, with 23% identity between domains 1 and 2 and 19% identity between 2 and 3. Intragenic homology of Gc was also analyzed using the ALIGN program. Domain 1 (residues 1–192) aligned with domain 2 (residues 193–378) produced a score of 11.61 SD units. An alignment score of >3 SD units is considered statistically significant (Barker and Dayhoff, 1982). Portions of domains 1 (residues 1–79) and 2 (residues 193–269) were compared to Gc's partial domain 3 (residues 379–458) to give scores of 5.25 SD and 6.18 SD units, respectively. Alignment scores for the domains indicated a high probability of a common ancestor (Yang *et al.*, 1985c).

Interesting evidence for the order of evolutionary events leading to the three-domain structure of mouse α-fetoprotein has been reported by Eiferman *et al.* (1981). These investigators found a direct correspondence between the threefold repeat of four exons and the three protein domains when the 14 exon junctions in the α-fetoprotein gene were determined using cloned genomic DNA. The nucleotide sequence was compared among the four exons of each domain to deduce the likely structure of the primordial domain and the order and mechanisms of its

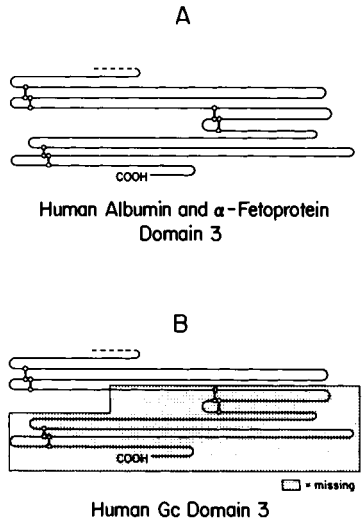


Fig. 21. The third domain of albumin and α-fetoprotein (A) compared to that of Gc (B). Gc has lost two-thirds of the coding sequence of domain 3 and only one of the three large double disulfide loops was retained.

triplication to form the ancestral gene from which the proteins in the albumin family arose. The predicted order of events are domain 3 duplicated by an unequal crossover to a two-domain gene, followed by a subsequent homologous recombination that gave rise to domains 1 and 2. This hypothesis is supported by close similarity of domains 1 and 2 in all three proteins of the albumin family.

When the cDNA sequences of the three human Gc domains were compared by statistical analysis to identify regions of extensive internal nucleotide sequence identity, three paired sequences, designated as homology blocks A, B, and C, reflected 40–44% identity in nucleotide sequence and 22–27% identity in amino acid sequence (Yang *et al.*, 1985c). The basis of the evolutionary constraints on the codons contained in the three homology blocks is unknown. Future studies that characterize the vitamin D binding site may clarify the basis for the conserved regions.

3. Homology of Gc, Albumin, and α-Fetoprotein

The question of evolutionary homology in Gc and albumin was raised earlier because of their genetic linkage (Weitkamp *et al.*, 1966), the low content of carbohydrate, and the unusually large number of half-cystinyl residues in each (Bowman, 1969). The amino acid sequence deduced from the sequence of the *Gc* cDNA confirmed the homology of the three proteins (Yang *et al.*, 1985a,c). The

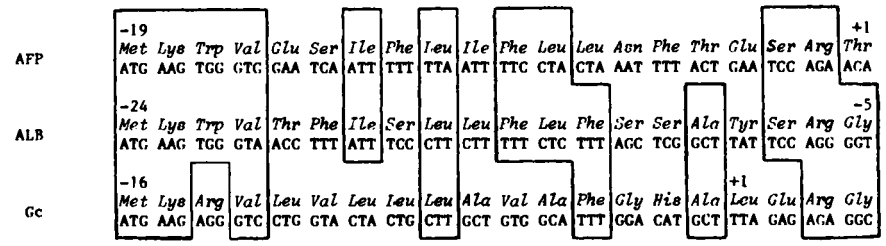


Fig. 22. Homology in the nucleotide and amino acids of the leader sequences of α-fetoprotein, albumin, and Gc (from Yang *et al.*, 1985c). Identical amino acid residues are boxed. There is 40% identity in amino acid sequences of Gc and albumin and 48% identity in respect to the nucleotide sequence in this region.

most conspicuous similarities in Gc, albumin, and α-fetoprotein were the conserved patterns of disulfide bridges forming the nine double loops that define the three structural domains (Fig. 20).

The leader and N-terminal sequences of Gc, albumin, and α-fetoprotein were found to be highly similar irrespective of where the N-terminal sites occurred in the three homologous proteins. Figure 22 compares the leader and N-terminal regions of the three proteins. There is 40% identity in amino acid sequences of Gc and albumin regions and 48% identity in respect to the nucleotide sequence. In an inherited albumin variant called proalbumin Christchurch, the altered prosequence Arg-Gly-Val-Phe-Arg-Gln is not removed before the nascent protein is released into the circulation (Brennan and Carrell, 1978). The amino terminus of the Gc sequence begins two residues earlier than proalbumin Christchurch but shares its Arg-Gly in the third and fourth residues.

It is of interest that one Cys-Cys sequence in position 58–59 in the first domain of Gc was not present in albumin or α-fetoprotein and would be available to form a disulfide bond with residue 13 and a disulfide bond with residue 67. This would correspond to the hypothetical disulfide bridge in the first subdomain of the albumin ancestor model proposed by Brown (1976). In this region albumin contains one disulfide bond, residues 53 and 62, and is missing the other ancestral disulfide bridge, whereas α-fetoprotein contains no disulfide bonds in its first subdomain. A comparison of the disulfide patterns of proteins in the albumin family is shown in Fig. 23.

The homology existing in the amino acid sequences of Gc, albumin, and α-fetoprotein is presented in Fig. 24. After the amino acid sequences are aligned according to the double disulfide bonds, 11 gaps in Gc's sequence and 6 gaps in the α-fetoprotein and albumin amino acid sequences were introduced to maximize homology. Gc and albumin were 23% identical; Gc and α-fetoprotein were 19% identical. In this alignment, albumin and α-fetoprotein were 35% identical. Using the program ALIGN, the amino acid sequence of Gc, including the leader

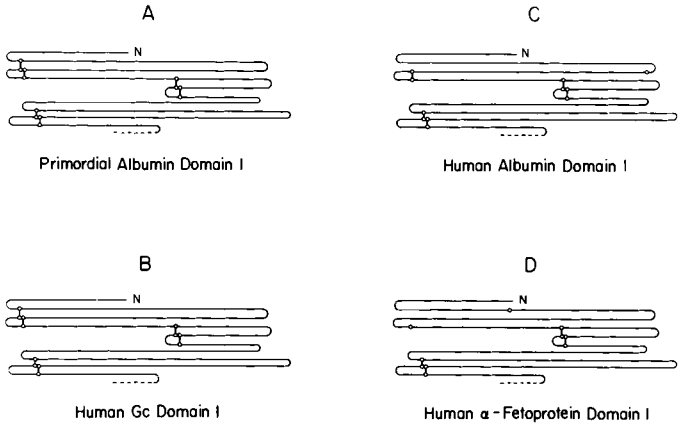


Fig. 23. Comparison of the double disulfide loops in the first domain of the albumin ancestor model proposed by Brown (1976) (A), of Gc (B), of albumin (C), and of α-fetoprotein (D). The one Cys-Cys sequence in position 58–59 in the first domain of Gc is available to form a disulfide bridge with residue 13 and a disulfide bridge with residue 67. This Cys-Cys sequence is not present in albumin or α-fetoprotein.

sequence, was aligned to portions of α-fetoprotein and albumin of comparable length. Residues −16 to 458 of Gc were aligned with residues −19 to 466 of α-fetoprotein to give an alignment score of 19.59 SD units and aligned with residues −18 to 461 of albumin for a score of 24.87 SD units (Yang *et al.*, 1985c). In general, an alignment score of >3 units may be considered statistically significant (Barker and Dayhoff, 1982). Thus the Gc, α-fetoprotein, and albumin alignments not only exceed the requirement for significance but also meet the definition proposed by Doolittle (1981) for grouping of protein families. Gc2 and α-fetoprotein each have one potential N-glycosylation site in their second domains while albumin has none (Dugaiczyk *et al.*, 1982). The potential glycosylation site in Gc2 occurs at residues 272–274 (Asn-Leu-Ser) (Yang *et al.*, 1985a). Previous studies (Svasti *et al.*, 1979) demonstrated that Gc1, but not Gc2, had an O-glycosylation site within the sequence of residues 412–424, probably at the Thr residue at position 420. Residue 420 is occupied by Lys in Gc2 (Yang *et al.*, 1985a). Viau *et al.* (1983) also reported glycosylation in Gc1, but not in Gc2.

The relative age of the *Gc* gene is difficult to estimate. Homology data indicated that albumin and α-fetoprotein evolved from an ancestral gene that was produced by an intragenic triplication 300–500 million years ago (Eiferman *et al.*, 1981). Two lines of evidence suggest that the *Gc* gene is older than the other two members of the albumin family. Albumin (Alb) and α-fetoprotein are more similar to each other than either is to Gc, indicating that the gene duplications

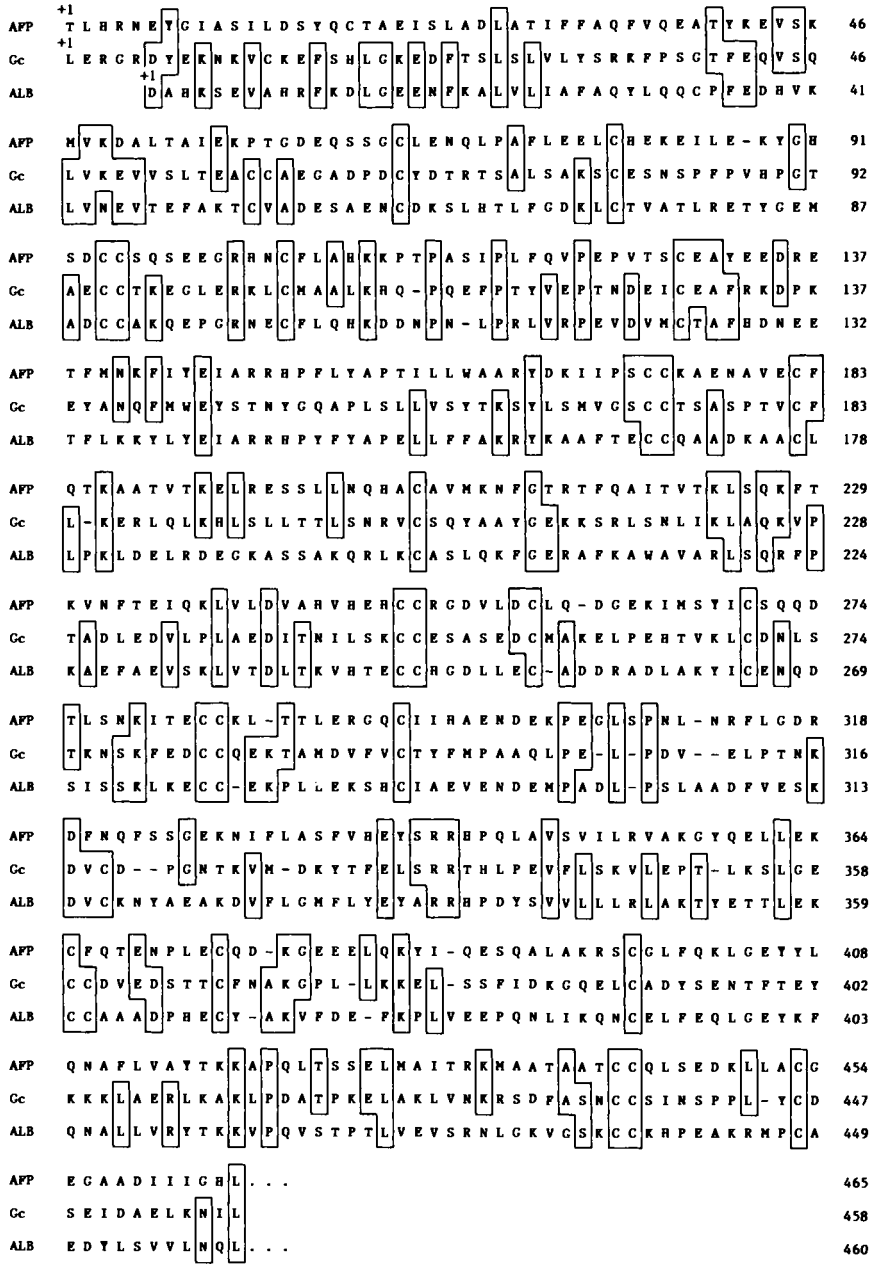


Fig. 24. Homology of amino acids in sequences of human α-fetoprotein, Gc, and albumin (from Yang *et al.*, 1985c). Numbers to the right of the figure indicate amino acid residue positions. Identical residues are boxed.

producing them may have occurred more recently. Furthermore, Gc, alone, contains four half-cystine residues at positions 13, 58, 59, and 67 that are capable of forming the double disulfide bond predicted by Brown (1976) to have occurred in the first domain of the triplicated ancestral gene. Whether or not the partial loss of Gc's third domain is a recent evolutionary event must await further studies of the *Gc* gene in other species.

4. Gene Mapping and Linkage

After analyzing segregation patterns in informative families, Weitkamp *et al.* (1966) discovered that the human *Gc* and *Alb* genes belonged to the same linkage group with a recombination fraction of 0.015. Therefore, the loci for *Gc* and *Alb* are approximately 1.5 centimorgans apart. Studies of members of a family having a deletion in chromosome 4 and incompatibility within the *Gc* allelic system demonstrated that the *Gc* gene was on the long arm of chromosome 4 at p11-p13 (Mikkelsen *et al.*, 1977). Utilization of *Gc* cDNA has confirmed its locus to chromosome 4 by somatic hybrid cell analysis (Yang *et al.*, 1985a; Naylor and Sakaguchi, 1985) (Fig. 2). *In situ* hybridization with radiolabeled *Gc* cDNA on human chromosomal spreads mapped *Gc* close to the centromere, at 4q13 (McCombs *et al.*, 1985).

The subregional mapping of albumin and α-fetoprotein was resolved by hybridization studies with DNA probes. The DNAs encoding human albumin and α-fetoprotein were used simultaneously by Harper and Dugaiczyk (1983) for hybridization with human metaphase chromosomal spreads; their results confirmed that these two genes were closely linked and both were located between 4q11-q13. Urano *et al.* (1984) demonstrated that the human genes encoding albumin and α-fetoprotein are present in tandem, in the same transcriptional orientation, with the albumin gene 14.5 kbp upstream from the α-fetoprotein gene. In the mouse the α-fetoprotein and albumin genes are also linked (D'Eustachio *et al.*, 1981); genes belonging to the same linkage group are often found segregating together in widely separated species.

Utilization of the *Gc* cDNA probe has also detected two DNA polymorphisms in panels of human lymphocyte DNA. The first polymorphism was detected with human lymphocyte DNA digested with *Bam*HI endonuclease. Two fragments, a 16-kbp and a 14-kbp fragment, were observed in 20% of the individuals examined (lanes 3, 4, and 6 of Fig. 25A), while the remaining 80% had a darkly stained 30-kbp fragment (lanes 1, 2, and 5 of Fig. 25A). The other polymorphism, a 12-kbp fragment detected with *Pvu*II, is shown in lane 6 of Fig. 25B (Yang *et al.*, 1985c). The detection of DNA polymorphisms with *Gc* cDNA may prove useful in identifying the defective gene causing dentinogenesis imperfecta, an autosomal dominant disease mapped on chromosome 4, approximately 7 centimorgans from the *Gc* gene (Ball *et al.*, 1982).

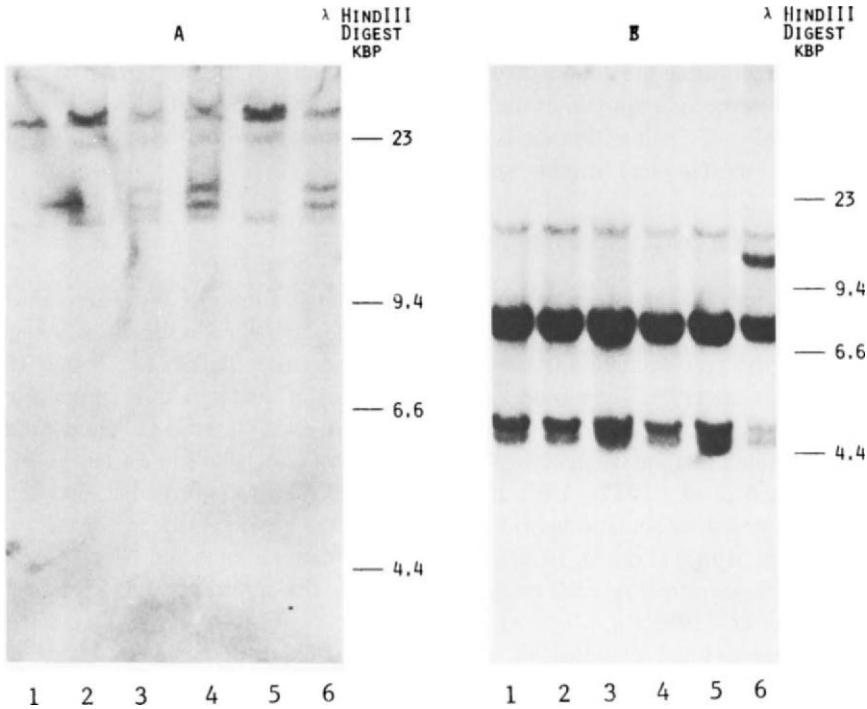


Fig. 25. Filter hybridization of human *Gc2* cDNA with human lymphocyte DNA (from Yang *et al.*, 1985c). Two DNA polymorphisms were detected in panels of human DNA digested with (A) *Bam*HI, where 20% of individuals tested had both 16-kbp and 14-kbp fragments (lanes 3, 4, and 6) and the remainder had a darkly stained 30-kbp fragment (lanes 1, 2, and 5), and (B) *Pvu*II, where one individual in 16 was found to have a 12-kbp fragment (lane 6).

Murray *et al.* (1984) have identified eight polymorphic sites in the human albumin locus. Two haplotypes were found in three races, Caucasoids, Blacks, and Asians, indicating that the DNA polymorphisms contributing to the haplotype predated human racial divergence. The origin of one haplotype predated human–African ape divergence. The smaller than expected haplotypic arrangements of the albumin DNA polymorphisms led Murray to suggest that the limited number of haplotypes at the chromosomal site of the albumin gene may be the result of decreased recombination due to the proximity of the centromere of chromosome 4 since centromeres are regions known to have low recombination frequency.

5. Expression of Genes in the Albumin Family

There have been no cases found in man where Gc is absent from the circulating plasma proteins. Inherited deficiencies of a wide range of plasma proteins

have been discovered; however, a severe deficiency of Gc appears to be incompatible with life.

Both humans and rats have been described that lack detectable concentrations of serum albumin because of a mutation affecting albumin synthesis or maturation. In analbuminemic rats the albumin gene is missing an intronic 7-bp sequence which must be required for the maturation of albumin mRNA (Esumi *et al.*, 1983). As a result of this mutation, albumin mRNA precursors were present in the nuclei of rat liver but albumin mRNA was absent in the cytoplasm of the same cells. Avery *et al.* (1983) analyzed the serum albumin gene from a human with analbuminemia. No gross structural rearrangement or deletion was found; however, the authors pointed out the mutation may result from an abnormality of the gene's fine structure, as was found in analbuminemic rats.

Experiments utilizing *in situ* hybridization with a radiolabeled *Gc* cDNA probe have confirmed that *Gc* mRNA transcription occurs in the liver; experiments are in progress to detect additional sites of transcription (Lum *et al.*, 1986). Gc has not yet been expressed *in vitro*. Expression studies with the albumin and α-fetoprotein genes, however, have provided interesting information about development and regulation.

Scott and Tilghman (1983) studied the expression of a mouse α-fetoprotein minigene in HeLa cells and observed the effects of deletions in the 5′ promotor region. In their studies transcription was unaffected by deletion of DNA upstream of the TATA box but was greatly affected by the distance between the viral control region and the 5′ end of the gene. Lawn *et al.* (1981) reported synthesis of human serum albumin by bacteria containing the mature protein coding sequence of the human albumin gene.

Krumlauf *et al.* (1985) introduced modified α-fetoprotein genes into fertilized mouse eggs and observed the time and tissue sites of α-fetoprotein expression in the transgenic mice. Their studies suggested that the microinjected genes contain the DNA sequence necessary to direct both tissue-specific expression and developmental regulation. This furnished additional information indicating that elements within the DNA sequence governed both tissue-specific activation and postnatal developmental regulation of the gene.

IV. Conclusions

Characterization of human plasma proteins has provided valuable information about the existence of families of related genes and their respective receptors, some of which have existed together on the same chromosome for millions of years and others which have been separated by chromosomal translocations and inversions. The extensive DNA polymorphisms detected by cDNA encoding plasma proteins will add to the molecular arsenal being employed to map the 3.3×10^6 kbp of DNA in the human genome. Polymorphisms detected by the

plasma protein probes will continue to be useful for prenatal diagnosis and chromosomal mapping of defective genes causing inherited diseases.

The haptoglobin, transferrin, and albumin families of genes are products of duplication and triplication events that occurred in evolution millions of years ago. The intragenic amplifications have refined and added to the functional sites of the protein products. For example, transferrin in vertebrates can bind two iron ions because of an intragenic duplication. The prochordate *Pyura,* on the other hand, has a primordial *Tf* gene that produces a mono-sited iron-binding polypeptide (Martin *et al.*, 1984). A few new, evolutionary experiments can be observed in contemporary organisms; crabs carry a triplicated transferrin gene that produces a protein able to bind only two, not three, ions of iron, despite its greater size (Huebers *et al.*, 1982). A few humans have the Hp^3 gene that produces a haptoglobin type with a triplicated α chain. Albumin, α-fetoprotein, and Gc are all triplicated structures. Yet no species has been identified that still carries single-domain proteins related to the albumin family.

Plasma proteins illustrate the results of "molecular tinkering" (Jacob, 1983) in which bits and pieces of several genes are melded together to produce genes encoding proteins with novel structures and diverse functions. The recent discovery of a virus element in the first intron of the haptoglobin-related (*Hpr*) gene may have important implications in regulation of plasma protein expression and may encourage extensive sequence analysis of intervening sequences within the genomes of other acute-phase reactants (Maeda, 1985). The likely origin of the β domain of the haptoglobin gene from a reverse transcript of chymotrypsinogen mRNA points out another evolutionary mechanism, reverse transcription, that has been utilized in molecular tinkering (Bowman *et al.*, 1985).

The human plasma protein genes offer excellent models for studying gene expression during ontogeny, acute-phase reaction, and malignancy. The exciting discoveries made after analyzing the tissue-specific expression and developmental regulation of injected albumin and transferrin genes in transgenic mice (Krumlauf *et al.*, 1985; McKnight *et al.*, 1983) promise direct means by which regulation, development, and expression can be observed. New understanding of the regulatory signals that specify tissue-specific expression and developmental variations is likely to be found by identifying and characterizing the genes of plasma proteins in different cell types throughout development.

Acknowledgments

The research in the laboratories of Drs. Bowman and Yang was supported in part by NIH Grants AM34130, HD16584, and GM33298, American Cancer Society Grant NP-470, and a Grant-in-Aid from the American Heart Association, with funds contributed by the Texas Affiliate. We are grateful to Betty Russell and Judith Pride for preparation of the manuscript.

References

Aisen, P., and Listowsky, I. (1980). *Annu. Rev. Biochem.* **49,** 357–393.

Allen, A., Mantle, M., and Pearson, J. P. (1980). *In* "Perspectives in Cystic Fibrosis" (J. M. Sturgess, ed.), pp. 102–112. Imperial Press, Mississauga, Canada.

Allen, F. M. (1974). *Vox Sang.* **27,** 382–384.

Alper, C. A. (1981). *In* "The Role of the Major Histocompatibility Complex in Immunobiology" (M. E. Dorf, ed.), pp. 173–220. Garland, New York.

Alper, C. A., Marcus, D., Raum, D., Peterson, B. H., and Spira, T. J. (1983). *J. Clin. Invest.* **72,** 1526–1531.

Anderson, M. L. M., Szajnert, M. F., Kaplan, J.-C., McColl, L., and Young, B. D. (1984). *Nucleic Acids Res.* **12,** 6647–6661.

Avery, R. A., Alpert, E., Weigard, K., and Dugaiczyk, A. (1983). *Biochem. Biophys. Res. Commun.* **116,** 817–821.

Ball, S. P., Cook, P. J. L., Mars, M., and Buckton, D. E. (1982). *Ann. Hum. Genet.* **46,** 35–40.

Barker, W. C., and Dayhoff, M. O. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 2836–2839.

Barnett, D. R., Lee, T.-H., and Bowman, B. H. (1970). *Nature (London)* **225,** 938–939.

Bell, G. I., Quinto, C., Quiroga, M., Valenzuela, P., Craik, C. S., and Rutter, W. J. (1984). *J. Biol. Chem.* **259,** 14265–14270.

Bensi, G., Raugei, G., Klefenz, H., and Cortese, R. (1985). *EMBO J.* **4,** 119–126.

Bishop, J. M. (1983). *Annu. Rev. Biochem.* **52,** 301–354.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). *Am. J. Hum. Genet.* **32,** 314–331.

Bowman, B. H. (1969). *Biochemistry* **8,** 4327–4335.

Bowman, B. H., and Kurosky, A. (1982). *Adv. Hum. Genet.* **12,** 189–261.

Bowman, B. H., Yang, F., Brune, J. L., Naylor, S. L., Barnett, D. R., McGill, J. R., Moore, C. M., Lum, J. B., and McCombs, J. (1985). *Protides Biol. Fluids* **33,** 15–20.

Boyd, D., Jain, S. K., Crampton, J., Barrett, K. J., and Drysdale, J. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 4751–4755.

Brennan, S. O., and Carrell, R. W. (1978). *Nature (London)* **274,** 908–909.

Brown, J. P., Hewick, R. M., Hellström, I., Hellström, K. E., Doolittle, R. F., and Dreyer, W. J. (1982). *Nature (London)* **296,** 171–173.

Brown, J. P., Rose, T. M., and Plowman, G. D. (1985). *In* "Proteins of Iron Storage and Transport" (G. Spik, J. Montreuil, R. R. Crichton, and J. Mazurier, eds.), pp. 39–46. Elsevier, Amsterdam.

Brown, J. R. (1976). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **35,** 2141–2144.

Broxmeyer, H. E., Lu, L., and Bognacki, J. (1983). *Blood* **62,** 37–50.

Brune, J. L., Yang, F., Barnett, D. R., and Bowman, B. H. (1984). *Nucleic Acids Res.* **12,** 4531–4538.

Camerino, G., Grzeschik, K. H., Jaye, M., De La Salle, H., Tolstoshev, P., Lecocq, J. P., Heilig, R., and Mandell, J. L. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 498–502.

Carroll, M. C., Campbell, R. D., Bentley, D. R., and Porter, R. R. (1984). *Nature (London)* **307,** 237–241.

Carson, S. D., Henry, W. M., and Shows, T. B. (1985). *Science* **229,** 991–993.

Caskey, C. T., and White, R. L., eds. (1983). "Banbury Report," No. 14. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Caskey, J. H., Jones, C., Miller, Y. E., and Seligman, P. A. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 482–486.

Cheung, P., Kao, F.-T., Law, M. L., Jones, C., Puck, T. T., and Chan, L. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 508–511.

Chow, V., Murray, R. K., Dixon, J. D., and Kurosky, A. (1983). *FEBS Lett.* **153,** 275–279.
Cleve, H., and Bearn, A. (1962). *Prog. Med. Genet.* **2,** 64–82.
Cleve, H., and Patutschnick, W. (1977). *Hum. Genet.* **38,** 289–296.
Constans, J., and Viau, M. (1977a). *Am. J. Hum. Genet.* **29,** 280–286.
Constans, J., and Viau, M. (1977b). *Science* **198,** 1070–1071.
Cooke, N. E., and David, E. V. (1985). *J. Clin. Invest.* **76,** 2420–2424.
Coppenhaver, D. H., Kueppers, F., Schidlow, D., Bee, D., Isenburg, J. N., Barnett, D. R., and Bowman, B. H. (1981). *Hum. Genet.* **57,** 399–403.
Coppenhaver, D. H., Sollenne, N. P., and Bowman, B. H. (1983). *Arch. Biochem. Biophys.* **226,** 218–223.
Cox, D. W., and Francke, U. (1985). *Hum. Genet.* **70,** 109–115.
Craik, C. S., Choo, Q.-L., Swift, G. H., Quinto, C., MacDonald, R. J., and Rutter, W. J. (1984). *J. Biol. Chem.* **259,** 14255–14264.
Croce, C. M., Shander, M., Martinis, J., Cicurel, L., D'Ancona, G. G., Dolby, T. W., and Koprowski, H. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 3416–3419.
Daiger, S. P., Schanfield, M. S., and Cavalli-Sforza, L. L. (1975). *Proc. Natl. Acad. Sci. U.S.A.* **72,** 2076–2080.
Daiger, S. P., Rummel, D. P., Wang, L., and Cavalli-Sforza, L. L. (1982). *Am. J. Hum. Genet.* **33,** 640–684.
Darlington, G. J., Astrin, K. H., Muirhead, S. P., Desnick, R. J., and Smith, M. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 870–873.
Dautry-Varsat, A., Ciechanover, A., and Lodish, H. F. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 2258–2262.
deGrouchy, J., Turleau, C., and Finaz, C. (1978). *Annu. Rev. Genet.* **17,** 289–328.
D'Eustachio, P., Ingram, R. S., Tilghman, S. M., and Ruddle, F. H. (1981). *Somatic Cell Genet.* **7,** 289–294.
Dixon, G. H. (1966). *Essays Biochem.* **2,** 147–204.
Doolittle, R. F. (1981). *Science* **214,** 149–159.
Doolittle, R. F. (1985). *Trends Biochem. Sci.* **10,** 233–237.
Dugaiczyk, A., Law, S. W., and Dennison, O. E. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 71–75.
Eiferman, F. A., Young, P. R., Scott, R. W., and Tilghman, S. M. (1981). *Nature (London)* **294,** 713–718.
Erikson, J., Martinis, J., and Croce, C. M. (1981). *Nature (London)* **294,** 173–175.
Esumi, H., Takahashi, Y., Sato, S., Nagase, S., and Sugimura, T. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 95–99
Ferguson-Smith, M. A., and Aitken, D. A. (1978). *Cytogenet. Cell Genet.* **22,** 513.
Francke, U., Brown, M. S., and Goldstein, J. L. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 2826–2830.
Fuller, G. M., Rasco, M. A., McCombs, M. L., Barnett, D. R., and Bowman, B. H. (1973). *Biochemistry* **12,** 253–258.
Giblett, E. R. (1964). *Cold Spring Harbor Symp. Quant. Biol.* **29,** 321–326.
Ginsburg, D., Handin, P. J., Bonthron, D. T., Donlon, T. A., Bruns, G. A. P., Latt, S. A., and Orkin, S. H. (1985). *Science* **228,** 1401–1406.
Gitshier, J., Dragna, D., Toddenham, E. G. D., White, R. L., and Lawn, R. M. (1985). *Nature (London)* **314,** 738–740.
Goodfellow, P. N., Jones, E. A., van Heyningen, V., Solomon, E., Bobrow, M., Miggiano, V., and Bodmer, W. F. (1975). *Nature (London)* **254,** 267–269.
Goodfellow, P. N., Banting, G., Sutherland, R., Greaves, M., Solomon, E., and Povey, S. (1982). *Somatic Cell Genet.* **8,** 197–206.
Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E.,

Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B., Shoulson, J., Bonilla, E., and Martin, J. B. (1983). *Nature (London)* **306,** 234–238.

Hanley, J. M., and Heath, E. C. (1985). *Arch. Biochem. Biophys.* **239,** 404–419.

Harper, M. E., and Dugaiczyk, A. (1983). *Am. J. Hum. Genet.* **35,** 565–572.

Harper, M. E., and Saunders, G. F. (1981). *Chromosoma* **83,** 431–439.

Haugen, T. H., Hanley, J. M., and Heath, E. C. (1981). *J. Biol. Chem.* **256,** 1055–1057.

Hecht, F., Blaine, T., Magenes, R. E., Kimberling, W. J., Wyandt, H., and Lovrien, E. W. (1971). *Nature (London)* **233,** 480.

Henry, J., Wzan, G., Weil, D., Nicolas, H., Kaplan, J. C., Marquerie, C., Kahn, A., and Junien, C. (1984). *Am. J. Hum. Genet.* **36,** 760–768.

Hill, A. V. S., Bowden, D. K., Flint, J., Whitehouse, D. B., Hopkinson, D. A., Oppenheimer, S. J., Serjeantson, S. W., and Clegg, J. B. (1985). *Am. J. Hum. Genet.* **38,** 382–389.

Hobart, M. J., Rabbits, T. H., Goodfellow, P. N., Solomon, E., Chambers, S., Spurr, N., and Povey, S. (1981). *Ann. Hum. Genet.* **45,** 331–335.

Honey, N. K., Sakaguchi, A. Y., Quinto, C., McDonald, R. J., Bell, G. I., Craik, C., Rutter, W. J., and Naylor, S. L. (1984). *Somatic Cell Genet.* **10,** 369–376.

Huebers, H. A., Csiba, E., Josephson, B., Huebers, E., and Finch, C. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 621–625.

Huebers, H. A., Huebers, E., Finch, C. A., and Martin, A. W. (1982). *J. Comp. Physiol.* **148,** 101–109.

Humphries, S. E., Berg, K., Gill, L., Cumming, A. M., Robertson, F. W., Stalenhoef, A. F. H., Williamson, R., and Børresen, A.-L. (1984a). *Clin. Genet.* **26,** 389–396.

Humphries, S. E., Williams, L., Myklebast, O., Stalenhoef, A. F. H., Demacker, P. N. M., Baggio, G., Crepaldi, G., Galton, D. J., and Williamson, R. (1984b). *Hum. Genet.* **67,** 151–155.

Iacopetta, B. J., Morgan, E. H., and Yeoh, G. C. T. (1983). *J. Histochem. Cytochem.* **31,** 336–344.

Jacob, F. (1983). *In* "Evolution from Molecule to Men" (D. S. Bendall, ed.), pp. 131–144. Cambridge Univ. Press, London and New York.

Jeltsch, J. M., and Chambon, P. (1982). *Eur. J. Biochem.* **122,** 291–295.

Kan, C. C., Solomon, E., Belt, K. T., Chain, A. C., Hidrons, L. R., and Fey, G. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 2282–2286.

Kant, J. A., Fornace, A. J., Saxe, D., Simon, M. I., McBride, O. W., and Crabtree, G. R. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 2344–2348.

Kao, F. T., Morse, H. G., Law, M. L., Lidsky, A., Chandra, T., and Woo, S. L. C. (1984). *Hum. Genet.* **67,** 34–36.

Karathanasis, S. K., McPherson, J., Zannis, V., and Breslow, J. L. (1983). *Nature (London)* **304,** 371–373.

Koch, G. A., Schoen, R. C., Klebe, R. J., and Shows, T. B. (1982). *Exp. Cell Res.* **141,** 293–302.

Krumlauf, R., Hammer, R. E., Tilghman, S. M., and Brinster, R. L. (1985). *Mol. Cell. Biol.* **5,** 1639–1648.

Kurnit, D. M., Philipp, B. W., and Bruns, G. A. P. (1982). *Cytogenet. Cell Genet.* **34,** 282–288.

Kurosky, A., Hay, R. E., and Bowman, B. H. (1979). *Comp. Biochem. Physiol.* **62B,** 339–344.

Kurosky, A., Barnett, D. R., Lee, T.-H., Touchstone, B., Hay, R. E., Arnott, M. S., Bowman, B. H., and Fitch, W. M. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 3388–3392.

Lachmann, P. J. (1982). *Immunol. Today* **3,** 27–28.

Lackrer, N. J. Law, S. W., Brewer, Jr., H. B., Sakaguchi, A. Y., and Naylor, S. L. (1984). *Biochem. Biophys. Res. Commun.* 122, 877-883.

Law, S. W., Lee, N., Monge, J. C., Brewer, H. B., Jr., Sakaguchi, A. Y., and Naylor, S. L. (1985). *Biochem. Biophys. Res. Commun.* **131,** 1003–1012.

Lawn, R. M., Adelman, J., Bock, S. C., Franke, A. E., Houck, C. M., Najarian, R. C., Seeburg, P. H., and Wion, K. L. (1981). *Nucleic Acids Res.* **9,** 6103–6114.

Le Beau, M. M., Diaz, M. O., Yang, F., Plowman, G. D., Brown, J. P., and Rowley, J. D. (1985). *Am. J. Hum. Genet.* **37,** A31, Abstr. 086.

Lum, J. B. (1986). *BioTechniques* **4,** 32–40.

Lum, J. B., Infante, A. J., and Bowman, B. H. (1985). *Protides Biol. Fluids* **33,** 87–90.

Lum, J. B., Infante, A. J., Makker, D. M., Yang, F., and Bowman, B. H. (1986). *J. Clin. Invest.* **77,** 841–849.

McBride, D. W., Battey, J., Hollis, G. F., Swan, D. C., Siebenlist, U., and Leder, P. (1982). *Nucleic Acids Res.* **10,** 8155–8170.

McClelland, A., Kühn, L. C., and Ruddle, F. H. (1984). *Cell (Cambridge, Mass.)* **39,** 267–274.

McCombs, J. L., Yang, F., Bowman, B. H., McGill, J. R., and Moore, C. M. (1986). *Cytogenet. Cell Genet.* **42,** 62–64.

McGill, J. R., Boyd, D., Barrett, K. J., Drysdale, J. W., and Moore, C. M. (1984a). *Am. J. Hum. Genet.* **36,** 146S.

McGill, J. R., Yang, F., Baldwin, W. D., Brune, J. L., Barnett, D. R., Bowman, B. H., and Moore, C. M. (1984b). *Cytogenet. Cell Genet.* **38,** 155–157.

MacGillivray, R. T. A., Mendes, E., Shewale, J. G., Sinha, S. K., Lineback-Zins, J., and Brew, K. (1983). *J. Biol. Chem.* **258,** 3543–3553.

McKnight, G. S., Hammer, R. E., Kuenzel, E. A., and Brinster, R. L. (1983). *Cell (Cambridge, Mass.)* **34,** 335–341.

Maeda, N. (1985). *J. Biol. Chem.* **260,** 6698–6709.

Maeda, N., Yang, F., Barnett, D. R., Bowman, B. H., and Smithies, O. (1984). *Nature (London)* **309,** 131–135.

Magenis, R. E., Hecht, F., and Lovrien, E. W. (1970). *Science* **170,** 85–87.

Magnusson, S., Peterson, T. E., Sottrup-Jensen, L., and Cloeys, H. (1975). *Cold Spring Harbor Conf. Cell Proliferation* **2,** 123–149.

Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M. A., Bentley, D. L., and Rabbitts, T. H. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 4957–4961.

Mantzouranis, E. C., Dowton, S. B., Whitehead, A. S., Edge, M. D., Bruns, G. A. P., and Colten, H. R. (1985). *J. Biol. Chem.* **260,** 7752–7756.

Martin, A. W., Huebers, H., Huebers, E., Webb, J. M., and Finch, C. A. (1984). *Blood* **64,** 1047–1052.

Metz-Boutigue, M.-H., Iolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montrevil, J., and Jollès, P. (1984). *Eur. J. Biochem.* **145,** 659–676.

Mevag, B., Rogde, S., Olaisen, B., Gedde-Dahl, T., Teisberg, P., and Tedesco, F. (1983). *Proc. Congr. Haematogenet., 10th.*

Mikkelsen, M., Jacobsen, P., and Henningsen, K. (1977). *Hum. Hered.* **27,** 105–107.

Miller, Y. E., Jones, C., Scoggin, C., Morse, H., and Seligman, P. (1983). *Am. J. Hum. Genet.* **35,** 573–583.

Murray, J. C., Mills, K. A., Demopulos, C. M., Hornung, S., and Motulsky, A. G. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 3486–3490.

Murray, J. C., Sadler, E., Eddy, R. L., Shows, T. B., and Buetow, K. H. (1985). *Cytogenet. Cell Genet.* **40,** 709.

Nakazato, M., Kangawa, K., Minamino, N., Tawara, S., Matsuo, H., and Araki, S. (1984). *Biochem. Biophys. Res. Commun.* **123,** 921–928.

Nance, W. E., and Smithies, O. (1963). *Nature (London)* **198,** 869–870.

Naylor, S. L., and Sakaguchi, A. Y. (1985). *Protides Biol. Fluids* **33,** 71–76.

Naylor, S. L., Elliott, R. W., Brown, J. A., and Shows, T. B. (1982). *Am. J. Hum. Genet.* **34,** 235–244.

Naylor, S. L., Yang, F., Cutshaw, S., Barnett, D. R., and Bowman, B. H. (1985). *Cytogenet. Cell Genet.* **40,** 711.

Newman, R., Schneider, C., Sutherland, R., Vodinelich, L., and Greaves, M. (1982). *Trends Biochem. Sci.* **7,** 397–400.

Nishiya, K., Chiao, J. W., and DeSousa, M. (1980). *Br. J. Haematol.* **46,** 235–245.

Olaisen, B., Teisberg, P., and Gedde-Dahl, T., Jr. (1982). *Hum. Genet.* **62,** 233–236.

Olaisen, B., Teisberg, P., Jonassen, P., Thorsby, E., and Gedde-Dahl, T. (1983). *Ann. Hum. Genet.* **47,** 285–292.

Olaisen, B., Gedde-Dahl, T., Teisberg, P., Thorsby, E., Siverts, A., Jonassen, R., and Wilhelmy, M. C. (1985). *Am. J. Hum. Genet.* **37,** 215–220.

Oliviero, S., DeMarchi, M., Bensi, G., Raugei, G., and Carbonara, A. O. (1985a). *Hum. Genet.* **70,** 66–70.

Oliviero, S., DeMarchi, M., Carbonara, A. O., Berrini, L. F., Bensi, G., and Raugei, G. (1985b). *Hum. Genet.* **71,** 49–52.

Park, I., Schaeffer, E., Sidoli, A., Baralle, F. E., Cohen, G. N., and Zakin, M. M. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 3149–3153.

Pearson, P. L., van der Kamp, J., and Veldtkamp, J. (1982). *Cytogenet. Cell Genet.* **32,** 309.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., and Collen, D. (1983). *Nature (London)* **301,** 214–221.

Petrini, M., Galbraith, R. M., Emerson, D. L., Nel, A. E., and Arnaud, P. (1985). *J. Biol. Chem.* **260,** 1804–1810.

Pfeiffer, R. A., Ott, R., Gilgenkrantz, S., and Alexandre, P. (1982). *Hum. Genet.* **62,** 358–360.

Plowman, G. D., Brown, J. P., Enns, C. A., Schröder, J., Nikinmaa, B., Sussman, H. H., Hellström, K. E., and Hellström, I. (1983). *Nature (London)* **34,** 70–72.

Putnam, F. W. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.), Vol. 4, pp. 96–101. Academic Press, New York.

Rabin, M., McClelland, A., Kühn, L., and Ruddle, F. H. (1985). *Am. J. Hum. Genet.* **37,** 1112–1116.

Rajput, B., Friezner Degen, S. J., Reich, E., and Shows, T. B. (1985). *Protides Biol. Fluids* **33,** 99–102.

Raugei, G., Bensi, G., Colantouni, V., Romano, V., Santoro, C., Costanzo, F., and Cortese, R. (1983). *Nucleic Acids Res.* **11,** 5811–5819.

Riccio, A., Verde, P., Grimaldi, G., Tripputi, P., and Blasi, F. (1985). *Protides Biol. Fluids* **33,** 67–70.

Rocchi, M., Roncuzzi, L., Santamaria, R., Sbarra, D., Mochi, M., Archidiacono, N., Covone, A., Cortese, R., and Romeo, G. (1985). *Cytogenet. Cell Genet.* **40,** 734 (abstract).

Rogde, S., Olaisen, B., Gedde-Dahl, T., Jr., and Teisberg, P. (1985). *Cytogenet. Cell Genet.* **40,** 734 (abstract).

Rose, T. M., Plowman, G. D., Teplow, D. B., Dreyer, W. J., Hellström, K. E., and Brown, J. P. (1986). *Proc. Natl. Acad. Sci. U.S.A.* **83,** 1261–1265.

Sakaguchi, A. Y., Naylor, S. L., Quinto, C., Rutter, W. J., and Shows, T. B. (1982). *Cytogenet. Cell Genet.* **32,** 313.

Sargent, T. D., Yang, M., and Bonner, J. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 243–246.

Schaeffer, E., Huchenq, A., Park, I., Lucero, M. A., Cohen, G. N., Zakin, M. M., and Constans, J. (1985). *Protides Biol. Fluids* **33,** 131–134.

Schamaun, O., Olaisen, B., Mevag, B., Gedde-Dahl, T., Jr., Ehnholm, C., and Teisberg, P. (1984). *Hum. Genet.* **68,** 181–184.

Schneider, C., Owen, M. J., Banville, D., and Williams, J. G. (1984). *Nature (London)* **311,** 675–678.

Scott, R. W., and Tilghman, S. M. (1983). *Mol. Cell. Biol.* **3,** 1295–1309.

Shows, T. B., and McAlpine, P. J. (1982). *Cytogenet. Cell Genet.* **32,** 221–245.

Smithies, O., and Walker, N. F. (1955). *Nature (London)* **176,** 1265–1266.

Smithies, O., Connell, G. E., and Dixon, G. H. (1962a). *Am. J. Hum. Genet.* **14,** 14–21.

Smithies, O., Connell, G. E., and Dixon, G. H. (1962b). *Nature (London)* **196,** 232–236.
Solomon, E., Skok, J., Griffin, J., and Reid, K. B. M. (1985). *Cytogenet. Cell Genet.* **40,** 734 (abstr.).
Soltys, H. D., and Brody, J. I. (1970). *J. Lab. Clin. Med.* **75,** 250–257.
Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Lonblad, P. B., Jones, C. M., Wierbicki, D. M., Magnusson, S., Domdey, H., Wetsel, R. A., Lundwall, A., Tack, B. F., and Fey, G. H. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 9–13.
Stoll, C., and Roth, M. P. (1980). *Hum. Genet.* **53,** 303–304.
Svasti, J., and Bowman, B. H. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **253,** 4188–4194.
Svasti, J., Kurosky, A., Bennett, A., and Bowman, B. H. (1979). *Biochemistry* **18,** 1611–1617.
Takahashi, N., Ortel, T. L., and Putnam, F. W. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 390–394.
Tata, F., Henry, I., Markham, A. F., Wallis, S. C., Weil, D., Grzeschik, K. H., Junien, C., Williamson, R., and Humphries, S. E. (1985). *Hum. Genet.* **69,** 345–349.
Thibodeau, S. N., Lee, D. C., and Palmiter, R. D. (1978). *J. Biol. Chem.* **253,** 3771–3774.
Urano, Y., Sakai, M., Watanabe, K., and Tamaoki, T. (1984). *Gene* **32,** 255–261.
Van Baelen, H., Bouillon, R., and De Moor, P. (1980). *J. Biol. Chem.* **255,** 2270–2272.
van der Straten, A., Herzog, A., Jacobs, P., Cabezón, J. T., and Bollen, A. (1983). *EMBO J.* **6,** 1003–1007.
van der Straten, A., Herzog, A., Cabezón, T., and Bollen, A. (1984). *FEBS Lett.* **168,** 103–107.
van der Straten, A., Résibois, A., Cabezón, T., Loriau, R., Falque, J. C., Herzog, A., and Bollen, A. (1985). *Protides Biol. Fluids* **33,** 151–155.
Vehar, G. A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., and Harkins, R. N. (1984). *Nature (London)* **312,** 337–342.
Viau, M., Constans, J., Debray, H., and Montreuil, J. (1983). *Biochem. Biophys. Res. Commun.* **117,** 324–331.
Wallace, M. R., Naylor, S. L., Kluve-Beckerman, B., Long, G., McDonald, L., Shows, T. B., and Benson, M. (1985). *Biochem. Biophys. Res. Commun.* **129,** 753–758.
Weitkamp, L. R. (1983). *Ann. Hum. Genet.* **47,** 293–297.
Weitkamp, L. R., Rucknagel, D. L., and Gershowitz, H. (1966). *Am. J. Hum. Genet.* **18,** 559–571.
Weitkamp, L. R., Renwick, J. H., Berger, J., Shreffler, D. C., Drachman, O., Wuhrmann, R., Braend, M., and Franglen, G. (1970). *Hum. Hered.* **20,** 1–7.
Whitehead, A. S., Bruns, G. A. P., Markham, A. F., Colten, H. R., and Woods, D. E. (1983). *Science* **221,** 69–71.
Williams, J. (1982). *Trends Biochem. Sci.* **7,** 394–397.
Williams, J., Elleman, T. C., Kingston, I. B., Wilkins, A. G., and Kuhn, K. A. (1982). *Eur. J. Biochem.* **122,** 297–303.
Woo, S. L. C., Kidd, V. J., Pam, Z.-K., Wallace, R. B., and Itakura, K. (1983). *Banbury Rep.* **14,** 105.
Yang, F., Brune, J. L., Baldwin, W. D., Barnett, D. R., and Bowman, B. H. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5875–5879.
Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., van Bragt, P. H., Baldwin, W. D., and Bowman, B. H. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 2752–2756.
Yang, F., Brune, J. L., Naylor, S. L., Cupples, R., Naberhaus, K. H., and Bowman, B. H. (1985a). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 7994–7998.
Yang, F., Lum, J. B., and Bowman, B. H. (1985b). *Protides Biol. Fluids* **33,** 31–34.
Yang, F., Luna, V. J., McAnelly, R. D., Naberhaus, K. H., Cupples, R. L., and Bowman, B. H. (1985c). *Nucleic Acids Res.* **13,** 8007–8017.
Yang, F., Naylor, S. L., Lum, J. B., Cutshaw, S., McCombs, J. L., Naberhaus, K. H., McGill, J. R., Adrian, G. S., Moore, C. M., Barnett, D. R., and Bowman, B. H. (1986). *Proc. Natl. Acad. Sci. U.S.A.* **83,** 3257–3261.
Yunis, J. J., and Prakash, O. (1982). *Science* **215,** 1525–1530.
Zardi, L., Cianfriglia, M., Balza, E., Carnemolla, B., Siri, A., and Croce, C. M. (1982). *EMBO J.* **1,** 929–933.

2 / Immunoglobulins: Structure, Function, and Genes

Frank W. Putnam

Department of Biology
Indiana University
Bloomington, Indiana 47405

The Plasma Proteins, Vol. V

I. Introduction

Because of intensive study of immunoglobulins, initially using myeloma proteins and recently through the revolutionary developments in the molecular biology of immunoglobulin genes, more has been learned about this family of proteins, and thus about antibodies, than about any other set of proteins. It would require a volume rather than just a chapter to give a complete summary and interpretation of the progress in a decade since the last review of immunoglobulins in this series (Putnam, 1977a,b,c). In this period the relationship of immunoglobulin structure to antibody specificity has been elucidated in fine detail. The biological effector functions of antibodies, such as complement activation and cytotropic properties, are beginning to be clarified. The cloning and sequencing of immunoglobulin genes has revolutionized our understanding of the relationship of immunoglobulin structure to genomic organization and has largely solved the antibody dilemma, that is, the genetic mechanism for generation of antibody diversity and specificity. However, for reasons described below, the focus of this chapter is on immunoglobulins as a family of plasma proteins, rather than on antibody specificity or immunoglobulin genes.

Although specific combination of antibody with antigen is the obligate first step in effecting the humoral immune response, the full repertoire of antibody activity is ultimately elicited only through the biological effector functions that are vested in the constant regions of the five classes of immunoglobulins. Numerous investigations summarized by frequent reviews have elucidated the nature of the antibody combining site and how it is formed by the variable (V) regions of light and heavy chains (Capra and Kehoe, 1975; Capra and Edmundson, 1977; Padlan, 1977a,b; Putnam, 1977b; Potter, 1977; Kabat, 1982; Kindt and Capra, 1984). The genetic mechanisms governing the origin, diversity, assembly, and expression of the V region genes have been elegantly deciphered by the methods of molecular biology (Rabbitts *et al.*, 1981; Leder *et al.*, 1981; Leder, 1982; Cushley and Williamson, 1982; Honjo, 1983; Hood *et al.*, 1983; Wall, 1983; Fudenberg *et al.*, 1984; Tonegawa, 1985; Honjo and Habu, 1985). However, the structure–function relationships of the C (constant) regions of antibodies are not yet well clarified. Thus, this chapter will focus on the relationship of structure to the biological activity of the C regions of the five classes of immunoglobulins. The emphasis will be on human immunoglobulins, but of necessity work on other species, notably the mouse, will be introduced when relevant. One advantage of this restricted focus is that the structures of only seven classes of polypeptide chains need be presented, compared to the unending series of V region sequences that continue to flood the data base.

In the author's main area of research—the primary structure of immunoglobulins—the pace of progress has been phenomenal. Indeed, as shown in

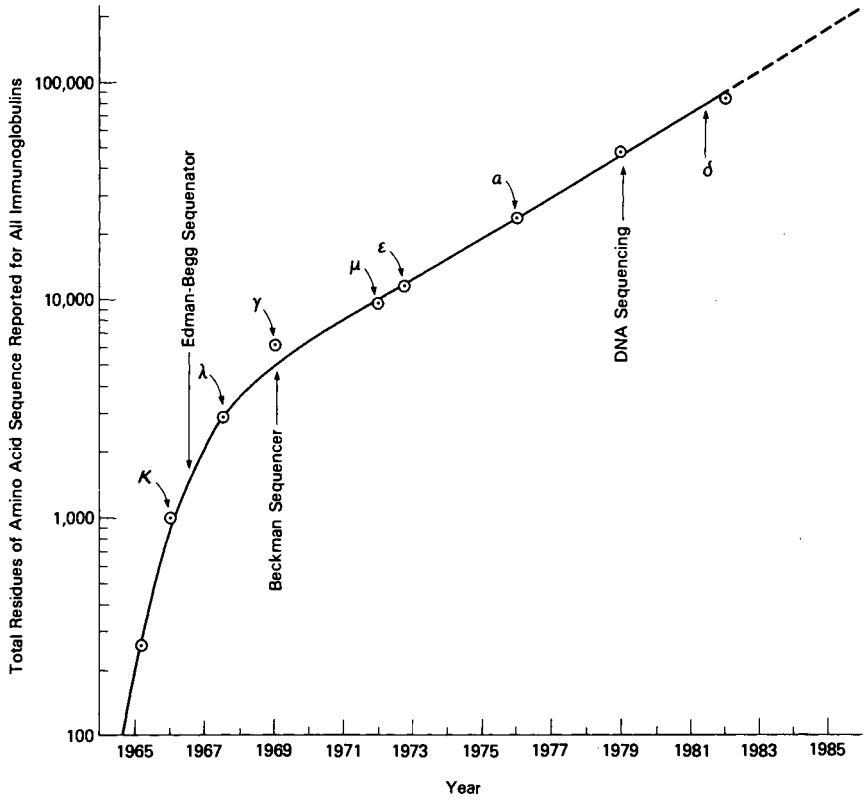


Fig. 1. Increase in amino acid sequence data for immunoglobulins of all species since the first reports for the human κ light chain in 1965. The Greek letters indicate the dates of the completion of the sequence of the human light (κ, λ) and heavy (γ, μ, ε, α, δ) chains beginning with the first report of the complete sequence of the κ chain (Putnam *et al.*, 1966) to the final report on the δ chain (Takahashi *et al.*, 1982). For additional references see the text. For a historical review see Putnam (1983), from which this figure has been taken and updated. The sources of data include the sequence compilations of Kabat *et al.* (1979, 1983), Dayhoff (1978), and the on-line computerized data base of the "Atlas of Protein Sequence and Structure" now designated the NBFR-PIR data base (Barker *et al.*, 1983, 1985). Note the logarithmic scale for the ordinate (updated from Putnam, 1984).

Fig. 1, the increase in sequence data has followed an exponential rate since 1969, and it shows no signs of decreasing since the advent of DNA sequencing in 1979. By 1983 more than 100,000 residues of amino acid sequence had been reported, Even today despite the numerous gene sequences being reported for viruses and bacterial proteins, immunoglobulin sequences still account for about one-third of

the total data in the protein sequence data base. Indeed, the retrieval and use of the data would be impossible without access to the new data bases and the powerful computer programs that various groups provide. For this review I have used the NBRF-PIR data base (formerly "Atlas of Protein Structure and Sequence") (Barker *et al.*, 1985) and also the earlier compilation of Kabat *et al.* (1983), which is restricted to immunoglobulin and related structures.

The unique features of immunoglobulin structure, genomic organization, and DNA and RNA rearrangement startled and fascinated the scientific world as they were discovered successively. Now the mechanism that governs the generation of antibody specificity is being found to be shared by other molecules that have the function of recognition, such as histocompatibility antigens, the T-cell receptor, and other cell surface receptors (Stott and Williamson, 1982; Hood *et al.*, 1983, 1985; Honjo and Habu, 1985). Such recognition molecules are being called members of the immunoglobulin superfamily. This chapter will largely be focused on secreted immunoglobulins, i.e., plasma proteins; however, the wealth of knowledge amassed on these proteins forms the principal basis for understanding the structure, assembly, and specificity of related recognition molecules in this newly identified superfamily of proteins.

This chapter is written from the point of view of a protein chemist and is directed primarily to biochemists, immunologists, and other medical scientists concerned with the structure, function, physiological role, and pathological significance of immunoglobulins. Certain subjects are already well reviewed elsewhere and will not be covered here in detail. Several books are available on the nature of the antibody molecule (Nisonoff *et al.*, 1975; Nezlin, 1977), the biology of idiotypes (Greene and Nisonoff, 1984), immunogenetics (Fudenberg *et al.*, 1984), and the enigma of antibody diversification (Kindt and Capra, 1984). Since there are many excellent reviews on the molecular biology of antibody genes (Rabbitts *et al.*, 1981; Honjo, 1983; Wall, 1983; Fudenberg *et al.*, 1984), this important and exciting subject will be introduced only as needed for understanding the genetic control of immunoglobulin structure and antibody specificity. Likewise, there are frequent symposia on the molecular aspects of antibodies, the preparation and use of monoclonal antibodies (Milstein, 1980; Yelton and Scharff, 1981; McMichael and Fabre, 1982; Haynes and Eisenbarth, 1983), and the protective, destructive, and regulatory role of antibodies in disease (Milgrom *et al.*, 1985). Also, there are numerous reviews on specialized subjects ranging from crystallographic structure (Amzel and Poljak, 1979; Davies and Metzger, 1983; Huber, 1984) to idiotypic specificities (Rudikoff, 1983; Huppi *et al.*, 1984). Again, these areas will be treated only in the context of our focus rather than to the degree that they might merit in themselves. The history of immunoglobulin research has been covered by Putnam (1977a, 1983), Swazey and Reeds (1978), and Kindt and Capra (1984).

II. Principles of Immunoglobulin Structure

A. Relationship of Physical Properties, Biological Activities, and the Prototype Structure

1. The Five Classes

"Immunoglobulins are proteins of animal origin endowed with known antibody activity and (also include) certain proteins related to them in chemical structure and hence antigenic specificity. Related proteins for which antibody activity has not been demonstrated are included—for example, myeloma proteins, Bence–Jones proteins, and naturally occurring subunits of immunoglobulins. Immunoglobulins are not restricted to the plasma but may be found in other body fluids or tissues such as urine, spinal fluid, lymph nodes, spleen, etc." (Subcommittee on Immunoglobulin Nomenclature, 1966). Thus, immunoglobulins are present in the sera of all vertebrates, where they represent the antibodies formed in response to the cumulative immunological memory of the individual. Normal immunoglobulins are heterogeneous in physical properties and in chemical structure. Rarely, homogeneous or monoclonal immunoglobulins of a particular class may be produced in large amounts to the virtual exclusion of all other classes, especially in lymphoproliferative diseases such as multiple myeloma or macroglobulinemia, which have been called monoclonal gammopathies (Kyle and Bayrd, 1976). In fact, until the recent widespread production of monoclonal antibodies, the myeloma proteins of man and the mouse were the principal models for study of antibody structure. Most of the structural data in the literature are still based on study of myeloma proteins; thus, except for discussion of defined specificities of monoclonal antibodies, no distinction need be made between the two model systems.

The predominant immunoglobulin class was historically designated γ-globulin because of its low electrophoretic mobility but is now called IgG. IgG is also traditionally characterized by its molecular weight (M_r of about 160,000) and its sedimentation coefficient (7 S) (Putnam, 1977a). By serological and sequence analysis five classes of immunoglobulins (IgG, IgA, IgM, IgD, IgE) have been identified in the sera of higher vertebrates. They are usually divided into the major (IgG, IgA, IgM) and minor classes (IgD, IgE) based on normal levels in human serum. IgG is normally the second most abundant protein in plasma and is thus readily identified in animal sera, whereas IgD and IgE are difficult to identify because of their extremely low concentration. In many species IgG exists as four subclasses and IgA as two. Much is now known about the structure and genetic control of all five classes of human immunoglobulins and for those of laboratory rodents such as rabbits, mice, and rats, but little is known about the immunoglobulins of other species. Thus, their comparative biochemistry will not

be considered here. However, much evidence indicates that the general principles of structure of immunoglobulins are the same in all species.

The physical, chemical, and biological properties of human immunoglobulins are summarized in Table I. Data on the biosynthesis, turnover rate, and circulating pool were established early (Rothschild and Waldmann, 1970; Morell and Riesen, 1982). Such data have much clinical relevance and are fully discussed elsewhere as are the methods of analysis for the proteins (Laurell, 1977). Determination of the serum levels of IgG, IgA, and IgM and of their variation in disease is one of the most common procedures of the clinical laboratory (Natelson *et al.*, 1978). Although it was this interest that originally spurred the research on immunoglobulins, it is not appropriate to the focus of this chapter. Yet it is interesting that measurements of serum IgG, IgA, and IgM are now also being done for pets and large domestic animals in some veterinary clinics.

2. The Prototype Four-Chain Structure

The cardinal characteristics of immunoglobulin structure are illustrated schematically in Fig. 2, which also shows the sites of antibody functions. This generalized model illustrates the now well-known four-chain polypeptide structure common to immunoglobulins of all classes, subclasses, and species. The prototype monomer consists of a pair of heavy (H) chains (γ, α, μ, δ, or ϵ) that determine the immunoglobulin class (IgG, IgA, IgM, IgD, and IgE, respectively) and a pair of light (L) chains (κ or λ) that can be linked to the H chains of any class. The chains can be dissociated by reduction of disulfide bonds and can be separated in the presence of denaturing agents by a variety of procedures.

The chief features of the prototype immunoglobulin structure are: (1) a monomer consisting of a pair of identical light chains ($M_r \simeq 23{,}000$) disulfide-bonded to a pair of heavy chains ($M_r \simeq 50{,}000$), which in turn are linked to each other by one or more disulfide bonds, (2) the division of each polypeptide chain into variable (V) and constant (C) regions, and (3) the domain structure typical of each chain, which includes the further division of the C region of the H chains into structural and functional domains (C_H1, etc.). Whereas the C region determines the class of the chain, the hypervariability of the amino acid sequences of the V regions of the H and L chains (V_H and V_L) governs the antibody specificity and the shape of the antigen combining site. Antibody specificity is a unique property of immunoglobulins and is vested in the Fab (antigen binding) fragment in two identical combining sites formed by the hypervariable (HV) or complementarity-determining (CDR) regions, three of which are present in each H and L chain. On the other hand, the biological effector functions such as complement fixation and cytotropic properties of antibodies are located in the C regions of the H chains and to some extent are compartmentalized in specific domains (Dorrington, 1979; Hahn, 1983; Unkeless and Wright, 1984; Burton, 1985). As first

TABLE I

Physical, Chemical, and Biological Properties of Human Immunoglobulin Classes[a]

Property	IgG	IgA	IgM	IgD	IgE
Usual molecular form	Monomer	Monomer, dimer, etc.	Pentamer	Monomer	Monomer
Molecular formula	$\kappa_2\gamma_2$ or $\lambda_2\gamma_2$	$(\kappa_2\alpha_2)_n$ or $(\lambda_2\alpha_2)_n$	$(\kappa_2\mu_2)$ or $(\gamma_2\mu_2)$	$\kappa_2\delta_2$ or $\lambda_2\delta_2$	$\kappa_2\epsilon_2$ or $\lambda_2\epsilon_2$
Other chains	—	J chain, SC	J chain	—	—
Subclasses	IgG1, IgG2, IgG3, IgG4	IgA1, IgA2	None established	None	None
Subclass heavy (H) chains	γ1, γ2, γ3, γ4	α1, α2	—	—	—
Heavy chain allotypes	Gm (ca. 20)	Am (3)	Mm (2)	—	—
Molecular weight	150,000	$(160{,}000)_n$	950,000	175,000	190,000
Sedimentation constant (s_{20})	6.6 S	7 S, 9 S, 11 S, 14 S	19 S	7 S	8 S
Carbohydrate content (%)	3	7	10	9	13
Serum level (mg/100 ml) (adult average)	1250 ± 300	210 ± 60	125 ± 50	≃4	≃0.03
Percentage of total serum level	75–85	7–15	5–10	≃0.3	≃0.003

(continued)

TABLE I (*continued*)

Property	IgG	IgA	IgM	IgD	IgE
Total circulating pool (mg/kg of body weight)	494.0	95.0	37.0	1.1	0.019
Half-life (days)	23.0	5.8	5.1	2.8	2.5
Rate of synthesis (mg/kg of body weight per day)	33.0	24.0	6.7	0.4	0.016
Paraaproteinemia	Myeloma	Myeloma	Macroglobulinemia	Myeloma	Myeloma
Antibody valence	2	2	5 or 10	2	2
Serological properties	Precipitation: toxin neutralization	Precipitation: virus neutralization	Agglutination: hemagglutination: hemolysis; virus neutralization	Virus neutralization	Not measureable
Complement fixation	IgG1, IgG2, and IgG3	—	Most IgMs	?	—
Complement activation via bypass	IgG4	IgA1, IgA2	—	?	IgE
Binding to cells	Macrophages, K cells, B cells, etc.	—	—	?	Mast cells
Other biological properties	Secondary Ab[b] response; placental transfer	Characteristic Ab in mucous secretions	Primary Ab response; rheumatoid factor	B-cell membrane receptor	Homocytotropic Ab; anaphylaxis; allergy

[a]Modified from Putnam (1977a).
[b]Antibody.

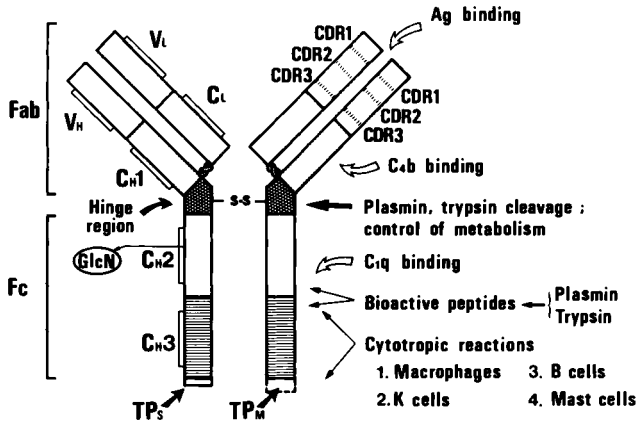


Fig. 2. Generalized schematic diagram of immunoglobulin structure and sites of antibody functions. Although based on human IgG1, the model is the monomeric prototype for all classes of human and animal antibodies. However, the latter will differ in primary structure and may differ in the number and location of disulfide bonds, size, degree of polymerization, presence of hinge region, number of C_H domains and oligosaccharides, and biological functions. The left side of the figure gives the approved notation of the domains of the light chain (V_L for variable, C_L for constant) and of the heavy chain (V_H for variable, C_H1, etc., for the C region domains). The right side identifies the sites of biological functions. TPs refers to a small extra segment (tailpiece) that is present at the C_γ terminus of the secreted immunoglobulin (sIG) and TPm to a different segment that replaces TPs in the membrane-bound form (mIg). These are the only structural differences of sIg and mIg. The domains are distinguished by individual shading. Each domain consists of about 110 amino acid residues and has an invariant interchain disulfide bond joining about 60 residues. In contrast, the number and location of intrachain disulfide bonds may differ characteristically for immunoglobulins of various classes and species (see Fig. 3) (modified from Putnam *et al.*, 1985). For rules on nomenclature see Subcommittee on Immunoglobulin Nomenclature (1966, 1969, 1972).

reported by Porter (1959) for rabbit IgG, some classes (IgG, IgD) are quite susceptible to rather specific proteolytic cleavage in an interdomain segment called the hinge region thus yielding Fab and Fc fragments; others are highly resistant, notably IgA. Furthermore, although the monomeric form is illustrated in Fig. 2, IgA and IgM generally are present in serum as polymers (IgA as the monomer, dimer, or tetramer; IgM as the pentamer or higher polymers; these are called poly-Igs, see Table I).

The general molecular characteristics illustrated in Fig. 2 are shared by immunoglobulins of all classes and species. However, a series of structural features differentiates the five classes, and each class also exhibits species differences. These differentiating characteristics largely are located in the H chains and include the length and number of domains in the H chains, the presence and unique structural aspects of the hinge region, the number and distribution of disulfide bonds, and the number, kinds, and sites of attachment of oligosac-

TABLE II

Number of Amino Acid Residues, Domains, and Oligosaccharides in the C Regions of Human Heavy Chains[a]

	Polypeptide			Residues		Oligosaccharides	
Class	Chain	Molecular weight	C_H domains	Hinge	C region	GalN (hinge)	GlcN (domains)
IgG1	γ1	36,106	3	15	330	0	1
IgG2	γ2	35,884	3	12	326	0	1
IgG3	γ3	(41,000)	3	62	(375)	0	2
IgG4	γ4	35,940	3	12	326	0	1
IgA1	α1	37,648	3	26	353	5	2
IgA2	α2	36,573	3	13	340	0	4, 5
IgM	μ	49,270	4		451		5
IgD	δ	42,243	3	64	384	4 or 5	3
IgE	ε	47,019	4		428		6

[a]The polypeptide molecular weight calculated from the amino acid sequence is given. This excludes the carbohydrate but includes the C-terminal lysine that is cleaved postsynthetically from the γ and δ chains. Values for the γ3 chain (in parentheses) are approximate, as are the estimates of the length of the hinge region. The μ and ε chains lack a hinge region but have an extra (fourth) C_H domain. The A2m(1) allotype of the α2 chain has four GlcN oligosaccharides, and the A2m(2) allotype has five. (Table revised from Putnam *et al.*, 1982.)

charides of the glucosamine (GlcN) and galactosamine (GalN) types. These differentiating features are summarized in Table II for the five classes of human immunoglobulins, for which they were first and best defined, and for the subclasses of IgG and IgA. Corresponding classes of other species have similar features (Fig. 3), but it is sometimes difficult to equate subclasses of IgG or IgA of different species, which suggests that they probably arose rather late in evolution.

3. *Variable (V) and Constant (C) Regions*

The one unique characteristic of immunoglobulins is the precise division of all heavy and light chains into variable (V) and constant (C) regions. The V region has a length of about 110 amino acid residues in light chains and about 120 in heavy chains, and the C region is characteristic of the kind of chain. This unprecedented principle was first discovered by sequence analysis of light chains, i.e., Bence–Jones proteins excreted by patients with multiple myeloma (Titani and Putnam, 1965; Titani *et al.*, 1965; Hilschmann and Craig, 1965; Putnam *et al.*, 1966, 1967a,b). This led Dreyer and Bennett (1965) to propose the then heretical doctrine of "two genes–one polypeptide chain." A similar divi-

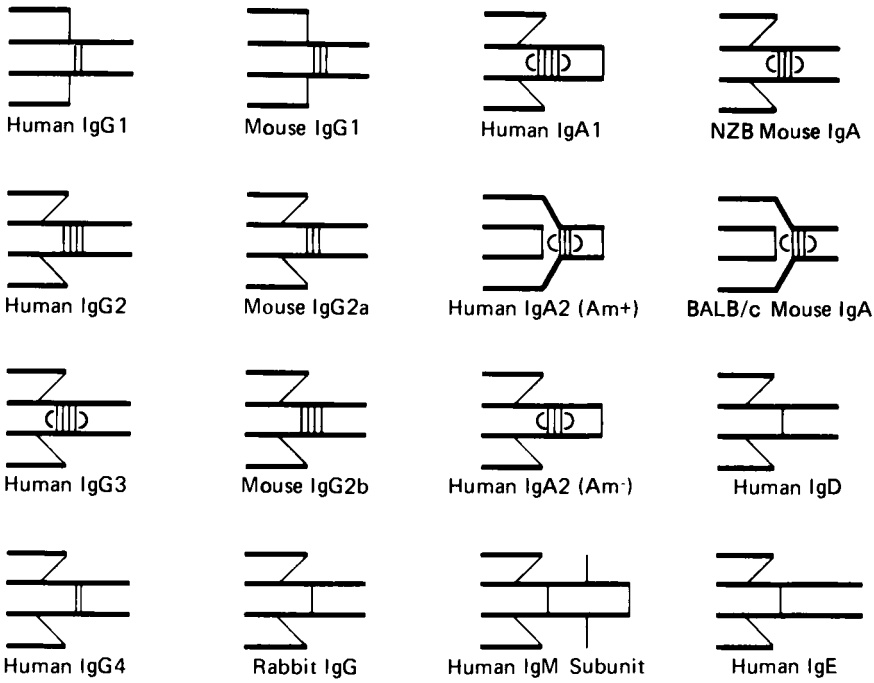


Fig. 3. Comparison of the four-chain model structures and the distribution of interchain disulfide bridges (thin lines) linking L and H chains (short and long thick lines) in different immunoglobulin classes and subclasses of several species. In some cases the disulfide bridges in the hinge region are bracketed where the exact number is not yet known. However, IgG3 probably has 11 interchain disulfide bridges because of its quadruplicated hinge (see Fig. 9) (from Fudenberg *et al.*, 1984).

sion of heavy chains into V and C regions was soon discovered, and the V regions of both heavy and light chains were shown to be associated with the combining site of antibodies.

The paradox was that the V region (the first domain) of a light or heavy chain could have many different amino acid sequences, whereas the remainder of the chain (the C region) had a constant sequence characteristic of the class, subclass, and allotype. This paradox was seized on by protein chemists, immunologists, and geneticists because it was recognized to be the key to solving the nature of antibody specificity and the origin of antibody diversity, what Kindt and Capra (1984) have called "the antibody enigma." After sequence analysis of many myeloma proteins and antibodies (see Fig. 1), coupled with serological and crystallographic analysis, the solution came through the brilliant new approaches of molecular biology. Molecular cloning and nucleotide sequencing of immunoglobulin genes revealed the genomic organization and mode of assembly of

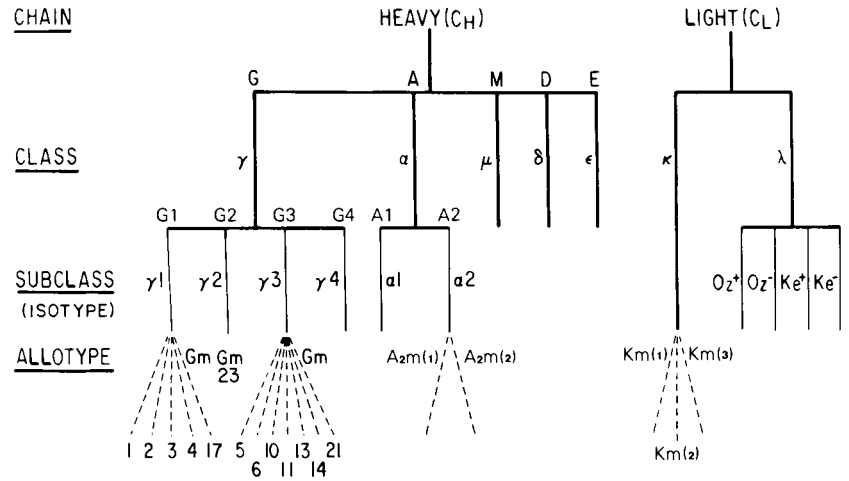


Fig. 4. Classes, subclasses, and allotypes of the heavy chain (C_H) and light chain (C_L) constant regions. Subclass and allotype designations are defined in the text. The abbreviations Ke^+ and Ke^- signify the $Kern^+$ and $Kern^-$ isotypes, and likewise for Oz^+ and Oz^-. The Inv allotype is now designated Km (modified from Putnam, 1977a).

multiple gene segments for both L and H chains. A similar mechanism is being elucidated for other members of the immunoglobulin superfamily such as histocompatibility antigens and the T-cell receptor (Hood *et al.*, 1985; Tonegawa, 1985). This field is changing rapidly, and the volume of literature is increasing exponentially. The mechanism of V–C gene recombination will be treated only briefly later, but it has been extensively reviewed elsewhere (Honjo, 1983; Wall, 1983).

4. Isotypes, Allotypes, and Idiotypes

Although there are only two kinds of light chains (κ and λ), there are five classes of heavy chains (γ, α, μ, δ, and ε). Any H chain can pair with either κ or λ in the tetrachain formula H_2L_2 to yield the ten possible molecular formulas listed in Table I. Because the C region determines the class of immunoglobulin, there are five classes, IgG, IgA, IgM, IgD, and IgE, defined by γ, α, μ, δ, and ε chains, respectively (Fig. 4). In humans there are four subclasses or isotypes of IgG (IgG1, IgG2, IgG3, IgG4) and two of IgA (IgA1, IgA2); all these are made by normal individuals, but a single isotype (and allotype) may be synthesized in great excess by patients with monoclonal diseases such as multiple myeloma. This phenomenon is called allelic exclusion. Subclasses of IgG differ in amino acid sequence by only about 5%, yet may differ in properties such as the ability

TABLE III

Amino Acid Interchanges Associated with Allotypic Differences in Human Immunoglobulins[a]

Chain	Domain	Residue number	Amino acid residues	Associated allotype
γ1	C_H3	356, 358	Asp, Leu	G1m(1)
			Glu, Met	G1m(1−)
γ1	C_H3	431	Gly	G1m(2)
			Ala	G1m(2−)
γ1	C_H1	214	Arg	G1m(3)
			Lys	G1m(3−)
γ3	C_H2	296, 436	Phe, Phe	G3m(5)
			Tyr, Tyr	G3m(21)
α2	Hinge	212, 221	Ser, Arg	A2m(2)
			Pro, Pro	A2m(1)
κ	C_L	153, 191	Val, Leu	Km(1)
		153, 191	Ala, Leu	Km(1,2)
		153, 191	Ala, Val	Km(3)

[a]Modified from Table 3.3 Fudenberg *et al.* (1984), which also lists some allotypes of animal immunoglobulins. Note that Asp-356 and Leu-358 present in the Gm(1) allotype of the γ1 chain are changed to Glu-356 and Met-358 in the Gm(1−) allotype, and so forth.

to fix complement. For example, human IgG1, IgG2, and IgG3 fix complement and initiate the classical pathway, but IgG4 does so poorly or not at all.

Allotypes are genetically determined antigenic differences in proteins that vary in different members of the same species (Fudenberg *et al.*, 1984). As shown in Table III, allotypes such as the Gm genetic markers in human IgG usually are associated with only one or a few amino acid substitutions that require only a single base change in the triplet of a codon. However, conformational features of the hinge region or elsewhere may contribute to the serological specificity of the allotype. Allotypes such as the Gm and Am markers are defined serologically. In many cases the differences in amino acid sequence of the Gm allotypes are still unknown, but the structural differences of the A2m allotypes have been determined (Torano and Putnam, 1978; Tsuzukida *et al.*, 1979). Unlike isotypes and allotypes, which are genetically determined characteristics of C regions, idiotypes are unique antigenic determinants of the V regions of monoclonal antibodies and myeloma proteins. Since idiotypes represent the antigenicity of the antigen binding site, many idiotypes are produced during the normal polyclonal response in contrast to the unique idiotypes characteristic of monoclonal immunoglobulins. Allotypes of immunoglobulins have been reviewed by Fudenberg *et*

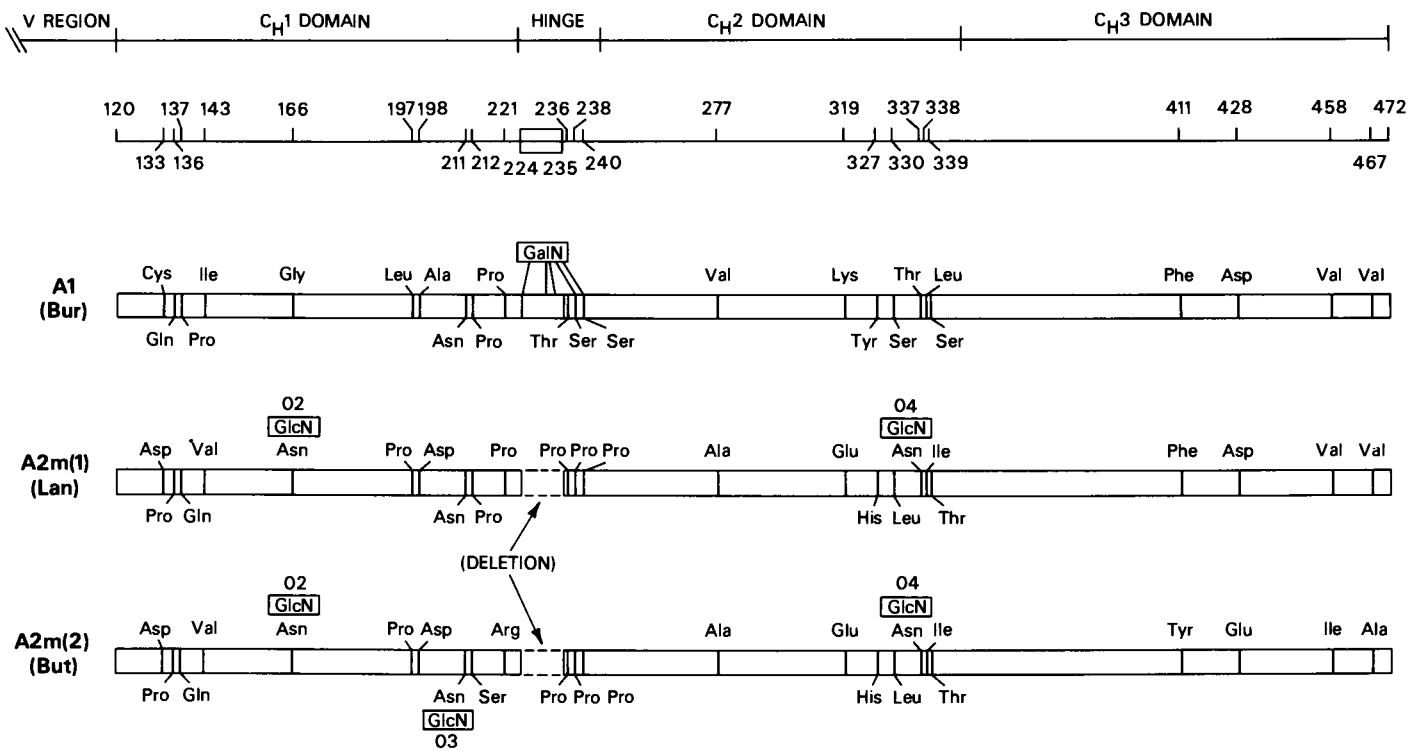


Fig. 5. Differences among the C regions of three human IgA α heavy chains of different subclasses (A1 and A2) and allotypes [$A_2m(1)$ and $A_2m(2)$]. The positions of the substitutions shown are indicated by the sequence numbers given at the top of the figure. Only those oligosaccharides are shown that differ among the three proteins (Bur, Lan, and But). From Tsuzukida *et al.* (1979).

al. (1984) and idiotypes by Greene and Nisonoff (1984) and Nisonoff and Gurish (1984). Table III illustrates the WHO nomenclature and lists amino acid interchanges that have been associated with some human Gm and Am allotypes. Figure 5 shows the characteristic structural differences of the human IgA1 and IgA2 subclasses and of the A2(m) allotypes of IgA2. Although the A2 allotypes were first determined serologically (Van Loghem *et al.*, 1976) and later established by sequence analysis (Tsuzukida *et al.*, 1979), they have recently been differentiated by restriction fragment length polymorphism (Lefranc and Rabbitts, 1984).

B. Characteristic Structural Features

1. The Domain Structure

A unique structural characteristic of all immunoglobulin polypeptide chains is their division into a series of structural and functional domains. These domains or homology regions contain about 110 amino acid residues of which about half are brought together by an intrachain disulfide loop. All domains of all classes of immunoglobulin chains of all species studied have at least 20% identity in primary structure. Although all domains are homologous in amino acid sequence and are similar in three-dimensional structure (the immunoglobulin fold), the only absolutely invariant residues are the two cysteines that form the intrachain disulfide bridge and a tryptophan that is nearby in the three-dimensional structure. As illustrated in Fig. 2, there are two such domains in κ and λ light chains (V_L and C_L); however, as shown in Table II, the number of domains varies with the class of the heavy chain and determines its length. The domains in the H chains are numbered with reference to the chain, e.g., Cγ1, Cγ2, Cγ3 for the γ chain and Cμ1 to Cμ4 for the μ chain. The precise division into domains and the homology among the domains are illustrated in Fig. 6 for the C regions of a human IgD protein. A model of the three-dimensional structure of an IgG antibody molecule is shown in Fig. 7.

Protein chemists first recognized the domains by three features: (1) their characteristic length, which varies from 110 to 120 amino acid residues; (2) the presence of the intrachain disulfide loop; and (3) their homology in amino acid sequence. Later, crystallographers showed that the domains were compact structural subunits each of which has a basic three-dimensional structure illustrated in Fig. 8 and known as the "immunoglobulin fold" (Poljak *et al.*, 1973). Recently, molecular biologists found that a separate gene segment encodes each domain. Given these facts, it is not surprising that immunologists have tried to identify specific biological effector functions with each C region domain, just as the antibody combining site was localized to the V_L and V_H domains. The extent to which this has been possible is one of the main themes of this chapter.

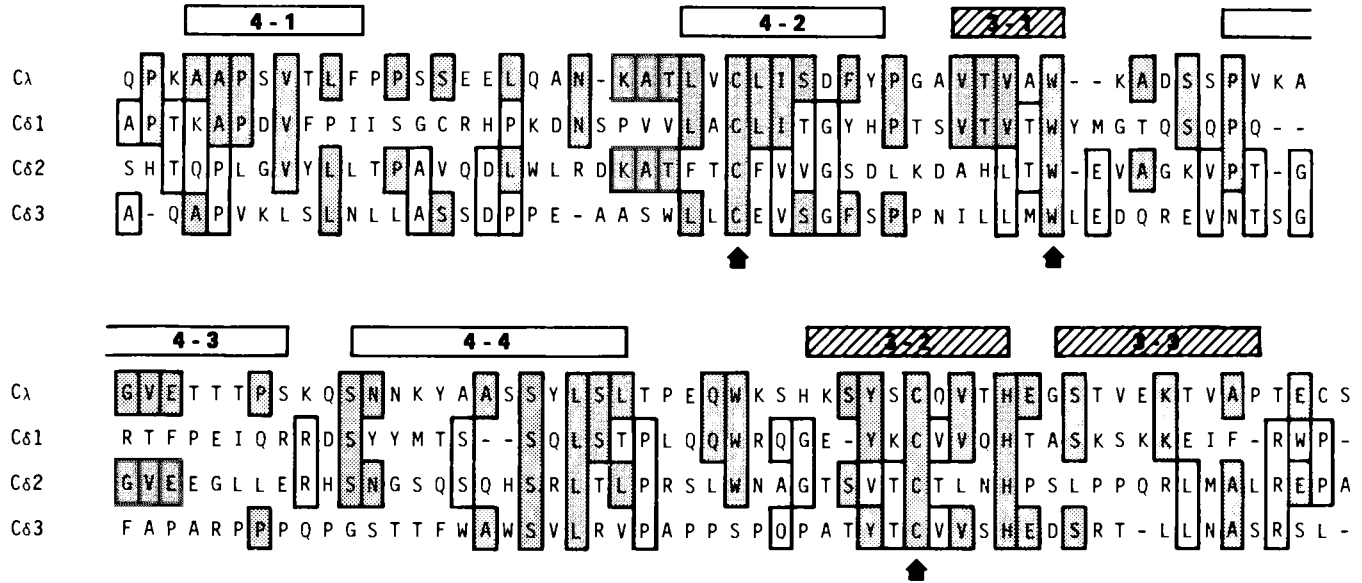


Fig. 6. Comparison of the amino acid sequences of the Cλ domain of the L chain and of the C_H domains (Cδ1, Cδ2, Cδ3) of the H chain of human IgD WAH. The one-letter notation for amino acids is given in Dayhoff (1978). Gaps have been inserted to maximize the homology. The three invariant cysteine (C) and tryptophan (W) residues in each domain were used to place the alignment in register, and they are indicated by arrows. Residues on the δ chain sequence that share identity with corresponding residues in the λ chain are outlined by shaded boxes; other residues that are identical in the δ chain domains are outlined in open boxes. The β strands are numbered according to Edmundson *et al.* (1975) with the four-stranded β-sheet elements in open bars and the three-stranded β elements in hatched bars. The hinge region of the δ chain is not shown. For the complete sequence of this IgD protein see Takahashi *et al.* (1982).

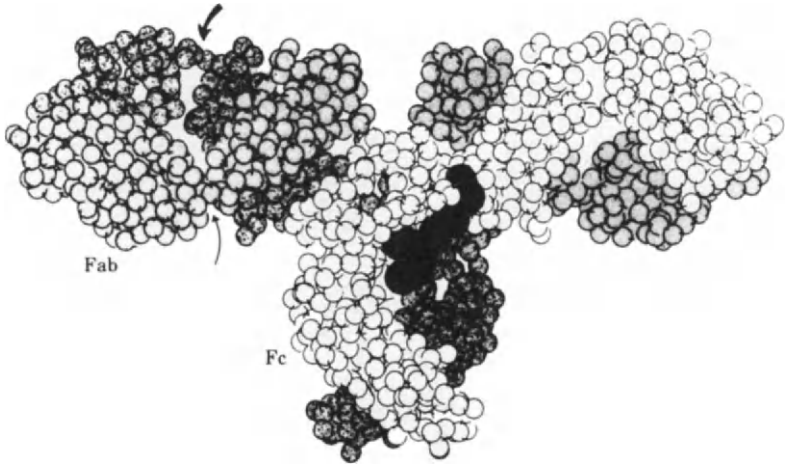


Fig. 7. Space-filling model of an IgG molecule. One complete heavy chain is white and the other heavy chain is dark gray and is speckled. The two light chains are lightly shaded. The large black spheres represent the individual hexose units of the complex GlcN carbohydrate. In this view, the twofold axis of symmetry is vertical. A crevasse is seen between the C_H2 of the white heavy chain and the C_L domain of the Fab on the left. This myeloma IgG molecule (Dob) lacks a hinge (from Silverton *et al.*, 1977).

2. Disulfide Bridge Pattern

Although the repeating pattern of interchain disulfide bridges is an invariant feature, the location of the interchain disulfide bridges frequently is the distinguishing characteristic of immunoglobulins of different classes and species (Fig. 3). Except for a few IgA proteins, most immunoglobulins have an L–H interchain bridge, but the bridge from the C-terminus of the L chain may link to the C_H1 domain or to the hinge region. This is not surprising since the two sites on the H chain are only a few angstroms apart in the three-dimensional structure (Davies *et al.*, 1975; Amzel and Poljak, 1979; Davies and Metzger, 1983). As shown previously in Fig. 3, the location of the L–H bridge differs within the human and also within the mouse IgG subclasses. Although the principal interchain H–H disulfide bridges are located in the hinge region, the number of such bridges also differs within both the human and the mouse IgG subclasses. This is not a trivial structural detail, because the stability of an antibody molecule and probably its sensitivity threshold for transmitting an induced response from the antigen combining site both depend on the conformational flexibility of the hinge region (Metzger, 1978; Pecht, 1982). Unfortunately, only one crystallographic structure exists for a complete hinge-containing immunoglobulin molecule, the IgG1 myeloma protein Kol (Deisenhofer, 1981). In addition to inter-H disulfide bridges like those in monomeric IgG, IgE, and IgD, the poly-Igs (IgA

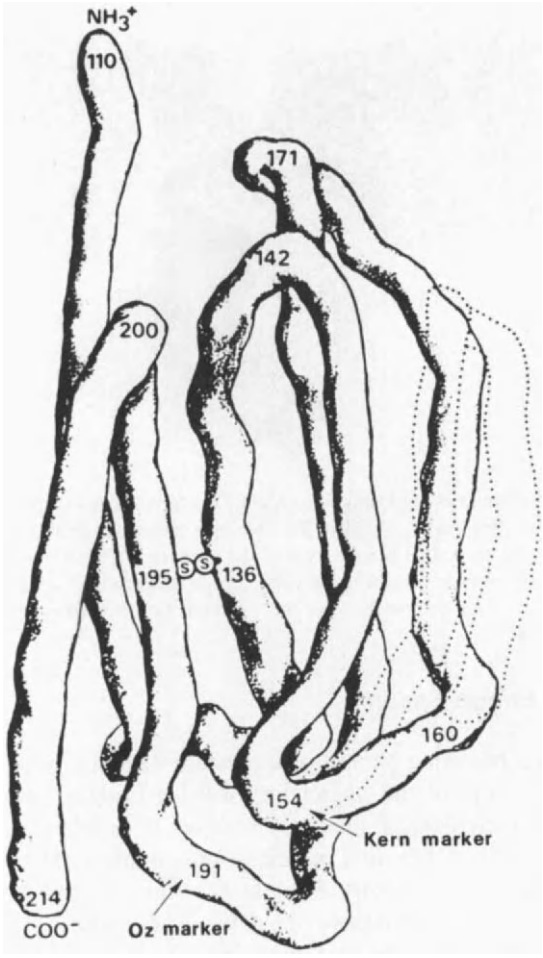


Fig. 8. Diagram of the basic "immunoglobulin fold." The solid trace shows the folding of the polypeptide chain in the constant subunit (C_L, C_H1, etc.). Numbers designate human λ light chain C region (Cλ) residues, beginning at NH_3^+, which corresponds to residue 110 for the λ chain. Broken lines indicate the additional loop of polypeptide chain characteristic of the V_L and V_H subunits (from Poljak *et al.*, 1973).

and IgM) contain disulfide bonds linking the monomer units and the J chain and possibly also have labile S–S bonds, but these are not well established.

3. *The Hinge Region and the Fab and Fc Fragments*

The hinge region is an unusual structure in the segment of the heavy chain that joins the Fd and Fc regions of immunoglobulin classes having only three C_H

domains (IgG, IgA, and IgD) but is absent in IgM and IgE, which have four C_H domains (Table II). The amino acid sequence of the hinge region is unique for each class, differs markedly even for subclasses, and appears to be unrelated to the rest of the H chain (Putnam, 1977a; Putnam *et al.*, 1985). This is illustrated in Fig. 9 for the hinge regions of human IgG, IgA, and IgD. In γ and α chains the hinge region is rich in proline and cysteine; thus, it probably has a random flexible structure that pivots on multiple interchain disulfide bridges conveying a segmental flexibility that is thought to transduce a signal from the antibody combining site to the biological effector domains of Fc (Metzger, 1978; Pecht, 1982). The discovery that the hinge region of IgG is encoded precisely by a separate exon (Sakano *et al.*, 1979) helped explain the frequency of deletions and duplications in the hinge regions of IgG and IgA subclasses. However, the γ and α hinge exons do not seem to fit into the evolutionary scheme linking all V and C region domains to a common ancestral gene that coded for a single domain (Putnam *et al.*, 1982). A previous figure (Fig. 3) illustrated the variation in the interchain disulfide bridging pattern of the hinge regions of a number of immunoglobulins of several species.

As an exposed flexible structure the hinge region itself (IgD) or its neighboring sequences (IgG) are readily subject to limited proteolytic cleavage to produce Fab and Fc fragments. This is not the case for IgA, which is very resistant to proteases because of its unusual proline-rich structure that is protected by GalN sugars in human IgA1 (Kratzin *et al.*, 1975; Liu *et al.*, 1976) and by a pentaproline sequence in IgA2 (Torano *et al.*, 1977; Torano and Putnam, 1978; Tsuzukida *et al.*, 1979). IgM and IgE lack a hinge, but have an extra domain. However, IgA1 is cleaved at a prolyl-threonyl bond in the hinge region by microbial IgA proteases, but IgA2, which lacks this bond, is not susceptible (Plaut *et al.*, 1975; Plaut, 1978). In IgM the extra domain is disintegrated by trypsin, which produces Fab fragments and a cyclic Fc fragment (Lin and Putnam, 1978). The Fab and Fc fragments of IgG have been invaluable for crystallographic analysis and for study of the antibody combining sites of Fab and the biological effector functions of Fc (Dorrington, 1979). Some of the fragments are very labile, notably the Fab of IgD (Figs. 10 and 11) (Lin and Putnam, 1979; Ishioka *et al.*, 1987). The maximum yields of Fab, Fc, and of two rather stable hinge peptides (Fig. 10) are obtained rapidly after incubation. Figure 11 shows the kinetics of tryptic cleavage at 37°C of a human myeloma IgD protein (WAH). Although the Fc of IgD is quite stable, the Fab is rapidly degraded proteolytically into a series of fragments. These results cast doubt on measurements of serum IgD made with antibodies only to the Fc fragment, such as most commercial antisera for IgD.

Under mild conditions of digestion proteolytic enzymes may also cut other interdomain sequences to yield a great variety of fragments. The most common of these are listed in Table IV. The fragments correspond to domains or to a

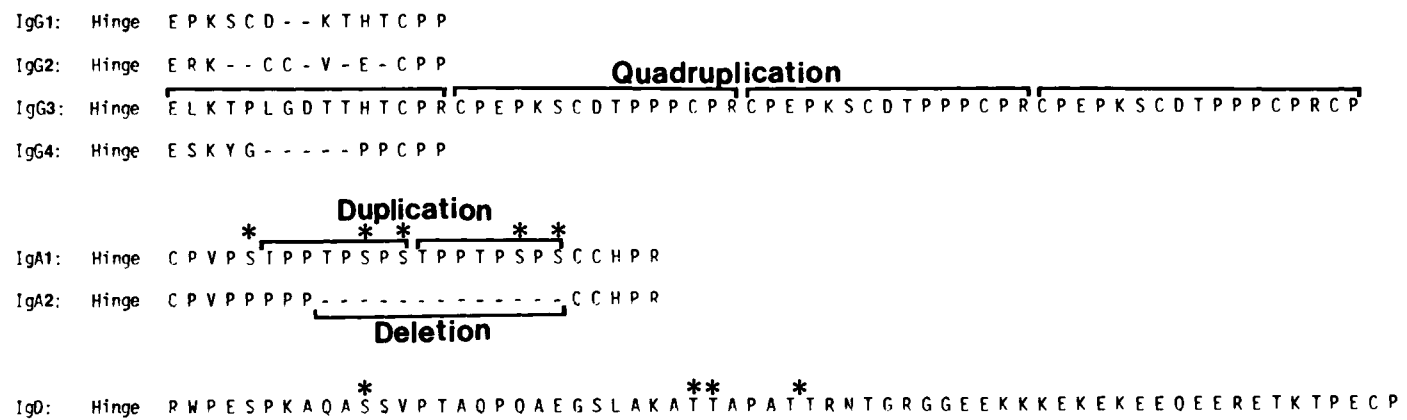


Fig. 9. Amino acid sequence of the hinge regions of human IgG, IgA, and IgD. Gaps have been inserted in the sequences for the IgG subclasses to show maximum homology. The primary structures of animal IgG hinges are similar; they range from 13 to 20 residues and are rich in cysteine and proline. See Kabat *et al.* (1983) for a computer printout comparing many hinge regions. Note the quadruplication in the human IgG3 hinge, the internal duplication in IgA1, and the deletion in IgA2. The hinge regions shown are based on protein sequence analysis and sites of proteolytic cleavage, but they correspond closely to the individual gene segments (exons) that code for the hinges. Aberrant recombination of immunoglobulin gene segments, as in heavy chain disease, may result in deletion of sequences preceding the hinge region or in deletion of the whole hinge or of one or more of the quadruplicated segments in the IgG3 hinge (Frangione and Franklin, 1979; Alexander *et al.*, 1982). Like the α1 hinge, the δ hinge consists of a protease-resistant polypeptide that contains a series of galactosamine oligosaccharides at sites marked by the asterisks. However, the δ hinge has a second segment that is highly charged and very susceptible to proteolytic cleavage (Putnam *et al.*, 1982). The two segments are encoded by separate exons (see Fig. 10).

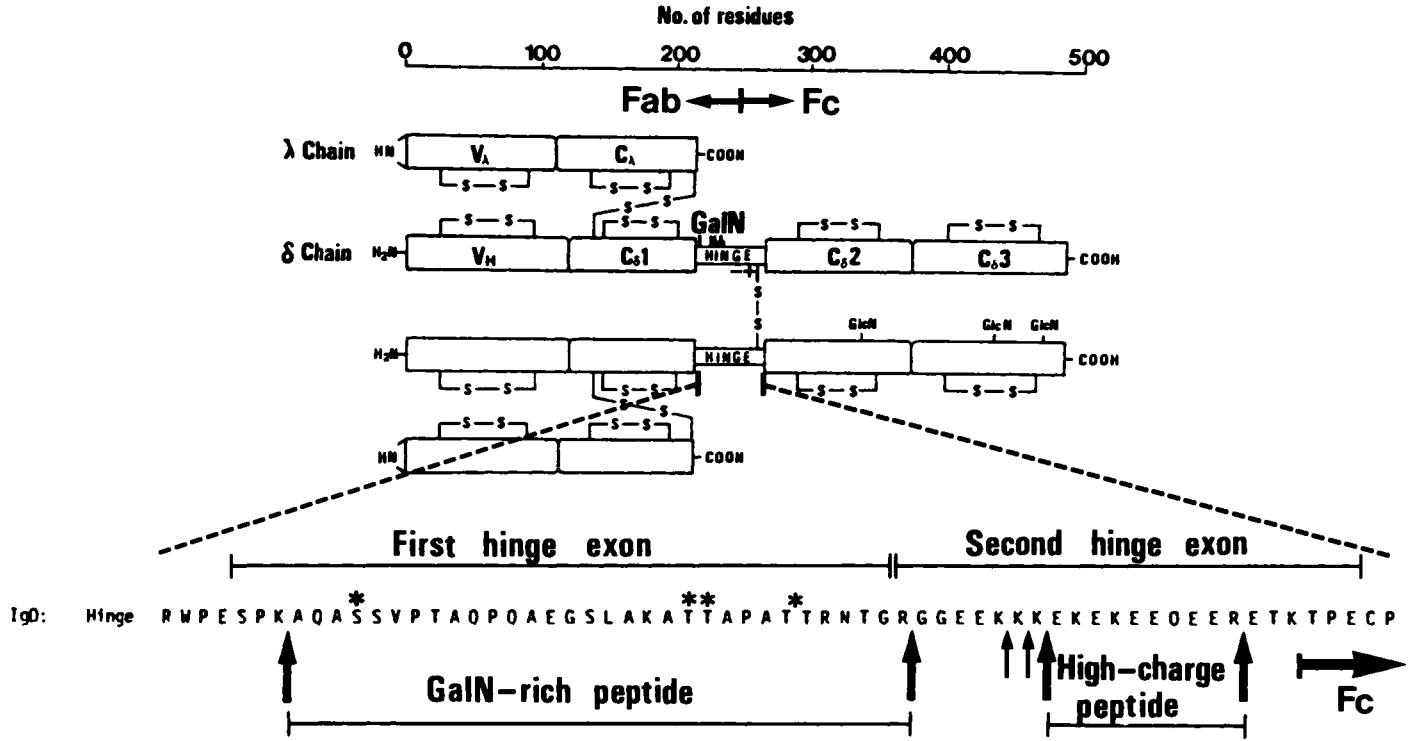


Fig. 10. Schematic molecular model of human IgD showing the amino acid sequence of the hinge region on an expanded scale. Large arrows indicate the sites of cleavage by trypsin to yield the GalN-rich peptide and the high-charge peptide referred to in Fig. 11. Small arrows indicate additional sites of tryptic cleavage within the hinge region, and asterisks denote sites of attachment of a GalN oligosaccharide (from Ishioka *et al.*, 1987).

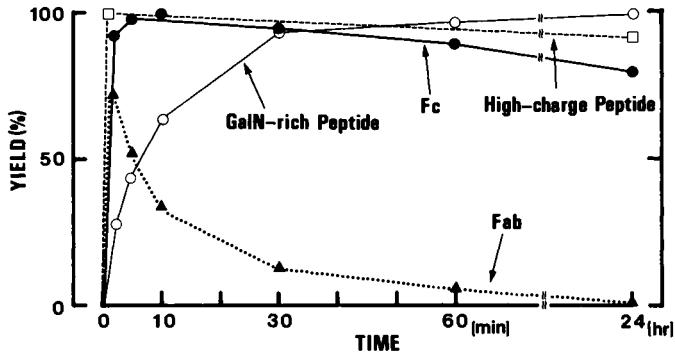


Fig. 11. Kinetics of tryptic cleavage at 37°C of a human IgD myeloma protein (WAH). Within 6 min after incubation of the undenatured IgD the maximum yields of Fab, Fc, and a high-charge peptide are obtained. The GalN-rich peptide approaches a maximum yield at 30 min, at which time Fab is largely degraded to a series of peptides derived from both the δ chain and the light chain. In contrast, the two hinge peptides decline little or not at all after proteolysis for 24 hr, and Fc is still present at about 75% yield (from Ishioka *et al.*, 1987).

TABLE IV

Typical Fragments of Immunoglobulins Obtained by Mild Proteolytic Digestion[a]

Class	Domains	Fragment	Enzyme
IgG	$(V_L–C_L)$ $(V_H–C_H1)$	Fab	Papain 37°
	$[(V_L–C_L)$ $(V_H–C_H1)]_2$	$(Fab')_2$	Pepsin 37°
	$(C_H2–C_H3)_2$	Fc	Papain 37°
	$(V–C_H1–C_H2)_2$	Fabc	Plasmin 37°
	$(C_H3)_2$	Fc′	Pepsin 37°
IgA1	$(V_L–C_L)$ $(V_H–C\alpha1)$	Fabα	IgA1 protease 37°
	$(C\alpha2–C\alpha3)_2$	Fcα	IgA1 protease 37°
	(V_L,V_H)	Fv	Pepsin 37°
IgM	(V_L,V_H)	Fv	Pepsin 4°
	$(V_L–C_L)$ $(V_H–C\mu1)$	Fabμ	Trypsin 60°
	$(C\mu3–C\mu4)_{10}$	$(Fc)_{5\mu}$	Trypsin 60°
	$(C\mu2)_4$	$F(c\mu2)_4$	Pepsin 4°

[a]For information about the preparation and properties of the fragments see Nisonoff *et al.* (1975) for IgG, Putnam *et al.* (1979) for IgA1, Lin and Putnam (1978) for IgM, and Lin and Putnam (1979) for IgD. To achieve specific limited cleavage the undenatured protein is incubated with enzyme under well-defined mild conditions for periods ranging from a few minutes (IgD) to up to 24 hours (IgG, IgA).

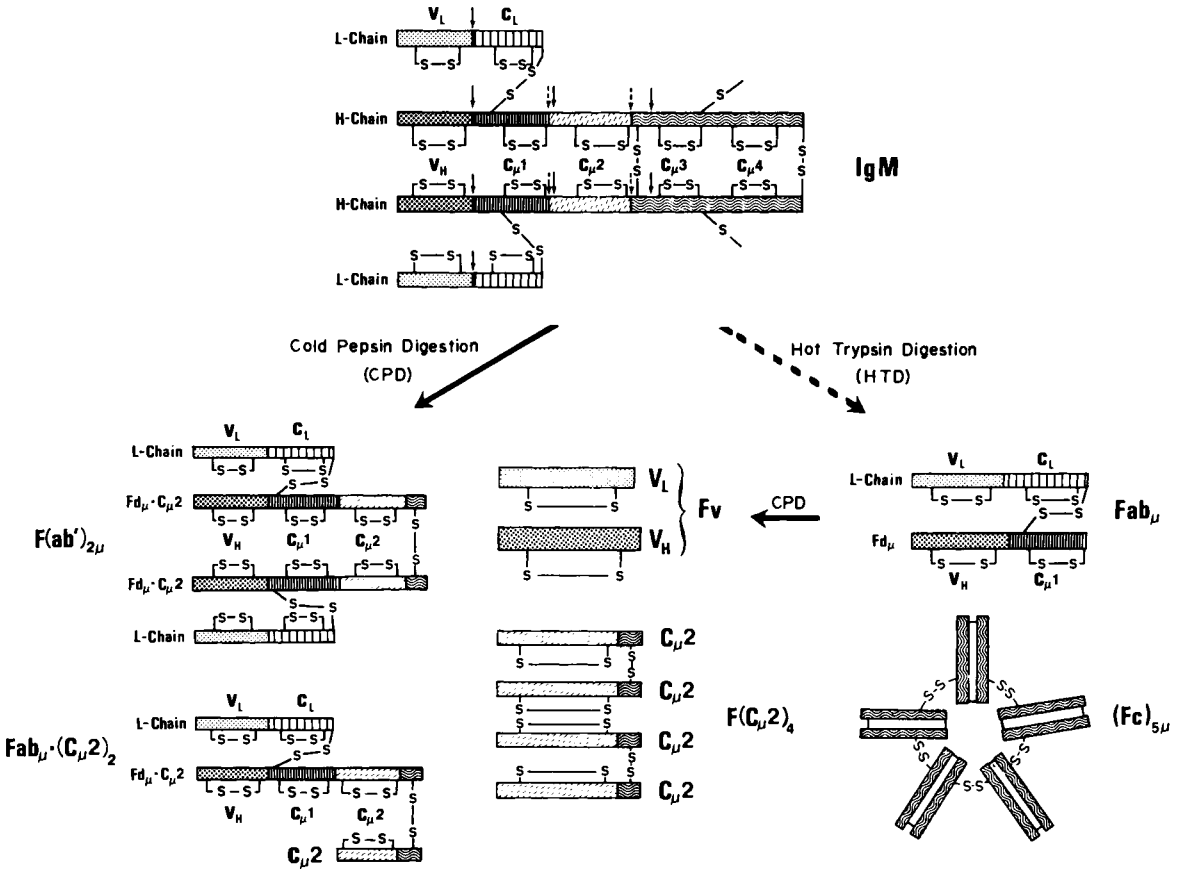
series of connected domains. Generally these are linked by disulfide bonds, but some are held together by noncovalent forces. The number and the kinds of fragments depend on many factors, such as the class, subclass, and species of the immunoglobulin, the protease, and the conditions of digestion such as pH and temperature. For example, although trypsin is used at 60°C ("hot trypsin") to prepare the Fabμ fragment and the pentameric Fc fragment $(Fc)_5\mu$ of IgM, pepsin is used at 4°C ("cold pepsin") to prepare the Fv fragment of human IgM, and also a series of other fragments (Fig. 12) (Lin and Putnam, 1978). Such fragments have been valuable in localizing the sites of biological activities.

4. J Chain and Other Structural Features

Other structural features are typical of certain immunoglobulin classes; these are carbohydrate, the J (joining) chain, and the secretory component (SC). Carbohydrate, which is described in a later section, is present to the extent of one or more polysaccharide units on all H chains. Covalent polymer formation is typical of IgA and IgM. As described below, the polymers are joined via the J chain and disulfide bonds (Koshland, 1985). The polymers may also be complexed with a very large polypeptide chain earlier called secretory piece or secretory component and now known to be part of the transepithelial membrane receptor for poly-Igs (poly-IgR) (Mostov *et al.*, 1984). As discussed in a later section, poly-IgR has a series of repeating domains that have some structural homology to immunoglobulins; however, neither poly-IgR nor the J chain have the division into V and C regions that is the signature of immunoglobulin chains. Nonetheless, J chain appears to have a β-pleated sheet structure similar to the immunoglobulin fold (Zikan *et al.*, 1985).

The J chain has been thoroughly reviewed by Koshland (1985), who emphasizes that it is a third immunoglobulin polypeptide that is often overlooked. Although J chain does not contribute to antibody specificity, it must participate in biological effector functions specific for polymeric IgA and IgM because it is covalently linked to Fc. The amino acid sequences of human and mouse J chains have been reported by Mole *et al.* (1977) and Cann *et al.* (1982), respectively, but the three-dimensional structures are unknown. Of the 137 residues in each J chain, 77% are identical, including the eight cysteines.

A two-domain model has been proposed for J chain by Cann *et al.* (1982). In this model two α chains of IgA (or two μ chains of IgM) are linked via an interchain disulfide bridge to the amino-terminal half of one J chain. The linkage is to the cysteine that is the penultimate residue in the carboxy-terminal tail of each heavy chain. The J chain fits well with either IgA or IgM because of the high degree of sequence homology of their carboxy-terminal domains. The stoichiometry is one J chain per polymer, and the hypothesis is that formation of a J chain-containing dimer facilitates polymer formation.


L-Chain
H-Chain
H-Chain
L-Chain
IgM
Cold Pepsin Digestion
(CPD)
Hot Trypsin Digestion
(HTD)
F(ab')2μ
Fab'μ·(Cμ2)2
Fv
CPD
Fabμ
F(Cμ2)4
(Fc)5μ

III. The Variable Region

A. Light Chains

1. Variability in Amino Acid Sequence

More sequence analysis has been done on the light chains of immunoglobulins than on any other group of proteins. Only computerized data bases such as those maintained by Barker *et al.* (1985) and Kabat *et al.* (1983) can keep track of the vast accumulation of data and the continuing flow of new sequences. Partial sequences of a series of peptides of human κ light chains (Bence–Jones proteins) were first reported by Titani and Putnam (1965) with more extensive data soon after from Titani *et al.* (1965) and Hilschmann and Craig (1965). Within a year the first complete sequence of any light chain (the human κ Bence–Jones protein Ag) was published (Putnam *et al.*, 1966). In the next year the complete sequences of three human λ proteins were determined (Putnam *et al.*, 1967a,b). Remarkably, the three proteins differed from one another at 40 to 50 positions in the amino-terminal half of the light chain (now called the variable region). but their sequences were identical in the carboxyl half of the chain (now called the constant region) (Fig. 13). There was no precedent for such variation in the sequence of homologous proteins from individuals of the same species, nor any parallel for the flood of structures that resulted from this work. Kabat *et al.* (1983) list 184 human κ light chains for which sequences have been published; of these 37 are complete or nearly so. The same compilation also lists 97 human λ sequences of which 36 are essentially complete. As a previous figure shows (Fig. 1), immunoglobulin sequence data have continued to increase at an exponential rate, doubling every few years. The advent of gene cloning and DNA sequencing has kept up the logarithmic pace. All this activity attests to the great interest in

---

Fig. 12. Enzymatic fragmentation of human IgM proteins and the structural characteristics of the proteolytic fragments. Cold pepsin digestion (CPD) degrades the Fc region and produces the Fv, the $F(ab')_{2\mu}$, the $Fab\mu{\cdot}(C\mu2)_2$, the Fab_μ (not shown), and the $F(C\mu2)_4$ fragments. Hot trypsin digestion (HTD) degrades the Cμ2 domain to form the Fab_μ and the $(Fc)_{5\mu}$ fragments (Florent *et al.*, 1974; Plaut and Tomasi, 1970). The Fv fragment can also be produced from the tryptic Fab_μ fragment by cold pepsin digestion. In the diagram, a schematic monomeric subunit is used to represent the actual pentameric structure of human IgM molecules. The solid arrows indicate the major peptic cleavage sites and the dashed arrow indicates the major tryptic cleavage sites. The diagrams for IgM, $F(ab')_{2\mu}$, and $Fab_\mu{\cdot}(C\mu2)_2$ are drawn to the same scale, those for Fv and $F(C\mu2)_4$ are magnified, and the diagram for $(Fc)_{5\mu}$ is reduced in scale. The individual domains of the H and L chains are identified by different shading (from Lin and Putnam, 1978).

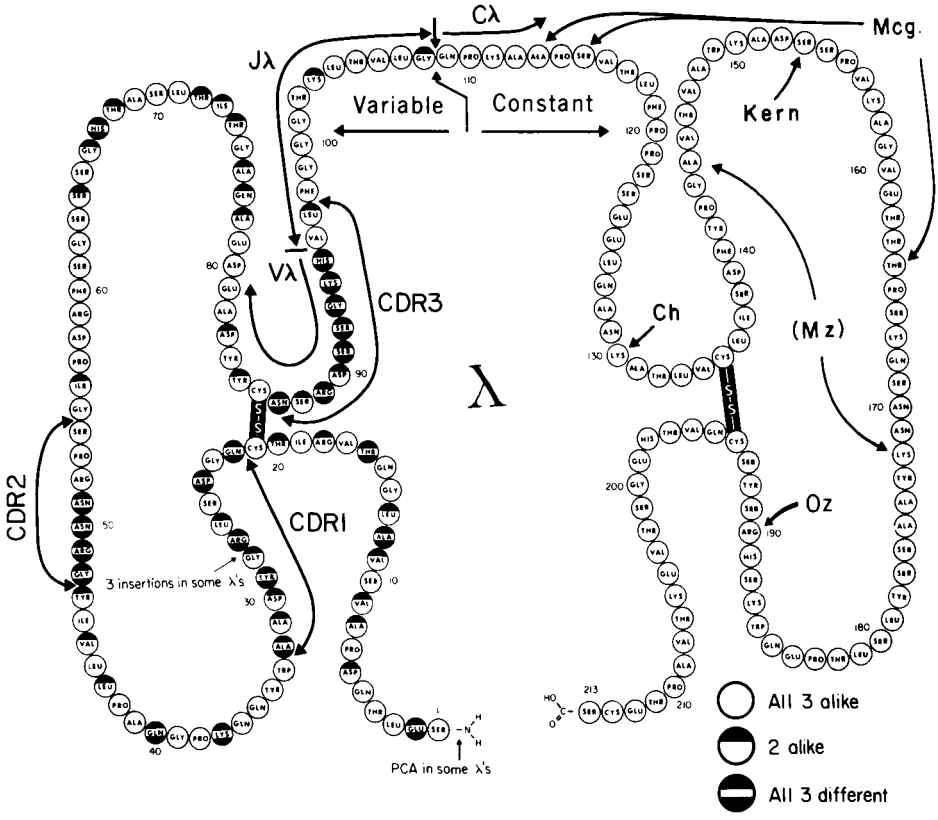


Fig. 13. Amino acid sequence of the human λ Bence–Jones protein Sh. Positions given in white circles are identical in the human λ light chains Ha and Bo. Where the circle is black at the top, two of the three λ chains have the same amino acid but differ from the third. All three proteins differ in positions where the circles are black at the top and bottom. The numbering system is for the Sh λ chain (the first to be sequenced) and differs by one residue from most other λ chains. In the C region the positions of amino acid replacement in the isotypes Oz, Kern, and Mcg, and also positions where replacements have been reported in two other proteins, Mz and Ch, are indicated by arrows. Parentheses around Mz indicate that only the amino acid composition was determined for peptides containing the substituted residues, rather than the complete amino acid sequence of the light chain as was done for the other isotypes. Brackets at the top of the figure identify the three gene segments (Vλ, Jλ, and Cλ) that were rearranged and combined to form the complete gene for the λ chain. Arrows point to the three hypervariable sections in the λ chain sequence that are now designated complementarity-determining regions CDR1, CDR2, and CDR3 (modified from Putnam, 1977a).

solving the enigma of antibody diversity, the key to which lay in the variability in protein sequence and the genetic mechanism by which this was generated.

The essential points of light chain structure were early established by study of Bence–Jones proteins from individual patients with multiple myeloma or mice with different clones of myeloma cells. The dramatic findings revolutionized thinking about the antibody problem and attracted many new entrants into the field. The salient discovery was that no human κ or λ chain from one individual had the same amino acid sequence as the light chain from any other individual; furthermore, the variability was precisely restricted to the N-terminal half of the light chain (Vκ or Vλ). As sequence data accumulated, it was recognized that both Vκ and Vλ sequences could be classified into a series of subgroups based on homology (e.g., Vκ I, Vκ II, Vκ III or κ subgroups I, II, III; also Vλ I, Vλ II, etc.). More important, it is easy to see by visual inspection of Fig. 13 that there are three hypervariable regions in Vλ. This is also true in Vκ. Computer analysis of many sequences localized the hypervariability to three fairly well defined segments, i.e., positions 24–34, 50–56, and 89–97 (Wu and Kabat, 1970). The position numbers vary a bit because of somewhat different lengths of the V regions of light chains. These hypervariable regions are now called complementarity-determining segments and are designated CDR1, CDR2, and CDR3, respectively (Kabat *et al.*, 1983). The remainder of the V region where the sequence is more conserved is called the framework and consists of four clusters or framework segments denoted FR1 to FR4. The order in the V region is FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4. As shown in Fig. 14 and discussed later, V regions of the heavy chains have a similar distribution of FR and CDR segments. It should be noted from Figs. 13 and 14 that only about one-third of the positions in the Vκ and Vλ sequences are hypervariable, and two-thirds of the residues change infrequently.

Although the general characteristics of Vκ and Vλ sequences are similar, including the location of CDR1, CDR2, and CDR3, and in homology with some of the framework residues, nonetheless there is one important distinction, namely, light chain V regions are joined only to C regions of the same class. That is, Vκ sequences are linked only to the Cκ sequence, and Vλ sequences are linked only to a Cλ sequence. Hybrid unions such as VκCλ cannot be formed because κ genes and λ genes are on different chromosomes, and the joining of V_L and C_L is governed at the nucleic acid level rather than at the protein level.

2. Structure and Rearrangement of Light Chain Genes

The solution to the problem of the genetic control of immunoglobulin sequence variability and thus of antibody diversity began in 1978 with the cloning and sequencing of the genes for mouse immunoglobulin light chains by Tonegawa's group, Leder's group, and many others (Bernard *et al.*, 1978; Brack *et*

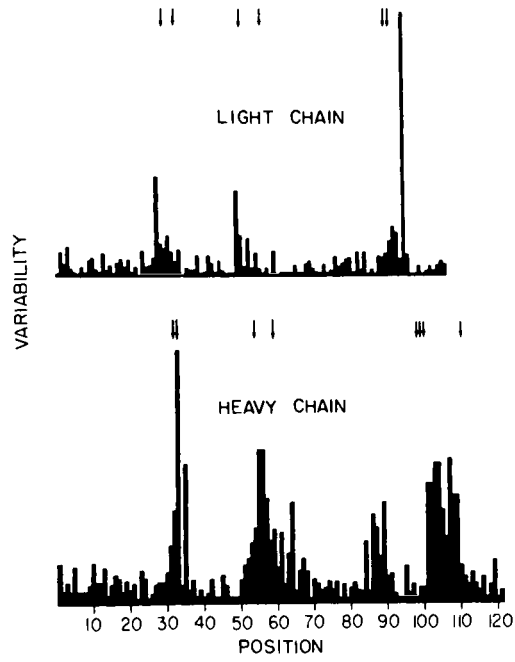


Fig. 14. Variability at different positions for immunoglobulin light and heavy chains of all species. The variability equals the number of different amino acids at a given position divided by the frequency of the most common amino acid at that position (Wu and Kabat, 1970). The peaks of greatest variability around positions 30, 55, and 95 (105 in the heavy chain) correspond to the three hypervariable or complementarity-determining regions CDR1, CDR2, and CDR3. The peaks also coincide with positions that have been localized in the combining site of antibodies by use of affinity labels (from Kindt and Capra, 1984).

al., 1978; Seidman *et al.*, 1978; Seidman and Leder, 1978). Numerous reports since then on the genes of light and heavy chains of mice and humans have been summarized in many reviews (Rabbitts *et al.*, 1981; Leder *et al.*, 1981; Leder, 1982; Honjo, 1983; Hood *et al.*, 1983; Wall, 1983; Fudenberg *et al.*, 1984; Korsmeyer and Waldmann, 1984; Tonegawa, 1985; Honjo and Habu, 1985). A large number of molecular biologists, who were impelled by the genetic aspects of the antibody problem, entered this research area. They built on the structural basis first established by amino acid sequence analysis and later pictured in fine detail by X-ray crystallography. The molecular biologists cloned the genes of mice with multiple myeloma, and also the embryonic genes, and later the human genes. They determined the cDNA sequence and later the genomic sequences and identified the coding elements (exons) and the intervening sequences (introns, IVS), and the rearrangement of the RNA to yield mRNA.

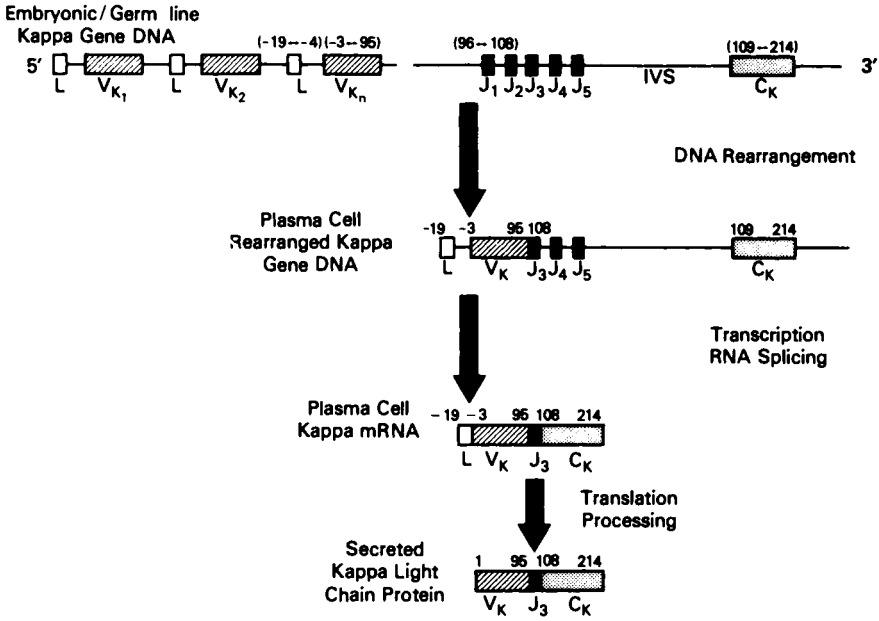


Fig. 15. Schematic representation of the human κ gene locus. Multiple germ-line variable (Vκ) regions exist, each accompanied by a leader (L) sequence. There are five alternating joining (Jκ) segments, each coding for amino acid positions 96–108. There is only one constant (Cκ) region per allele. DNA rearrangement joins a single Vκ and Jκ segment. The remaining intervening sequences (IVS) are removed by RNA splicing (from Korsmeyer and Waldmann, 1984).

The results of gene cloning and nucleotide sequencing confirmed the early prediction of Dreyer and Bennett (1965) made on the basis of amino acid sequence analysis that the V and C region domains of light chains are encoded by separate segments of DNA (exons) that are rearranged and joined. Three major new findings were made: (1) The presence of multiple V genes—up to 100—was established. (2) A new genetic element, the J minigene, was discovered. (3) the genomic arrangement of the genetic elements and the mechanism of DNA and RNA rearrangement were elucidated. The exons are separated by noncoding segments of DNA (introns) which are removed during transcription into pre-RNA and during its further processing into mRNA prior to translation of the mRNA into protein. This process is illustrated in Fig. 15 for κ light chains.

The same principles of gene recombination apply for λ light chains except for an important difference in germ-line arrangement. In both mice and humans the Cλ locus is more complex than Cκ. Whereas there may be allelic forms of Cκ, there is only one Cκ gene in both species. However, there are at least four mouse Cλ genes (Blomberg *et al.*, 1981; Miller *et al.*, 1981), and there are six human λ-

like constant region genes, of which at least three are expressed (Hieter *et al.*, 1981). Furthermore, each mouse Cλ gene has its own J region, whereas there are four or five Jκ regions available for recombination with the multiple Vκ genes and the single Cκ gene in both species. The organization of the Jλ segments in humans has yet to be clarified, but it is probably similar to that in mice.

Current research is directed toward elucidating the fine controls on the genetic mechanism for rearrangement of κ and λ genes and is outside the scope of this chapter. Recent studies have focused on regulatory nucleotide sequences (enhancers) that affect the recombination and expression of κ genes (Bergman *et al.*, 1984; Potter *et al.*, 1984; Lewis *et al.*, 1985). The number of V region genes in the light chain genomic repertoire has not yet been established; however, evidence is mounting that there may be more than 100 but less than 300 Vκ segments in the mouse genome (Nishi *et al.*, 1985).

B. Heavy Chains

1. Sharing of Variable Region (V_H) Sequences by Heavy Chains of Different Classes

The general characteristics of the V_H region of heavy chains are similar to those of the V_L region of light chains in most respects but differ in one fundamental way. The similarities include: (1) the restriction of the variability to the first (i.e., V_H) domain, but this is about 10 residues longer than the V_L domain because of the addition of a D_H diversity segment; (2) the localization of hypervariability to three segments, also called CDR1, CDR2, and CDR3, but CDR3 in H chains is longer and even more variable than in L chains (see Fig. 14); (3) the classification of V_H region sequences into a number of subgroups (V_{HI}, V_{HII}, V_{HIII} or I, II, III) based on sequence homology within the subgroups; and (4) the presence of a framework structure similar to that of V_L but characteristic for V_H.

Although there are significant differences in the number, size, and structure of C_H regions compared to C_L regions, the fundamental difference between light and heavy chains is that the same V_H subgroups are shared by all classes of heavy chains, which was discovered by Kohler *et al.* (1970); in contrast, Vκ subgroups are joined only to Cκ, and Vλ subgroups are joined only to Cλ. This critical difference is due to the fact that the H chain genes are all clustered together on one chromosome (chromosome 14 in humans), whereas κ chain genes and λ chain genes are each on a separate chromosome (chromosomes 2 and 22, respectively, in humans). The clustering of the H chain gene locus is the basis for "class switching" whereby the first antibodies formed in response to an antigen may be IgM and later ones may be IgG, but both classes have the same specificity and may have the same light chains. Thus, the class character of heavy chains is expressed only in the constant region, whereas idiotypic dif-

ferences—and thus antigen-specific differences in primary structure—are restricted to the V_H region.

2. Structure and Rearrangement of Heavy Chain Genes

The complex arrangement of the immunoglobulin heavy chain gene (Igh) locus is illustrated in Figs. 16 and 17. The arrangement is similar in mice and humans except for a distribution of the human C_H genes into two clusters suggestive of gene duplication (Flanagan and Rabbitts, 1982). However, some aspects of the human Igh locus are yet to be worked out (Section IV,B,1). Soon after the discovery of the mechanism for light chain gene assembly, a similar mechanism was found to control the Igh genes (Sakano *et al.*, 1979; Early *et al.*, 1979; Tucker *et al.*, 1979). Sakano *et al.* (1979) cloned the gene coding for the entire C region of the mouse γ1 chain gene, determined the DNA sequence, inferred the amino acid sequence, and showed that the exon/intron structure correlated precisely with the protein domains and the hinge region. Although the exon/domain correlation is omitted in Figs. 16 and 17 in order to simplify the

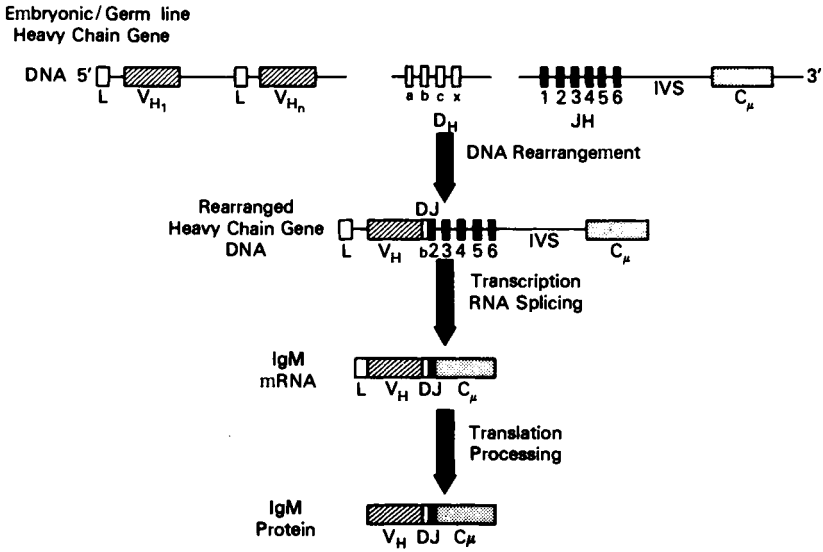


Fig. 16. Schematic model of the organization and assembly of the human heavy chain gene. In addition to multiple variable (V_H) regions with leader (L) sequences, there are six functional joining (J_H) segments and families of diversity (D_H) segments. Single V_H, D_H, and J_H regions are recombined at the DNA level. RNA splicing later removes the residual intervening sequences (IVS). To simplify the diagram, the exon/intron structure of Cμ is not shown (from Korsmeyer and Waldmann, 1984).

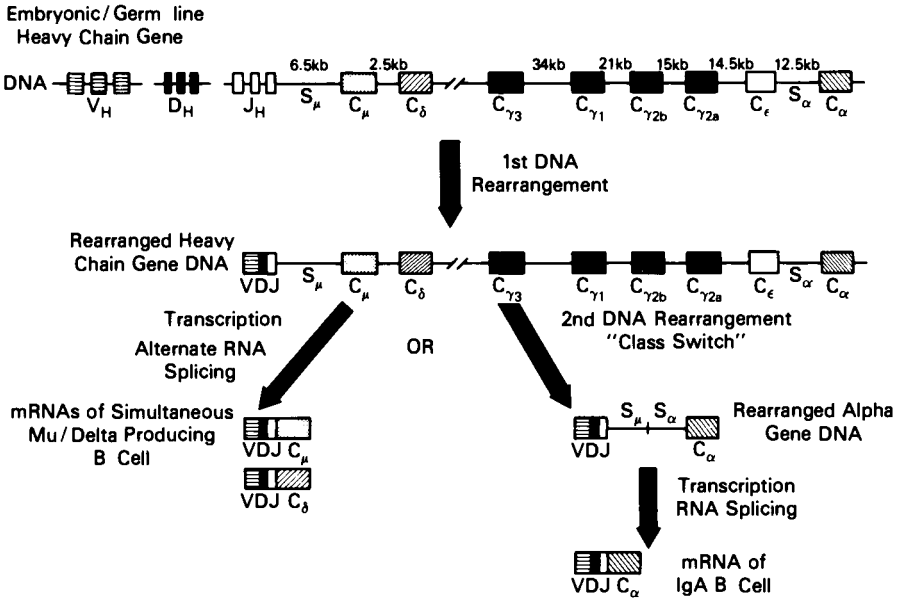


Fig. 17. Schematic diagram of the mouse heavy chain gene locus, revealing the constant region gene order and spacing. Following the initial DNA rearrangement recombining a V_H, D_H, and J_H region, a B cell can utilize alternative sites of RNA splicing to simultaneously produce IgM and IgD. Alternatively, such a B cell can further differentiate and switch to production of another heavy chain class. For example, a second DNA recombination at the highly homologous switch sites (Sμ and Sα) in front of the Cμ and Cα genes would result in IgA production. Similar homologous switch sites (not shown here) are found in front of each of the constant regions. To simplify the diagram the exon/intron structure of the C_H genes is not shown (from Korsmeyer and Waldmann, 1984).

diagrams, it is shown in Fig. 18. Intervening sequences were also found to divide the mouse Cμ gene into segments, each of which encoded a domain (Gough *et al.*, 1980). The successive events involved in the somatic rearrangement and recombination of the V and C region genes for the μ and α chains were next identified (Davis *et al.*, 1980; Early *et al.*, 1980; Cory and Adams, 1980; Cory *et al.*, 1980; Newell *et al.*, 1980; Maki *et al.*, 1980; Rabbitts *et al.*, 1981).

Recombinant DNA research thus showed that an immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D_H, and J_H (Schilling *et al.*, 1980; Alt and Baltimore, 1982). As indicated in Fig. 16, the three kinds of segments are represented in the genome (1) by a large but unknown number of V_H exons ($n \simeq 100$), each with its own leader sequence, (2) by several families of diversity (D_H) segments, and (3) by a small cluster of J_H joining elements. In the first recombination step the heavy chain gene DNA is rearranged so that single V_H, D_H, and J_H DNA segments are selected and are

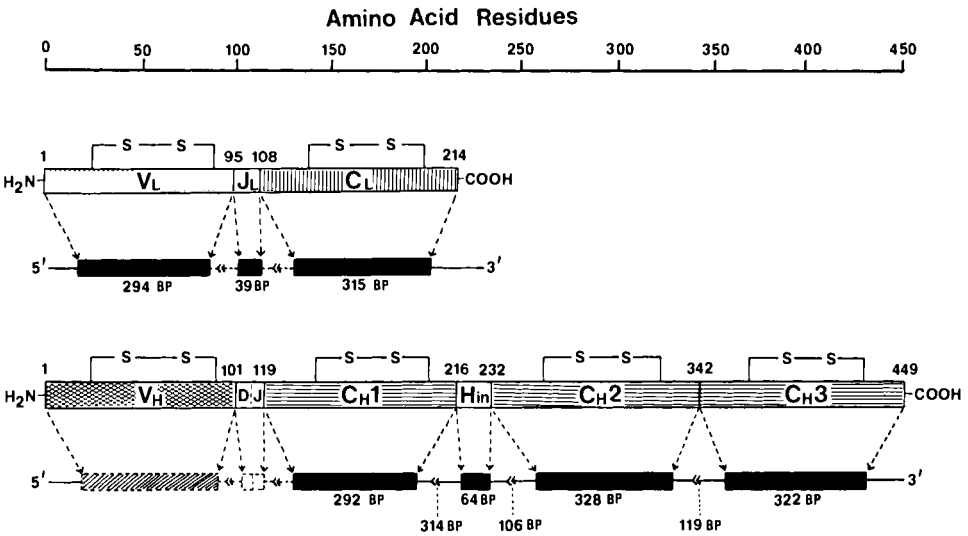


Fig. 18. Schematic summary of the correlation of protein structure of an IgG antibody molecule with the genomic structure of exons and introns for light and heavy chains. The structural domains of the light and heavy chains are identified by differences in shading. Smaller joining segments are denoted as J_L for the light chain and D, J, and Hin (hinge) for the heavy chain, but the C-terminal tailpieces (Ts and Tm) and their exons are omitted for the H chain. The numbers above polypeptide structures give the approximate residue positions. The numbers for the polynucleotides give the length of the exons or introns in base pairs (BP) (modified from Putnam, 1983).

joined to form a complete V gene, which is still separated from the C region gene. The choice of V_H gene segment may not be random and may shift as the B-cell population matures (Yancopoulos *et al.*, 1984). After transcription of the rearranged DNA the remaining introns are removed by splicing, which joins the four coding elements (V_H, D_H, J_H, and C_H) to form the mRNA. The mRNA is translated and secreted with removal of the leader sequence. Section VII,A discusses the mechanism for synthesis of the secreted and membrane forms of immunoglobulins (sIg and mIg). This involves selection and splicing of DNA segments for the corresponding tailpieces (TPs and TPm).

3. Class Switching

The process of exon shuffling to generate a complete heavy chain gene is coupled to the class switch, the mechanism by which a B-cell clone shifts from production of IgM to another class such as IgG. This second DNA arrangement is illustrated in Fig. 17 for the switch from IgM to IgA. The beauty of this process is that the same specificity region (VDJ) is retained while the effector

region of the antibody molecule (C_H) is changed. Class switching is facilitated by a segment of DNA known as the S region (Kataoka *et al.*, 1981). The S region (switch site or switch sequence) is class-specific, and there is one S region upstream in the intron preceding each Igh gene. In order to simplify the diagram, only Sμ and Sα are shown in Fig. 17, but homologous S regions are present in front of each Igh gene. Honjo (1983) has reviewed the structure of the S regions and the role that they play in facilitating the class switch.

Much current research deals with the role of other nucleotide sequences known as enhancers because they promote and amplify the expression of immunoglobulin genes. For example, there is an enhancer present in the J–C intron of the human and mouse κ gene that is necessary for efficient expression of the cloned gene (Potter *et al.*, 1984; Bergman *et al.*, 1984; Falkner and Zachau, 1984). A tissue-specific transcription enhancer is located in the intron in front of Sμ in the rearranged Igh genes (Gillies *et al.*, 1983).

4. Aberrant Recombinations: Chromosomal Translocations and Heavy Chain Disease Proteins

Because the immunoglobulin gene loci have such a propensity for rearrangement, it is not surprising that they are sometimes translocated in human B-cell tumors. In one chromosomal translocation characteristic of Burkitt's lymphoma, the oncogene c-*myc* on chromosome 8 translocates to the Sμ switch just 5′ to the Cμ gene on chromosome 14; there is a reciprocal translocation of the V_H–D_H–J_H region to chromosome 8 (Klein and Klein, 1985). The translocated c-*myc* oncogene is transcriptionally active. If an enhancer element is present, the oncogene may be activated because an enhancer is active whether it is upstream of the gene, or within it, or in both orientations (Bergman *et al.*, 1984). A similar translocation has been associated with T-cell tumors (Baer *et al.*, 1985), and an unprecedented rearrangement of immunoglobulin V_H gene segments occurred. The V_H gene was joined with the JαCα gene segment of the T-cell receptor α chain. This novel V_H–JαCα rearrangement was productive at the genomic level and could encode a hybrid immunoglobulin/T-cell receptor polypeptide.

Heavy chain disease (HCD) proteins are probably another example of aberrant recombination. These defective proteins produced by rare patients lack light chains and have truncated heavy chains in which one or more domains or the hinge region may be missing or shortened. The most common feature of HCD proteins is an internal deletion of most of the V region and of the entire C_H1 domain and sometimes the hinge, but the normal C region sequence initiates after the deletion. As reviewed by Franklin and Frangione (1975) and in Volume III of this treatise (Putnam, 1977a), a number of HCD proteins have been partially sequenced by protein sequencing. The gene sequence has been analyzed for one IgG3 HCD protein by Alexander *et al.* (1982), who suggest that the protein

abnormality results from a partial gene deletion rather than defective splicing. The genetic mechanism accounting for a precise immunoglobulin domain deletion in a variant of mouse myeloma cells was attributed to a frameshift followed by a premature termination (Kenter and Birshtein, 1979). However, the complexity of the deletions in four human γ3 HCD proteins led Frangione and Franklin (1979) to conclude that generalizations about the genetic mechanism are not possible. In summary, HCD proteins could result from (1) misalignment or defective splicing of gene segments with loss of some exons, (2) frameshifts followed by premature termination, or (3) deletions of exons.

One puzzling aspect of the structure of HCD proteins is that only small segments of the N-terminal part of the V region are often joined directly to the hinge. Also, the incomplete V regions are often unusual in sequence, frequently are heterogeneous, and sometimes contain unexpected carbohydrate. In fact, this was true of the first two HCD proteins discovered, CRA and ZUC (Franklin and Frangione, 1975), but only a partial sequence had been reported for ZUC. Recently, amino acid sequence analysis of the ZUC protein has shown that there are two structural forms, a monomer and a dimer (Takahashi *et al.*, 1985b). The dimer has a V region of only 18 residues and contains a GalN oligosaccharide; in this form the shortened V region is joined to part of the hinge and thus a dimer can be formed as in intact γ3 proteins. However, the monomer has an even shorter V region corresponding to residues 10–18 only; it lacks the entire hinge and thus cannot form a dimer. It is uncertain if the absence of the first nine residues in the monomer is due to a second error at the level of biosynthesis or if it is due to a posttranslational event. However, the fact that this HCD protein has two different polypeptide structures suggests that the protein abnormality results from two different partial gene deletions, or mismatches in splicing, or else from a two-step partial gene deletion.

C. Genetic Origin of Variability and the Antibody Combining Site

1. Recombination Mechanisms

The preceding discussion makes clear that a number of mechanisms contribute to the genetic origin of the variable region sequences and thus to the diversity and specificity of the antibody combining site. The four major factors that have been discussed so far are: (1) combinatorial association, i.e., the apparent lack of restriction on the association of the two types of light chains and the nine isotypes of heavy chains; (2) the presence in the genome of multiple V region genes for κ, λ, and H chains—at least 100 genes of each kind; (3) the three families of J region minigenes, Jκ, Jλ, and J_H; and (4) the occurrence of families of D_H minigenes for H chains. The last three factors together probably have the major

role in generating sequence variation and antibody diversity, and it is difficult to separate their relative contributions. By itself, DNA rearrangement of the various families of antibody gene segments (V, J, D, C) on three different chromosomes could produce a very large number of light and heavy chain genes, each of which could be expressed as a different polypeptide. Depending on the number of V region genes assumed, the repertoire of antibody molecules that could be produced by free association of the resultant polypeptide chains has been estimated by various authors to be from 1×10^7 (Crews *et al.*, 1981; Honjo, 1983) to 1×10^{10} (Kindt and Capra, 1984). This might seem sufficient; however, the recombinatorial mechanisms do not explain all the sequence variation, particularly that within the framework regions or the extreme diversity of the CDR3 region.

Two additional mechanisms to generate variability have been implicated. One of these is flexibility of joining at the recombination sites and the other is somatic mutation. Both exert their greatest effect in the CDR3 region of light and heavy chains. Flexibility of joining refers to the fact that the splicing of genetic elements does not occur at precise recognition sequences. Several different nucleotide sequences can be generated at the V–J junction of light chain genes because of the palindromic recognition sequences. For the same reason, flexibility of joining exists at the V–D and D–J junctions of heavy chains. Although this may affect only a single codon at each junction, short extra nucleotide sequences may also be inserted; this explains the differences in length of V regions.

2. Somatic Mutation

Somatic mutation or hypermutation refers to the alteration of V genes after their assembly. This leads to further diversity in the antibody repertoire. Although this was first demonstrated for Vλ light chains from mouse myeloma cells, much recent work has been centered on cDNA clones from mouse hybridomas that make specific antibodies (Crews *et al.*, 1981; Bothwell *et al.*, 1981; Selsing and Storb, 1981; Rudikoff *et al.*, 1982; Sims *et al.*, 1982; Diamond and Scharff, 1984). In one frequently quoted experiment Hood's group followed the immune response of BALB/c mice to phosphorylcholine (Crews *et al.*, 1981). In these mice the entire phosphorylcholine response is derived from a single V_H gene segment designated T15, which yields the prototype T15 sequence. Nineteen V_H regions from myeloma and hybridoma immunoglobulins that bound phosphorylcholine were completely sequenced (Gearhart *et al.*, 1981). Nine V_H protein variants were observed in the 19 completely sequenced V_H regions. Crews *et al.* (1981) cloned and sequenced most, if not all, of the germ-line V_H genes that coded for these variants and found that each variant V_H protein sequence differed from all of the V_H gene segments of the T15 family. Extensive somatic mutation was found in the framework regions as well as in the CDR segments. Although the somatic mutation mechanism has been correlated

with the class switching process (Gearhart *et al.*, 1981; Hood *et al.*, 1984), it affects only the V regions and not the C regions.

IV. Primary Structure of the Constant Regions of Immunoglobulin Polypeptide Chains

A. Light Chains

1. Proteins and Genes of the Cκ and Cλ Regions

The principles of the primary structure of immunoglobulins were first established by complete amino acid sequence analysis of human κ and λ Bence–Jones proteins (Putnam *et al.*, 1966, 1967a,b). The cardinal principle is the division into a V and a C region, which was illustrated previously for the human λ chain (Fig. 13, Section III,A). The same principles were found to apply to the light chains of other species such as the mouse (Dreyer *et al.*, 1967). The comparative biochemistry of the C region sequences of human and animal light chains was reviewed in Volume III of this series (Putnam, 1977c), and the sequences are given in Volume IV (Barker and Putnam, 1984). Although much more data have been amassed in recent years, particularly by gene cloning and DNA sequencing, no major new principles have emerged. To be sure, the normal distribution of the κ and λ light chains varies greatly among species. All vertebrate species studied have two classes of light chains identifiable by homology to human κ and λ. However, the proportion of κ to λ in the normal Ig classes of different species varies greatly. The mouse and rat have at least 95% κ chains, whereas in other species λ chains predominate (>95%), e.g., the cow, goat, sheep, horse, chicken, and turkey (Putnam, 1977c). The fact that all these animals are immunocompetent signifies that κ and λ light chains are about equally effective in contributing to antibody function.

In contrast to the large number of Vκ and Vλ genes that have been shown to be present in the normal genome by use of gene hybridization and cloning techniques, only one copy of the Cκ gene has been identified in humans but this exists in three allelic (Km) forms (Table III). Likewise, the mouse has only one Cκ gene (Honjo, 1983). In contrast, the genomic organization of Cλ genes is more complex in both species. In humans the λ light chain locus unexpectedly contains six λ-like C region genes arranged in tandem on chromosome 22 (Hieter *et al.*, 1981). Three of these are nonallelic and correspond to three λ chain isotypes for which the protein sequences were earlier determined: Mcg, Ke^- Oz^-, and Ke^-Oz^+. For the differences in sequence see Fig. 13 and also the Appendix of Volume IV (Barker and Putnam, 1984). Three other unlinked λ-like genes were sequenced, but whether these represent coding sequences or pseu-

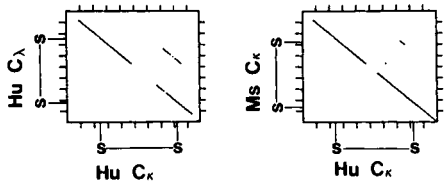


Fig. 19. Graphic matrix plots generated by the DOTMATRIX computer program of comparisons of the amino acid sequences of the human (Hu) Cλ and Cκ regions and of the human and mouse (Ms) Cκ regions. Identical sequences would give a single diagonal (45°) line. In both graphs the identity is greatest around the disulfide bridge sequences. Human Cκ and mouse Cκ are more closely related than human Cκ and Cλ.

dogenes was not established. The mouse λ chain organization is also complex; Blomberg *et al.* (1981) have cloned four λ-like C region genes that occur in two clusters. Because there are few differences in sequence in the human Cλ genes, it is thought that the clusters arose by duplication with subsequent point mutations.

2. Homology of Cκ and Cλ Protein Sequences

There is a high degree of homology in the C region sequences of light chains both within a species and between species. However, light chains of the same type are more alike from different species (such as human and mouse κ) than are light chains of different types (κ and λ) within the same species. An alignment comparing the C region sequences of κ and λ chains of humans and the mouse has been presented in Volume III (Putnam, 1977c). The same comparison can be presented in simplified graphic form by use of the DOTMATRIX computer program of Barker *et al.* (1985). Figure 19 shows that human Cκ and mouse Cκ are more closely related than human Cκ and Cλ. In evolutionary terms this indicates that the Cκ and Cλ genes diverged prior to the divergence of the two species. Inspection of Fig. 19 reveals that the homology in all cases is greatest around the disulfide bridge sequences. This accords with the absolute requirement for the integrity of the disulfide bridge in every immunoglobulin domain.

3. Biological Functions of the Light Chain C Regions

No specific biological functions of the light chain C regions have been identified other than their role of interacting laterally with the C_H1 region of heavy chains to stabilize the molecular conformation. Although some degree of preferential association of λ chains with α and δ chains has been observed, there are no apparent structural constraints for free combinatorial association of both types of light chains and all classes of heavy chains.

The allotypic Km substitutions in human κ chains and the substitutions in human λ isotypes do not appear to contribute to biological properties of antibodies and impose no structural constraints. In the λ chain the Ke (Kern) marker at position 154 and Oz marker at position 191 each represent single amino acid substitutions that are on the exterior of the three-dimensional structure (Fig. 8) and do not affect the basic immunoglobulin fold (Poljak *et al.*, 1973). The Km allotypic substitutions are at exactly homologous positions in the κ chain (positions 153 and 191, Table III) and thus are equally without effect.

B. Heavy Chains

1. Organization of the Heavy Chain Genes

Probably because of late evolutionary expansion the genomic organization and the polymorphism of heavy chain genes differ in various species. The result is that C_H protein subclasses in one species may be difficult to correlate with those in another species. The nine human heavy chain isotypes that normally are functionally expressed (Fig. 4) are the product of nine C_H genes (μ, δ, γ1, γ2, γ3, γ4, ε, α1, and α2). In the mouse the four γ isotypes are clustered together on chromosome 12 in the order 5′-μ-δ-γ3-γ1-γ2b-γ2a-ε-α-3′ (see Fig. 17). In the rabbit only two Cγ genes have been identified so far, and they appear to be allelic genes rather than isotypes (Martens *et al.*, 1984). In fact, current evidence suggests that the rabbit genome has a single Cγ gene, one Cμ gene, and as many as ten Cα genes, of which four have been cloned (Knight *et al.*, 1985). In humans there are five Cγ genes, three Cε genes, and two Cα genes, and these are arranged on chromosome 14 in two clusters suggestive of gene duplication. The proposed order is 5′-μ-δ-γ3-γ1-ε2-α1-ψγ-γ2-γ4-ε1-α2-3′ (Ueda *et al.*, 1985). Only one of the Cε genes (ε1) is expressed; one ε pseudogene (ε2) is truncated by recombination, and the other (ε3) is processed and translocated from chromosome 14 to 9 (Nishida *et al.*, 1982; Battey *et al.*, 1982). Thus, there are at least three human C_H pseudogenes that are inactive: ψγ, ε2, and ε3. The mechanism of class switching was discussed in Section III,B,3; it should be recalled that Cμ is the first C_H gene to be expressed on the surface of the B lymphocyte. However, because of the proximity of the Cμ and Cδ genes, both may be expressed simultaneously on the B lymphocyte.

The C_H genomic arrangement is very complex, so it is not surprising that errors in rearrangement occasionally occur. Not only are there nine tandemly arranged active genes and three pseudogenes, but each active gene is divided into a series of exons that encode the several C_H domains, the hinge region if there is one, and the two possible tailpieces (Figs. 17 and 18). Although many B cell-related immunodeficiencies are known, their total incidence in the general population is very low. Selective immunodeficiencies of a single isotype are quite

rare, as are multliple gene deletions. Only two examples of gene cluster deletions were found in a study of 11,000 individuals by use of the method of restriction fragment length polymorphism (Migone *et al.*, 1984). Likewise the incidence of heavy chain disease (HCD) is quite rare. In this disorder there appears to be an aberrant rearrangement of the exons leading to a deletion of one or more gene segments (Alexander *et al.*, 1982) (Section III,B,4).

2. Primary Structure of the Heavy Chain C Regions

The complete amino acid sequences of all the human H chains have been determined, and the C region sequences are given in the Appendix of Volume IV (Barker and Putnam, 1984). Table II lists characteristic properties of the nine H chains, such as the length, number of domains, presence of a hinge region, and number and type of oligosaccharides. The structures of all the oligosaccharides of human H chains have been determined; they have been reviewed by Baenziger (1984) and are discussed later. In some instances the allotypic sites and substitutions have been determined; these are listed in Table III. Crystal structure data are available only for human IgG, and the structure is illustrated later.

3. Homology in the Primary Structure of the C Regions of Human Heavy Chains

The C region of each heavy chain isotype displays internal homology in amino acid sequence and also exhibits homology to all the other isotypes. An example of the internal homology was illustrated in Fig. 6 for the δ chain domains. Figure 6 also showed that light chain C_L domains are about as homologous to C_H domains as the latter are to each other.

Of all C_H domains the C_H1 domains are most alike in amino acid sequence and presumably also in the three-dimensional structure, and for good reason—they all have a similar function, which is to interact with the C_L domains of the Fab. Figure 20 compares the amino acid sequences of the C_H1 domains of the five classes of human immunoglobulins. As in Fig. 6, the invariant cysteine and tryptophan residues are used to place the alignment in register. This figure brings out the fact that the conservation of structure is greatest around the intrachain disulfide bridge.

The Fc region, in which specific biological effector functions are vested, is made up of the last two domains in all five classes of H chains. The C_H2 domains of the larger μ and ε chains take the place of the hinge region in γ, α, and δ chains. Also, the C_H2 domain of the μ chain is rapidly degraded during tryptic cleavage to form Fab and Fc fragments (Fig. 12) (Lin and Putnam, 1978). Figure 21 compares the amino acid sequence of the Fc regions of the five classes of human immunoglobulins. The distribution of homologous residues among the

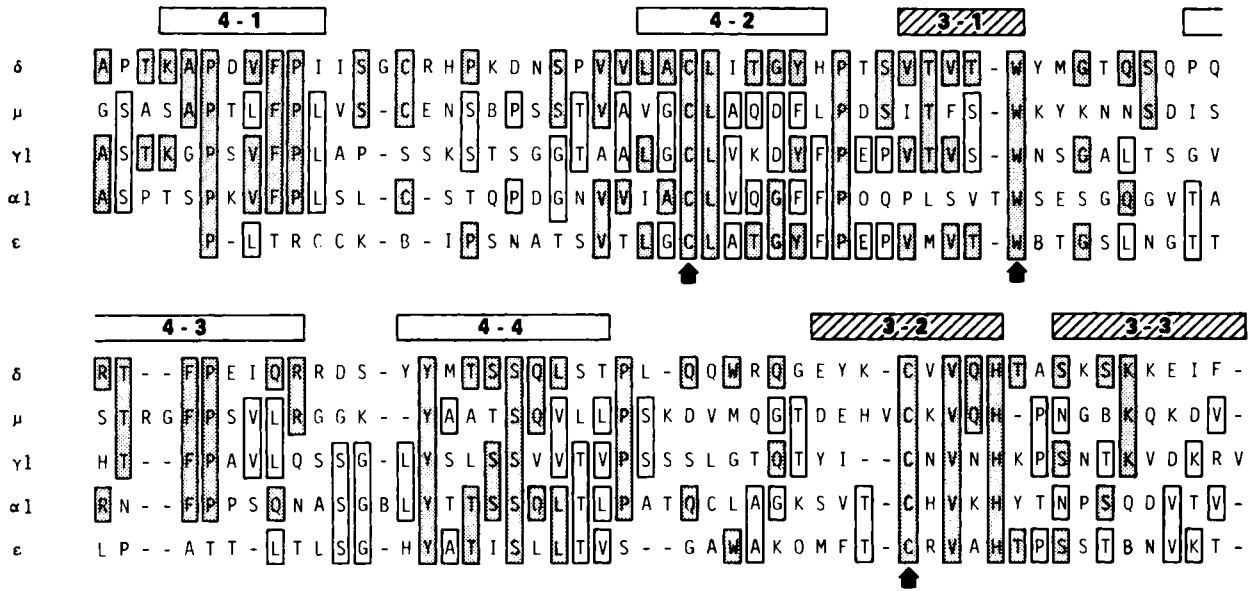


Fig. 20. Comparison of the amino acid sequences of the C_H1 domains of the five classes of human immunoglobulins. Arrows denote the three invariant cysteine (C) and tryptophan (W) residues in each domain that were used to align the sequences; gaps were placed to maximize the homology. Residues in the μ, γ, α, and ε sequences that are identical with corresponding residues in the δ chain are in shaded boxes; residues that are identical in sequences other than the δ chain are in open boxes. The β strands are numbered according to Edmundson *et al.* (1975) with the four-stranded β elements in open bars and the three-stranded β-pleated sheets in hatched bars (from Putnam *et al.*, 1981).

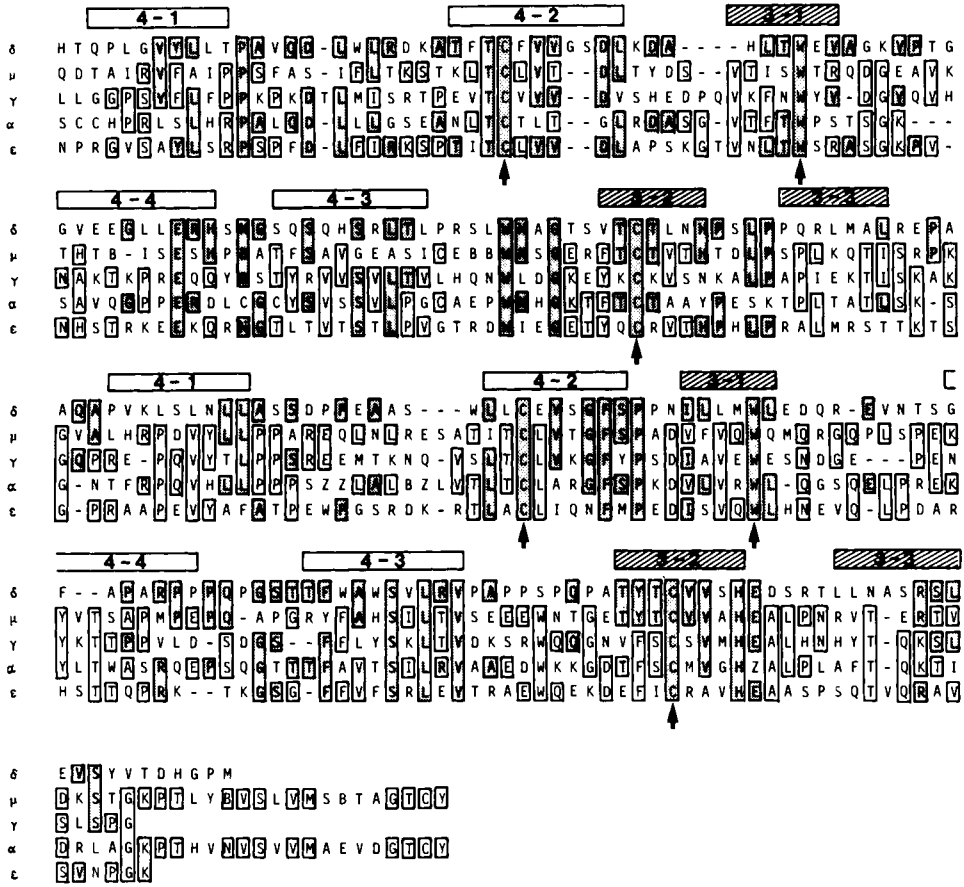


Fig. 21. Comparison of the amino acid sequence of the Fc regions of the five classes of human immunoglobulins. As in Fig. 20, the arrows denote the invariant cysteine and tryptophan residues used for the alignment. All other residues in the μ, γ, α, and ε chains that are identical with the δ chain are in shaded boxes. Open boxes outline identities that do not involve the δ chain. The bars have the same meaning as in Fig. 20. The alignment is so arranged that the top two and the bottom three rows represent, respectively, amino acids encoded by the exons for the last two constant domains of each heavy chain class (C_H2 and C_H3 for δ, γ, and α and C_H3 and C_H4 for μ and ε). The extra segments after the lysine (K) in the μ and α chains are the tailpieces of the secreted forms of IgM and IgA, which are encoded by a separate exon (from Lin and Putnam, 1981).

five chains is not random. Clearly there are two sets of homologous domains in the five chains. This alignment, which predates the gene cloning and sequencing of H chains, corresponds almost exactly to the gene structure later defined for the H chains. In other words, each genomic exon precisely encodes a structural domain in the protein, as was first predicted by sequence analysis and later established by protein crystallography (see Fig. 18).

Taken together, the striking homology in primary structure of the C_H domains of all classes of H chains and the similar genomic organization of the gene segments that encode the C_H domains both demonstrate that there has been strong evolutionary pressure to conserve the basic three-dimensional structural unit of all antibody classes—the immunoglobulin fold. Unfortunately, the crystallographic structure has only been determined for IgG. Hence, the folding structure of the Fc of IgA, IgM, IgD, and IgE can only be estimated by modeling methods, and these have inherent errors as past experience has shown. Despite this caveat it is useful to try to fit the Fc sequences to the known folding patterns of immunoglobulin chains. In Figs. 20 and 21 the β-pleated sheet structure is superimposed above the sequences following criteria proposed by Beale and Feinstein (1976). The fit is fairly good. Residues highly conserved among all five chains are found mainly, but not necessarily, around segments occupied by the β-pleated sheets. The exceptions are strands 4-4 and 3-3 of both Fc domains (Fig. 21); presumably this is because these β strands occupy a marginal position in each β-pleated sheet (Fig. 22), so they are less critical in providing hydrogen bonds that are crucial for maintaining the overall domain stability.

Figure 22 should be considered a schematic drawing rather than a model based on exact coordinates; however, it is useful to illustrate both the structural constraints that are imposed on Fc by the similarities and differences in sequence of

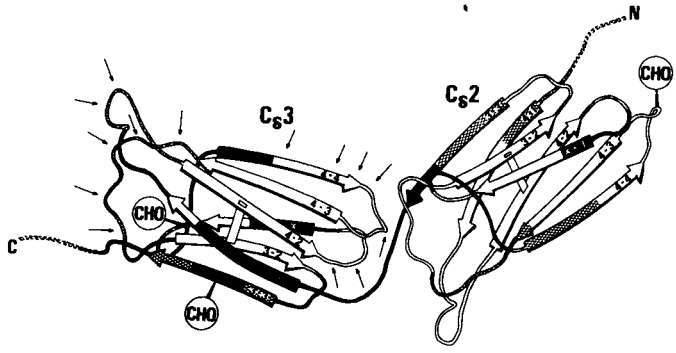


Fig. 22. Schematic spatial model of the IgD Fc region showing its sequence relationships to Fc regions of other classes (see Fig. 21). The schematic diagram for the α-carbon backbone of the Cδ2 and Cδ3 domains is adapted from a drawing of the Cλ domain of the Mcg Bence–Jones dimer (Edmundson *et al.*, 1975). The lengths of the β-sheet strands (broad segments) and the connecting segments are adjusted as described by Beale and Feinstein (1976). The shading on the backbone indicates the extent of sequence homology between the δ chain and the other four heavy chains as follows: open, highly conserved among all five chains; crosshatched, high divergence among all five chains; solid, high homology among all heavy chains except the δ chain. The three glucosamine oligosaccharides attached to the Fc of IgD are designated CHO, enclosed in circles. Arrows pointing to the Cδ3 domain indicate the clustering of proline residues at the carboxy terminus (from Lin and Putnam, 1981).

the five Ig classes. The most highly conserved areas are inside the barrel structure and are clustered around the two invariant cysteines of the intradomain disulfide bridge and the nearby tryptophan. These three residues apparently form the domain nucleation center during the folding of nascent immunoglobulin chains. As might be expected, the most striking differences in sequence appear to be on the exterior of the β-barrel structure and to cluster at the longitudinal surfaces. This is especially the case for IgD, which has many proline residues unique to the δ chain (see arrows in Fig. 22). Because the proline residue tends to change the course of the polypeptide backbone, this clustering of proline residues will necessarily impart to the Cδ3 domain a surface conformation significantly different from that of other carboxyl-terminal domains. The effect of multiple carbohydrates on the Fc spatial structure must also be considered. Figure 22 shows only the three GlcN oligosaccharides in the Fc of IgD. The one in Cδ2 is homologous to the single oligosaccharide in the γ chain, but the two in Cδ3 have no counterpart in other immunoglobulin classes.

4. Computer-Generated Comparisons of the C Region Sequences of Human Heavy and Light Chains

In the preceding sections the amino acid sequences of the C regions of H and L chains have been compared directly to illustrate that all immunoglobulin chains are composed of a series of tandem homologous domains each of which shares the basic polypeptide structure characteristic of all antibody molecules. Recently a computer program called DOTMATRIX has been developed that provides a rapid comparison of all possible segments of a specified size (e.g., 20 residues) for two polypeptide sequences (Barker *et al.*, 1985). The program generates a graphic matrix plot in which homologous relationships based either on identities or on a mutation data matrix are plotted as a diagonal line. Comparison of two identical sequences would give a single 45° diagonal line.

Figure 23 illustrates application of the DOTMATRIX program to comparison of the human Cγ sequence with the sequences of the C regions of human light chains (Cκ and Cλ) and with the other human heavy chains (Cδ, Cα, Cμ, and Cε). From Fig. 23 it is evident that in general the C_H1 domains are most homologous to each other, the C_H2 domains of Cγ, Cδ, and Cα are most homologous to each other and to the C_H3 domains of Cμ and Cε, and the carboxyl-terminal domains of all five chains are most homologous to each other. However, the hinge regions are unrelated. The graphs bring out the fact that in all instances the homology is greatest around the disulfide bridges and usually is least in the middle of the domain, which is located in a link that connects two exterior β strands of the barrel structure. It is noteworthy that the Cκ and Cλ regions are homologous to all three Cγ domains and are more so than are many of the domains of other H chains.

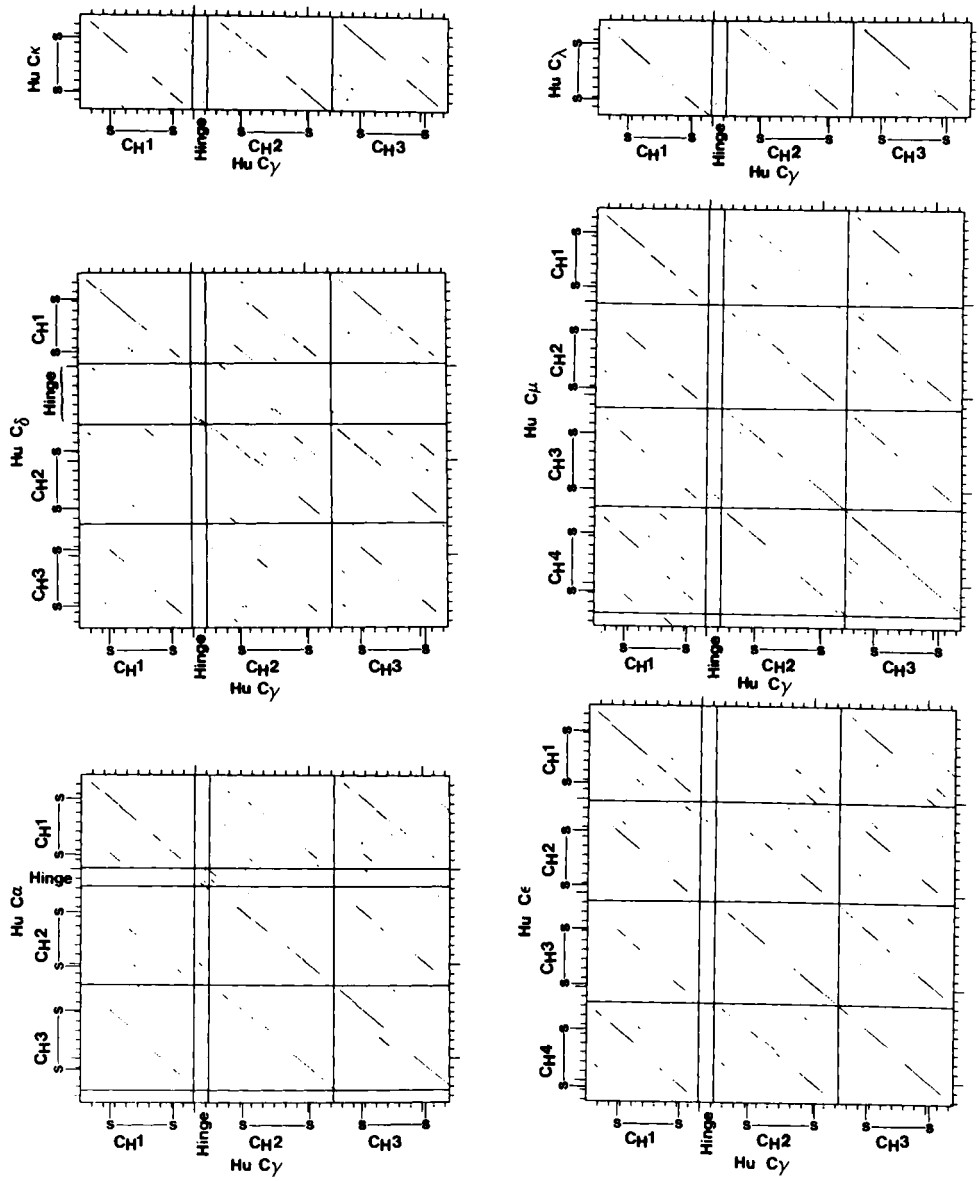


Fig. 23. Graphic matrix plots generated by the DOTMATRIX computer program of comparisons of the amino acid sequences of the C regions of human light chains (Cκ and Cλ) and heavy chains (Cγ, Cδ, Cα, Cμ, and Cε). In all cases the Cγ sequence is on the abscissa, and the C region domains, hinges, and intrachain disulfide bridges are aligned and plotted to scale. A perfect match of two chains would give a single 45° diagonal line. The slanting lines in each domain indicate sequence similarity of two domains to each other. The hinge regions of Cγ, Cδ, and Cα cause a proportionate displacement from the diagonal. The Cμ and Cε sequences lack a hinge but have an extra domain.

5. Interspecies Homology of the Heavy Chain Constant Regions

Just as light chains of the same type are more alike from different species (such as human and mouse κ) than are light chains of different type (κ and λ) within the same species (Section IV,A,2), so are heavy chains of the same class more alike from different species than are heavy chains of different classes within the same species. This is illustrated in Table V. On the basis of the homology alignment of the human Fc regions for the five H chain classes shown previously in Fig. 21, Lin and Putnam (1981) calculated that the overall homology of Fcδ to other Fc regions is Fcα (25.6%) > Fcε (23.8%) > Fcδ (22.8%) > Fcμ (22.5%). In other words, the homology of the Fc regions of the five human Ig classes, and of their constituent domains, is of the order of 25%. However, Table V shows that if the hinge region is excluded, the C_H3 regions of the α chains of man, mouse, and the rabbit have a much higher degree of homology than is found among the Fc regions of the five human Ig classes. The same is true if the C_H3 domains of the δ chains of the three species are compared. However, the C_H1 domains of human and rodent IgD are only about 25% alike. In fact, the human and rodent IgD proteins exhibit an unprecedented structural difference in that the Cδ2 domain present in human IgD is missing in mouse and rat IgD (Putnam *et al.*, 1982; Sire *et al.*, 1982; Blattner and Tucker, 1984; White *et al.*, 1985). Gilliam *et al.* (1984) conclude that "the dramatic difference seen in the structures of IgD in mouse and man suggests a major evolutionary divergence of Cδ in the two species. This divergence may also be reflected in the noncoding regions between Cμ and Cδ."

One final point is that the highest degree of homology identified by this interspecies comparison is displayed by the last domain when the same class of chain is compared for unrelated species. That is, the last domains are more like each other than any one of them is like any other domain in heavy chains. This suggests that each C_H domain is mutating at its own rate determined by its function. Independent evolution of each C_H domain within any class of H chain is possible because separate exons code for each domain.

V. Structure and Function of Oligosaccharides

A. Carbohydrate Structure and Linkage Sites

1. Types of Carbohydrate Structure and Structural Heterogeneity

All immunoglobulins are glycoproteins. The carbohydrate may comprise from 3% of the molecular mass (IgG) to up to 13% (IgE); yet, many immunologists

TABLE V

Percentage Homology of the C Regions of IgA and IgD Immunoglobulins of Different Species

	IgA[a]				IgD		
Domain	Human α1 / Human α2	Human α1 / Mouse α	Human α1 / Rabbit αg	Rabbit αg / Mouse α	Human δ / Mouse δ	Human δ / Rat δ	Mouse δ / Rat δ
C_H1	90	51	35	30	25	26	48
Hinge	50	18	9	5	Low[b]	Low[b]	57
C_H2	93	55	50	60	0	0	—
C_H3	98	69	70	59	53	50	78
TP	89	61	68	44	None	None	86

[a]Data are given for the A2m(2) allotype of the human α_2 chain. Human α_1 has a duplication of the hinge region (26 residues) compared to human α_2 (13 residues).

[b]The mouse hinge (36 residues) has similarity to the GalN-rich half of the human δ hinge (64 residues) but is only half the size.

and molecular biologists overlook the significance and possible role of the carbohydrate. Indeed, all immunoglobulin classes except IgG have multiple specific sites of attachment of carbohydrate. Sugar (oligosaccharide, glycan) is always linked at one or more points to the C region of the H chain; however, rarely and due to the chance occurrence of a signal peptide sequence, carbohydrate may be attached fortuitously to a V region or to an L chain. The oligosaccharides may differ in structure and site from one immunoglobulin class to another and also from species to species for the same class. There are two major kinds of carbohydrate, and each of these has subsets. One kind (M_r 2500–3000) is composed of a series of sugar units (mannose, glucosamine, galactose, fucose, sialic acid) that radiate from a glucosamine (GlcN) that is N-linked to an asparagine in the peptide backbone. The GlcN glycans may be dibranched (biantennary) or tribranched or finished or unfinished with respect to postsynthetic processing and metabolic degradation (Kornfeld and Kornfeld, 1985). In the other kind, galactosamine (GalN) is O-linked to the peptide chain, but it has only one or two units of galactose and sialic acid attached and thus has a lower molecular weight (ca. 750). The GlcN is always N-linked to asparagine in the obligate acceptor sequence Asn-X-Thr/Ser, in which X may be any amino acid, though rarely proline, and the third residue is either serine or threonine. Perhaps owing to conformational inaccessibility not all such sites are occupied or fully occupied in all glycoproteins. However, with the exception of an Asn-Pro-Ser Sequence in the α1 chain and in the A2m(1) allotype of α2, all Asn-X-Ser/Thr sequences in the C region of the human α, μ, γ, δ, and ϵ primary structures have GlcN carbohydrate attached. Galactosamine is O-linked to serine or threonine. No acceptor signal for GalN oligosaccharides has been clearly identified, but the sequence around the site is often rich in proline and must have a conformation accessible for glycosylation. Most other plasma proteins are glycoproteins and have oligosaccharides that are similar in structure (Baenziger, 1984).

Until recently structural analysis of oligosaccharides was slow, tedious, and difficult; however, advances in techniques such as the use of 350- and 500-MHz ^{1}H NMR have made it possible to determine large carbohydrate structures rapidly. Consequently the structures of the oligosaccharides of most human and mouse immunoglobulins are now known (Baenziger, 1984; Kornfeld and Kornfeld, 1985).

Baenziger (1984) has reviewed the synthesis, structure, and function of the oligosaccharides of plasma glycoproteins with emphasis on immunoglobulins as a model system. Their use as a model is possible because the site of attachment and the structure of the carbohydrate is now known for all five classes of human immunoglobulins (Baenziger, 1984) and also for immunoglobulins of several animal species (Rearick *et al.*, 1983). References for determination of the carbohydrate structure of the human proteins are: IgG (Kornfeld *et al.*, 1971), IgA

(Baenziger and Kornfeld, 1974a,b; Pierce-Cretel *et al.*, 1984), IgM (Chapman and Kornfeld, 1979a,b), IgD (Mellis and Baenziger, 1983a,b), and IgE (Baenziger and Kornfeld, 1974c,d).

Although the basic type structures of the GlcN and GalN oligosaccharides of immunoglobulins have been well established, the structures of the GlcN oligosaccharides are heterogeneous and fall into three classes: high mannose, hybrid, and complex. Typical structures are given by Baenziger (1984). Heterogeneity of structure also occurs within these classes depending on such factors as the number of branches (two to four) and the number, location, and linkage of the peripheral sugars such as sialic acid and fucose. Consequently, immunoglobulins exhibit microheterogeneity in electrophoretic behavior (Anderson *et al.*, 1984). It has recently been shown that the microheterogeneity results not only from variation in the carbohydrate structure but also because of incomplete glycosylation. For example, by use of HPLC to isolate the glycopeptide at a single site in the sequence of IgD, Takahashi *et al.* (1984) found that only about half the molecules of IgD were glycosylated at this site (Asn-345 in the δ chain). Likewise, some sites in heavy chain disease proteins are only partly glycosylated (Takahashi *et al.*, 1985b).

The causes of carbohydrate heterogeneity (which is characteristic of all plasma glycoproteins) include incomplete glycosylation and incomplete processing of the added carbohydrate during postsynthetic steps, and also loss of sialic acid during purification (Baenziger, 1984; Kornfeld and Kornfeld, 1985). In fact, Parekh *et al.* (1985) have listed at least 30 different complex-type biantennary GlcN oligosaccharide structures that they determined in human IgG, which has a single site for the GlcN carbohydrate at about position 300 in each γ chain. Thus, the most striking characteristic of the GlcN carbohydrate of IgG is its remarkable diversity in structure.

2. Sites of Attachment

The GlcN and GalN oligosaccharides are normally present only in the C region of immunoglobulins, and the number, type, and site of attachment differ for and are characteristic of each class. Figure 24 compares the specific sites and types for human heavy chains. The γ3 chain is omitted, but it has a second GlcN at the position where Lys-275C is changed to asparagine to yield the acceptor sequence Asn-Thr-Thr. Altogether there are 25 GlcN and 9 GalN oligosaccharides shown in Fig. 24, and the structure and the site have been determined in each case. Thus, more is probably known about the structure and linkage of the carbohydrate in this family of proteins than in any other. Although the human IgG1 subclass and the IgG of most animal species have just one GlcN oligosaccharide on each γ chain (at Asn-297 in the human γ1 chain), many different

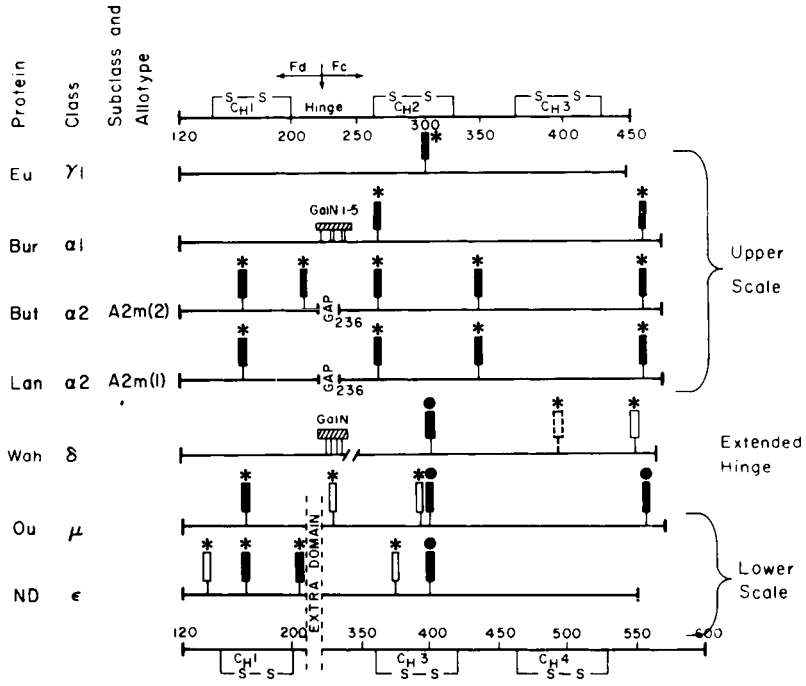


Fig. 24. Oligosaccharides of human heavy chains. Vertical rectangles denote glucosamine oligosaccharides. Shading indicates that these have homologous positions in two or more chains. Solid circles denote the mannose-rich type and asterisks the complex type of GlcN glycan. In the δ chain the dashed rectangle shows where the oligosaccharide is present on only about half the IgD molecules. The numbers in the upper and lower scales give the residue positions in the chains, but the extra domain (Cμ2 and Cε2, respectively) has been omitted in the μ and ε chains (from Putnam *et al.*, 1985).

subpopulations exist with respect to the γ chain carbohydrate structure (Mizuochi *et al.*, 1982; Parekh *et al.*, 1985). No doubt the same holds true for the multiple carbohydrate units present in other classes of H chains.

Heavy chains of human IgA, IgM, IgD, and IgE and of homologous proteins in other species have multiple sites of attachment for carbohydrate, ranging from two to five, according to the class of chain. Most of the glycans are of the GlcN type, and most of these have a complex structure. However, the human δ and ε chains have one mannose-rich glycan, and the μ chain has two. Although some GlcN glycans are at homologous positions in several chains, others have no counterpart in other chains. GlcN glycans may differ in structure even when at homologous positions in several chains. For example, the single complex-type GlcN glycan in the γ1 chain is at a position homologous to a mannose-rich glycan in the δ, μ, and ε chains (Fig. 24). Since the mannose-rich glycan is

processed postsynthetically, the polypeptide bearing it may fold more rapidly, thus making the newly synthesized glycan less accessible to the enzymes of the processing system. In fact, the two GlcN oligosaccharides in IgG1 are in an asymmetric paired conformation and bridge the two C_H2 domains (see later figure) (Deisenhofer, 1981).

In human immunoglobulins, all the GlcN oligosaccharides are linked to the C region domains, whereas the GalN oligosaccharides are confined to the hinge regions of IgA1 and IgD. The hinge of the human α1 chain has sites for up to five GalN glycans (Liu *et al.*, 1976), but fewer may be present in some instances. In contrast, the hinge of the human α2 chain has a pentaproline sequence with no sites for GalN; yet both A2m allotypes of the α2 chain have two GlcN oligosaccharides that are absent in the α1 chain, and the A2m(2) allotype has a fifth GlcN oligosaccharide (Torano *et al.*, 1977; Tsuzukida *et al.*, 1979). These α chains lack the GlcN that is homologous in the four other chains (i.e., at Asn-297 in γ1) because a disulfide bond with a highly strained conformation replaces the GlcN acceptor sequence.

Similar differences occur in the IgD class of different species. For example, the hinge region of the human δ chain has four or five GalN glycans in IgD WAH (Takahashi *et al.*, 1982), and possibly up to seven in the IgD NIG-65 (Takayasu *et al.*, 1982), whereas there is a site for GlcN in the mouse δ hinge but no apparent GalN (Tucker *et al.*, 1980). Likewise, the distribution of the five GlcN oligosaccharides reported in the shortened mouse δ chain (Dildrop and Beyreuther, 1981) is very different from that of the three GlcN in the human δ chain (Putnam *et al.*, 1982).

In addition to the highly conserved sites for GlcN and GalN linkage in the C regions of H chains, carbohydrate may occasionally be attached to acceptor sites that occur by chance owing to the hypervariability of the V regions of both H and L chains. One example is the presence of a GlcN glycan in the V region of the α1 chain of the IgA1 protein Bur (Liu *et al.*, 1976). However, such unexpected glycosylation occurs most often in the unusual sequences characteristic of the N-terminal segments of heavy chain disease (HCD) proteins that have structural deletions. One example is the work of Takahashi *et al.* (1985b), who found that Ser-17 in the N-terminus of the V region of an HCD γ3 chain (Zuc) was glycosylated with GalN. They also noted only partial glycosylation with GlcN at the second C region site in this γ3 chain.

There are a number of early reports of carbohydrate in L chains, but until recently there was little information on the characterization of the carbohydrate and protein structure. However, Garver *et al.* (1981) determined the structure of a GlcN glycan located on Asn-25 and of a GalN glycan at Ser-21 of the V region of human λ light chain Sm, which has a deletion of about 80 residues. Likewise, Ohkura *et al.* (1985) elucidated the structure of a GlcN oligosaccharide linked to Asn-93 in the third hypervariable (CDR3) region of two λ light chains (Wh and

Nei), and Savvidou *et al.* (1984) did so for a GlcN glycan at Asn-107 in the J region of a κ light chain.

3. Role of Carbohydrate in the Three-Dimensional Structure of IgG

IgG is one of the very few glycoproteins of any kind of which a high-resolution structure has been obtained by crystallographic analysis. The three-dimensional structure of the carbohydrate has been solved for the single GlcN oligosaccharide on the γ chain of the Fc of the human myeloma IgG protein Kol (Deisenhofer, 1981) and the Fc of rabbit IgG (Sutton and Phillips, 1983), but there are no published structures for the carbohydrate in other Ig classes. The carbohydrate is linked to Asn-297 in the C_H2 domain of IgG at a sharp bend between strands 3 and 4 of layer 1; its interactions with C_H2 are largely hydrophobic in nature. The C_H2 domains are unlike the others because they have little lateral contact owing to their separation by the carbohydrate, which forms a weak bridge between them. The carbohydrate appears to stabilize the three-dimensional structure of the IgG, and Huber (1984) suggests that "the origin of the altered functional properties of carbohydrate-free antibodies lies in the structural destabilization." The crystalline order of the carbohydrate is better in one C_H2 domain than the other. In rabbit IgG the carbohydrate is also asymmetric with respect to the two halves of the molecule. However, in contrast to human Fc the carbohydrate chains in rabbit Fc make direct contact with each other. As illustrated in a later figure (Section VI,B), stereo drawings of the α-carbon backbone of human Fc show the carbohydrate occupying the space between the two C_H2 domains, whereas the binding site for C1q is on the outside of C_H2 and that for *Staphylococcus aureus* protein A is at the exterior junction of C_H2 and C_H3 (Davies and Metzger, 1983). These observations suggest that the carbohydrate in IgG does act as a spacer between domains and stabilizes the Fc structure. It would be desirable to have similar three-dimensional structures for immunoglobulins that have multiple carbohydrates.

B. Biological Roles of the Carbohydrate

Evolutionary preservation of the tripeptide acceptor sequence for GlcN, as in the IgG of many species and at homologous positions in IgM, IgD, and IgE, suggests that carbohydrate has both structural and biological roles. However, the nature of these roles continues to be elusive. Many functions have been ascribed (Putnam, 1977a; Putnam *et al.*, 1985). These include: (1) facilitating secretion; (2) increasing solubility; (3) acting as spacers between domains and as bridges between chains (as described above); (4) affecting transduction of signals from the antigen combining site in Fab to the effector regions in Fc; (5) involvement in

biological effector functions such as the binding of C1q to initiate the complement cascade and also in cytotropic interactions with lymphoid cells; and (6) regulation of catabolism both by conferring protection against proteases and by governing uptake of immunoglobulins by hepatocytes. Baenziger (1984) has reviewed these and other possible functions and concluded that it is easier to state what the function is *not* than what it is. He suggests that for some immunoglobulins but not others glycosylation may be necessary (1) for attaining and/or maintaining a conformation essential for assembly and secretion; (2) for protection against proteolytic degradation, and (3) for uptake by the reticuloendothelial receptor in the case of complex GlcN oligosaccharides, but not for the mannose-rich class. He also suggests that some GlcN oligosaccharides may have a critical but still unidentified role in the effector functions of the immune system.

Several recent studies with deglycosylated monoclonal IgG antibodies have indicated that carbohydrate is required for certain biological activities of the Fc region but not for specific antigen binding, which is vested in the Fab region. The carbohydrate depletion was achieved by culturing mouse hybridoma cells in the presence of tunicamycin, an inhibitor of glycosylation. The carbohydrate-deficient antibodies lost several biological properties such as the ability to activate complement and to bind to Fc receptors on macrophages (Nose and Wigzell, 1983) and also were not effective in suppressing the humoral immune response (Heymann *et al.*, 1985).

It has been suggested by Parekh *et al.* (1985) that the varying pattern of glycosylation of IgG and the resultant heterogeneity in structure may contribute to the autoimmune component of rheumatoid arthritis and primary osteoarthritis. This binational group undertook a massive study that required evaluation of the primary sequences of about 1400 oligosaccharides from 46 individual normal and patient samples. They identified some 30 related GlcN structures. The results indicated an association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of the total serum IgG. Although neither disease was associated with novel oligosaccharide structures, the IgG isolated from the normal group and from the patients had different distributions of the biantennary complex type of GlcN oligosaccharide. Parekh *et al.* (1985) suggested several mechanisms by which the altered glycosylation pattern could affect the molecular conformation of the Fc, for example, by exposing new determinants, which could induce an effect similar to the autoimmune response. This novel idea could explain the autoimmune response in rheumatoid arthritis, where IgM and IgA antibodies are produced against a patient's own IgG—a phenomenon earlier attributed to an altered conformation of denatured IgG. However, it should be pointed out that carbohydrate does not normally contribute to the antigenic response in heterologous species because of the ubiquity of similar oligosaccharide structures on many glycoproteins.

A comparative structural study of the carbohydrate moieties of normal IgM and of the monoclonal IgM from two patients with macroglobulinemia was made by Cahour *et al.* (1983) and the GlcN structures were later determined precisely (Cahour *et al.*, 1984a,b). A predominance of multiantennary structures was observed in the more heavily glycosylated IgM from the two patients. This variation was attributed to differences in the biosynthetic processing pathway of the carbohydrate units or enhanced expression by the clone of pathological cells. These careful analytical studies and those described above should warn molecular biologists that antibodies prepared from hybridomas, though having well-defined antigen specificities, may differ in whatever biological effector functions to which carbohydrate may contribute.

VI. Three-Dimensional Structure and Binding Sites

A. Three-Dimensional Structure of Myeloma Proteins and of Antibody–Antigen Complexes

1. Early Findings and New Directions

The fundamental features of the three-dimensional structure of antibodies were established a decade ago by crystallographic analysis of antibody analogues such as Bence–Jones proteins and myeloma globulins. The results have been amply reviewed (Davies *et al.*, 1975; Nisonoff *et al.*, 1975; Putnam, 1977a; Padlan, 1977a,b; Capra and Edmundson, 1977; Poljak, 1978; Amzel and Poljak, 1979; Marquart and Deisenhofer, 1982; Hahn, 1983; Davies and Metzger, 1983; Huber, 1984). In fact, there are more detailed reviews of the stuctures than there are structures of different proteins. This confirms the importance attached to understanding the three-dimensional structure of the antibody combining site and the attendant interest in elucidating the mechanism whereby combination with antigen facilitates biological effector functions. Other reasons for emphasis on crystallographic study of immunoglobulins include: (1) the availability of a series of crystalline myeloma proteins and of crystallizable fragments such as Fab, Fc, and light chains, (2) the greater ease of solving a large structure such as IgG by study of its fragments, (3) the increased facility for solving homologous structures, such as a series of Bence–Jones proteins, once the first is done, and (4) the availability of so much sequence data on immunoglobulins.

Despite the incentives listed above the structures of only two intact immunoglobulin molecules are yet known at high resolution (the myeloma proteins IgG1 Kol and Dob) (Marquart *et al.*, 1980; Sarma and Laudin, 1982), and crystal structures of the antibody–antigen complex are just beginning to emerge (Amit *et al.*, 1985a,b). The reasons are the arduous effort required to determine the

TABLE VI

X-Ray Crystallographic Analyses of Immunoglobulins[a]

Protein	Resolution (Å)	References
IgG immunoglobulins		
IgG1(λ) Kol	3.0	Marquart *et al.* (1980)
IgG1(κ) Dob	4.0	Silverton *et al.* (1977); Sarma and Laudin (1982)
IgG1(λ) Mcg	6.5	Rajan *et al.* (1983)
Other immunoglobulin classes IgM, IgA, IgD, IgE	—	None
Fab fragments		
IgG1(λ) Newm	2.0	Poljak *et al.* (1973, 1974); Saul *et al.* (1978)
IgG1(λ) Kol	1.9	Marquart *et al.* (1980)
IgA2(κ) McPC603 (mouse)	3.1	Segal *et al.* (1974)
IgA2(κ) J539 (mouse)	4.5	Navia *et al.* (1979)
IgG2(κ) Zie	—	Ely *et al.* (1978)
Monoclonal antibody (mouse)	6.0	Amit *et al.* (1985a,b)
Fc fragments		
Fc(IgG) and protein A	2.8	Deisenhofer (1981)
Fc(IgG) (rabbit)	2.7	Sutton and Phillips (1983)
Bence–Jones proteins		
Vκ dimer Rei	2.0	Epp *et al.* (1974, 1975)
Vκ dimer Au	2.0	Fehlhammer *et al.* (1975)
Vλ dimer Rhe	1.6	Furey *et al.* (1983)
Light chain (λ) dimer Mcg	2.3	Edmundson *et al.* (1974); Abola *et al.* (1980)
Light chain (λ) hybrid	3.5	Ely *et al.* (1985)
Light chain (λ) dimer Loc	3.0	Chang *et al.* (1985)
Class I MHC antigens (bovine), β_2-microglobulin (light chain)	2.9	Becker and Reeke (1985)

[a]Unless specified, all proteins listed are human myeloma immunoglobulins.

structures of such large molecules as IgA and IgM and the unavailability until recently of crystalline antibody–antigen complexes. Table VI summarizes the current status of X-ray crystallographic study of immunoglobulins. The conspicuous omissions are human IgM, IgA, IgD, and IgE.

The major findings of the early studies are: (1) the division of light and heavy chains into a series of connected domains that have a similar conformation and have lateral and longitudinal interactions; (2) the presence in each domain of the characteristic conformation known as the immunoglobulin fold, which has a sandwichlike structure consisting of two layers of antiparallel segments in β-pleated sheet configuration (see Fig. 8, Section II,B,1); and (3) the topography

of the combining site, which consists of a pocket or cavity lined by the CDR segments of the V_H and V_L regions.

The new directions include: (1) preparation, crystallization, and preliminary crystallographic study of Fab fragments of monoclonal antibodies and of their complexes with antigens (Mariuzza *et al.*, 1984; Amit *et al.*, 1985a,b); (2) determination at high resolution of the structure of Fc from normal IgG and of its complex with protein A from *Staphylococcus aureus* (Deisenhofer, 1981); (3) determination of the three-dimensional structure of the carbohydrate of the Fc of human IgG (Deisenhofer, 1981) and of rabbit IgG (Sutton and Phillips, 1983); (4) molecular model building and computerized molecular graphics display of antibody molecules and their complexes with antigen based on crystallographic coordinates (Feldmann *et al.*, 1981; Stanford and Wu, 1981; Hahn, 1983; Davies and Metzger, 1983; Novotny *et al.*, 1983; Huber, 1984; Novotny and Haber, 1985; Tonegawa, 1985); (5) extension of the molecular graphics approach to other members of the immunoglobulin supergene family such as the T-cell receptor (Patten *et al.*, 1984); and (6) determination of the three-dimensional structure of the β_2-microglobulin, the immunoglobulin-like light chain of the major histocompatibility complex class I antigens (Becker and Reeke, 1985).

2. *Crystallographic Study of the Antigen Binding Sites of Monoclonal Antibodies*

Whereas the early crystallographic work focused mainly on the binding sites of Bence–Jones proteins and myeloma Fab fragments for chance ligands such as dinitrophenyl derivatives, several investigations of crystalline complexes of monoclonal antibodies with specific antigens are now in progress (Amit *et al.*, 1985a,b; Mariuzza *et al.*, 1984; Rose *et al.*, 1983; Amit *et al.*, 1983). Earlier, the crystal structures of four Fab fragments had been reported. Three of these bound ligands, i.e., human Fab IgG1(λ) Newm, which binds a derivative of vitamin K_1 (Poljak *et al.*, 1974; Saul *et al.*, 1978; Amzel and Poljak, 1979), and two murine IgA2(κ) Fab fragments, one of which (McPC603) binds phosphorylcholine (Segal *et al.*, 1974) and the other (J539) binds galactans (Navia *et al.*, 1979). The specificity of the fourth Fab fragment (Kol) is unknown (Marquart *et al.*, 1980). The first three-dimensional structure of an actual antigen–antibody complex at 6 Å resolution has been reported by Amit *et al.* (1985a), and higher-resolution data of this and other antigen–antibody complexes are anticipated soon from several sources.

Amit *et al.* (1985a) determined the structure of the complex between hen egg-white lysozyme and the Fab of a monoclonal anti-lysozyme antibody. Figure 25 shows the preliminary three-dimensional model of the antibody–lysozyme complex as a stereo diagram. The electron density data could readily be interpreted using the known α-carbon backbone structure for Fab Newm (Poljak *et al.*,

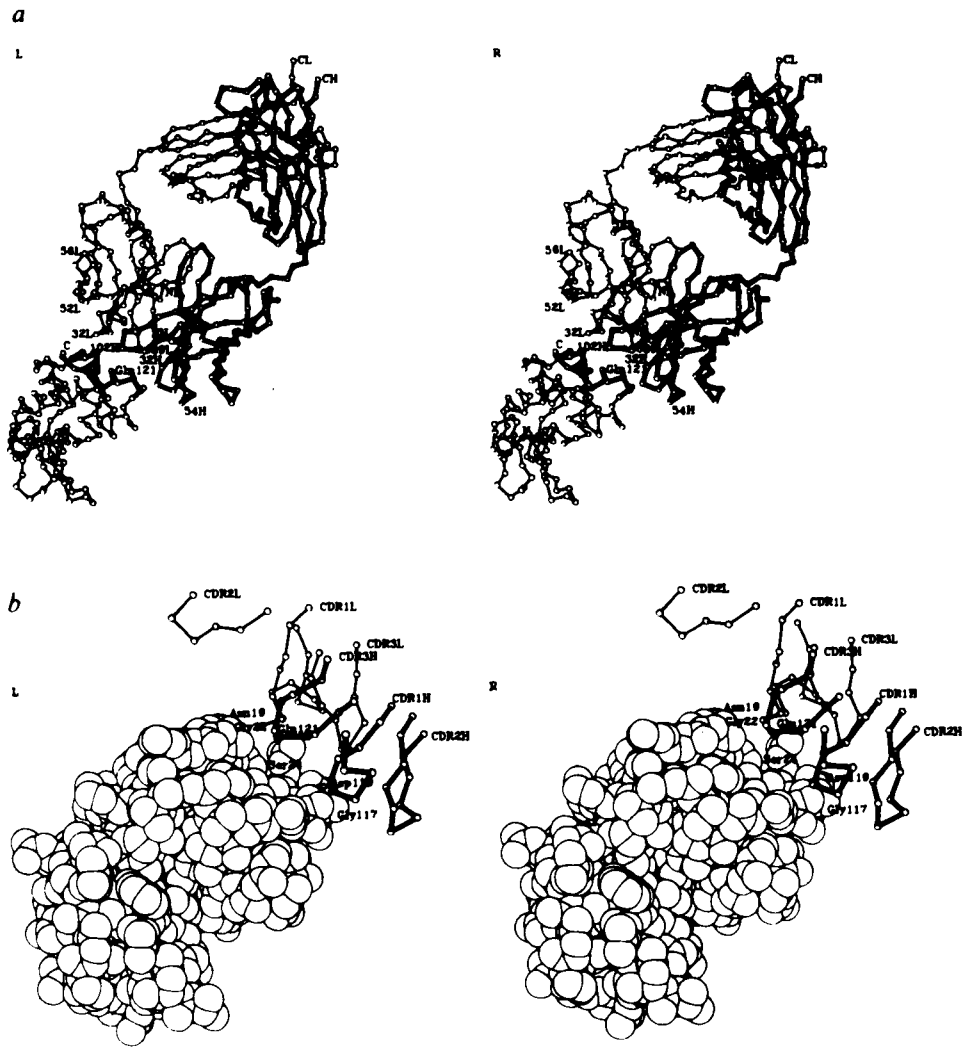


Fig. 25. Preliminary three-dimensional stereo models of the monoclonal antibody–lysozyme complex. (a) Stereo diagram of the α-carbon skeleton obtained by fitting the structure of lysozyme and Fab Newm to the electron density map. (b) Stereo diagram of the space-filling model of lysozyme in contact with the CDR segments of Fab Newm. (Reprinted by permission from Amit *et al.*, 1985a. Copyright © 1985 Macmillan Journals Limited.)

1973) and for lysozyme (Blake *et al.*, 1965). Although the main conclusions of studies of complexes of myeloma Fab fragments and ligands are supported, the monoclonal Fab appears to be a more flexible molecule. Also the combining site is not just a cleft lined by all six CDR segments but appears to be a larger area extending farther out. This suggests that the antigen binding sites of antibodies may differ significantly in topography from one to another and from the earlier models based on myeloma Fab fragments and light chain dimers. The recent crystallographic analysis of the λ-chain Bence–Jones protein dimer Loc even suggests that there may be protrusions in the binding site (Chang *et al.*, 1985). Other crystallographic studies now in progress, such as that of a complex of the Fab from a monoclonal anti-*p*-azophenylarsonate antibody and its hapten (Mariuzza *et al.*, 1984), may help resolve the question of whether a conformational change occurs when an antibody combines with an antigen. Also, the work of Amit *et al.* (1985a,b) is helping to define the determinant sites (epitopes) of protein antigens such as lysozyme.

B. Three-Dimensional Structure of the Hinge Region, Fc Fragment, and Correlation with Binding Functions

1. The Hinge Region

The hinge region is the segment of the heavy chain that covalently links the Fab and Fc of IgG, IgA, and IgD and also crosslinks the two heavy chains in the monomeric molecule. As discussed previously (Section II,B,3), interest has focused on the hinge region because it is postulated to transduce a signal from the antibody combining site to the biological effector domains of Fc (Metzger, 1978; Pecht, 1982) and to confer flexibility on the molecule (Huber, 1984). As shown earlier in Fig. 9, the hinge region of each immunoglobulin class and subclass is unique in its primary structure (and presumably in its spatial structure); furthermore, each hinge region is encoded by one or more exons that are unrelated to the exons coding for the characteristic V and C domains of immunoglobulins.

Because of the flexibility and apparent disorder of the hinge segment in Fab fragments, it has been difficult to determine the three-dimensional structure of the hinge region by crystallographic analysis. In fact, the conformation of the hinge is known for only one intact protein, IgG1 Kol (Marquart *et al.*, 1980). The two other immunoglobulins for which spatial structures have been reported (Dob and Mcg) are abnormal proteins in which the hinge region is missing; this no doubt facilitated their crystallographic analysis. The hinge segment of IgG1 Kol forms a short poly(L-proline) double helix from Cys-226 to Cys-230. Both the segment preceding and following the polyproline helix are flexible in the Fc fragment crystals (Deisenhofer *et al.*, 1976). This conclusion is supported by a

proton nuclear magnetic resonance study of an intact human IgG1 and of IgG1 Dob, which has a hinge deletion (Endo and Arata, 1985). Thus, the rigid polyproline helix acts as a pivot and is flanked on both sides by flexible segments that would allow independent movement of the Fab arms and the Fc part (Huber, 1984). These findings give support to the view that flexible segments at the V/C switch region and in the hinge allow conformational movements that facilitate binding of antibodies to multivalent antigens.

Although the three-dimensional structure of the hinge region is not known for any other immunglobulin molecule, Marquart *et al.* (1980) have predicted that the whole repeating unit in the quadruplicated hinge of human IgG3 exists in a polyproline double helix conformation 48 residues long. They also emphasize that the switch peptides and elbow angle at the V/C junction in both light and heavy chains are somewhat like the hinge region in conveying flexibility that may allow movement of the Fab arms. Detailed predictions of the hinge structure of IgA and IgD have not been made. The high proline content of the $\alpha 1$ and $\alpha 2$ hinges suggests that they too are in a polyproline conformation. However, the human δ chain hinge is divided into two parts. The first half is the GalN-rich segment, and calculations suggest it has a random structure, whereas the second half appears to have an ideal α-helical structure (Putnam *et al.*, 1982). These observations illustrate the danger of generalizing conclusions on the biological implications of antibody conformation on the basis of a single subclass (IgG1); at the same time they highlight the need for crystal structure data for IgM, IgA, IgD, and IgE.

2. The Fc Region

Three-dimensional structures of Fc have been reported for the intact proteins or the Fc fragments of several human myeloma IgG proteins (Kol, Dob, Mcg) and for the Fc of human and rabbit IgG prepared from normal pooled serum (Table VI). In all cases the basic structural features of the polypeptide backbone are similar although the orientation of the carbohydrate may differ somewhat. In Fab there are strong lateral interactions between V_L and V_H to form the Fv module and also between C_L and C_H1; in Fc similar lateral interactions occur between the two C_H3 domains, but not between the C_H2 domains. As shown in Fig. 26, the C_H2 domains interact with but are separated by the single complex GlcN carbohydrate on each chain. Although the carbohydrate probably helps stabilize the first half of Fc, the C_H2–C_H2 module may be more subject to conformational changes because of its proximity to the hinge and its weak lateral interactions. This conformability may be associated with the binding of complement.

Although there is a single complex GlcN oligosaccharide N-linked to Asn-297 in human and animal γ chains, the branching carbohydrate structure may be quite

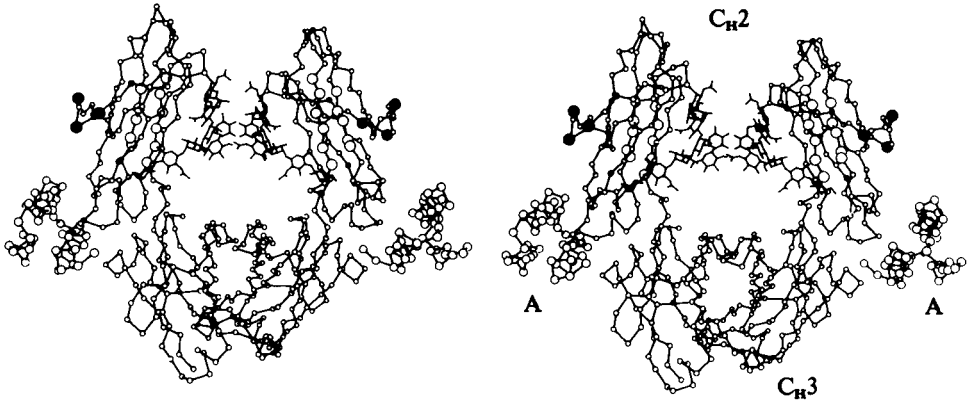


Fig. 26. Stereo drawing of the α-carbon backbone of the Fc of normal pooled human IgG bound to fragment B of *Staphylococcus aureus* protein A. This model from Davies and Metzger (1983) is based on the coordinates of Deisenhofer (1981) taken from the Brookhaven Protein Data Bank and includes additional data on the binding site of C1q. The complex GlcN carbohydrate occupies the space between the two C_H2 domains. Fragment B of protein A is shown with clusters of open circles in the region between the C_H2 and C_H3 domains. The symbols ●, •, and ○ on the Fc are the α carbons of the residues that Prystowsky *et al.* (1981), Bragado *et al.* (1982), and Burton *et al.* (1980), respectively, have proposed to bind C1q. (Reproduced, with permission, from the *Annu. Rev. Immunol.* **1**, © 1983 by Annual Reviews Inc.)

heterogeneous (Parekh *et al.*, 1985). The carbohydrate contributes some disorder and asymmetry to the C_H2 domains, and four hexose units are too mobile to be visualized. The role of the carbohydrate in the three-dimensional structure of IgG was discussed in Section V,A,3. The multiple carbohydrates present in IgM, IgA, IgD, and IgE will add to the difficulty of crystallographic analysis of these immunoglobulins. Complete removal of the GlcN carbohydrate by an enzyme such as *N*-glycanase might help solve this problem.

3. Binding Sites for Protein A and for Complement

Whereas the antigen combining sites of antibodies have long been investigated by a variety of approaches, in particular by study of the primary and three-dimensional structure of myeloma proteins, little is known about the structure of other functional sites of antibodies. These other biological properties are often called biological effector functions because they are involved in producing the ultimate biological effects of the humoral immune response, e.g., complement activation, phagocytosis, and certain cytotoxic reactions that follow binding to cells with Fc receptors. It should be kept in mind that although most biological effector activities such as C1q binding are associated with Fc, Fab may also be

involved at some stage. For example, C4b binding to antibody–antigen aggregates occurs to Fab probably via a reactive thiol ester, but not to Fc (Campbell *et al.*, 1980).

Only one binding site in Fc has been established by crystallographic analysis, i.e., the site for fragment B of *Staphylococcus aureus* protein A (Deisenhofer, 1981). All other functional sites have been studied by a series of indirect approaches such as enzymatic preparation of fragments of IgG, chemical modification of IgG, or competitive cell-binding experiments. In these the first objective has been to identify the domain in which a cytotropic function is vested rather than to localize the sequence site or stereochemical patch in IgG that is responsible for the function. Burton (1985) and Dwek *et al.* (1984) have reviewed the many attempts to characterize the functional sites of the C region of IgG, and these will be referred to later in Section VII,B.

Deisenhofer (1981) determined the three-dimensional structure at 2.8 Å resolution of a human Fc fragment of IgG and of its complex with fragment B of *S. aureus* protein A (SPA). SPA, a component of the cell wall of *S. aureus*, has a single polypeptide chain that contains four homologous domains in the order D, A, B, and C, each of which is a monovalent Fc binding unit about 60 residues in length. Tryptic cleavage of SPA produces the active fragments plus a C-terminal segment that does not bind to Fc. The biological role of SPA is unknown. However, SPA has been a very useful immunological reagent because it binds with high affinity to the Fc of the IgG of many species and thus can be used in affinity chromatography. SPA binds to human IgG1, IgG2, and IgG4 and to their Fc fragments, but not to IgG3.

It had earlier been shown that SPA binds to the intact Fc fragment but not to the isolated C_H2 and C_H3 domains, so both domains or their junction must be involved. In fact Deisenhofer (1981) demonstrated that FB (the B fragment of SPA) binds at the C_H2–C_H3 contact in each γ chain (Fig. 26). The Fc contacts include residues 251–254 and 309–315 in C_H2 and 430–436 in C_H3. The failure of human IgG3 to bind protein A was explained by the substitution of arginine in IgG3 for His-435 in IgG1, IgG2, and IgG4. Model building with a display system showed that because of the size and charge of arginine a place could not be found for it in the FB–Fc complex (Deisenhofer, 1981). With the exception of small changes at the contact site, the C_H3 domain of the FB–Fc fragment is identical with C_H3 in Fc lacking FB. However, the C_H2 domain of the FB–Fc complex is even more disordered than C_H2 in Fc alone. This accords with the view that the C_H2 domain is flexible and may be distorted or undergo a conformational change due to antigen binding in Fab (Huber, 1984; Burton, 1985; Klein *et al.*, 1981).

Although the binding of the complement subcomponent C1q has not yet been visualized by crystallographic analysis, knowledge of the three-dimensional structure of IgG and of the FB–Fc complex has contributed greatly to under-

standing the nature of the C1q binding site. C1q is a large protein molecule (M_r 410,000) that has a peculiar shape likened to a "bunch of flowers" because of its six protruding stalks, each having a globular head. Each head will bind to IgG or IgM or to their Fc fragments, so C1q is hexavalent. Effective fixation of C1q to initiate activation of the complement cascade requires two IgG molecules or one IgM. Human IgG1, IgG2, and IgG3 and also IgM fix complement, but IgG4 binds weakly or not at all. IgA1 and IgA2 do not bind C1q. It is questionable whether IgD and IgE fix C1q, but it also is physiologically irrelevant because of their low normal concentrations. Aggregation of IgG increases the binding of C1q.

Many studies, reviewed by Burton (1985) and others, have established that C1q binds to isolated Fc fragments of IgG. Colomb and Porter (1975) found that C1q binds to the plasmin-derived Facb fragment, which lacks the C_H3 domain, and Yasmeen *et al.* (1976) found that proteolytically prepared C_H2 domains of IgG bound C1q with affinity similar to that of Fc. Other attempts to localize the C1q binding site were based on three approaches: (1) comparison of the sequences of immunoglobulins of different subclasses and species that fix C1q with those that bind weakly or not at all (Brunhouse and Cebra, 1979); (2) reduction and chemical substitution of IgG (Vivanco-Martinez *et al.*, 1980; Bragado *et al.*, 1982); and (3) study of the inhibition of C1q binding by peptides corresponding to sequences in C_H2 of IgG, e.g., a 62-residue fragment (Kehoe *et al.*, 1974), and the peptides from Lys-274 to Gly-281 (Johnson and Thames, 1977; Boackle *et al.*, 1979) or from the region Phe-275 to Lys-290 (Prystowsky *et al.*, 1981; Lukas *et al.*, 1981). Comparison of the primary structures did not identify any single residue or sequence that uniquely correlated with the ability to bind C1q but did appear to implicate the sequence from Lys-290 to Glu-295 (Brunhouse and Cebra, 1979).

Studies of the inhibitory effect of the peptides from the C_H2 region are reviewed in detail by Burton (1985) in relation to the proposal of Burton *et al.* (1980) that the C1q site involves residues in the sequence from Gly-316 to Lys-322 and from Thr-335 to Lys-338. The latter approach focuses attention on two adjacent regions of C_H2 that have a high charge density and are located on the last two β strands of the C_H2 domain. Thus, at least three sites in the C_H2 domain have been proposed for attachment of C1q. Actually these are not far apart in the three-dimensional structure, and because of its large size C1q may bind at several sites on C_H2.

Although the studies described above have not been conclusive, the availability of the crystal structures of Fc and of FB–Fc has permitted spatial identification of the sequences that various workers have proposed to be critical for the binding of C1q (Fig. 26). The binding of protein A does not interfere with complement binding so the C1q binding site must be in the surface area of C_H2 not covered by FB. The large size of C1q would prevent its attachment to the inner side of the

C_H2 domain, where the carbohydrate occupies much of the lateral interdomain space. Thus, it is likely that C1q binds to the upper exterior of C_H2 in the area implicated by the studies described above.

4. Immunoglobulins of Other Classes and Species

As is evident from Table VI, most of the high-resolution X-ray analyses have been on human IgG myeloma proteins or human κ or λ Bence–Jones proteins. However, the Fc structures were determined on fragments of normal pooled human or rabbit IgG. The structure of the pFc′ fragment of guinea pig IgG1 has also been reported (Phizackerly *et al.*, 1979). Although the structures of the Fab fragments of two mouse IgA myeloma proteins have been determined, data on intact IgA proteins and on other immunoglobulin classes are lacking. The paucity of data on IgA, IgM, IgD, and IgE reflects two circumstances: (1) the widespread conviction that the fundamental spatial features of antibody structure have been solved by crystallographic analysis of IgG and its Fab and Fc fragments and (2) the difficulty of crystallization and structural study of the other four classes. All four are multiply glycosylated, several are polymeric (IgA and IgM), and the remaining either are not abundant (IgD) or are rare (IgE).

There is no doubt that the basic features of the antigen binding site have been established, and the structure of Fv is rather independent of the class of the H or L chain. However, recent studies of monoclonal antibodies referred to above indicate that the Fv binding sites have much more topographical variation than had been suspected just from studies of myeloma proteins. Likewise, the immunoglobulin fold may be the basic three-dimensional structure of all C region domains regardless of class or species. Nonetheless, there are significant differences in the functional activities of the C regions of IgG, IgM, IgA, and IgE and in their cell membrane receptors. The disparity in these properties may be attributable to the structural dissimilarity in the hinge regions of IgG and IgA, or to the replacement of the hinge by an extra domain in IgM and IgE, or to the additional oligosaccharides, as well as to differences in primary structure. Crystallographic study of IgA, IgM, and IgE should contribute much to understanding the basis of their characteristic biological effector functions, in particular their differences in C1q binding and in specific cytotropic functions. To date the conformation of IgM has only been investigated by small-angle X-ray scattering (Wilhelm *et al.*, 1984) and other physical techniques.

At present the only approach to estimation of the three-dimensional structures of immunoglobulins other than human IgG1 is by comparison of amino acid sequences and trying to fit these into the structure of human IgG1. This can be done by use of interactive molecular graphics displays or by alignment of many sequences. Beale (1984), following up an earlier study (Beale and Feinstein, 1976), used the latter approach for an extensive comparison of all available

amino acid sequences of different classes of immunoglobulins and histocompatibility C region domains from various mammalian species. Correlations between conservation of sequence and three-dimensional structure were sought. In addition to the invariant disulfide bond and tryptophan residue, all immunoglobulin C domains shared several conserved features that could be related to the crystallographic structure of immunoglobulin fold. However, these related mainly to individual positions in the sequence rather than to segments of the sequence.

VII. Cytotropic Functions of the Fc Region

A. Membrane-Bound Immunoglobulins as Antigen Receptors

The B-cell receptor for antigen has six unique characteristics: (1) It is present only on the surface of B lymphocytes, the only cells that have the potential for antibody production. (2) The membrane-bound receptor (mIg) is identical in structure to the secreted antibody (sIg) except for the presence of a short carboxy-terminal tailpiece that differs for mIg and sIg. (3) Though it is usually IgM and/or IgD, the receptor can be any of the five classes, and a class switch may occur during B-cell differentiation. (4) Though B cells as a class may exhibit a very large number of receptors, each clone of B cells has but a single type of receptor, and that receptor is characterized by its specificity and affinity for a particular antigenic determinant (epitope). (5) During differentiation of a clone of B cells having a particular antigen specificity, that specificity does not change although the class of the mIg may switch. (6) The specificity of the antigen receptor is vested in the V regions of the light and heavy chains and is the same for both mIg and sIg.

Elucidation of the structure of the B-cell antigen receptor was finally achieved by the methods of molecular biology after more than a decade of study by biochemical and immunochemical approaches. The early research, which focused on IgM and IgD, is reviewed by Dickler (1978) and Leslie and Martin (1978). Subsequent work emphasized the role of IgD (Goding, 1980; Pollock and Mescher, 1980). The relative role of IgM and IgD as receptors has been reevaluated by Vitetta (1982), and a symposium on IgD has a series of papers on this subject (Thorbecke and Leslie, 1982). Though IgD is a minor serum component in humans, it is a major Ig receptor on the B-cell surface, where IgM and IgD are often coexpressed. IgD appears later than IgM and is considered to be a characteristic of the mature or "virgin" resting B cell (Blattner and Tucker, 1984). One suggestion is that IgD is a triggering receptor, i.e., that interaction with antigen and T cells results in B-cell differentiation (Vitetta and Uhr, 1975;

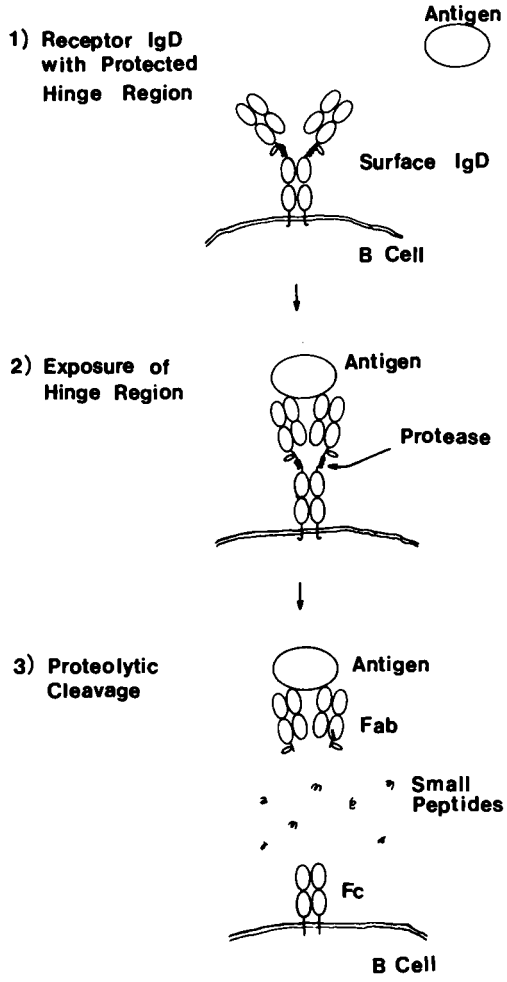


Fig. 27. Hypothetical scheme for the activation of B-cell function triggered by the proteolytic cleavage of receptor IgD (from Putnam *et al.*, 1982).

Vitetta, 1982). One hypothesis is that after interaction with antigen, IgD is proteolyzed and the Fc fragment is endocytosed within the B cell. This is illustrated schematically in Fig. 27.

The solution to the mystery of the B-cell antigen receptor came through gene cloning. The existence of mIg somewhat larger than sIg but similar in structure except for the carboxyl terminus had been suggested by the experiments referenced above and by many others (Kehry *et al.*, 1980; Oi *et al.*, 1980; Singer *et*

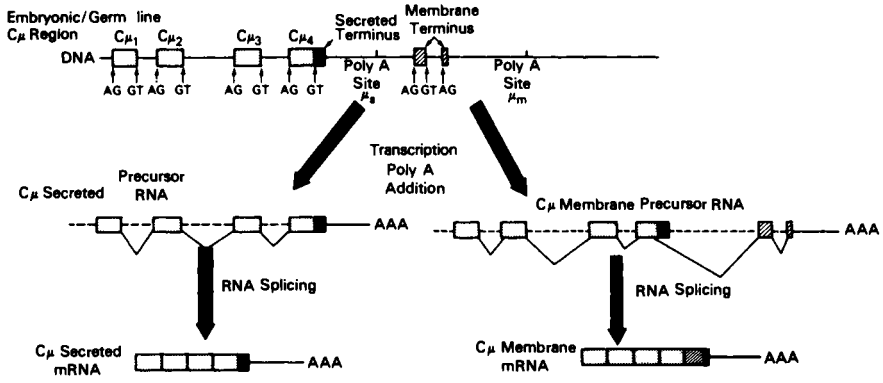


Fig. 28. Schematic model for creating distinct secreted and membranous forms of IgM from a single constant region locus. Donor (GT) and acceptor (AG) splice sites for RNA splicing border the four separated Cμ domains. Alternative sites of poly(A) addition and RNA splicing result in different mRNAs containing either the secreted or the membrane terminus (from Korsmeyer and Waldmann, 1984).

al., 1980; Alt *et al.*, 1980; Vassalli *et al.*, 1980). The approach changed when it was shown by cloning the cDNAs that two mRNAs can be produced from a single mouse immunoglobulin μ gene by alternative RNA processing pathways and that the two mRNAs encoded membrane-bound and secreted forms of the μ chain (Rogers *et al.*, 1980; Early *et al.*, 1980). The two μ chains differed in their carboxy terminus. The μ_s chain had a 20-residue hydrophilic C-terminal segment after the Cμ4 domain; in the μ_m chain this was replaced by a 41-residue C-terminal segment containing a hydrophobic sequence that appeared to be embedded in the cell membrane. This process is illustrated in Fig. 28.

In rapid succession the membrane and secreted forms of other mouse C_H genes were identified: Cγ (Rogers *et al.*, 1981; Tyler *et al.*, 1982), Cα (Tucker *et al.*, 1981), Cε (Ishida *et al.*, 1982), and Cδ (Cheng *et al.*, 1982). White *et al.* (1985) showed that the human δ gene also has a δs and δm form.

The structures of the membrane and secreted forms of the murine isotypes have been summarized by Blattner and Tucker (1984), who have also discussed the expression and role of the two forms. All the membrane domains are apparently encoded by two exons. The transmembranal (hydrophobic) segment contains 26 amino acids in a highly conserved sequence, but the hydrophilic cytoplasmic segment varies from 3 residues (IgM) to 28 residues in length (IgG isotypes and IgE). Honjo (1983) suggests that the hydrophobic sequence is conserved because it may be anchored to a common membrane protein, whereas the cytoplasmic segment may differ from class to class because it transmits a different signal.

B. Fc Cell Receptors

1. The Fc Receptor for IgG

With respect to the Fc receptor (FcR) there are a number of questions to be asked: (1) What is the cell receptor? (2) How many kinds are there, and how do they differ from one cell type to another? (3) What is the receptor site on the Ig molecule, and how does it differ from Ig class to class? (4) How does the interaction of Fc and FcR prompt a biological response? None of these questions can yet be answered positively. Although information on the first three is developing rapidly, the answer to the fourth is not yet in sight. Most of the research deals with cell receptors for IgG (Dickler, 1978), but progress is also being made for IgE because of interest in the specificity of the response to binding of IgE by mast cells and basophils (Ishizaka, 1985; Metzger, 1983). Because the study of receptors is more the purview of cellular rather than molecular immunology, the subject will be discussed only briefly here.

A variety of lymphoid cell types bind IgG in either the monomeric or aggregated form. These include monocytes, macrophages, polymorphonuclear leukocytes (PMNs), and lymphocytes. Platelets and certain placental cells also have Fc receptors. IgG binding by killer cells and macrophages is an important protective mechanism of the cellular immune response leading to phenomena such as antibody-dependent cell-mediated cytotoxicity and phagocytosis.

In a comprehensive review Burton (1985) has summarized recent studies of Fcγ receptors on cells of various types. In a series of tables he has listed the results of much research on: (1) binding of human IgGs to human monocytes, (2) binding of IgG to homologous and heterologous macrophages, (3) binding of IgG to human PMNs, (4) binding of IgG to lymphocyte Fc receptors, and (5) binding of IgG to placental Fc receptors. From this comparison several generalizations emerged: (1) With respect to binding of monomeric IgG1 by human cells, monocytes bind IgG1 very avidly, trophoblasts bind more weakly, and lymphocytes, neutrophils, and platelets bind even more weakly. (2) Most cell types bind aggregated IgG with much greater affinity than monomeric IgG. (3) Human cell types exhibit a similar subclass specificity pattern in binding affinity (IgG1, IgG3 > IgG2, IgG4). From the latter finding, together with studies of hinge-deleted and aglycosylated proteins, Burton (1985) concludes: "There is no good evidence for the involvement of the Cγ3 domain in IgG binding to any human cell type." However, he regards the data on the interaction of Cγ2 and Cγ3 domains with Fc receptors on cells of other species as contradictory and inconclusive.

2. IgG Binding Sites for FcR

Burton's conclusion (1985) that the Cγ3 domain is not involved in binding to FcR of human cells contrasts with earlier suggestions that the Cγ3 domain plays an important if not the dominant role (Dorrington and Painter, 1974; Ciccimarra *et al.*, 1975; Ellerson *et al.*, 1976; Yasmeen *et al.*, 1976; Dorrington, 1979; Diamond *et al.*, 1981; Phizackerly *et al.*, 1979). These workers primarily studied the mediation or inhibition of binding to FcR by proteolytic fragments lacking the Cγ3 domain (Facb) or equivalent to the Cγ3 domain (pFc'). More recent work has tended to assign a greater role to the Cγ2 regions (Dorrington and Klein, 1982, 1984; Hofstaetter *et al.*, 1984). However, the proximal part of the Cγ3 domain has recently been reported to be the most probable site for binding of the placental FcR (Tønder and Matre, 1985).

One of the problems in identifying the binding sites for FcR is that such a variety of cell types have been studied with either homologous or heterologous IgG that it is difficult to generalize the results. For example, Stanworth (1984) concludes that monocyte binding of human IgG involves a site within the Cγ2 domain, whereas a site within the Cγ3 domain is responsible for heterologous macrophage binding. Thus, while it is clear that the biological effector actions of antibodies are mediated through the Fc region, the molecular nature of the sites and of their interaction with FcRs is still obscure.

3. The IgE Receptor and IgE-Binding Factors

Mast cells, basophils, and certain rodent tumor cell lines have a receptor that binds IgE with very high affinity but has little affinity for other Ig isotypes. This high affinity and the specific nature of the biochemical response when two bound IgE antibodies are bridged by an allergen have prompted intensive effort to characterize the IgE receptor. Furthermore, the IgE receptor system may serve as a general model for study of other Fc cell receptors (Ishizaka, 1985).

Metzger (1983) has reviewed the many studies that have been made of the IgE receptor, most of them on a rat basophilic leukemia (RBL) cell line that expresses an IgE receptor. Although much descriptive information has been adduced, as yet only a schematic structural model can be drawn. The IgE receptor is a heterodimer consisting of an α chain (M_r 50,000), which is glycosylated, and a β chain (M_r 30,000 to 35,000), which is phosphorylated. The α chain by itself can bind IgE and thus has the receptor site and is assumed to be on the cell surface. The α chain binds IgE mole for mole with high affinity and specificity and interacts with the β chain buried in the cell membrane. The bridging of cell-bound IgE antibodies by multivalent antigen or by divalent anti-IgE results in clustering of receptors, which triggers the release of mediators of anaphylaxis

such as histamine (Ishizaka *et al.*, 1981). The bridging of the IgE receptors induces transmembrane activation of adenylate cyclase, activation of cAMP-dependent protein kinase, and secretion of granules with all the sequelae of immediate-type anaphylactic hypersensitivity.

Study of proteolytic fragments of a human myeloma IgE protein has suggested that the site to which the receptor binds is on the Cε3 or Cε4 domains or both (Dorrington and Bennich, 1978). Similar fragments have not been obtained from rat IgE, but other approaches suggest that the receptor interacts with the rodent IgE well up into the Fc region (Metzger, 1982). In a novel approach Geha *et al.* (1985) studied the inhibition of the Prausnitz–Kustner reaction in human skin by a human ε chain fragment that had been synthesized by cloning in *E. coli*. The fragment from the Cε2–4 region (residues 218–547) completely blocked the passive sensitization by serum containing IgE antibody to ragweed. The inhibitory activity was about 30% that of intact IgE on a molar basis.

Several IgE-binding factors (IgE-BF) that regulate the IgE response have been characterized. These are produced by rodent T lymphocytes and specifically bind to IgE. Ishizaka (1985) has reviewed the developments in the study of these factors, one of which is called IgE-potentiating factor and the other IgE-suppressive factor. Martens *et al.* (1985) have cloned and sequenced these factors. DNA sequence analysis of an IgE-BF cDNA indicated the presence of a 556-amino acid (62-kDa) protein coding region with two potential N-glycosylation sites and four potential proteolytic cleavage sites. It is suggested that the IgE-potentiating and IgE-suppressive factors share common precursor polypeptides and may differ chiefly in their glycosylation and their posttranslational processing. A different IgE-binding protein was identified by molecular cloning by Liu *et al.* (1985). This IgE-BF has an M_r of 31,000 but the coding sequence of the DNA was obtained for only about the carboxyl-terminal half of the protein. The significance of this newly identified IgE-BF and its relationship to IgE receptors have yet to be determined. However, the studies just described are rapidly advancing knowledge of the mechanism of the binding of IgE Fc to receptors and to lymphokines and of the immunoregulatory role of this interaction. The IgE system may well prove to be the model for study of the cytophilic interaction and immunoregulation of IgG, IgM, and IgA.

4. Future Prospects

In summary, although understanding of the structure and topography of antibody binding sites for antigen is well advanced, little is known or can even be conjectured about the molecular nature of the binding of Fcγ to cell receptors though progress has been rapid for Fcε. Despite the advances in recombinant DNA technology, most work on Fc receptors is still at the descriptive level. However, cell receptors for lipoproteins, transferrin, hormones, etc., have re-

cently been isolated, cloned, and sequenced so that they can be studied in purified form. Progress in this field is rapid and early success in the purification of Fc receptors can confidently be predicted. When this occurs, we should learn whether different lymphoid cells have different receptors for IgG and how receptors may differ for the several classes of immunoglobulins.

The availability of purified receptors of known structure will greatly facilitate study of their interaction with Fc, and this will contribute to understanding the mechanism whereby cells are triggered to evoke a specific immune response. One possibility is that receptors may have structures homologous to immunoglobulins. Indeed, the receptor for transepithelial transport of IgA and IgM (poly-IgR) contains immunoglobulin-like domains (Mostov *et al.*, 1984), and so does the T-cell receptor for antigen (Sim *et al.*, 1984; Patten *et al.*, 1984; Hannum *et al.*, 1984; Arden *et al.*, 1985; Tunnacliffe *et al.*, 1985; Hood *et al.*, 1985; Robertson, 1985). Thus, these receptors and other cell surface recognition molecules have been grouped in the immunoglobulin supergene family and are discussed later in Section VIII.

C. Biologically Active Peptides Derived from Fc

Although most of the attempts to identify the sites of Fc effector functions have involved fragmentation of IgG by various proteolytic enzymes, increasing effort is being directed toward testing specific peptides derived from Fc or synthesized chemically. An assumption implicit in this approach is that the biological activity being assayed is associated with specific amino acid sequences rather than being dependent on three-dimensional structure and conformation. An alternative concept is that although peptides may assume many conformations in solution, a few of these may be effective because they are similar to the natural topography of the peptide sequence in the intact antibody molecule. However, in that case the activity would be greatly attenuated, which is generally the case in studies of the biological activity of Fc peptides. Cognizant of this problem Stanworth (1984) has pointed out both the opportunity and the limitations in the use of synthetic peptides to probe the structural basis of Fc activities.

A series of recent reports as well as a symposium (Najjar and Fridkin, 1983) have focused attention on several peptides proteolytically derived from the Fc region of human IgG that are reported to have biological activity and may have an immunoregulatory role. The most widely studied are two tetrapeptides named tuftsin (after the institution of the discoverer) (Najjar and Fridkin, 1983) and rigin (after the city Riga) (Veretennikova *et al.*, 1981). A larger 23-residue peptide from IgG1 (designated p23) has also been the subject of a series of papers (Morgan *et al.*, 1982). In literature available only in abstract form, Russian workers have reported on biologically active fragments of immunoglobulins G, M, E, and A—

mainly pentapeptides—that they call immunopoietins because they are said to have an immunomodulating effect (Cipens *et al.*, 1984). These are tantalizing reports because if widely validated they will not only help elucidate the mechanism of biological effector functions but might also lead to useful therapeutic agents in cases of immunodeficiency.

Tuftsin has the sequence Thr-Lys-Pro-Arg corresponding to residues 289–292 of the human γ1 chain. Tuftsin can be obtained by tryptic digestion from the C_H2 domain of all four subclasses of human IgG. Surprisingly, a computer search shows that this simple tetrapeptide could not have been obtained by tryptic digestion of any other protein whose sequence was listed in the 1984 protein data base. Tuftsin is reported to be a natural activator of the macrophage by stimulating both its phagocytic and kinetic activity; it is also said to enhance B-cell proliferation and to have antitumor activity. These activities have been the focus of a recent symposium (Najjar and Fridkin, 1983). Rigin, also heralded as a phagocytosis-stimulating tetrapeptide, is derived proteolytically from the first four residues of the C_H3 domain of all four subclasses of human IgG (Veretennikova *et al.*, 1981). Its sequence is Gly-Gln-Pro-Arg corresponding to positions 340–343 in the γ1 chain. Analogs of both tuftsin and rigin have been tested for a variety of immunogenic activities. Rigin is part of several other biologically active peptides derived from human IgG; the relationship of these peptides to the structure of Fc is diagrammed in Fig. 29.

A 24-residue peptide that includes the rigin sequence and has immunoregulatory activity has been isolated by CNBr cleavage of a plasmin digest of IgG1 by Morgan *et al.* (1982) (residues 335–358 in γ1). Other papers describe the immunoregulatory effects of the synthetic peptide p23 representing residues 335 to 357 (Morgan *et al.*, 1983; Hobbs *et al.*, 1985). The p23 peptide induces murine B cells to secrete immunoglobulins. Martinez *et al.* (1983) synthesized an IgG decapeptide (residues 335–344) that begins just after the plasmin cleavage and ends with rigin. The decapeptide stimulates phagocytosis by polymorphonuclear leukocytes. Thus, there is increasing evidence that at least some of the immunoregulatory activities ascribed to Fc are vested in peptides that may be released from Fc produced by macrophage cleavage of IgG.

In the three-dimensional model of the structure of Fc the p23 peptide of Morgan *et al.* (1982) begins at the plasmin cleavage site between the C_H2 and C_H3 domains (Fig. 29). Limited digestion with trypsin cleaves human IgG1, after Lys-335 and also after Arg-344; this could release a hexapeptide ending with rigin, which itself begins the C_H3 domain. It has been suggested that antibody molecules immobilized on a cell surface by attachment to an Fc receptor or linked via Fab to a cellular antigen may undergo a conformational change that renders the connecting link between the C_H2 and C_H3 domains susceptible to cleavage by one of the many proteases present in plasma and phagocytic cells (Putnam *et al.*, 1985). Once the link is broken, further proteolysis could release

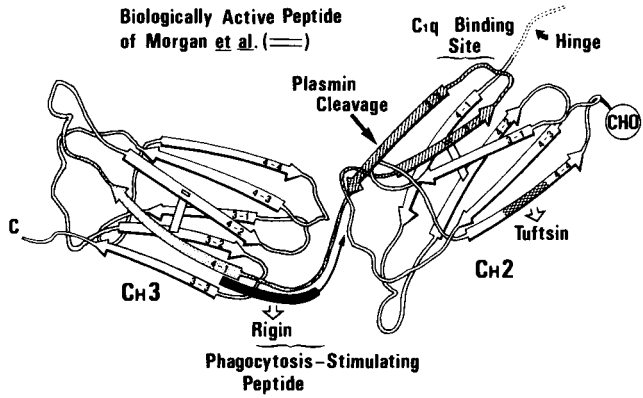


Fig. 29. Model showing the β-barrel structure of the Fc fragment of human IgG1 (Davies and Metzger, 1983) and the sites of biologically active peptides derived from Fc. The numbering in the sequence of the γ1 chain is: tuftsin, 289–292; C1q binding site, 316–340; biologically active peptide of Morgan *et al.* (1982), 335–358; phagocytosis-stimulating peptide, 335–344; rigin, 340–343. Note that the C1q binding site shown here is the one proposed by Burton *et al.* (1980) and Dwek *et al.* (1984) and differs from that shown earlier in Fig. 26 (from Putnam *et al.*, 1985).

biologically active peptides such as p23 and rigin, which could exert their immunostimulatory effects. Figure 27 in Section VII,A illustrates a speculative mechanism for this hypothesis using B-cell receptor and immunoregulatory IgD as an example. One conclusion to be drawn from the provocative studies described above is that more effort should be directed toward understanding the biological activities of the Fc region and of the fragments derived from it.

VIII. Structure and Evolution of the Immunoglobulin Supergene Family

A. Evolution of the Immunoglobulins

Many previous articles and reviews have discussed the evolution of the immunoglobulins and have displayed computer-generated phylogenetic or genealogical trees that depict relationships of light and heavy chains within a species or of the same chain from species to species. Noteworthy references include Putnam (1977c), Barker *et al.* (1978), Kindt and Capra (1984), and Fudenberg *et al.* (1984).

Although various evolutionary schemes may differ in detail, there is general agreement on the chief principles. These are: (1) All immunoglobulin genes descend from an ancestral gene encoding a primordial protein of about 110

amino acids. This corresponds to the present-day exon, which encodes a single protein domain having the structure of the immunoglobulin fold. (2) Tandem internal gene duplication produced the primordial light chain genes, and further duplication and elongation resulted in the primitive heavy chain genes. (3) The separation of V and C genes and development of the mechanism for rearrangement of V genes and joining of V and C genes occurred early in vertebrate evolution, perhaps contemporary with or after the separation of light and heavy chain genes. (4) Because the principal elements for producing antibody diversity were established well before mammalian radiation, both κ and λ light chains and probably all five types of heavy chains are present in mammals; however, some isotypes may be hard to identify in lower species because of their divergence or low concentration. (5) The development of subclasses is a more recent phenomenon because subclasses such as those of IgG and IgA are species-specific. (6) On an evolutionary time scale, allotypes are a recent innovation; for example, the human Gm and Am allotypic markers are distributed nonuniformly throughout the world population.

The principles of immunoglobulin evolution were first recognized by visual comparison of the homology of amino acid sequences and were later put on a firm basis by use of computer programs. This approach led to the construction of evolutionary trees for immunoglobulins such as the one given in Fig. 30, which shows a hypothetical pathway for the evolutionary development of the genes for the C regions of the five heavy chains. Information gained from cloning Ig genes has strongly supported this hypothesis and has greatly extended understanding of the genetic basis. For example, the equivalence of protein domains and exons was established and the mechanism for genetic control of the rearrangement of the V regions and the joining of V and C was discovered, as was the way in which the class switch of heavy chains is mediated. Discovery of the evolutionary duplication of a segment containing γ, α, and ϵ genes in the chromosomal arrangement of the human C_H genes explains the existence of species-specific subclasses of IgG and IgA (Flanagan and Rabbitts, 1982).

Some aspects of immunoglobulin evolution are still unclear. One of the mysteries is the evolutionary origin of the hinge region; this differs considerably in structure from class to class and even among the subclasses of IgG and IgA (Fig. 9, Section II,B,3), yet it is absent in IgM and IgE. The structure of each hinge region is unique and appears unrelated to the prototype immunoglobulin sequence. The hinge is also the most mutable region in Ig chains, as shown by the frequency of deletions and duplications. Furthermore, the human γ and δ hinges are encoded by distinct genetic elements; in fact, the quadruplicated γ3 hinge is encoded by four similar but separate exons (Krawinkel and Rabbitts, 1982) and the δ hinge by two disparate and separate exons (White *et al.*, 1985). An exception is the hinge region of the mouse α chain, which is encoded by the C_H2 domain (Tucker *et al.*, 1981).

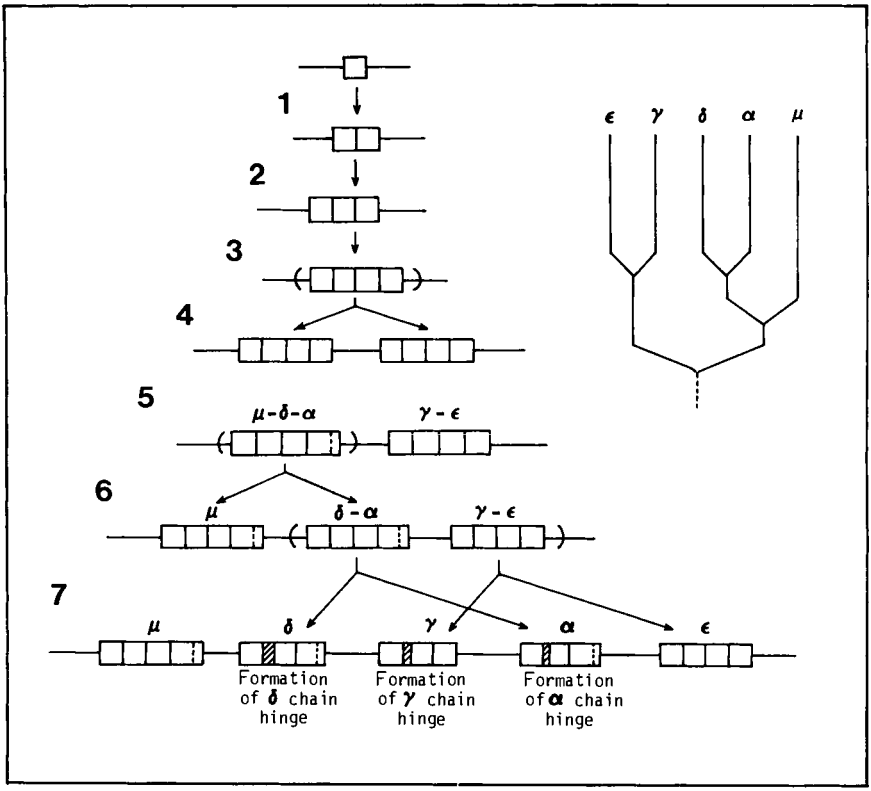


Fig. 30. Hypothetical pathways depicting the possible genetic events that might lead to the origin of the five H chain genes. Exons coding for the immunoglobulin C (constant) domains are boxed, with the tailpiece separated from the last C domain by a broken line. The noncoding DNA segments are represented by a thin horizontal line. For simplicity, introns between the domain exons are deleted from the diagram. The hinge region for different heavy chains is assumed to have evolved independently from the second C domain by an unknown genetic mechanism. Unbranched arrows in the pathway represent events of internal duplication that lengthened the C gene; branched arrows represent events of discrete duplication that created new C genes (parentheses on the DNA segment cover the range of discrete duplication). The evolutionary tree depicted by the pathway is shown as an inset at the top (from Lin and Putnam, 1981).

One possibility is that the hinge gene may have arisen by truncation of a C_H2 exon in the ancestral genome. In a computer search of the entire protein sequence data base the highest score for similarity to the first 30-residue segment of the human hinge was given by a segment of the Cμ2 region of the human and canine μ chains (Putnam *et al.*, 1981). Although the first and second segments of the human δ hinge have no apparent sequence similarity, the two may have a common evolutionary origin, for both score very high when compared to the same

segment of the Cμ2 domain. On the basis of these results, Putnam *et al.* (1981) predicted that the human δ hinge was encoded by two separate exons and suggested that the two δ hinge exons may have arisen by duplication and subsequent mutation of a common ancestral exon derived from the primordial gene for the Cμ2 domain. The proximity of the μ and δ genes in the H chain genomic arrangement makes this hypothesis more plausible. The prediction of two δ hinge exons was confirmed when White *et al.* (1985) determined the genomic structure of the human δ gene, and the exons accorded closely with the two predicted. Still to be explained, however, is the origin of the tailpieces of the heavy chains, Ts and Tm, each of which is encoded by a distinct exon.

Since few references have been cited above, it is pertinent here to summarize some of the information already presented that bears on the evolution of immunoglobulins. Figure 4 in essence is a genealogical tree of the human Ig C regions, and Fig. 5 shows the divergence of the subclasses and allotypes of human IgA. The internal homology of an immunoglobulin molecule is illustrated in Fig. 6 for IgD. The mechanism of the rearrangement of Ig genes is depicted in Figs. 15–17, and Fig. 18 demonstrates the equivalence of the structural domains of the proteins and the exons of the genes. The sequence homology of light chains and heavy chains is illustrated in several different ways in Figs. 19–22. All these examples support the hypothesis illustrated for the C regions in Fig. 30.

Obviously, the evolution of the V genes—though driven by the same process of gene duplication—followed a somewhat different pathway that resulted in 100 or more separate V genes and other genetic elements such as the J and D minigenes. The same process also occurred but to a more limited degree for other members of the immunoglobulin supergene family.

B. Other Members of the Immunoglobulin Supergene Family

1. β_2-Microglobulin

The term immunoglobulin superfamily (or supergene family) has been coined because of increasing awareness that a variety of proteins involved in the immune response that have a recognition function also have domain structures that exhibit homology in amino acid sequence to immunoglobulins. The members of this superfamily include a number of cell surface polypeptides such as class I and class II major histocompatibility antigens (MHC), lymphocyte antigens such as Thy-1 and Lyt-2, the poly-Ig receptor, and the recently identified T-cell receptor (Hood *et al.*, 1985). Molecular graphics displays show that the sequences can be more or less accommodated to the three-dimensional structure of IgG C domains (Patten *et al.*, 1984). Beale (1984) has compared many such sequences and has obtained evidence for conservation of structural features that correlate with the

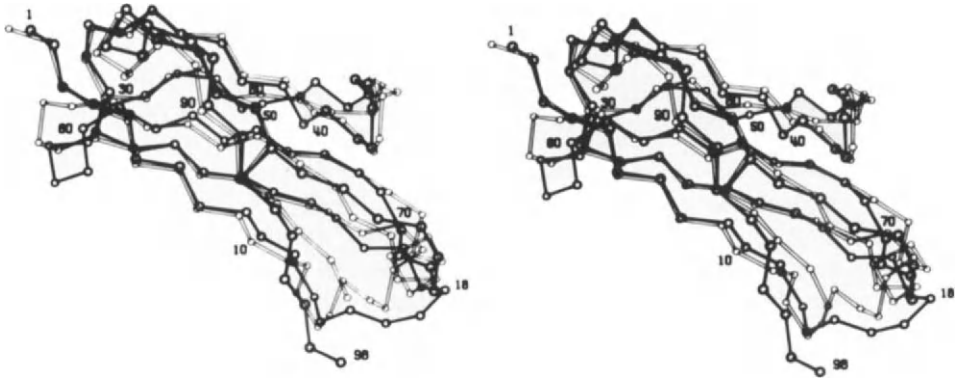


Fig. 31. Stereo comparison of the three-dimensional structure of the polypeptide chains of bovine β_2m and the C_H3 domain of human IgG. The β_2m chain is shown as the heavy line with sequence numbers and C_H3 as the light line without sequence numbers. From Becker and Reeke (1985).

crystallographic structure of IgG. However, β_2-microglobulin (β_2m) is the only example for which the three-dimensional structure has been solved and fitted to that of an IgG C domain (Becker and Reeke, 1985).

The light chain of the major histocompatibility complex class I antigen was first identified as a normal serum and urinary protein called β_2-microglobulin. β_2m is excreted in increased amounts in the urine of patients with myeloma and other forms of cancer and also by transplant patients during a graft rejection crisis (Poulik, 1975). β_2m has a molecular weight of 11,800, contains 99 amino acids, and its sequence is strongly homologous to Ig C domains. The three-dimensional structure of β_2m determined at 2.9 Å resolution strongly resembles Ig C domains in polypeptide folding and overall spatial structure (Becker and Reeke, 1985). This is illustrated in Fig. 31, which compares the polypeptide chain of bovine β_2m with the C_H3 domain of human Fc. The β_2m molecule is folded into the prototype β-barrel conformation made up of two antiparallel pleated sheets, one containing three strands and the other four. Figure 31 demonstrates that β_2m is indeed a member of the immunoglobulin superfamily. Furthermore, these results suggest that the MHC antigens associate through domain interactions similar to those of immunoglobulins and that both MHC domains and Ig domains evolved from a common ancestral gene.

2. *The Polyimmunoglobulin Receptor and Secretory Protein*

Understandable emphasis on the light and heavy polypeptide chains tends to obscure the fact that two other polypeptide chains, the J chain and the secretory

component (SC), may be associated with immunoglobulins. Neither has the characteristic division into V and C regions, and neither is associated with the monomeric classes IgG, IgD, and IgE. Whereas the main function of the J chain is to initiate polymerization of IgM and IgA (Koshland, 1985), SC is the proteolytically cleaved, extracellular portion of poly-IgR, the receptor for transepithelial transport of IgA and IgM (Mostov *et al.*, 1984). Thus, SC is derived from a larger transmembrane precursor on the surface of glandular epithelial cells.

The original interest in SC dealt with its association with dimeric or polymeric IgA in the form of secretory IgA (SC-IgA) in which J chain links the IgA monomers covalently. In mucosal fluids secretory IgA is the predominant form of IgA but not in serum. SC may also be present in the free form. SC is synthesized in epithelial cells, whereas IgA is made in plasma cells; hence, IgA myeloma proteins produced in large amounts in plasmacytomas lack SC. As a result, little was known about the structure of SC until the recent report of the complete amino acid sequence of human SC (Eiffert *et al.*, 1984). However, there was much interest in the secretory immune system, which has been the subject of several symposia (e.g., see McGhee and Mestecky, 1983). The chief function attributed to SC was protection of IgA from proteolytic degradation in the mucosal fluids, where secretory IgA functions as the first line of humoral immune defense. However, little is known about the possible influence of SC on the biological effector functions of IgA. SC also reacts reversibly with IgM and with its $(Fc)_{5\mu}$ fragment and does so with high affinity (Goto and Aki, 1984).

The focus changed once it was recognized that SC is not synthesized as a secreted protein, but rather is derived by proteolytic cleavage of the receptor for transepithelial transport (Brandtzaeg, 1981; Mostov and Blobel, 1983; Kuhn and Kraehenbuhl, 1983; Kuhn *et al.*, 1983; Mostov *et al.*, 1984). The cloning and sequencing of the cDNA for the rabbit poly-IgR showed that this receptor contains five extracellular domains that are homologous to each other and to immunoglobulin V regions (Mostov *et al.*, 1984). The complete poly-IgR precursor consists of 773 amino acids as follows: an 18-residue signal peptide, the extracellular poly-Ig binding portion containing 629 amino acids divided into five homologous domains of 110–115 amino acids, and a sixth domain that is less related to the others and that includes the transmembrane segment. The exact site of cleavage of rabbit poly-IgR to yield SC is unknown, but it is probably in the sixth (transmembrane) domain. Conversely, the structure of human SC is known but not that of human poly-IgR.

The extracellular portion of poly-IgR contains five homologous repeating units each composed of 100–115 amino acids, including a pair of cysteines that could form an intradomain disulfide bond. The five domains exhibit statistically significant sequence homology to each other and also to V regions of immunoglobulin light and heavy chains (Mostov *et al.*, 1984). These multiple immu-

noglobulin-like domains also exhibit sequence homology to the Thy-1 lymphocyte surface antigen. Hence, the poly-Ig receptor has been classified as a member of the immunoglobulin supergene family described in the next section. Mostov *et al.* (1984) propose that poly-IgR may be a prototype for other immunoglobulin receptors such as various leukocyte Fc receptors and the receptor for placental transport of IgG.

The amino acid sequence and the disulfide bond arrangement of human SC were determined by the methods of protein chemistry by Eiffert *et al.* (1984). Free SC is a single polypeptide chain containing 558 residues and seven GlcN oligosaccharides (M_r 86,000). As might be expected, human secretory component shows striking homology to rabbit poly-IgR (Fig. 32) and also exhibits internal homology indicative of the presence of five immunoglobulin-like domains. Just as in immunoglobulins, the most highly conserved segments in both rabbit poly-IgR and human SC are the sequences flanking the cysteine residues.

Unlike the case for rabbit poly-IgR, the disulfide bonding structure of human SC is known (Eiffert *et al.*, 1984). Each repeating domain of human SC has an intradomain disulfide bond linking approximately 65–70 residues; each corresponding domain of rabbit poly-IgR has a pair of cysteines that could form a homologous disulfide bond (Fig. 32). Human SC has four additional intradomain disulfide bonds, and three of these could be formed at homologous positions in rabbit poly-IgR. Secretory component forms two disulfide bonds with some classes of IgA. The sites in human IgA are predicted to be Cys-299 and/or Cys-301 in the C_H2 domain (Knight *et al.*, 1984). The fifth domain of SC is probably the site of covalent linkage to IgA (Cunningham-Rundles and Lamm, 1975; Eiffert *et al.*, 1984). This suggests that in its receptor form poly-IgR may be covalently linked to its ligand, IgA. The homologous domain structures of poly-IgR and IgA probably interact noncovalently to strengthen the binding of the ligand by the receptor.

The development of sequence data bases and of computer programs for retrieving and comparing sequences has greatly facilitated the search process for identifying unexpected structural relationships. One example is the recent discovery that α_1B-glycoprotein, a human plasma protein of unknown function, exhibits sequence similarity to poly-IgR and SC (Ishioka *et al.*, 1986). Like these receptor proteins α_1B also shows statistically significant homology to variable regions of immunoglobulins. Thus, α_1B is yet another member of the immunoglobulin supergene family though its biological role has still to be discovered.

3. Cell Surface Recognition Molecules

Like antibodies, the T-cell antigen receptors, the histocompatibility antigens, and certain lymphocyte surface antigens such as T8, Lyt-2, and Thy-1 also function as recognition molecules. They act through structures that are homolo-

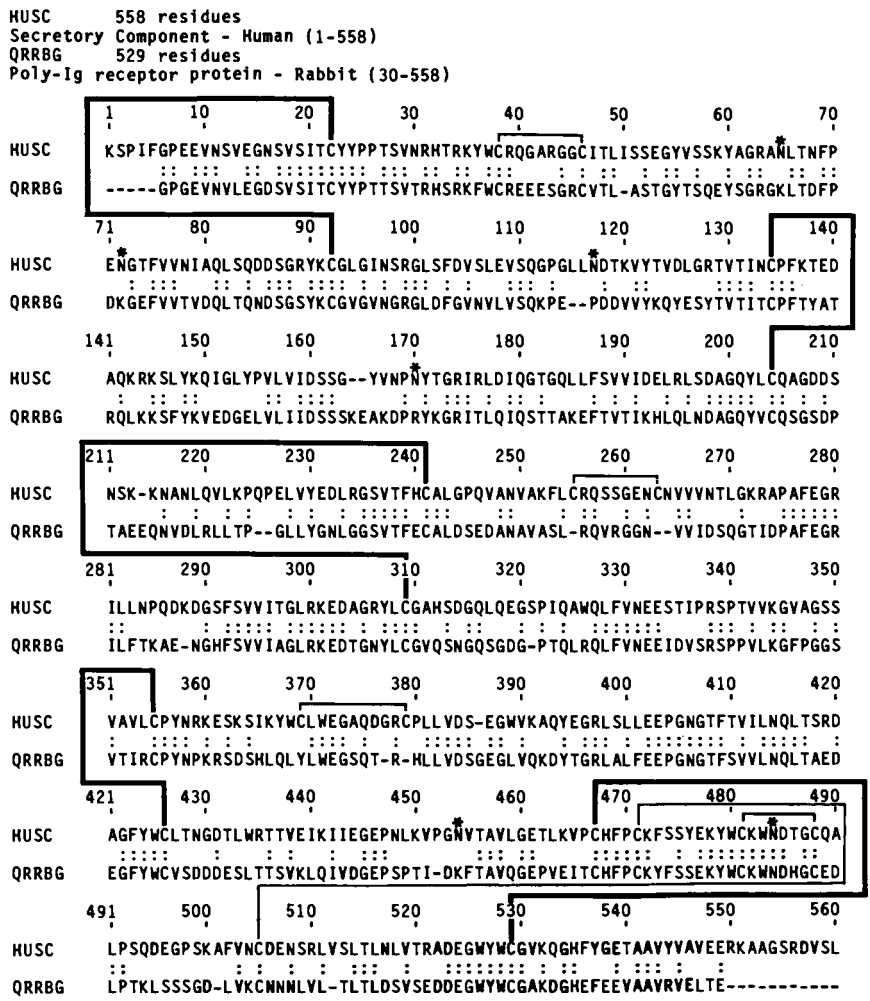


Fig. 32. Comparison of the amino acid sequence of human secretory component (HUSC) (Eiffert *et al.*, 1984) with the amino acid sequence deduced from the cDNA sequence of the rabbit poly-Ig receptor (QRRBG) (Mostov *et al.*, 1984). The human SC sequence (numbered residues) is compared with the corresponding segment of the extracellular domain portion of rabbit poly-IgR residues 30–558 (not numbered). Because the computer counts gaps, the numbering of the positions for SC is slightly off. The poly-IgR sequence shown corresponds approximately to the first five domains. The two sequences have been aligned by a computer program; identical residues are connected by dots. Intrachain disulfide bonds in human SC that are homologous to those in immunoglobulin domains are connected by heavy lines; other disulfide bonds are linked by light lines. Asterisks indicate the location of asparagine-linked glucosamine oligosaccharides in human SC. The carbohydrate structure of the GlcN oligosaccharides of HUSC has been determined (Mizoguchi *et al.*, 1982).

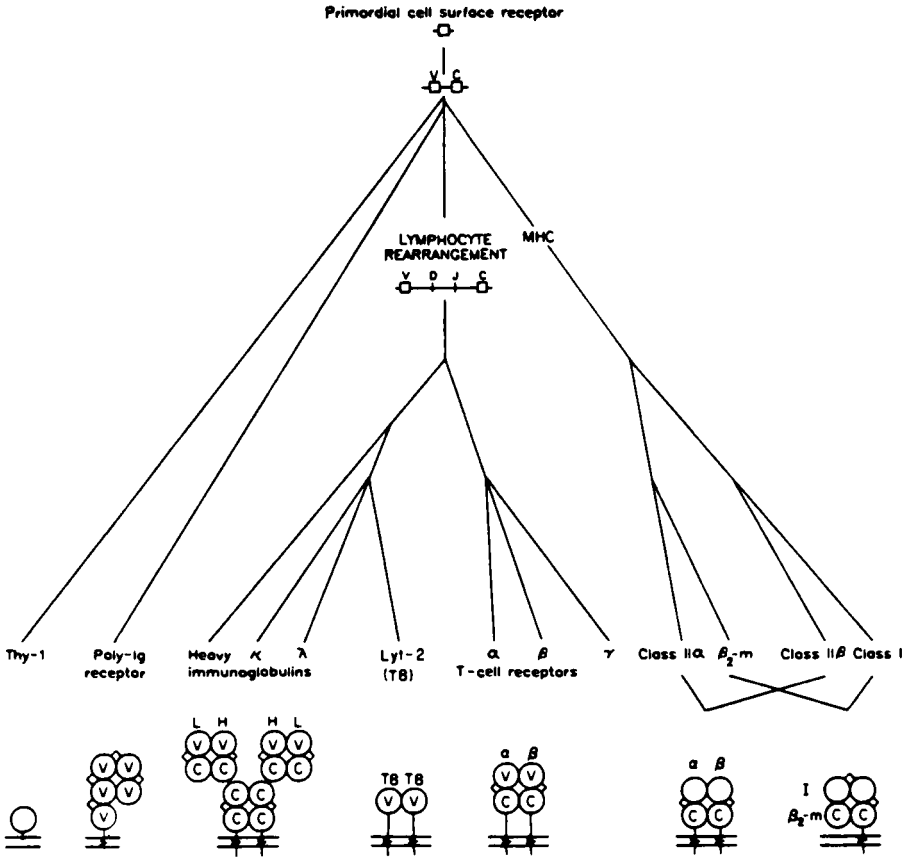


Fig. 33. Evolution of the genes of the immunoglobulin supergene family. This genealogical tree was constructed on the assumption that evolutionary relatedness correlates with the degree of sequence similarity among the members. Other features such as the exon/intron structure and the ability of DNA to rearrange were taken into account in this subjective assessment of the relative divergence times of the gene families (reprinted with permission from Hood *et al.*, 1985. Copyright © 1985 Massachusetts Institute of Technology).

gous to immunoglobulin domains (Hedrick *et al.*, 1984; Patten *et al.*, 1984; Hannum *et al.*, 1984; Chien *et al.*, 1984; Sim *et al.*, 1984; Tunnacliffe *et al.*, 1985). The homology units may resemble V domains or C domains, or intermediate structures. Since these proteins are surface antigens, each chain also will have a short C-terminal transmembrane domain and a cytoplasmic domain. The precursor chain has a signal sequence of about 20 residues. The immunoglobulin-like domain is extracellular and consists of about 100 residues and has an actual or a putative intrachain disulfide bridge around which the sequences are

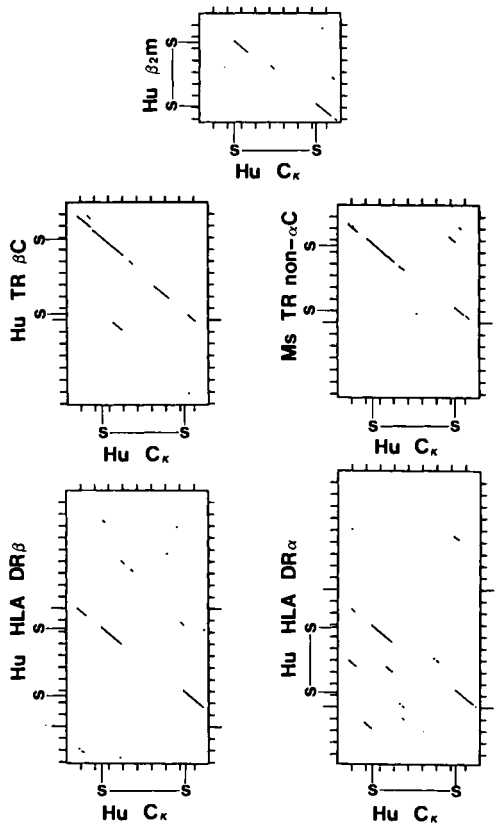


Fig. 34. Graphic matrix plot generated by the DOTMATRIX program of comparison of the amino acid sequence of the human Cκ region with the sequences of members of the immunoglobulin supergene family. The human (Hu) constant region sequence of the kappa light chain (Cκ) is compared with β_2-microglobulin (β_2m), with the C regions of the α and β chains of the human class II histocompatibility antigens (HLA DR), and also with the human T-cell receptor β chain C region domain (TR βC) and the mouse (MS) non-α chain C domain. The N-terminus of Cκ is at the left of the abscissa, the N-terminus of the other proteins is at the top, and the scale is marked in segments of ten residues. Disulfide bonds that are known are indicated. Cysteine residues that appear homologous to those in Cκ are indicated by S; however, additional cysteine residues are present in some of the other chains. The HLA and DR protein sequences were deduced from nucleotide sequences and may contain signal sequences and transmembrane and cytoplasmic domains because they are cell surface antigens. Thus, homology to Cκ (indicated by diagonal lines) is evident only for the extracellular domains of the HLA and TR structures, and is greatest in regions of sequence at homologous cysteine residues that represent actual or putative intrachain disulfide bonds. References for the sequences are: Cκ (Putnam *et al.*, 1966); β_2m (Suggs *et al.*, 1981); TR βC (Yanagi *et al.*, 1984); TR non-α C (Saito *et al.*, 1984); HLA DRβ (Long *et al.*, 1983); HLA DRα (Das *et al.*, 1983).

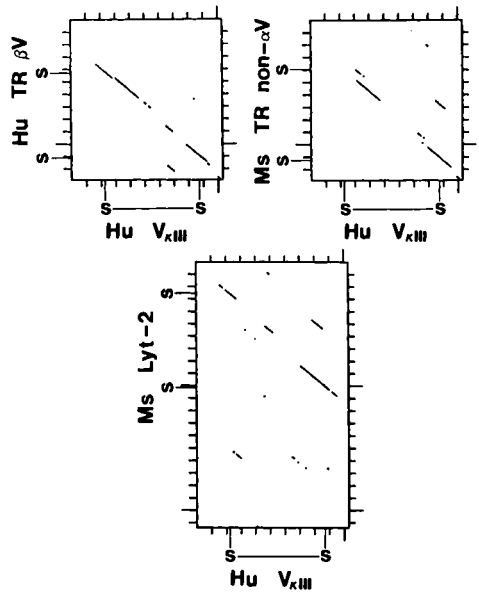


Fig. 35. Graphic matrix plot of the comparison of the V region sequence of a human κ light chain of subgroup II (Vκ III) with the sequence of the human T-cell receptor β chain V region (Hu TR βV) (Yanagi *et al.*, 1984), with the mouse T-cell receptor non-α V region (Ms TR non-α V) (Saito *et al.*, 1984), and with the Lyt-2 lymphocyte surface antigen of the mouse (Ms Lyt-2) (Nakauchi *et al.*, 1985). See Fig. 34 for the method of plotting and the significance of the diagonal lines.

most conserved in all members of the immunoglobulin supergene family. This is followed by a transmembrane domain and a cytoplasmic domain. As shown in Fig. 33 from Hood *et al.* (1985), the polypeptide chains may consist of one or more V- or C-like domains, and these domains interact laterally and longitudinally, as in immunoglobulins.

These newly recognized members of the immunoglobulin supergene family exhibit homology to both light and heavy chains. Figure 34 illustrates the structural relationship of the human Cκ domain with β_2-microglobulin, and with the C region domains of the polypeptide chains of the T-cell receptor and HLA class II histocompatibility antigens. Because these structures were all deduced by gene sequencing, they include the signal sequence and the transmembrane and cytoplasmic domains, which are unrelated to Cκ. However, the extracellular C domains of these cell surface molecules exhibit a striking sequence similarity to Cκ. In some cases the homology is as great as that of β_2m to Cκ.

The variable domains of immunoglobulins and T-cell antigen receptors are also similar in structure (Patten *et al.*, 1984; Arden *et al.*, 1985). Figure 35 shows a DOTMATRIX comparison of a human κ V region domain with the V

region domains of the T-cell receptor and a lymphocyte surface antigen (Lyt-2). By reference to a previous figure comparing the C region domains of human immunoglobulins (Fig. 23, Section IV,B,4), it is evident that some domains of the cell surface antigens and receptors are as homologous to immunoglobulin domains as the latter are to each other.

On the basis of the sequence similarity illustrated in Figs. 34 and 35 a genealogical tree has been constructed for members of the immunological supergene family by Hood *et al.* (1985). This tree (Fig. 33) suggests a possible pathway for evolution of this set of recognition molecules from an ancestral gene coding for a primordial cell surface receptor. Other factors supporting this pathway include the exon/intron arrangement of the genes and the similarity in the mechanism of rearrangement of their DNA (Hedrick *et al.*, 1984; Patten *et al.*, 1984). The genetic elements of these proteins encode functional and structural domains that have the property of self-association, which facilitates molecular assembly. Other recognition molecules that belong to this family will probably soon be discovered, and they likely will include receptors for Fc.

Acknowledgments

The author thanks Nobuhiro Takahashi, Yoko Takahashi, and Noriaki Ishioka for helpful comments and discussion and for help in preparing illustrations. He is indebted to Ann Scales for patient and invaluable support in preparation of the manuscript. The work presented in this review that was carried out in the author's laboratory was supported by NIH Grant CA08497 and American Cancer Society Grant IM-21.

References

Abola, E. E., Ely, K. R., and Edmundson, A. B. (1980). *Biochemistry* **19,** 432–439.

Alexander, A., Steinmetz, M., Barritault, D., Frangione, B., Franklin, E. C., Hood, L., and Buxbaum, J. N. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 3260–3264.

Alt, F. W., and Baltimore, D. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 4118–4122.

Alt, F. W., Bothwell, A. L. M., Knapp, M., Siden, E., Mather, E., and Koshland, M. (1980). *Cell (Cambridge, Mass.)* **20,** 293–301.

Amit, A. G., Mariuzza, R. A., Saludjian, P., Poljak, R. J., Lamoyi, E., and Nisonoff, A. (1983). *J. Mol. Biol.* **169,** 637–638.

Amit, A. G., Mariuzza, R. A., Phillips, S. E. V., and Poljak, R. J. (1985a). *Nature (London)* **313,** 156–158.

Amit, A. G., Boulot, G., Comarmond, M. B., Harper, M., Mariuzza, R. A., Phillips, S. E. V., Saludjian, P., Saul, F. A., Souchon, H., Tougard, P., Poljak, R. J., Conger, J. D., and Nisonoff, A. (1985b). *Ann. Immunol. (Paris)* **136C,** 121–129.

Amzel, L. M., and Poljak, R. J. (1979). *Annu. Rev. Biochem.* **48,** 961–997.

Anderson, N. L., Tracy, R. P., and Anderson, N. G. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 4, pp. 221–270. Academic Press, Orlando, Florida.

Arden, B., Klotz, J. L., Siu, G., and Hood, L. E. (1985). *Nature (London)* **316,** 783–787.

Baenziger, J. U. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 4, pp. 271–315. Academic Press, Orlando, Florida.

Baenziger, J. U., and Kornfeld, S. (1974a). *J. Biol. Chem.* **249,** 7260–7269.

Baenziger, J. U., and Kornfeld, S. (1974b). *J. Biol. Chem.* **249,** 7270–7281.

Baenziger, J. U., and Kornfeld, S. (1974c). *J. Biol. Chem.* **249,** 1889–1896.

Baenziger, J. U., and Kornfeld, S. (1974d). *J. Biol. Chem.* **249,** 1897–1903.

Baer, R., Chen, K.-C., Smith, S. D., and Rabbitts, T. H. (1985). *Cell (Cambridge, Mass.)* **43,** 705–713.

Barker, W. C., and Putnam, F. W. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 4, pp. 361–399. Academic Press, New York.

Barker, W. C., Ketcham, L. K., and Dayhoff, M. O. (1978). *In* "Atlas of Protein Sequence and Structure" (M. O. Dayhoff, ed.), Vol. 5, Suppl. 3, Chapter 12, pp. 197–227. Natl. Biomed. Res. Found., Washington, D.C.

Barker, W. C., Hunt, L. T., Orcutt, B. C., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Johnson, G. C., and Dayhoff, M. O. (1983). "Protein Sequence Database of the Protein Identification Resource." Natl. Biomed. Res. Found., Washington, D.C.

Barker, W. C., Hunt, L. T., Orcutt, B. C., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Johnson, G. C., and Dayhoff, M. O. (1985). "Protein Sequence Database of the Protein Identification Resource." Natl. Biomed. Res. Found., Washington, D.C.

Battey, J., Max, E. E., McBride, W. O., Swan, D., and Leder, P. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 5956–5960.

Beale, D. (1984). *Comp. Biochem. Physiol. B* **77B,** 399–412.

Beale, D., and Feinstein, A. (1976). *Q. Rev. Biophys.* **9,** 135–180.

Becker, J. W., and Reeke, G. N., Jr. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 4225–4229.

Bergman, Y., Rice, D., Grosschedl, R., and Baltimore, D. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 7041–7045.

Bernard, O., Hozumi, N., and Tonegawa, S. (1978). *Cell (Cambridge, Mass.)* **15,** 1133–1144.

Blake, C. C. F., Koening, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1965). *Nature (London)* **206,** 757–761.

Blattner, F. R., and Tucker, P. W. (1984). *Nature (London)* **307,** 417–422.

Blomberg, B., Traunecker, A., Eisen, H., and Tonegawa, S. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 3765–3769.

Boackle, R. J., Johnson, B. J., and Caughman, G. B. (1979). *Nature (London)* **282,** 742–743.

Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K., and Baltimore, D. (1981). *Cell (Cambridge, Mass.)* **24,** 625–637.

Brack, C., Hirama, M., Lenhard-Schuller, R., and Tonegawa, S. (1978). *Cell (Cambridge, Mass.)* **15,** 1–14.

Bragado, R., Lopez de Castro, J. A., Jaurez, C., and Ortiz, F. (1982). *Immunol. Lett.* **5,** 239–245.

Brandtzaeg, P. (1981). *Clin. Exp. Immunol.* **44,** 221–232.

Brunhouse, R., and Cebra, J. J. (1979). *Mol. Immunol.* **16,** 907–917.

Burton, D. R. (1985). *Mol. Immunol.* **22,** 161–206.

Burton, D. R., Boyd, J., Brampton, A. D., Easterbrook-Smith, S. B., Emanuel, E. J., Novotny, J., Rademacher, T. W., van Schravendijk, M. R., Sternberg, M. J. E., and Dwek, R. A. (1980). *Nature (London)* **288,** 338–344.

Cahour, A., Debeire, P., Hartmann, L., and Montreuil, J. (1983). *Biochem. J.* **211,** 55–63.

Cahour, A., Debeire, P., Hartmann, L., Montreuil, J., Van Halbeek, H., and Vliegenthart, J. F. G. (1984a). *FEBS Lett.* **170,** 343–349.

Cahour, A., Debeire, P., Parente, J. P., Hartmann, L., and Montreuil, J. (1984b). *Biochim. Biophys. Acta* **802,** 188–196.

Campbell, R. D., Dodds, A. W., and Porter, R. R. (1980). *Biochem. J.* **189,** 67–80.
Cann, G. M., Zaritisky, A., and Koshland, M. E. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 6656–6660.
Capra, J. D., and Edmundson, A. B. (1977). *Sci. Am.* **236,** 50–59.
Capra, J. D., and Kehoe, J. M. (1975). *Adv. Immunol.* **20,** 1–40.
Chang, C.-H., Short, M. T., Westholm, F. A., Stevens, F. J., Wang, B.-C., Furey, W., Jr., Solomon, A., and Schiffer, M. (1985). *Biochemistry* **24,** 4890–4897.
Chapman, A., and Kornfeld, R. (1979a). *J. Biol. Chem.* **254,** 816–823.
Chapman, A., and Kornfeld, R. (1979b). *J. Biol. Chem.* **254,** 824–828.
Cheng, H.-L., Blattner, F. R., Fitzmaurice, L., Mushinski, J. F., and Tucker, P. W. (1982). *Nature (London)* **296,** 410–415.
Chien, Y.-H., Gascoigne, N. R. J., Kavaler, J., Lee, N. E., and Davis, M. M. (1984). *Nature (London)* **309,** 322–336.
Ciccimarra, F., Rosen, F. S., and Merler, E. (1975). *Proc. Natl. Acad. Sci. U.S.A.* **72,** 2081–2083.
Cipens, G., Ancans, J., Zarins, P., Osis, L., Afanas'eva, G. A., and Zalitis, G. (1984). *Latv. PSR Zinat. Akad. Vestis, Kim. Ser.* **6,** 752–753.
Colomb, M., and Porter, R. R. (1975). *Biochem. J.* **145,** 177–183.
Cory, S., and Adams, J. M. (1980). *Cell (Cambridge, Mass.)* **19,** 37–51.
Cory, S., Adams, J. M., and Kemp, D. J. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 4943–4947.
Crews, S., Griffin, J., Huang, H., Calame, K., and Hood, L. (1981). *Cell (Cambridge, Mass.)* **25,** 59–66.
Cunningham-Rundles, C., and Lamm, M. E. (1975). *J. Biol. Chem.* **250,** 1987–1991.
Cushley, W., and Williamson, A. R. (1982). *Essays Biochem.* **18,** 1–39.
Das, H. K., Lawrance, S. K., and Weissman, S. M. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 3543–3547.
Davies, D. R., and Metzger, H. (1983). *Annu. Rev. Immunol.* **1,** 87–117.
Davies, D. R., Padlan, E. A., and Segal, D. M. (1975). *Annu. Rev. Biochem.* **44,** 639–667.
Davis, M. M., Calame, K., Early, P. W., Livant, D. L., Joho, R., Weissman, I. L., and Hood, L. (1980). *Nature (London)* **283,** 733–739.
Dayhoff, M. O., ed. (1978). "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3, pp. 73–93. Natl. Biomed. Res. Found., Washington, D.C.
Deisenhofer, J. (1981). *Biochemistry* **20,** 2361–2370.
Deisenhofer, J., Colman, P. M., Epp, U., and Huber, R. (1976). *Hoppe-Seyler's Z. Physiol. Chem.* **347,** 1421–1434.
Diamond, B., and Scharff, M. D. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 5841–5844.
Diamond, B., Campbell, R., Schneck, J., and Birshtein, B. K. (1981). *Protides Biol. Fluids* **29,** 415–418.
Dickler, H. B. (1978). *Adv. Immunol.* **24,** 167–214.
Dildrop, R., and Beyreuther, K. (1981). *Nature (London)* **292,** 61–63.
Dorrington, K. J. (1979). *Mol. Immunol.* **16,** 841–978.
Dorrington, K. J., and Bennich, H. H. (1978). *Immunol. Rev.* **41,** 3–25.
Dorrington, K. J., and Klein, M. H. (1982). *Mol. Immunol.* **19,** 1215–1221.
Dorrington, K. J., and Klein, M. H. (1984). *Prog. Immunol., Int. Congr, Immunol., 5th, 1983,* pp. 37–46.
Dorrington, K. J., and Painter, R. H. (1974). *Prog. Immunol., Int. Congr. Immunol., 2nd, 1974,* Vol. 1, p. 75–84.
Dreyer, W. J., and Bennett, J. C. (1965). *Proc. Natl. Acad. Sci. U.S.A.* **54,** 864–869.
Dreyer, W. J., Gray, W. R., and Hood, L. (1967). *Cold Spring Harbor Symp. Quant. Biol.* **32,** 353–367.
Dwek, R. A., Sutton, B. J., Perkins, S. J., and Rademacher, T. W. (1984). *Biochem. Soc. Symp.* **49,** 123–126.

Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N., and Hood, L. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 857–861.

Early, P. W., Rogers, J., Davis, M. M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). *Cell (Cambridge, Mass.)* **20,** 313–319.

Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M., Westholm, F. A., Fausch, M. D., and Deutsch, H. F. (1974). *Biochemistry* **13,** 3816–3827.

Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer, M., and Panagiototopoulos, N. (1975). *Biochemistry* **14,** 3953–3961.

Eiffert, H., Quentin, E., Decker, J., Hillemeir, S., Hufschmidt, M., Klingmuller, D., Weber, M. H., and Hilschmann, N. (1984). *Hoppe-Seyler's Z. Physiol. Chem.* **365,** 1489–1495.

Ellerson, J. R., Yasmeen, D., Painter, R. H., and Dorrington, K. J. (1976). *J. Immunol.* **116,** 510–517.

Ely, K. R., Colman, P. M., Abola, E. E., Hess, A. C., Peabody, D. S., Parr, D. M., Connell, G. E., Laschinger, C. A., and Edmundson, A. B. (1978). *Biochemistry* **17,** 820–823.

Ely, K. R., Wood, M. K., Rajan, S. S., Hodsdon, J. M., Abola, E. E., Deutsch, H. F., and Edmundson, A. B. (1985). *Mol. Immunol.* **22,** 93–100.

Endo, S., and Arata, Y. (1985). *Biochemistry* **24,** 1561–1568.

Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M., Huber, R., and Palm, W. (1974). *Eur. J. Biochem.* **45,** 513–524.

Epp, O., Lattman, E. E., Schiffer, M., Huber, R., and Palm, W. (1975). *Biochemistry* **14,** 4943–4952.

Falkner, F. G., and Zachau, H. G. (1984). *Nature (London)* **310,** 71–74.

Fehlhammer, H., Schiffer, M., Epp, O., Colman, P. M., and Lattman, E. E. (1975). *Biophys. Struct. Mech.* **1,** 139–146.

Feldmann, R. J., Potter, M., and Glaudemans, C. P. J. (1981). *Mol. Immunol.* **18,** 683–698.

Flanagan, J. G., and Rabbitts, R. H. (1982). *Nature (London)* **300,** 709–713.

Florent, G., Lehman, D., Lockhart, D., and Putnam, F. W. (1974). *Biochemistry* **13,** 3372–3381.

Frangione, B., and Franklin, E. C. (1979). *Nature (London)* **281,** 600–603.

Franklin, E. C., and Frangione, B. (1975). *Contemp. Top. Mol. Immunol.* **4,** 89–126.

Fudenberg, H. H., Pink, J. R. L., Wang, A.-C., and Ferrara, G. B. (1984). "Basic Immunogenetics," 3rd ed. Oxford Univ. Press, London and New York.

Furey, W., Jr., Wang, B. C., Yoo, C. S., and Sax, M. (1983). *J. Mol. Biol.* **167,** 661–692.

Garver, F. A., Chang, L. S., Kiefer, C. R., Medicino, J., Chandrasekaran, E. V., Isobe, T., and Osserman, E. F. (1981). *Eur. J. Biochem.* **115,** 643–652.

Gearhart, P. J., Johnson, N. D., Douglas, R., and Hood, L. (1981). *Nature (London)* **291,** 29–34.

Geha, R. S., Helm, B., and Gould, H. (1985). *Nature (London)* **315,** 577–578.

Gilliam, A. C., Shen, A., Richards, J. E., Blattner, F. R., Mushinski, J. F., and Tucker, P. W. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 4164–4168.

Gillies, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. (1983). *Cell (Cambridge, Mass.)* **33,** 717–728.

Goding, J. W. (1980). *J. Immunol.* **124,** 2082–2088.

Goto, Y., and Aki, K. (1984). *Biochemistry* **23,** 6736–6744.

Gough, N. M., Kemp, D. J., Tyler, B. M., Adams, J. M., and Cory, S. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 554–558.

Greene, M. I., and Nisonoff, A., eds. (1984). "The Biology of Idiotypes." Plenum, New York.

Hahn, G. S. (1983). *Clin. Physiol. Biochem.* **1,** 117–144.

Hannum, C. H., Kappler, J. W., Trowbridge, I. S., Marrack, P., and Freed, J. H. (1984). *Nature (London)* **312,** 65–67.

Haynes, B. F., and Eisenbarth, G. S., eds. (1983). "Monoclonal Antibodies." Academic Press, New York.

Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I., and Davis, M. M. (1984). *Nature (London)* **308,** 153–158.

Heymann, B., Nose, M., and Weigle, W. O. (1985). *J. Immunol.* **134,** 4018–4023.

Hieter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldmann, T. A., and Leder, P. (1981). *Nature (London)* **294,** 536–540.

Hilschmann, N., and Craig, L. C. (1965). *Proc. Natl. Acad. Sci. U.S.A.* **53,** 1403–1409.

Hobbs, M. V., Morgan, E. L., and Weigle, W. O. (1985). *J. Immunol.* **134,** 2847–2852.

Hofstaetter, T., Guthoehrlein, G., Kanzy, E. J., Zilg, H., and Seiler, F. R. (1984). *Behring Inst. Mitt.* **76,** 75–87.

Honjo, T. (1983). *Annu. Rev. Immunol.* **1,** 499–528.

Honjo, T., and Habu, S. (1985). *Annu. Rev. Biochem.* **54,** 803–830.

Hood, L. E., Steinmetz, M., and Malissen, B. (1983). *Annu. Rev. Immunol.* **1,** 529–568.

Hood, L. E., Weissman, I. L., Wood, W. B., and Wilson, J. H. (1984). "Immunology," 2nd ed. Benjamin/Cummings, Menlo Park, California.

Hood, L. E., Kronenberg, M., and Hunkapiller, T. (1985). *Cell (Cambridge, Mass.)* **40,** 225–229.

Huber, R. (1984). *Behring Inst. Mitt.* **76,** 1–14.

Huppi, K., Julius, M., Staudt, L., Gerhard, W., and Weigert, M. (1984). *Ann. Immunol. (Paris)* **135C,** 181–185.

Ishida, N., Ueda, S., Hayashida, H., Miyata, T., and Honjo, T. (1982). *EMBO J.* **1,** 1117–1123.

Ishioka, N., Takahashi, N., and Putnam, F. W. (1986). *Proc. Natl. Acad. Sci. U.S.A.* **83,** 2363–2367.

Ishioka, N., Takahashi, N., and Putnam, F. W. (1987). *Proc. Natl. Acad. Sci. U.S.A.* **84,** (in press).

Ishizaka, K. (1985). *J. Immunol.* **135,** i–x.

Ishizaka, T., Hirata, F., Sterk, A. R., Ishizaka, K., and Axelrod, J. A. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6812–6816.

Johnson, B. J., and Thames, K. E. (1977). *J. Pharm. Sci.* **66,** 427–428.

Kabat, E. A. (1982). *Pharm. Rev.* **34,** 23–38.

Kabat, E. A., Wu, T. T., and Bilofsky, H. (1979). "Sequences of Immunoglobulin Chains," Natl. Inst. Health, Bethesda, Maryland.

Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M., and Perry, H. (1983). "Sequences of Proteins of Immunological Interest." Natl. Inst. Health, Bethesda, Maryland.

Kataoka, T., Miyata, T., and Honjo, T. (1981). *Cell (Cambridge, Mass.)* **23,** 357–368.

Kehoe, J. M., Bourgois, A., Capra, J. D., and Fougereau, M. (1974). *Biochemistry* **13,** 2499–2504.

Kehry, M., Ewald, S., Douglas, R., Sibley, C., Raschke, W., Fambrough, D., and Hood, L. (1980). *Cell (Cambridge, Mass.)* **21,** 393–406.

Kenter, A. L., and Birshtein, B. K. (1979). *Science* **206,** 1307–1309.

Kindt, T. J., and Capra, J. D. (1984). "The Antibody Enigma." Plenum, New York.

Klein, G., and Klein, E. (1985). *Nature (London)* **315,** 190–195.

Klein, M., Haeffner-Cavaillon, N., Isenman, D. E., Rivat, C., Navia, M. A., Davies, D. R., and Dorrington, K. J. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 524–528.

Knight, K. L., Martens, C. L., Stoklosa, C. M., and Schneiderman, R. D. (1984). *Nucleic Acids Res.* **12,** 1657–1670.

Knight, K. L., Burnett, R. C., and McNicholas, J. M. (1985). *J. Immunol.* **134,** 1245–1250.

Kohler, H., Shimizu, A., Paul, C., Moore, V., and Putnam, F. W. (1970). *Nature (London)* **227,** 1318–1320.

Kornfeld, R., and Kornfeld, S. (1985). *Annu. Rev. Biochem.* **54,** 631–664.

Kornfeld, R., Keller, J., Baenziger, J. U., and Kornfeld, S. (1971). *J. Biol. Chem.* **246,** 3259–3268.

Korsmeyer, S. J., and Waldmann, T. A. (1984). *In* "Basic and Clinical Immunology" (D. P. Stites, J. D. Stobo, H. H. Fudenberg, and J. V. Wells, eds.), 5th ed., pp. 43–54. Lange Med. Publ., Los Altos, California.

Koshland, M. E. (1985). *Annu. Rev. Immunol.* **3,** 425–453.
Kratzin, H., Altevogt, P., Ruban, E., Kortt, A., Staroscik, K., and Hilschmann, N. (1975). *Hoppe-Seyler's Z. Physiol. Chem.* **356,** 1337–1342.
Krawinkel, U., and Rabbitts, T. H. (1982). *EMBO J.* **1,** 403–407.
Kuhn, L. C., and Kraehenbuhl, J.-P. (1983). *Ann. N.Y. Acad. Sci.* **409,** 751–759.
Kuhn, L. C., Kocher, H. P., Hanly, W. C., Cook, L., Jaton, J. C., and Kraehenbuhl, J.-P. (1983). *J. Biol. Chem.* **258,** 6653–6659.
Kyle, R. A., and Bayrd, E. D. (1976). "The Monoclonal Gammopathies." Thomas, Springfield, Illinois.
Laurell, C.-B., ed. (1977). "Electrophoretic and Electroimmunochemical Analysis of Proteins." University Park Press, Baltimore, Maryland.
Leder, P. (1982). *Sci. Am.* **246,** 102–103.
Leder, P., Max, E. E., Seidman, J. G., Kwan, S. P., Scharff, M., Nau, M., and Norman, B. (1981). *Cold Spring Harbor Symp. Quant. Biol.* **45,** 859–865.
Lefranc, M. P., and Rabbitts, T. H. (1984). *Nucleic Acids Res.* **12,** 1303–1311.
Leslie, G. A., and Martin, L. N. (1978). *Contemp. Top. Mol. Immunol.* **7,** 1–49.
Lewis, S., Gifford, A., and Baltimore, D. (1985). *Science* **228,** 677–785.
Lin, L.-C., and Putnam, F. W. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 2649–2653.
Lin, L.-C., and Putnam, F. W. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 6572–6576.
Lin, L.-C., and Putnam, F. W. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 504–508.
Liu, F.-T., Albrandt, K., Mendel, E., Kulczycki, A., Jr., and Orinda, N. K. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 4100–4101.
Liu, Y.-S., V., Low, T. L. K., Infante, A., and Putnam, F. W. (1976). *Science* **193,** 1017–1020.
Long, E.-O., Gorski, J., Rollini, P., Make, C. T., Strubin, M., Rabourdin-Combe, C., and Mach, B. (1983). *Hum. Immunol.* **8,** 113–121.
Lukas, T. J., Munoz, H., and Erickson, B. W. (1981). *J. Immunol.* **127,** 2555–2560.
McGhee, J. R., and Mestecky, J., eds. (1983). "The Secretory Immune System," *Ann. N.Y. Acad. Sci., Vol. 409.* N.Y. Acad. Sci., New York.
McMichael, A. J., and Fabre, J. W., eds. (1982). "Monoclonal Antibodies in Clinical Medicine." Academic Press, London.
Maki, R., Traunecker, A., Sakano, H., Roeder, W., and Tonegawa, S. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 2138–2142.
Mariuzza, R. A., Amit, A. G., Boulot, G., Saludjian, P., Saul, F. A., Tougard, P., Poljak, R. J., Conger, J., Lamoyi, E., and Nisonoff, A. (1984). *J. Biol. Chem.* **259,** 5954–5958.
Marquart, M., and Deisenhofer, J. (1982). *Immunol. Today* **3,** 160–166.
Marquart, M., Deisenhofer, J., Huber, R., and Palm, W. (1980). *J. Mol. Biol.* **141,** 369–391.
Martens, C. L., Currier, S. J., and Knight, K. L. (1984). *J. Immunol.* **133,** 1022–1027.
Martens, C. L., Huff, T. F., Jardieu, P., Trounstine, M. L., Coffman, R. L., Ishizaka, K., and Moore, K. W. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 2460–2464.
Martinez, J., Laur, J., and Winternitz, F. (1983). *Pept., Proc. Eur. Pept. Symp., 17th, 1982,* pp. 265–268.
Mellis, S. J., and Baenziger, J. U. (1983a). *J. Biol. Chem.* **258,** 11546–11556.
Mellis, S. J., and Baenziger, J. U. (1983b). *J. Biol. Chem.* **258,** 11557–11563.
Metzger, H. (1978). *Contemp. Top. Mol. Immunol.* **7,** 119–152.
Metzger, H. (1983). *Contemp. Top. Mol. Immunol.* **9,** 115–145.
Migone, N., Oliviero, S., De Lange, G., Delacroix, D. L., Boschis, D., Altruda, F., Silengo, L., DeMarchi, M., and Carbonara, A. O. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 5811–5815.
Milgrom, F., Abeyounis, C. J., and Albini, B., eds. (1985). "Antibodies: Protective, Destructive, and Regulatory Role," 9th Int. Convoc. Immunol. Karger, Basel.
Miller, J., Bothwell, A., and Storb, U. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 3829–3833.

Milstein, C. (1980). *Sci. Am.* **243,** 66–74.
Mizoguchi, A., Mizouchi, T., and Kobata, A. (1982). *J. Biol. Chem.* **257,** 9612–9617.
Mizouchi, T., Taniguchi, T., Shimizu, A., and Kobata, A. (1982). *J. Immunol.* **129,** 2016–2020.
Mole, J. E., Bhown, A. S., and Bennett, J. C. (1977). *Biochemistry* **16,** 3507–3513.
Morell, A., and Riesen, W. (1982). *In* "Immunochemotherapy: A Guide to Immunoglobulin" (U. E. Nydegger, ed.), pp. 17–26. Academic Press, London.
Morgan, E. L., Hugli, T. E., and Weigle, W. O. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 5388–5391.
Morgan, E. L., Shields, J. E., Campbell, C. S., Barton, R. L., Koppel, G. A., and Weigle, W. O. (1983). *J. Exp. Med.* **157,** 947–956.
Mostov, K. E., and Blobel, G. (1983). *Ann. N.Y. Acad. Sci.* **409,** 441–451.
Mostov, K. E., Friedlander, M., and Blobel, G. (1984). *Nature (London)* **308,** 37–43.
Najjar, V. A., and Fridkin, M., eds. (1983). "Antineoplastic, Immunogenic and Other Effects of the Tetrapeptide Tuftsin," *Ann. N.Y. Acad. Sci., Vol. 419.* N.Y. Acad. Sci., New York.
Nakauchi, H., Nolan, G. P., Hsu, C., Haung, H. S., Kavathas, P., and Herzenberg, L. A. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 5126–5130.
Natelson, S., Pesce, A., and Dietz, A., eds.) (1978). "Clinical Immunochemistry." Am. Assoc. Clin. Chem., Washington, D.C.
Navia, M. A., Segal, D. M., Padlan, E. A., Davies, D. R., Rao, N., Rudikoff, S., and Potter, M. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 4071–4074.
Newell, N., Richards, J. E., Tucker, P. W., and Blattner, F. R. (1980). *Science* **209,** 1128–1132.
Nezlin, R. S. (1977). "Struktur und Biosynthese der Antikorper." Fischer, Jena.
Nishi, M., Kataoka, T., and Honjo, T. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 6399–6403.
Nishida, Y., Miki, T., Hisajima, H., and Honjo, T. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 3833–3837.
Nisonoff, A., and Gurish, M. F. (1984). *In* "New Horizons in Microbiology" (A. Sanna and G. Morace, eds.), pp. 119–128. Elsevier, Amsterdam.
Nisonoff, A., Hopper, J. E., and Spring, S. B. (1975). "The Antibody Molecule." Academic Press, New York.
Nose, M., and Wigzell, H. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 6632–6636.
Novotny, J., and Haber, E. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 4592–4596.
Novotny, J., Bruccoleri, R., Newell, J., Murphy, D., Haber, E., and Karplus, M. (1983). *J. Biol. Chem.* **258,** 14433–14437.
Ohkura, T., Isobe, T., Yamashita, K., and Kobata, A. (1985). *Biochemistry* **24,** 503–508.
Oi, V. T., Bryan, V. M., and Herzenberg, L. A. (1980). *J. Exp. Med.* **151,** 1260–1274.
Padlan, E. A. (1977a). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 2551–2555.
Padlan, E. A. (1977b). *Q. Rev. Biophys.* **10,** 35–65.
Parekh, R. B., Dwek, R. A., Sutton, B. J., Fernandes, D. L., Leung, A., Stanworth, D., and Rademacher, T. W. (1985). *Nature (London)* **316,** 452–457.
Patten, P., Yokota, T., Rothbard, J., Chien, Y.-H., Arai, K.-I., and Davis, M. M. (1984). *Nature (London)* **312,** 40–46.
Pecht, I. (1982). *Antigens* **6,** 1–68.
Phizackerly, R. P., Wishner, B. C., Bryant, S. H., Amzel, L. M., Lopez De Castro, J. A., and Poljak, R. J. (1979). *Mol. Immunol.* **16,** 841–850.
Pierce-Cretel, A., Debray, H., Montreuil, J., Spik, G., Van Halbeek, H., Mutsaers, H. G. M., and Vliegenthart, J. F. G. (1984). *Eur. J. Biochem.* **139,** 337–349.
Plaut, A. G. (1978). *N. Engl. J. Med.* **298,** 1459–1463.
Plaut, A. G., and Tomasi, T. B., Jr. (1970). *Proc. Natl. Acad. Sci. U.S.A.* **65,** 318–322.
Plaut, A. G., Gilbert, J. V., Artenstein, M. S., and Capra, J. D. (1975). *Science* **190,** 1103–1105.
Poljak, R. J. (1978). *CRC Crit. Rev. Biochem.* **5,** 45–84.

Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P., and Saul, F. (1973). *Proc. Natl. Acad. Sci. U.S.A.* **70,** 3305–3310.

Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerly, R. P., and Saul, F. (1974). *Proc. Natl. Acad. Sci. U.S.A.* **71,** 3440–3444.

Pollock, R. R., and Mescher, M. F. (1980). *J. Immunol.* **124,** 1668–1674.

Porter, R. R. (1959). *Biochem. J.* **73,** 119–127.

Potter, H., Weir, L., and Leder, P. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 7161–7165.

Potter, M. (1977). *Adv. Immunol.* **25,** 141–211.

Poulik, M. D. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 433–454. Academic Press, New York.

Prystowsky, M. B., Kehoe, J. M., and Erickson, B. W. (1981). *Biochemistry* **20,** 6349–6356.

Putnam, F. W. (1967). *Nobel Symp. 3,* 45–70.

Putnam, F. W. (1977a). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 3, pp. 1–153. Academic Press, New York.

Putnam, F. W. (1977b). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 3, pp. 155–221. Academic Press, New York.

Putnam, F. W. (1977c). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 3, pp. 223–284. Academic Press, New York.

Putnam, F. W. (1983). *Clin. Physiol. Biochem.* **1,** 63–91.

Putnam, F. W. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.) 2nd ed., Vol. 4, pp. 1–44. Academic Press, Orlando, Florida.

Putnam, F. W., Titani, K., and Whitley, E., Jr. (1966). *Proc. R. Soc. London, Ser. B* **166,** 124–137.

Putnam, F. W., Titani, K., Wikler, M., and Shinoda, T. (1967a). *Cold Spring Harbor Symp. Quant. Biol.* **32,** 9–28.

Putnam, F. W., Shinoda, T., Titani, K., and Wikler, M. (1967b). *Science* **157,** 1050–1053.

Putnam, F. W., Liu, Y.-S. V., and Low, T. L. K. (1979). *J. Biol. Chem.* **254,** 2865–2874.

Putnam, F. W., Takahashi, N., Tetaert, D., Debuire, B., and Lin, L.-C. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6168–6172.

Putnam, F. W., Takahashi, N., Tetaert, D., Lin, L.-C., and Debuire, B. (1982). *Ann. N.Y. Acad. Sci.* **399,** 41–68.

Putnam, F. W., Takahashi, N., and Ishioka, N. (1985). *In* "Antibodies: Protective, Destructive, and Regulatory Role" (F. Milgrom, C. J. Abeyounis, and B. Albini, eds.), 9th Int. Convoc. Immunol., pp. 26–36. Karger, Basel.

Rabbitts, T. H., Bentley, D. L., and Milstein, C. P. (1981). *Immunol. Rev.* **59,** 69–91.

Rajan, S. S., Ely, K. R., Abola, E. E., Wood, M. K., Colman, P. M., Athay, R. J., and Edmundson, A. B. (1983). *Mol. Immunol.* **20,** 787–799.

Rearick, J. I., Kulczycki, A., Jr., and Kornfeld, S. (1983). *Arch. Biochem. Biophys.* **220,** 95–105.

Robertson, M. (1985). *Nature (London)* **317,** 768–771.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R. (1980). *Cell (Cambridge, Mass.)* **20,** 303–312.

Rogers, J., Choi, E., Souza, L., Carter, C., Word, C., Kuehl, M., Eisenberg, D., and Wall, R. (1981). *Cell (Cambridge, Mass.)* **26,** 19–27.

Rose, D. R., Seaton, B. A., Petsko, G. A., Novotny, J., Margolies, M. N., Locke, E., and Haber, E. (1983). *J. Mol. Biol.* **165,** 203–206.

Rothschild, M. A., and Waldmann, T., eds. (1970). "Plasma Protein Metabolism: Regulation of Synthesis, Distribution and Degradation." Academic Press, New York.

Rudikoff, S. (1983). *Contemp. Top. Mol. Immunol.* **9,** 169–209.

Rudikoff, S., Giusti, A. M., Cook, W. D., and Scharff, M. D. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 1979–1983.

Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N., and Tonegawa, S. (1984). *Nature (London)* **309,** 757–762.

Sakano, H., Rogers, J. H., Jueppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R., and Tonegawa, S. (1979). *Nature (London)* **277,** 627–633.

Sarma, R., and Laudin, A. G. (1982). *J. Appl. Crystallogr.* **15,** 476–481.

Saul, F., Amzel, L. M., and Poljak, R. J. (1978). *J. Biol. Chem.* **253,** 585–597.

Savvidou, G., Klein, M., Grey, A. A., Dorrington, K. J., and Carver, J. P. (1984). *Biochemistry* **23,** 3736–3740.

Schilling, J., Clevinger, B., Davie, J. M., and Hood, L. (1980). *Nature (London)* **283,** 35–40.

Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M., and Davies, D. R. (1974). *Proc. Natl. Acad. Sci. U.S.A.* **71,** 4298–4302.

Seidman, J. G., and Leder, P. (1980). *Nature (London)* **286,** 779–783.

Seidman, J. G., Leder, A., Nau, M., Norman, B., and Leder, P. (1978). *Science* **202,** 11–17.

Selsing, E., and Storb, U. (1981). *Cell (Cambridge, Mass.)* **25,** 47–58.

Silverton, E. W., Navia, M. A., and Davies, D. R. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5140–5144.

Sim, G. K., Yague, Y., Nelson, J., Marrack, P., Palmer, E., Augustin, A., and Kappler, J. (1984). *Nature (London)* **312,** 771–775.

Sims, J., Rabbitts, T. H., Estess, P., Slaughter, C., Tucker, P. W., and Capra, J. D. (1982). *Science* **216,** 309–311.

Singer, P. A., Singer, H. H., and Williamson, A. R. (1980). *Nature (London)* **285,** 294–300.

Sire, J., Auffray, C., and Jordan, B. R. (1982). *Gene* **20,** 377–386.

Stanford, J. M., and Wu, T. T. (1981). *J. Theor. Biol.* **88,** 421–439.

Stanworth, D. R. (1984). *Biochem. Soc. Trans.* **12,** 747–754.

Stott, D. I., and Williamson, A. R. (1982). *Compr. Biochem.* **19B,** 189–329.

Subcommittee on Immunoglobulin Nomenclature (1966). *Bull. W. H. O.* **35,** 953.

Subcommittee on Immunoglobulin Nomenclature (1969). *Bull. W. H. O.* **41,** 975–977.

Subcommittee on Immunoglobulin Nomenclature (1972). *Biochemistry* **11,** 3311–3312.

Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H., and Itakura, K. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6613–6617.

Sutton, B., and Phillips, D. C. (1983). *Biochem. Soc. Trans.* **11,** 130–132.

Swazey, J. P., and Reeds, K. (1978). "Today's Medicine, Tomorrow's Science." Natl. Inst. Health, Bethesda, Maryland.

Takahashi, N., Tetaert, D., Debuire, B., Lin, L.-C., and Putnam, F. W. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 2850–2854.

Takahashi, N., Takahashi, Y., Ortel, T. L., Lozier, J. N., Ishioka, N., and Putnam, F. W. (1984). *J. Chromatogr.* **317,** 11–26.

Takahashi, N., Ishioka, N., Takahashi, Y., and Putnam, F. W. (1985a). *J. Chromatogr.* **326,** 407–418.

Takahashi, N., Takahashi, Y., Ishioka, N., Heiny, M. E., and Putnam, F. W. (1985b). *Protides Biol. Fluids* **33,** 541–544.

Takayasu, T., Suzuki, S., Kametani, F., Takahashi, N., Shinoda, T., Okuyama, T., and Munekata, E. (1982). *Biochem. Biophys. Res. Commun.* **105,** 1066–1071.

Thorbecke, G. J., and Leslie, C. A., eds. (1982). "Immunoglobulin D: Structure and Function," Ann. N.Y. Acad. Sci., Vol. 399. N.Y. Acad. Sci., New York.

Titani, K., and Putnam, F. W. (1965). *Science* **147,** 1304–1305.

Titani, K., Whitley, E., Jr., Avogardo, L., and Putnam, F. W. (1965). *Science* **149,** 1090–1092.

Tønder, O., and Matre, R. (1985). *In* "Antibodies: Protective, Destructive and Regulatory Role" (F. Milgrom, C. J. Abeyounis, and B. Albini, eds.), 9th Int. Convoc. Immunol., pp. 234–240. Karger, Basel.

Tonegawa, S. (1985). *Sci. Am.* **253,** 122–131.
Torano, A., and Putnam, F. W. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 966–969.
Torano, A., Tsuzukida, Y., Liu, Y.-S. V., and Putnam, F. W. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 2301–2305.
Tsuzukida, Y., Wang, C.-C., and Putnam, F. W. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 1104–1108.
Tucker, P. W., Marcu, K. B., Newell, N., Richards, J., and Blattner, F. R. (1979). *Science* **206,** 1303–1306.
Tucker, P. W., Liu, C.-P., Mushinski, J. F., and Blattner, F. R. (1980). *Science* **209,** 1353–1360.
Tucker, P. W., Slightom, J. J., and Blattner, F. R. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 7684–7688.
Tunnacliffe, A., Kefford, R., Milstein, C., Forster, A., and Rabbitts, T. H. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 5068–5072.
Tyler, B. M., Cowman, A. F., Gerondakis, S. D., Adams, J., and Bernard, O. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 2008–2012.
Ueda, S., Takenaka, O., and Honjo, T. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 3712–3715.
Unkeless, J. C., and Wright, S. D. (1984). *Contemp. Top. Immunobiol.* **14,** 171–187.
Van Loghem, E., de Lange, G., and Koistinen, J. (1976). *Scand. J. Immunol.* **5,** 161–164.
Vassalli, P., Tartakoff, A., Pink, J. R. L., and Jaton, J. C. (1980). *J. Biol. Chem.* **255,** 11822–11827.
Veretennikova, N. I., Chipens, G. I., Nikiforovich, G. V., and Betinsh, Y. R. (1981). *Int. J. Pept. Protein Res.* **17,** 430–435.
Vitetta, E. S. (1982). *Ann. N.Y. Acad. Sci.* **399,** 255–264.
Vitetta, E. S., and Uhr, J. W. (1975). *Science* **189,** 964–969.
Vivanco-Martinez, F., Bragado, R., Albar, J. P., Jaurez, C., and Ortiz-Masllorens, F. (1980). *Mol. Immunol.* **17,** 327–336.
Wall, R. (1983). *Annu. Rev. Immunol.* **1,** 393–422.
White, M. B., Shen, A. L., Word, C. J., Tucker, P. W., and Blattner, F. R. (1985). *Science* **228,** 733–737.
Wilhelm, P., Pilz, I., Schwarz, E., Mihaesco, C., and Mihaesco, E. (1984). *Int. J. Biol. Macromol.* **6,** 273–276.
Wu, T. T., and Kabat, E. A. (1970). *J. Exp. Med.* **132,** 211–250.
Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I., and Mak, T. W. (1984). *Nature (London)* **308,** 145–149.
Yancopoulos, G. D., Disiderio, S. V., Paskind, M., Kearney, J. F., Baltimore, D., and Alt, F. W. (1984). *Nature (London)* **311,** 727–733.
Yasmeen, D., Ellerson, J. R., Dorrington, K. J., and Painter, R. H. (1976). *J. Immunol.* **116,** 518–522.
Yelton, D. E., and Scharff, M. D. (1981). *Annu. Rev. Biochem.* **50,** 657–680.
Zikan, J., Novotny, J., Trapane, T. L., Koshland, M. E., Urry, D. W., Bennett, J. C., and Mestecky, J. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 5905–5909.

3 Plasma Apolipoproteins: Gene Structure, Function, and Variants

Angelo M. Scanu

Departments of Medicine, Biochemistry, and Molecular Biology and the Lipoprotein Study Unit
University of Chicago, Pritzker School of Medicine
Chicago, Illinois 60637

The Plasma Proteins, Vol. V

I. Introduction

At the time of our previous review (Scanu *et al.*, 1975) a main focus of the research on plasma lipoproteins was on the description of the amino acid sequence of their apolipoproteins and on the structural information that could be derived from that knowledge. The relative abundance of α-helical amphiphilic repeats occurring in many of these apolipoproteins was recognized and also the general relevance of these repeats in protein–lipid interactions. The main concept emerging from those studies was that apolipoproteins are important determinants of the supramolecular organization of plasma lipoproteins by influencing their assembly and remodeling and, in some instances, their targeting to specific membrane receptors. Important new information is now being gathered from the application of the techniques of molecular biology, which have permitted studies of the gene structure and of the posttranscriptional and posttranslational events attending the process of maturation of each major plasma apolipoprotein. The purpose of this review is to describe these advances and discuss them in the context of newly discovered genetic variants present either in normal phenotypes

or in dyslipoproteinemic states associated with a high incidence of cardiovascular disease. Our discussion will examine apoA-I, apoA-II, apoA-IV, apoB, apoC-I, apoC-II, apoC-III, and apoE, apolipoproteins for most of which significant information on their structure and function is available (Table I). Recent reviews on the subject have appeared (Scanu *et al.*, 1982, 1984; Galton *et al.*, 1983; Mahley *et al.*, 1984; Gordon *et al.*, 1985a; Breslow. 1985, 1986; Weisgraber, 1985).

II. ApoA-I

A. General Properties, Biosynthesis, and Processing

ApoA-I (Scanu *et al.*, 1982, 1984; Gordon *et al.*, 1986; Breslow, 1985, 1986), the most abundant apolipoprotein of human plasma high-density lipoproteins (HDL) is represented in its monomeric form by a single polypeptide chain 243 residues long which has been shown to self-associate in aqueous buffers, interact with amphiphilic surfaces, and activate lecithin-cholesterol acyltransferase (LCAT; phosphatidylcholine-sterol acyltransferase), an enzyme which is responsible for the generation in the plasma of cholesteryl ester from unesterified cholesterol and lecithin. The liver and intestine are the two main sites of synthesis of this apolipoprotein; in both cases, the primary translation product is 24 amino acids longer than the mature apoA-I owing to the attachment to the NH_2-terminus of an 18-residue segment (presegment or signal peptide) and a 6-residue-long propeptide. Based on *in vitro* and *in vivo* studies, the presegment is cleaved intracellularly by a signal peptidase, whereas the prosegment undergoes cleavage at a Gln-Asp bond extracellularly. The protease responsible for this change is metal-dependent, inhibited by EDTA and *o*-phenanthroline, has an affinity for chylomicrons, VLDL, and HDL, and is dissociated from the surface of these lipoproteins by high salts (Edelstein *et al.*, 1983). The source and molecular properties of this enzyme have not yet been established. Its participation in proapoA-I to apoA-I conversion suggests that it may play a role in the process of HDL assembly and maturation.

B. Properties of the cDNA Clone

The examination of the cDNA clones (Fig. 1) has confirmed that apoA-I is synthesized as a prepropeptide with an mRNA 893 bp long (Shoulders and Baralle, 1982; Cheung and Chan, 1983; Law and Brewer, 1984; Karathanasis *et al.*, 1983a; Sharpe *et al.*, 1984). This includes a 5′ untranslated region of 35 bp, an 801-bp coding region, a termination codon, TGA, and a 3′ untranslated region of 54 bp followed by a poly(A) tail. The cDNA sequence specified an amino acid sequence of mature apoA-I of 243 amino acids similar to the protein

TABLE I

General Properties of Apolipoproteins of Normal Human Plasma

Apolipoprotein	Concentration in plasma (mg/dl)	Molecular weight	Isoelectric point, p*I*	Physiological role	Lipoprotein association
A-I	100–200	28,016	A-I_2 = 5.85 A-I_3 = 5.74 A-I_4 = 5.65 A-I_5 = 5.52 A-I_6 = 5.40	LCAT activator	Chyl, HDL_2, HDL_3, VHDL
A-II	30–40	17,440	4.9	Unknown	Chyl, HDL_2, HDL_3
A-IV	16–20	46,000	5.5	Transport	VLDL, HDL_2
B-100	90–110	~500,000	—	Cholesterol carrier; ligand apoB,E receptor	VLDL, IDL, LDL
B-48	0	250,000	—	Unknown	Chyl
C-I	4–6	6,630	7.5	LCAT activator	VLDL, IDL, HDL_2
C-II	3–5	8,824	4.9	Lipoprotein lipase activator	VLDL, IDL, HDL_2
C-III	12–14	8,764	C-III-0 = 5.0[a] C-III-1 = 4.85 C-III-2 = 4.65	Inhibition remnant uptake	VLDL, IDL, HDL_2
E	3–6	34,145	E-2 = 5.89 E-3 = 6.02[b] E-4 = 6.18	Ligand for apoE and apoB,E receptor	VLDL, IDL, HDL_c

[a]C-III-0, C-III-1, and C-III-2 each contain 0, 1, and 2 moles of sialic acid per mole protein, respectively.
[b]Disialylation of E-3 changes the p*I* of this isoprotein: E-$3_{s\text{-}1}$ = 5.89; E-$3_{s\text{-}2}$ = 5.78; E-$3_{s\text{-}3}$ = 5.68; s is the degree of sialylation in moles per mole protein.

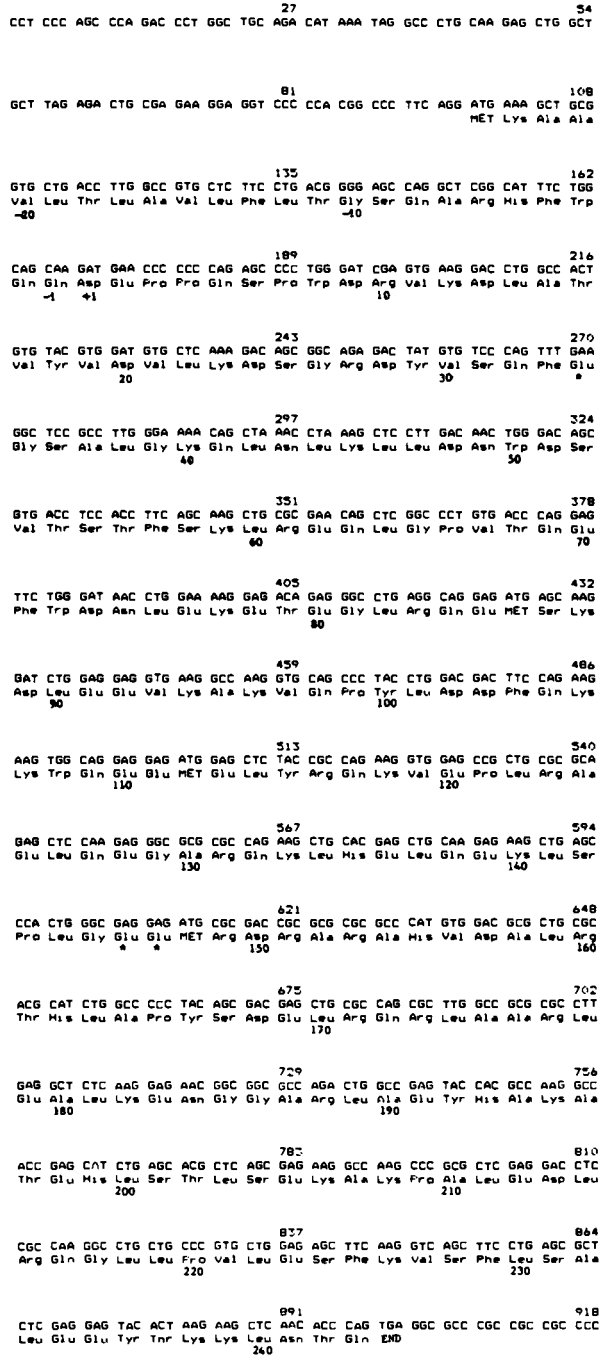


Fig. 1. Nucleotide sequence of human apoA-I, cDNA. The derived amino acid sequence is also reported (Karathanasis *et al.*, 1983a, 1986). An asterisk indicates the difference between nucleotide and amino acid sequences (see text).

sequence previously reported by Brewer *et al.* (1978) except that Glu and not Gln is the amino acid in position 34. The examination of the cDNA clone has also confirmed the notion derived from amino acid sequence data that between residues 99 to 230, there are six 22 α-helical amphiphilic tandem repeats, five of which begin with proline, an α-helix breaker. Oligopeptides obtained by chemical synthesis to mimic the 22-amino acid repeats of apoA-I have been found to exhibit solubility and LCAT-activation properties similar to apoA-I (Scanu *et al.*, 1982). More recently, the studies by Nakagawa *et al.* (1985) have provided evidence that the fundamental functional unit of apoA-I is a 44-amino acid structure prepared by linking two identical 22-residue segments through a proline residue.

C. ApoA-I Gene

The apoA-I gene (Karathanasis *et al.*, 1983a; Shoulders *et al.*, 1983; Sharpe *et al.*, 1984) has been reported to be 1863 bp in length and to contain three introns and four exons. Of the three introns, IVS-1, 197 bp long, is located in the 5′ untranslated region between bases 20 and 21 upstream of the Met initiator codon. IVS-2, 186 bp long, interrupts the codon specifying amino acid 10 in the presegment of apoA-I. IVS-3, 588 bp long, interrupts the codon specifying amino acid 43 of mature apoA-I. Of the four exons, exon 2 contains most of the prepeptide segment, exon 3 the propeptide and the NH_2-terminal sequence of mature apoA-I, and exon 4, 200 amino acids in the COOH-terminal portion of apoA-I and 66-bp repeats. The analysis of several apoA-I cDNA clones has also made it possible to identify the transcription initiation site and, upstream of it, the potential promoter region, 7 bp long, and rich in AT, the "TATA box." By somatic cell hybridization (Bruns *et al.*, 1984; Law *et al.*, 1984; Cheung *et al.*, 1984), the apoA-I gene has been found to occupy a single locus in the long arm of chromosome 11 in the proximity of the apoC-III and apoA-IV genes.

D. Protein Polymorphism

By the systematic application of the technique of isoelectric focusing and by partial sequence analyses, seven apoA-I genetic variants have been identified and characterized (Menzel *et al.*, 1983; Mahley *et al.*, 1984; Breslow, 1985, 1986) by a difference in a single + or − charge unit (normal isoprotein, p*I* 5.64). The names of these variants are those of the city of discovery (see Table II). No true pathological states have been associated with these variants except for a slight decrease in the plasma HDL levels in some of them (see also Section XIIA). The subjects with these variants are heterozygotes having one normal structural allele and a second abnormal one with one charge shift. These allelic variants have been observed in only one out of 1000 subjects examined; however, this frequen-

TABLE II

Human ApoA-I Variants

Variant	Charge relative to normal apoA-I	Mutation site	Abnormality[a]
Milano	−1	Arg^{173}→Cys	Low HDL
Munster-2(A) (Marburg)	−1	Lys^{107}→deletion	Low HDL, decreased
Munster-2(B)	−1	Ala^{158}→Glu	LCAT activation
Munster-3(A)	+1	Asp^{103}→Asn	N
Munster-3(B)	+1	Pro^{4}→Arg	N
Munster-3(C)	+1	Pro^{3}→His	N
Munster-3(D)	+1	Asp^{213}→Gly	N
Giesssen	+1	Pro^{143}→Arg	N
Munster-4	+2	Glu^{198}→Lys	N
Norway	+2	Glu^{136}→Lys	N

[a]N, none.

cy may increase when a larger-scale population is systematically screened utilizing both nucleotide and amino acid sequence techniques. It should be noted that isoelectric focusing alone will not detect mutations involving substitutions of neutral amino acids.

III. ApoA-II

A. General Properties, Biosynthesis, and Processing

ApoA-II (Scanu *et al.*, 1982; Mahley *et al.*, 1984; Breslow, 1985, 1986) is the second major protein constituent of human HDL: it consists of two 8700-Da proteins covalently linked by a single disulfide bond between the cysteine residue at position 6. The presence in small quantities of single-chain monomers has been reported; however, it is uncertain whether they are of natural occurrence or are artifactual. Synthesis of apoA-II has been reported to take place in the liver and the intestine in the form of a precursor protein. *In vitro* translation studies have shown that the apoA-II mRNA encodes a 100-amino acid-long protein comprising an 18-amino acid prepeptide, a 5-amino acid propeptide, and a 77-amino acid polypeptide representing the mature protein (Gordon *et al.*, 1983). The cleavage of the prepeptide is a cotranslational event; in turn the propeptide is cleaved posttranslationally. According to the results obtained in the human hepatoma cell line HepG2, this cleavage occurs after secretion and not intracellularly as it has been reported to occur in prosegments terminating with two

positively charged arginine residues. Recent studies by Gordon *et al.* (1985) have shown that the enzyme involved in the proapoA-II to apoA-II conversion is a thiol protease. This conclusion has been based on the following observations: (1) activity of the enzyme blocked by antipain, leupeptin, and Ala-Lys-Arg-chloromethyl ketone, all thiol protease inhibitors; (2) affinity labeling of a 52-kDa protease by ^{125}I-iodotyrosylated Ala-Lys-Arg-chloromethyl ketone, an affinity probe for cathepsin B, a thiol protease; (3) immunoprecipitation of the affinity-probe-labeled 54-kDa extracellular protease by a monospecific antiserum raised against human liver cathepsin B; and (4) inhibition of the extracellular conversion of proapoA-II to apoA-II by a cathepsin antibody. Based on these findings and molecular weight estimates on SDS gels, Gordon *et al.* (1985) have proposed that the enzyme responsible for the extracellular conversion of proapoA-II to apoA-II is a procathepsin B-like protease secreted from the cell as an uncleaved proenzyme. It is of interest that the participation of a procathepsin B enzyme in proprotein processing has been recently suggested by Docherty *et al.* (1982) for the intracellular conversion of proinsulin to insulin.

B. Properties of the cDNA Clone

The isolation of the cDNA clone has been reported by Sharpe *et al.* (1984), Moore *et al.* (1984), and Knott *et al.* (1984a). In the 5′ untranslated sequence, 22 bp have been identified (Fig. 2) but not the transcription initiation site. The rest of the DNA sequence, 300 bp long, specifies a region coding for 100 amino acids, a termination codon, TGA, and a 3′ untranslated region of 112 bp. In agreement with the results of the primary translation studies, the translated 100-amino acid region has been found to comprise an 18-amino acid presegment cleaved cotranslationally by microsomal membranes, a 5′ amino acid NH_2-terminal prosegment (Ala-Leu-Val-Arg-Arg) cleaved posttranslationally, and a polypeptide, 77 amino acids long, representing the mature protein. The amino acid sequence of mature apoA-II which has been derived from the cDNA studies is similar to that obtained by protein sequencing (Brewer *et al.*, 1972) except for residue 35, where the DNA sequence predicts Glu instead of Gln.

C. ApoA-II Gene

The gene for human ApoA-II has been isolated from a human genomic DNA library (Lackner *et al.*, 1985a) and located in chromosome 1 (Moore *et al.*, 1984; Lackner *et al.*, 1984). The cloned fragment was approximately 14 kb long and extended about 9.0 kb upstream as well as 3.5 kb downstream from the apoA-II gene, which was contained within a 3.1-kb *Hin*dIII fragment of human DNA. From the complete nucleic acid sequence it was established that the apoA-II gene contains four exons interrupted by three introns of 182, 293, and 395 bp. The first exon is located between nucleotides 34 and 35 of the 5′ untranslated region,

```
                                        27                                  54
ATA CCC GAG GAC AGA GAT GTT GGT TAG GCC GCC CTC CCC ACT GTT ACC AAC ATG
                                                                    MET

                                        81                                  108
AAG CTG CTC GCA GCA ACT GTG CTA CTC CTC ACC ATC TGC AGC CTT GAA GGA GCT
Lys Leu Leu Ala Ala Thr Val Leu Leu Leu Thr Ile Cys Ser Leu Glu Gly Ala
     -20                                            -10

                                                    a
                                        135         |                       162
TTG GTT CGG AGA CAG GCA AAG GAG CCA TGT GTG GAG AGC CTG GTT TCT CAG TAC
Leu Val Arg Arg Gln Ala Lys Glu Pro Cys Val Glu Ser Leu Val Ser Gln Tyr
            -1  +1                                  10

                                        189                                 216
TTC CAG ACC GTG ACT GAC TAT GGC AAG GAC CTG ATG GAG AAG GTC AAG AGC CCA
Phe Gln Thr Val Thr Asp Tyr Gly Lys Asp Leu MET Glu Lys Val Lys Ser Pro
                    20                                  30

                                        243                                 270
GAG CTT CAG GCC GAG GCC AAG TCT TAC TTT GAA AAG TCA AAG GAG CAG CTG ACA
Glu Leu Gln Ala Glu Ala Lys Ser Tyr Phe Glu Lys Ser Lys Glu Gln Leu Thr
                 *          40                                          50

                                        297                                 324
CCC CTG ATC AAG AAG GCT GGA ACG GAA CTG GTT AAC TTC TTG AGC TAT TTC GTG
Pro Leu Ile Lys Lys Ala Gly Thr Glu Leu Val Asn Phe Leu Ser Tyr Phe Val
                                       60

                                                    b
                                        351         |                       378
GAA CTT GGA ACA CAG CCT GCC ACC CAG TGA AGT GTC CAG CAC CAT TGT CTT CCA
Glu Leu Gly Thr Gln Pro Ala Thr Gln END
     70

                                        405                                 432
ACC CCA GCT GGC CTC TAG AAC ACC CAC TGG CCA GTC CTA GAG CTC CTG TCC CTA

                                        459
CCC ACT CTT TGC TAC AAT AAA TGC TGA ATG AAT CC
```

Fig. 2. Nucleotide sequence of apoA-II, cDNA. The derived amino acid sequence is also reported (Knott *et al.*, 1984a; Karathanasis *et al.*, 1986). An asterisk indicates the difference between nucleotide and amino acid sequences (see text).

the second is located between the first and second nucleotide representing the codon for amino acid -6 of the prepeptide segment (separating it from the rest of the apoA-II protein sequence), and the third one is located between the second and third base of the codon specifying amino acid 39 of the mature apolipoproteins. From the studies of Lackner *et al.* (1985a), it has emerged that the structure of the apoA-II gene is quite similar to that of apoA-I. Both apoA-I and

apoA-II have a short first exon of about 30 bp in the 5′ untranslated region of the mRNA. In either case, the second exon codes for most of the signal peptide. The third exon contains the propeptide segment as well as the first 38 and 42 amino acids of apoA-II and apoA-I, respectively. The fourth exon contains the remainder of the mRNA. The similarity in genetic structure between apoA-I and apoA-II has been taken to suggest that the two genes have derived from a common ancestral gene (see Section XI for details). Because of the difference in chromosomal localization (chromosome 1 for apoA-II and 11 for apoA-I), a genetic linkage between these two major apolipoproteins is not expected.

D. Protein Polymorphism

At this time genetic variants of apoA-II have not been reported. However, polymorphic forms have been detected in human plasma and thoracic duct lymph (Lackner *et al.*, 1985b). The major isoform identified had a p*I* of 4.9, and proapoA-II had a p*I* of 6.79. The other minor isoforms reported by Lackner *et al.* had p*I* values of 5.17, 4.68, 4.42, and 4.20, respectively, all occurring at very small concentrations. Several of these isoforms exhibited an apparent molecular weight higher than the major isoform. Since these high molecular forms were no longer present after neuraminidase treatment, they were considered to represent sialylated species; no significant differences in amino acid composition among them have been reported.

IV. ApoA-IV

A. General Properties, Biosynthesis, and Processing

ApoA-IV (Mahley *et al.*, 1984; Breslow, 1985, 1986), a 46,000-Da protein, is a relatively minor component of human plasma. Contrary to the rat, in which this apolipoprotein was first discovered (Swaney *et al.*, 1974) and is mostly carried in the chylomicrons and HDL fractions, in human plasma it is to a large extent unassociated with plasma lipoproteins (Beisiegel and Utermann, 1979; Fidge, 1980). Whether this lack of association represents an ultracentrifugal artifact or is related to the intrinsic structural properties of this protein is not clearly established. It has been proposed that the capacity of apoA-IV to self-associate in aqueous solution may partially prevent its affiliation with the lipoprotein surface (Weinberg and Spector, 1985). Both the liver and the intestine have been shown to produce apoA-IV. In the intestine fat feeding doubles the synthesis of this apolipoprotein (Gordon *et al.*, 1982); this observation has led to the suggestion that apoA-IV plays a role in synthesis and/or secretion of triglyceride-rich lipoproteins. Recent *in vitro* studies have shown that apoA-IV can activate LCAT (Steinmetz and Utermann, 1985). Moreover, in the rat it has been shown

that the LCAT reaction may affect the plasma distribution of this protein; thus, the association of apoA-IV with HDL may be related to the cholesteryl ester content of this particle (DeLamatre *et al.*, 1983). These proposed relationships between LCAT activity, apoA-IV, and HDL structure deserve further exploration. The RNA from human intestinal mucosa has been isolated and translated in wheat germ lysates and the primary translation product found to contain a prepeptide 20 amino acids long but not a propeptide. Similar findings have been reported in the rat. A 55% homology has been found between the human and rat prepeptide.

B. Properties of the cDNA Clone

A full-length apoA-IV clone of 1423 bp has been isolated from a rat intestinal cDNA library (Boguski *et al.*, 1984). From sequence analysis (Fig. 3) it has been shown that the 1173-nucleotide coding region specifies a protein of 391 amino acids, which includes a 20-amino acid signal peptide. The portion of the cDNA sequence representing the mature plasma protein contains a 66-nucleotide sequence that is repeated at least 13 times. Nine of these thirteen 22-amino acid repeat units of the derived protein sequences have been found to represent α-helices according to the Chou–Fasman rule and to be interrupted by proline residues. When hydration potentials of individual residues were taken into account, the α-helices were found to be amphiphilic.

C. ApoA-IV Gene

Starting from oligonucleotides as primers, the sequence of 2687 bp of the rat apoA-IV gene and flanking regions has been determined by Boguski *et al.* (1986). The apoA-IV gene consists of three exons of 126, 142, and 1157 bp and two introns 277 and 673 bp long. The results of these studies have uncovered some discrepancies between the sequence of the apoA-IV cDNA and that derived from the apoA-IV gene. Two of these discrepancies relate to base substitutions in exon D. Moreover, the codon beginning at nucleotide 2027 (residue 253) in the preapoA-IV region specifies a glutamine instead of histidine. According to Boguski *et al.* (1986), this substitution may represent a true variant rather than a cloning artifact. The transcription initiation site is at either nucleotide 229 or 230 and the putative promoter sequences CAAC and TTTAAA are 81 and 31 nucleotides upstream from this site. Boguski *et al.* (1985) also made the interesting observation that rat apoA-IV gene lacks the first intron described in the 5′ nontranslated regions of the mRNA of the genes of human apoA-I, apoC-III, and apoE. In the rat apoA-IV gene the 5′ nontranslated region, which also contains 16 out of the 20 codons comprising the signal peptide domain, is located within the first exon, whose length is approximately equal to the sum of the first and

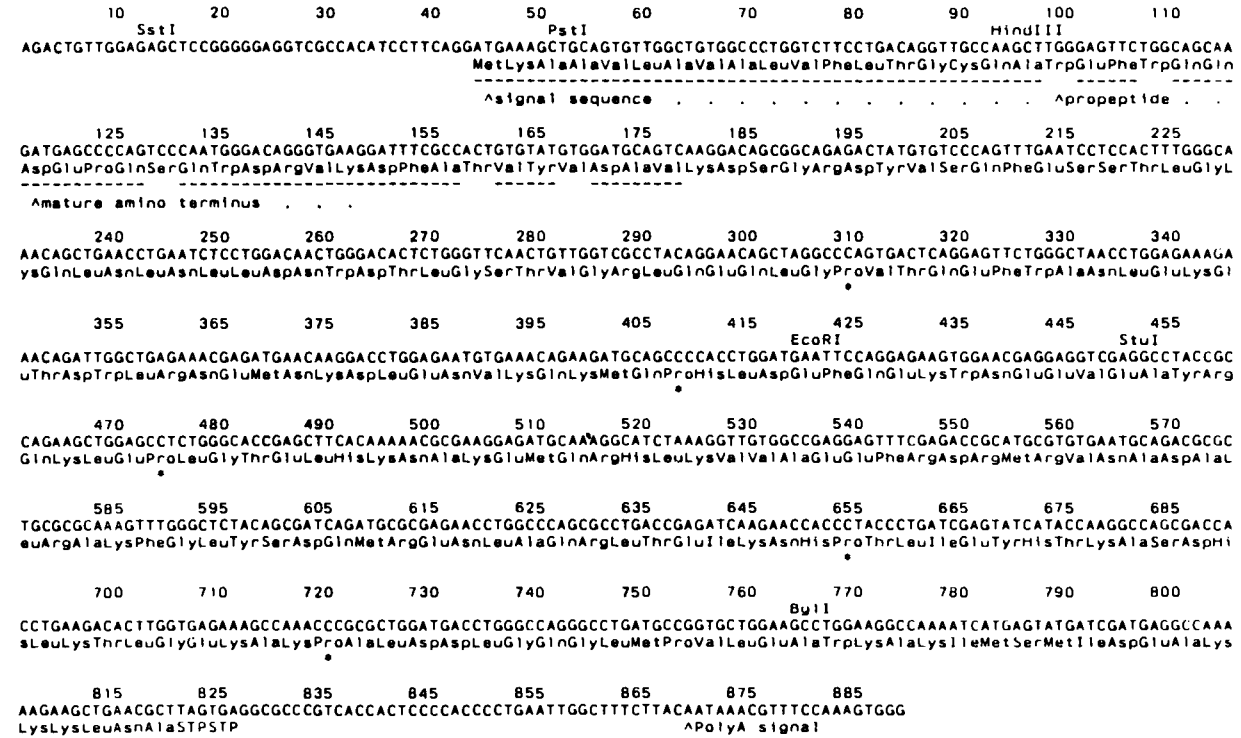


Fig. 3. Nucleotide sequence of rat apoA-IV, cDNA. The derived amino acid sequence is also reported (Boguski *et al.*, 1984).

second exons of the human apoA-I, apoC-III, and apoE genes. Thus, the first exon of the rat apoA-IV gene may have originated from the fusion of the first and second exons of other apolipoprotein genes. According to this interpretation the first intron of the rat apoA-IV gene would correspond to the second intron of the human apoA-I, apoC-III, and apoE genes. The second exon of the rat apoA-IV gene contains codons which specify the last three amino acids of the signal peptide and the first 39-amino acid segment of the mature apolipoprotein. This is the segment where the first of the 13 amphiphilic 22-residue repeats characteristic of the mature apoA-IV sequence is present. The third exon of the apoA-IV gene contains codons for the remaining 332 amino acids and the 3′ nontranslated region 158 nucleotides long.

Boguski *et al.* (1986) have used Southern blot analyses and the rat genomic apoA-IV clone as a probe to survey the genomic DNA of apoA-IV of several animal species. The rat apoA-IV gene hybridized only weakly to human DNA, suggesting dissimilarities between these two genes. This observation has been corroborated by Elshourbagy *et al.* (1986), who have reported the nucleotide and amino acid sequence of human apoA-IV (Fig. 4). A comparison of the sequences of human and rat apoA-IV revealed a 79% identity in the remainder of the sequences with the human protein containing 5 extra residues near the carboxyl terminus (Fig. 5).

D. Protein Polymorphism

Several genetic variants of apoA-IV have been described in the literature. In studies by Utermann *et al.* (1982) and by Menzel *et al.* (1982a) conducted in a normal German population, the frequency of the major isoprotein having a p*I* of 5.50 was 85.6%. Of the remainder of the subjects, 13.8% exhibited this isoprotein as well as another differing in one basic charge unit. The additional 0.6% of the subjects studied exhibited the more basic isoprotein. Genetic analyses of this population indicated that the frequency of the major allele specifying the acidic isoform of apoA-IV was 92.5%, whereas the frequency of the minor allele specifying the more basic isoform was 7.5%. Taken together, the studies were considered to be compatible with a model consisting of a single genetic locus and two alleles (Menzel *et al.*, 1982b).

V. ApoB

A. General Properties, Biosynthesis, and Processing

ApoB (Scanu *et al.*, 1975, 1982; Kane, 1983) represents the main protein component of the low-density lipoproteins (LDL); it is also an important constituent of chylomicrons and VLDL. Besides its lipid-carrier property apoB is

```
GAG GTC AGT GCT GAC CAG CTG GCC ACA GTG ATG TGG GAC TAC TTC AGC CAG CTG AGC AAC AAT GCC AAG GAG GCC GTG GAA CAT CTC CAG
Glu Val Ser Ala Asp Gln Val Ala Thr Val Met Trp Asp Tyr Phe Ser Gln Leu Ser Asn Asn Ala Lys Glu Ala Val Glu His Leu Gln
1                                  30                                  60                                  90
AAA TCT GAA CTC ACC CAG CAA CTC AAT GCC CTC TTC CAG GAC AAA CTT GGA GAA GTG AAC ACT TAC GCA GGT GAC CTG CAG AAG AAG CTG
Lys Ser Glu Leu Thr Gln Gln Leu Asn Ala Leu Phe Gln Asp Lys Leu Gly Glu Val Asn Thr Tyr Ala Gly Asp Leu Gln Lys Lys Leu
                                  120                                 150                                 180
GTG CCC TTT GCC ACC GAG CTG CAT GAA CGC CTG GCC AAG GAC TCG GAG AAA CTG AAG GAG GAG ATT GGG AAG GAG CTG GAG GAG CTG AGG
Val Pro Phe Ala Thr Glu Leu His Glu Arg Leu Ala Lys Asp Ser Glu Lys Leu Lys Glu Glu Ile Gly Lys Glu Leu Glu Glu Leu Arg
                                  210                                 240                                 270
GCC CGG CTG CTG CCC CAT GCC AAT GAG GTG AGC CAG AAG ATC GGG GAC AAC CTG CGA GAG CTT CAG CAG CGC CTG GAG CCC TAC GCG GAC
Ala Arg Leu Leu Pro His Ala Asn Glu Val Ser Gln Lys Ile Gly Asp Asn Leu Arg Glu Leu Gln Gln Arg Leu Glu Pro Tyr Ala Asp
                                  300                                 330                                 360
CAG CTG CGC ACC CAG GTC AAC ACG CAG GCC GAG CAG CTG CGG CGC CAG CTG ACC CCC TAC GCA CAG CGC ATG GAG AGA CTG CTG CGG GAG
Gln Leu Arg Thr Gln Val Asn Thr Gln Ala Glu Gln Leu Arg Arg Gln Leu Thr Pro Tyr Ala Gln Arg Met Glu Arg Val Leu Arg Glu
                                  390                                 420                                 450
AAC GCC GAC AGC CTG CAG GCC TCG CTG AGG CCC CAC GCC GAC GAG CTC AAG GCC AAG ATC GAC CAG AAC GTG GAG GAG CTC AAG GGA CGC
Asn Ala Asp Ser Leu Gln Ala Ser Leu Arg Pro His Ala Asp Glu Leu Lys Ala Lys Ile Asp Gln Asn Val Glu Glu Leu Lys Gly Arg
                                  480                                 510                                 540
CTT ACG CCC TAC GCT GAC GAA TTC AAA GTC AAG ATT GAC CAG ACC GTG GAG GAG CTG CGC CGC AGC CTG GCT CCC TAT GCT CAG GAC ACG
Leu Thr Pro Tyr Ala Asp Glu Phe Lys Val Lys Ile Asp Gln Thr Val Glu Glu Leu Arg Arg Ser Leu Ala Pro Tyr Ala Gln Asp Thr
                                  570                                 600                                 630
CAG GAG AAG CTC AAC CAC CAG CTT GAG GGC CTG ACC TTC CAG ATG AAG AAG AAC GCC GAG GAG CTC AAG GCC AGG ATC TCG GCC AGT GCC
Gln Glu Lys Leu Asn His Gln Leu Glu Gly Leu Thr Phe Gln Met Lys Lys Asn Ala Glu Glu Leu Lys Ala Arg Ile Ser Ala Ser Ala
                                  660                                 690                                 720
GAG GAG CTG CGG CAG AGG CTG GCG CCC TTG GCC GAG GAC GTG CGT GGC AAC CTG AGG GGC AAC ACC GAG GGG CTG CAG AAG TCA CTG GCA
Glu Glu Leu Arg Gln Arg Leu Ala Pro Leu Ala Glu Asp Val Arg Gly Asn Leu Arg Gly Asn Thr Glu Gly Leu Gln Lys Ser Leu Ala
                                  750                                 780                                 810
GAG CTG GGT GGG CAC CTG GAC CAG CAG GTG GAG GAG TTC CGA CGC CGG GTG GAG CCC TAC GGG GAA AAC TTC AAC AAA GCC CTG GTG CAG
Glu Leu Gly Gly His Leu Asp Gln Gln Val Glu Glu Phe Arg Arg Arg Val Glu Pro Tyr Gly Glu Asn Phe Asn Lys Ala Leu Val Gln
                                  840                                 870                                 900
CAG ATG GAA CAG CTC AGG ACG AAA CTG GGC CCC CAT GCG GGG GAC GTG GAA GGC CAC TTG AGC TTC CTG GAG AAG GAC CTG AGG GAC AAG
Gln Met Glu Gln Leu Arg Thr Lys Leu Gly Pro His Ala Gly Asp Val Glu Gly His Leu Ser Phe Leu Glu Lys Asp Leu Arg Asp Lys
                                  930                                 960                                 990
GTC AAC TCC TTC TTC AGC ACC TTC AAG GAG AAA GAG AGC CAG GAC AAG ACT CTC TCC CTC CCT GAG CTG GAG CAA CAG CAG GAA CAG CAT
Val Asn Ser Phe Phe Ser Thr Phe Lys Glu Lys Glu Ser Gln Asp Lys Thr Leu Ser Leu Pro Glu Leu Glu Gln Gln Gln Glu Gln His
                                 1020                                1050                                1080
CAG GAG CAG CAG CAG GAG CAG GTG CAG ATG CTG GCC CCT TTG GAG AGC TGA GCTGCCCCTG GTGCACTGGC CCCACCCTCG TGGACACCTG
Gln Glu Gln Gln Gln Glu Gln Val Gln Met Leu Ala Pro Leu Glu Ser ***
                                 1110                                1141

  CCCTGCCCTG CCACCTGTCT GTCTGTCCCA AAGAAGTTCT GGTATGAACT TGAGGACACA TGTCCAGTGG GAGGTGAGAC CACCTCTCAA TATTCAATAA
                               1201                                1231                                1261

  AGCTGCTGAG AATCTAGCCT C - Poly(A)
                    1291
```

```
EVSADQVATVMW   human
::  :::: :::
EVTSDQVANVMW   rat

13                                40
DYFSQLSNNAKEAVEHLQKSELTQQLN ALFQDKLGEVNTYAGDLQKKLV   human
::: :::::::::::: :::   :::::  :::::::  :::: ::: :::
DYFTQLSNNAKEAVEQLQKTDVTQQLN TLFQDKLGNINTYADDLQNKLV   rat

62                                      84
PFATELHERLAKDSEKLKEEIG                  KELEELRARLL   human
:::  :   : :  :   :::                   :::: :::
PFAVQLSGHLTKETERVREEIQ                  KELEDLRANMM   rat

95                                117
PHANEVSQKIGDNLRELQQRLE      PYADQLRTQVNTQAEQLRRQLT   human
:::: :::  :::   ::  :       :::  :  : : :     ::::
PHANKVSQMFGDNVQKLQEHLR      PYATDLQAQINAQTQDMKRQLT   rat

139                               161
PYAQRMERVLRENADSLQASLR      PHADELKAKIDQNVEELKGRLT   human
:: :::      :   :: :       : : ::: :  :: : ::: ::
PYIQRMQTTIQDNVENLQSSMV      PFANELKEKFNQNMEGLKGQLT   rat

183                               205
PYADEFKVKIDQTVEELRRSLA      PYAQDTQEKLNHQLEGLTFQMK   human
: : : :  :::  : ::  ::      : :   ::::::: ::: ::::
PRANELKATIDQNLEDLRSRLA      PLAEGVQEKLNHQMEGLAFQMK   rat

227                               249
KNAEELKARISASAEELRQRLA      PLAEDVRGNLRGNTEGLQ       human
::::::    :     :   ::      :: :::   : :::::::
KNAEELHTKVSTNIDQLQKNLA      PLVEDVQSKLKGNTEGLQ       rat

267                               289
KSLAELGGHLDQQVEEFRRRVE      PYGENFNKALVQQMEQLRTKLG   human
:::  :   :::::: ::: ::      : :  :: :::::::  :  ::
KSLEDLNKQLDQQVEVFRRAVE      PLGDKFNMALVQQMEKFRQQLG   rat

311
PHAGDVEGHLSFLEKDLRDKVN                               human
   :::: ::::::: :: ::
SDSGDVESHLSFLEKNLREKVS                               rat

SFFSTFKEKESQDKTLSLPELEQQQEQHQEQQQEQVQMLAPLES        human
:: ::   : : :  : ::  :: ::: ::: :       ::::
SFMSTLQKKGSPDQPLALPLPEQVQEQVQEQVQPK-----PLES        rat
```

Fig. 5. Alignment of the amino acid sequences of human and rat apoA-IV. Data from Elshourbagy *et al.* (1986).

Fig. 4. Nucleotide and amino acid sequence of human apoA-IV. The sequence begins with the first amino acid of the mature plasma protein. The translation termination codon is indicated by asterisks (Elshourbagy *et al.*, (1986).

involved in the secretion into the plasma of newly synthesized triglyceride-rich particles and also as a specific ligand for the high-affinity membrane receptor responsible for the uptake and degradation of LDL. The knowledge of the physicochemical properties of this protein is still limited largely due to its poor solubility in aqueous media. This insolubility has been variably attributed to either intrinsic properties of the apoprotein, self-association, or oxidative events. ApoB is a glycoprotein containing 8–10% carbohydrate by weight (galactose, mannose, glucosamine, and neuraminic acid). Two forms of this glycoprotein have been described, apoB-100 with molecular weight between 400,000 and 500,000 and apoB-48 with molecular weight of about 250,000 (Kane, 1983). The relationship between these two apolipoproteins is not clearly established although they are immunologically related and one may represent half of the other. In spite of many efforts, standard protein sequencing methods have failed to produce substantive information on the primary amino acid sequence of apoB. Recently, however, partial sequence information has been obtained for this apolipoprotein by LeBoeuf *et al.* (1984). *Staphylococcus aureus* protease was employed to produce large peptides which were then isolated in a relatively pure form by preparative gel electrophoresis. Two of these peptides had the following sequences: -Ala-Leu-Val-Gly-Ile-Asn-Gly-Glu-Ala-Asn-Leu-Asp-Phe-Leu-Asn-Ile-Pro-Leu-Arg-Ile-Pro-Pro-Met-Arg- and -Leu-Val-Ala-Lys-Pro-Ser-Val-Ser-Val-Glu-Phe-Val-Thr-Asn-Met-Gly-Ile-Ile-Pro-Lys-Phe-Ala-Arg-. These two sequences, which had no obvious homologies with those reported for other plasma apolipoproteins, were used to construct oligonucleotide probes and these in turn were utilized toward the elucidation of the structure of cDNA clones (see below). Sequence analyses were also carried out by Knott *et al.* (1985) on fragment T2 of the four obtained by digesting human apoB with thrombin. The NH_2-terminus yielded a single sequence: Ala-Val-Ser-Met-Pro-Ser-Phe-Ser-Ile-Leu-Gly-Ser-Asp-Val-X-Val-Pro-Ser-Tyr-Thr-Leu-Ile-Leu-Pro-Ser-Leu-Glu-Leu-Pro. Since residue X could not be unequivocally identified, the oligonucleotide probe synthesized for use in screening the cDNA libraries was based on only the first 14 amino acid residues (see below).

In terms of biosynthesis, Bell-Quint *et al.* (1981) have reported that both apoB-100 and apoB-48 are produced by cultured rat hepatocytes although only apoB-48 was identified in the medium. Evidence for the production of both apoB species in the liver has also been reported by Wu and Windmueller (1979), Sparks *et al.* (1981), and others (see Kane, 1983). In addition, evidence is emerging that the two apolipoproteins may be under different regulation. Several studies have also shown that apoB is synthesized and secreted by the cells of the intestinal mucosa and that apoB-48 is the preferred species. However, Wu and Windmueller (1979), studying the mesenteric lymph of lymph-diverted rats, have shown that 5% of the apoB is of the 100 type, suggesting that a limited synthesis of this hepatic species of apoB can also occur in the intestine. The

interpretation of these findings is made difficult by our current limited understanding of the structural relationship between apoB-100 and apoB-48.

Wettsten *et al.* (1985) have used pulse–chase experiments to study the synthesis of apoB in a human hepatoma cell line, $HepG_3$. A 2-min pulse with [^{35}S]methionine was followed by a 5- to 90-min chase period during which time a protein with a molecular mass of 312 ± 41 kDa was immunoprecipitated from the cells by either a monoclonal or a polyclonal antibody raised against human apoB-100. The synthesized species appeared in the medium after 30–35 min of chase and had a complement of lipid compatible with that of an LDL particle. The results of these studies were taken to suggest that apoB is synthesized as a large polypeptide and rapidly secreted following synthesis at least by the transformed cells used in those experiments. In previous studies using an *in vitro* translation system the same laboratory reported that an 80-kDa protein was precipitated by a monoclonal antibody against apoB (Boström *et al.*, 1984); based on the more recent pulse–chase studies, it is apparent that only a partial translation of the apoB mRNA was achieved.

B. Properties of the cDNA Clone

Lusis *et al.* (1985) have reported the cloning of the apoB cDNA from rat liver. Fifteen putative clones were identified by antibody screening of a rat liver cDNA library in the λgt 11 expression vector. All of these clones proved to be identical to each other by immunological studies and to contain sequences found only in the high-molecular-weight form of rat liver apoB. Blotting studies showed that the clones hybridize to a single 20-kb liver mRNA species sufficiently large to encode the entire apoB protein estimated to be 400 kDa in size. The apoB mRNA was abundant in liver, present in the rat intestine, and absent in the other tissues examined, i.e., spleen, kidney, heart, and brain. One clone, corresponding to a 240-base segment of the apoB mRNA, exhibited homology with a short region of rat apoE mRNA. The secondary structure of this protein segment was not rich in the amphiphilic α-helical structures noted in many of the apolipoproteins described thus far. The studies by Lusis *et al.* (1985) have also shown that apoB cDNA clones from human HepG2 cells identified by screening expression libraries with polyclonal antibodies to intact apoB hybridize to an mRNA which is identical in size with that observed in the rat product. These studies, indicating that the cDNA clone from either rat or human liver codes for a high-molecular-weight peptide corresponding to the size of apoB, have received support from the work of Deeb *et al.* (1985). Human liver cDNA library was screened for sequences coding for apoB using as a hybridization probe a mixture of synthetic oligonucleotides 26 bases long coding for one of the apoB peptides sequenced by LeBoeuf *et al.* (1984). Deeb *et al.* (1985) identified a clone with a cDNA insert of 593 bp containing sequences coding for the 24-residue peptide that had been

isolated from products of the limited proteolysis of apoB. The entire nucleotide sequence of the cDNA insert was found to consist of an open reading frame coding for 197 amino acids. The amino acid sequence derived from bases 5 to 29 was found by computer search to have no significant similarity to any of the other plasma apolipoproteins reported thus far. An internal homology was noted between the peptide segment coded by bases 9 to 35 (Phe-Pro-Asp-Ser-Val-Asn-Lys-Ala-Leu) and that coded by bases 54 to 80 (Val-Pro-Asp-Gly-Val-Ser-Lys-Val-Leu). Neither of the two had amphiphilic α-helix properties. In additional studies, apoB-related RNAs were found in a human hepatoma cell line (HepG2) and in baboon liver but not in placenta, simian virus 40-transformed fibroblasts, or a lymphoblastoid cell line. The length of the mature apoB in RNA was estimated to be 18 kb, enough to code for a protein with a molecular weight of approximately 500,000.

Knott *et al.* (1985) have reported the primary structure of the carboxyl-terminal 30% (1455 amino acids) of human apoB as deduced from the nucleotide sequence of the complementary DNA representing the 3′ end of apoB-100 amino acids of the COOH-terminus (Fig. 6). From Chou–Fasman analysis this region was found to be highly ordered with approximately 30% α-helical and 25% β-structure. Hydrophobic plots also predicted a balance of hydrophilic/hydrophobic regions characteristic of integral nonexchangeable proteins and thus distinct from the other apolipoproteins. The potential domains for apoB,E receptor and heparin binding were identified as well as the sites for CHO linkage.

C. ApoB Gene

At the time of this writing no information on the apoB gene is available in either human or other animal species. In humans it has been located in chromosome 2 by Knott *et al.* (1985).

D. Protein Polymorphism

The modest information on the primary structure of apoB has not made it possible to explore whether a protein polymorphism is associated with either normolipidemic or dyslipoproteinemic states. Most of the information thus far is based on immunological studies except for the Lp(a) variant for which important new information is available (see Section VI) and the Ag system described by Allison and Blumberg (1961) in subjects undergoing multiple transfusions. Based on binding affinities, Schumaker *et al.* (1984) have identified three phenotypes of apoB having strong, intermediate, and weak binding for monospecific antibodies for apoB. These findings were taken to suggest that these three phenotypes have a single genetic locus and two alleles specifying strong and weak binding forms of apoB. According to this interpretation, strong and weak binding

would represent homozygosity for each allele whereas the intermediate pattern would represent heterozygosity. These observations, although attributed to differences in amino acid sequences, require additional studies. Genetic variants of apoB attended by pathological states will be discussed in Section XII.

VI. Lp(a)

A. General Properties

Lipoprotein(a) (Berg, 1983; Fless and Scanu, 1986) is a variant of LDL in which apoB is linked by disulfide bridge(s) to an apoprotein, named apo(a). Unlike the other lipoproteins, whose nomenclature is usually derived from their buoyant density, Lp(a) was named by the immunogeneticist Berg (1963) to designate a factor or antigen detected in the plasma of some individuals using antisera from rabbits hyperimmunized with LDL of different human subjects. Subsequent studies showed that this antigen is associated with a lipoprotein having pre-β electrophoretic mobility on agarose gels and a hydrated density ranging between 1.05 and 1.12gm/ml.

By classical ultracentrifugal procedures Lp(a) is difficult to separate from LDL because of the small differences in size and density. Moreover, a small fraction of LDL is isopycnic with Lp(a) and therefore when density gradient ultracentrifugation is used, the two particles band in the same position. Two methods have been developed to isolate Lp(a) free of LDL (Fless *et al.*, 1984). One makes use of heparin Sepharose chromatography and takes advantage of the differential affinity of Lp(a) and LDL for heparin; the other is chromatofocusing, by which the particles are separated according to charge or isoelectric point. Purified Lp(a) from different individuals has been found to range in density between *d* 1.047 to 1.100 gm/ml. Moreover, the same human subject can have more than one Lp(a) species differing in density and sometimes also in electrophoretic behavior. Fless *et al.* (1984) found that the dense Lp(a) has a molecular weight larger than the low-density form due to the higher apo(a) mass. When Lp(a) is analyzed by SDS gel electrophoresis in the presence of 2-mercaptoethanol, apo(a) dissociates from apoB. Under these conditions, the apo(a) of the lower-density Lp(a) is smaller and the apo(a) of the higher-density Lp(a) is larger than apoB-100. Some individuals appear to have a third kind of Lp(a) with an intermediate density; on reduction it yields apo(a) and apoB-100, both exhibiting equal mobility but differing from each other by immunological criteria. When the overall properties of Lp(a) and LDL are compared, the molecular weight and density of the Lp(a) species are generally larger than those of LDL. Since the molar content of the lipid classes is almost the same in both lipoproteins it follows that the difference must relate to the protein moiety. From structural analyses Fless *et al.* (1984)

1 20 40
AAC ATT ATG GAG GCC CAT GTA GGA ATA AAT GGA GAA GCA AAT CTG GAT TTC TTA AAC ATT CCT TTA ACA ATT CCT GAA ATG CGT CTA CCT TAC ACA ATA ATC ACA ACT CCT CCA CTG AAA
Asn Ile Met Glu Ala His Val Gly Ile Asn Gly Glu Ala Asn Leu Asp Phe Leu Asn Ile Pro Leu Thr Ile Pro Glu Met Arg Leu Pro Tyr Thr Ile Ile Thr Thr Pro Pro Leu Lys

60 80
GAT TTC TCT CTA TGG GAA AAA ACA GGC TTG AAG GAA TTC TTG AAA ACG ACA AAG CAA TCA TTT GAT TTA AGT GTA AAA GCT CAG TAT AAG AAA AAC AAA CAC AGG CAT TCC ATC ACA AAT
Asp Phe Ser Leu Trp Glu Lys Thr Gly Leu Lys Glu Phe Leu Lys Thr Thr Lys Gln Ser Phe Asp Leu Ser Val Lys Ala Gln Tyr Lys Lys Asn Lys His Arg His Ser Ile Thr Asn

100 120
CCT TTG GCT GTG CTT TGT GAG TTT ATC AGT CAG AGC ATC AAA TCC TTT GAC AGG CAT TTT GAA AAA AAC AGA AAC AAT GCA TTA GAT TTT GTC ACC AAA TCC TAT AAT GAA ACA AAA ATT
Pro Leu Ala Val Leu Cys Glu Phe Ile Ser Gln Ser Ile Lys Ser Phe Asp Arg His Phe Glu Lys Asn Arg Asn Asn Ala Leu Asp Phe Val Thr Lys Ser Tyr Asn Glu Thr Lys Ile

140 160
AAG TTT GAT AAG TAC AAA GCT GAA AAA TCT CAC GAC GAG CTC CCC AGG ACC TTT CAA ATT CCT GGA TAC ACT GTT CCA GTT GTC AAT GTT GAA GTG TCT CCA TTC ACC ATA GAG ATG TCG
Lys Phe Asp Lys Tyr Lys Ala Glu Lys Ser His Asp Glu Leu Pro Arg Thr Phe Gln Ile Pro Gly Tyr Thr Val Pro Val Val Asn Val Glu Val Ser Pro Phe Thr Ile Glu Met Ser

180 200
GCA TTC GGC TAT GTG TTC CCA AAA GCA GTC AGC ATG CCT AGT TTC TCC ATC CTA GGT TCT GAC GTC CGT GTG CCT TCA TAC ACA TTA ATC CTG CCA TCA TTA GAG CTG CCA GTC CTT CAT
Ala Phe Gly Tyr Val Phe Pro Lys Ala Val Ser Met Pro Ser Phe Ser Ile Leu Gly Ser Asp Val Arg Val Pro Ser Tyr Thr Leu Ile Leu Pro Ser Leu Glu Leu Pro Val Leu His

220 240
GTC CCT AGA AAT CTC AAG CTT TCT CTT CCA CAT TTC AAG GAA TTG TGT ACC ATA AGC CAT ATT TTT ATT CCT GCC ATG GGC AAT ATT ACC TAT GAT TTC TCC TTT AAA TCA AGT GTC ATC
Val Pro Arg Asn Leu Lys Leu Ser Leu Pro His Phe Lys Glu Leu Cys Thr Ile Ser His Ile Phe Ile Pro Ala Met Gly Asn Ile Thr Tyr Asp Phe Ser Phe Lys Ser Ser Val Ile

260 280
ACA CTG AAT ACC AAT GCT GAA CTT TTT AAC CAG TCA GAT ATT GTT GCT CAT CTC CTT TCT TCA TCT TCA TCT GTC ATT GAT GCA CTG CAG TAC AAA TTA GAG GGC ACC ACA AGA TTG ACA
Thr Leu Asn Thr Asn Ala Glu Leu Phe Asn Gln Ser Asp Ile Val Ala His Leu Leu Ser Ser Ser Ser Ser Val Ile Asp Ala Leu Gln Tyr Lys Leu Glu Gly Thr Arg Leu Thr

300 320
AGA AAA AGG GGA TTG AAG TTA GCC ACA GCT CTG TCT CTG AGC AAC AAA TTT GTG GAG GGT AGT CAT AAC AGT ACT GTG AGC TTA ACC ACG AAA AAT ATG GAA GTG TCA GTG GCA AAA ACC
Arg Lys Arg Gly Leu Lys Leu Ala Thr Ala Leu Ser Leu Ser Asn Lys Phe Val Glu Gly Ser His Asn Ser Thr Val Ser Leu Thr Thr Lys Asn Met Glu Val Ser Val Ala Lys Thr

340 360
ACA AAA GCC GAA ATT CCA ATT TTG AGA ATG AAT TTC AAG CAA GAA CTT AAT GGA AAT ACC AAG TCA AAA CCT ACT GTC TCT TCC TCC ATG GAA TTT GAT TTC AAT TCT TCA ATG
Thr Lys Ala Glu Ile Pro Ile Leu Arg Met Asn Phe Lys Gln Glu Leu Asn Gly Asn Thr Lys Ser Lys Pro Thr Val Ser Ser Ser Met Glu Phe Lys Tyr Asp Phe Asn Ser Ser Met

380 400
CTG TAC TCT ACC GCT AAA GCA GCA GTT GAC CAC AAG CTT AGC TTG GAA AGC CTC ACC TCT TAC TTT TCC ATT GAG TCA TCT ACC AAA GCA GAT GTC AAG GGT TCG GTT CTT TCT CGG GAA
Leu Tyr Ser Thr Ala Lys Gly Ala Val Asp His Lys Leu Ser Leu Glu Ser Leu Thr Ser Tyr Phe Ser Ile Glu Ser Ser Thr Lys Gly Asp Val Lys Gly Ser Val Leu Ser Arg Glu

420 440
TAT TCA GGA ACT ATT GCT AGT GAG GCC AAC ACT TAC TTG AAT TCC AAG AGC ACA CGG TCT TCA GTG AAG CTG CAG GGC ACT TCC AAA ATT GAT GAT ATC TGG AAC CTT GAA GTA AAA GAA
Tyr Ser Gly Thr Ile Ala Ser Glu Ala Asn Thr Tyr Leu Asn Ser Lys Ser Thr Arg Ser Ser Val Lys Leu Gln Gly Thr Ser Lys Ile Asp Asp Ile Trp Asn Leu Glu Val Lys Glu

Fig. 6. Nucleotide sequence of cDNA clones representing the 3′ end of apoB-100 mRNA and the derived amino acid sequence of the COOH-terminal 30% of apoB-100. The arrow indicates the thrombin cleavage site that gives rise to fragment T2. The lines under the derived amino acid sequence indicate those regions found from peptide sequencing. The broken lines indicate those amino acid residues that could not be assigned with certainty. The dotted lines indicate the peptide R2-5 of LeBoeuf *et al.* (1984). Circles indicate potential N-glycosylation sites. Asterisks indicate the termination codon from Knott *et al.* (1985). (*Continued.*)

```
                                                                          780                                                                       800
TTC TCT GTA CCT GCT GGA ATT GTC ATT CCT TCC TTT CAA GCA CTG ACT GCA CGC TTT GAG GTA GAC TCT CCC GTG TAT AAT GCC ACT TGG AGT GCC AGT TTG AAA AAC AAA GCA GAT TAT
Phe Ser Val Pro Ala Gly Ile Val Ile Pro Ser Phe Gln Ala Leu Thr Ala Arg Phe Glu Val Asp Ser Pro Val Tyr Asn Ala Thr Trp Ser Ala Ser Leu Lys Asn Lys Ala Asp Tyr

                                                                          820                                                                       840
GTT GAA ACA GTC CTG GAT TCC ACA TGC AGC TCA ACC GTA CAG TTC CTA GAA TAT GAA CTA AAT GTT TTG GGA ACA CAC AAA ATC GAA GAT GGT ACG TTA GCC TCT AAG ACT AAA GGA ACA
Val Glu Thr Val Leu Asp Ser Thr Cys Ser Ser Thr Val Gln Phe Leu Glu Tyr Glu Leu Asn Val Leu Gly Thr His Lys Ile Glu Asp Gly Thr Leu Ala Ser Lys Thr Lys Gly Thr

                                                                          860                                                                       880
CTT GCA CAC CGT GAC TTC AGT GCA GAA TAT GAA GAA GAT GGC AAA TTT GAA GGA CTT CAG GAA TGG GAA GGA AAA GCG CAC CTC AAT ATC AAA AGC CCA GCG TTC ACC GAT CTC CAT CTG
Leu Ala His Arg Asp Phe Ser Ala Glu Tyr Glu Glu Asp Gly Lys Phe Glu Gly Leu Gln Glu Trp Glu Gly Lys Ala His Leu Asn Ile Lys Ser Pro Ala Phe Thr Asp Leu His Leu

                                                                          900                                                                       920
CGC TAC CAG AAA GAC AAG AAA GGC ATC TCC ACC TCA GCA GCC TCC CCA GCC GTA GGC ACC GTG GGC ATG GAT ATG GAT GAA GAT GAC GAC TTT TCT AAA TGG AAC TTC TAC TAC AGC CCT
Arg Tyr Gln Lys Asp Lys Lys Gly Ile Ser Thr Ser Ala Ala Ser Pro Ala Val Gly Thr Val Gly Met Asp Met Asp Glu Asp Asp Asp Phe Ser Lys Trp Asn Phe Tyr Tyr Ser Pro

                                                                          940                                                                       960
CAG TCC TCT CCA GAT AAA AAA CTC ACC ATA TTC AAA ACT GAG TTG AGG GTC CGG GAA TCT GAT GAG GAA ACT CAG ATC AAA GTT AAT TGG GAA GAA GAG GCA GCT TCT GGC TTG CTA ACC
Gln Ser Ser Pro Asp Lys Lys Leu Thr Ile Phe Lys Thr Glu Leu Arg Val Arg Glu Ser Asp Glu Glu Thr Gln Ile Lys Val Asn Trp Glu Glu Glu Ala Ala Ser Gly Leu Leu Thr

                                                                          980                                                                      1000
TCT CTG AAA GAC AAC GTG CCC AAG GCC ACA GGG GTC CTT TAT GAT TAT GTC AAC AAG TAC CAC TGG GAA CAC ACA GGG CTC ACC CTG AGA GAA GTG TCT TCA AAG CTG AGA AGA AAT CTG
Ser Leu Lys Asp Asn Val Pro Lys Ala Thr Gly Val Leu Tyr Asp Tyr Val Asn Lys Tyr His Trp Glu His Thr Gly Leu Thr Leu Arg Glu Val Ser Ser Lys Leu Arg Arg Asn Leu

                                                                         1020                                                                      1040
CAG AAC AAT GCT GAG TGG GTT TAT CAA GGG GCC ATT AGG CAA ATT GAT GAT ATC GAC GTG AGG TTC CAG AAA GCA GCC AGT GGC ACC ACT GGG ACC TAC CAA GAG TGG AAG GAC AAG GCC
Gln Asn Asn Ala Glu Trp Val Tyr Gln Gly Ala Ile Arg Gln Ile Asp Asp Ile Asp Val Arg Phe Gln Lys Ala Ala Ser Gly Thr Thr Gly Thr Tyr Gln Glu Trp Lys Asp Lys Ala

                                                                         1060                                                                      1080
CAG AAT CTG TAC CAG GAA CTG TTG ACT CAG GAA GGC CAA GCC AGT TTC CAG GGA CTC AAG GAT AAC GTG TTT GAT GGC TTG GTA CGA GTT ACT CAA AAA TTC CAT ATG AAA GTC AAG CAT
Gln Asn Leu Tyr Gln Glu Leu Leu Thr Gln Glu Gly Gln Ala Ser Phe Gln Gly Leu Lys Asp Asn Val Phe Asp Gly Leu Val Arg Val Thr Gln Lys Phe His Met Lys Val Lys His

                                                                         1100                                                                      1120
CTG ATT GAC TCA CTC ATT GAT TTT CTG AAC TTC CCC AGA TTC CAG TTT CCG GGG AAA CCT GGG ATA TAC ACT AGG GAG GAA CTT TGC ACT ATG TTC ATA AGG GAG GTA GGG ACG GTA CTG
Leu Ile Asp Ser Leu Ile Asp Phe Leu Asn Phe Pro Arg Phe Gln Phe Pro Gly Lys Pro Gly Ile Tyr Thr Arg Glu Glu Leu Cys Thr Met Phe Ile Arg Glu Val Gly Thr Val Leu

                                                                         1140                                                                      1160
TCC CAG GTA TAT TCG AAA GTC CAT AAT GGT TCA GAA ATA CTG TTT TCC TAT TTC CAA GAC CTA GTG ATT ACA CTT CCT TTC GAG TTA AGG AAA CAT AAA CTA ATA GAT GTA ATC TCG ATG
Ser Gln Val Tyr Ser Lys Val His Asn Gly Ser Glu Ile Leu Phe Ser Tyr Phe Gln Asp Leu Val Ile Thr Leu Pro Phe Glu Leu Arg Lys His Lys Leu Ile Asp Val Ile Ser Met

                                                                         1180                                                                      1200
TAT AGG GAA CTG TTG AAA GAT TTA TCA AAA GAA GCC CAA GAG GTA TTT AAA GCC ATT CAG TCT CTC AAG ACC ACA GAG GTG CTA CGT AAT CTT CAG GAC CTT TTA CAA TTC ATT TTC CAA
Tyr Arg Glu Leu Leu Lys Asp Leu Ser Lys Glu Ala Gln Glu Val Phe Lys Ala Ile Gln Ser Leu Lys Thr Thr Glu Val Leu Arg Asn Leu Gln Asp Leu Leu Gln Phe Ile Phe Gln
```

```
                                                                                                                    1220                                                                                                                    1240
CTA ATA GAA GAT AAC ATT AAA CAG CTG AAA GAG ATG AAA TTT ACT TAT CTT ATT AAT TAT ATC CAA GAT GAG ATC AAC ACA ATC TTC AAT GAT TAT ATC CCA TAT GTT TTT AAA TTG TTG
Leu Ile Glu Asp Asn Ile Lys Gln Leu Lys Glu Met Lys Phe Thr Tyr Leu Ile Asn Tyr Ile Gln Asp Glu Ile Asn Thr Ile Phe Asn Asp Tyr Ile Pro Tyr Val Phe Lys Leu Leu

                                                                                                                    1260                                                                                                                    1280
AAA GAA AAC CTA TGC CTT AAT CTT CAT AAG TTC AAT GAA TTT ATT CAA AAC GAG CTT CAG GAA GCT TCT CAA GAG TTA CAG CAG ATC CAT CAA TAC ATT ATG GCC CTT CGT GAA GAA TAT
Lys Glu Asn Leu Cys Leu Asn Leu His Lys Phe Asn Glu Phe Ile Gln Asn Glu Leu Gln Glu Ala Ser Gln Glu Leu Gln Gln Ile His Gln Tyr Ile Met Ala Leu Arg Glu Glu Tyr

                                                                                                                    1300                                                                                                                    1320
TTT GAT CCA AGT ATA GTT GGC TGG ACA GTG AAA TAT TAT GAA CTT GAA GAA AAG ATA GTC AGT CTG ATC AAG AAC CTG TTA GTT GCT CTT AAG GAC TTC CAT TCT GAA TAT ATT GTC AGT
Phe Asp Pro Ser Ile Val Gly Trp Thr Val Lys Tyr Tyr Glu Leu Glu Glu Lys Ile Val Ser Leu Ile Lys Asn Leu Leu Val Ala Leu Lys Asp Phe His Ser Glu Tyr Ile Val Ser

                                                                                                                    1340                                                                                                                    1360
GCC TCT AAC TTT ACT TCC CAA CTC TCA AGT CAA GTT GAG CAA TTT CTG CAC AGA AAT ATT CAG GAA TAT CTT AGC ATC CTT ACC GAT CCA GAT GGA AAA GGG AAA GAG AAG ATT GCA GAG
Ala Ser Asn Phe Thr Ser Gln Leu Ser Ser Gln Val Glu Gln Phe Leu His Arg Asn Ile Gln Glu Tyr Leu Ser Ile Leu Thr Asp Pro Asp Gly Lys Gly Lys Glu Lys Ile Ala Glu
                ●
                                                                                                                    1380                                                                                                                    1400
CTT TCT GCC ACT GCT CAG GAA ATA ATT AAA AGC CAG GCC ATT GCG ACG AAG AAA ATA ATT TCT GAT TAC CAC CAG CAG TTT AGA TAT AAA CTG CAA GAT TTT TCA GAC CAA CTC TCT GAT
Leu Ser Ala Thr Ala Gln Glu Ile Ile Lys Ser Gln Ala Ile Ala Thr Lys Lys Ile Ile Ser Asp Tyr His Gln Gln Phe Arg Tyr Lys Leu Gln Asp Phe Ser Asp Gln Leu Ser Asp

                                                                                                                    1420                                                                                                                    1440
TAC TAT GAA AAA TTT ATT GCT GAA TCC AAA AGA TTG ATT GAC CTG TCC ATT CAA AAC TAC CAC ACA TTT CTG ATA TAC ATC ACG GAG TTA CTG AAA AAG CTG CAA TCA ACC ACA GTC ATG
Tyr Tyr Glu Lys Phe Ile Ala Glu Ser Lys Arg Leu Ile Asp Leu Ser Ile Gln Asn Tyr His Thr Phe Leu Ile Tyr Ile Thr Glu Leu Leu Lys Lys Leu Gln Ser Thr Thr Val Met

                                                                                                                1455
AAC CCC TAC ATG AAG CTT GCT CCA GGA GAA CTT ACT ATC ATC CTC TAA   TTTTTTAAAA GAAATCTCAT TATCTCTTTC CAATGAACTT CACATAGCAC AGAAAAAATC AAACTGCCTA TATTGATAAA
Asn Pro Tyr Met Lys Leu Ala Pro Gly Glu Leu Thr Ile Ile Leu ***

ACCATACAGT GAGCCAGCCT TGCAGTAGGC AGTAGACTAT AAGCAGAAGC ACATATGAAC TGGACCTGCA CCAAAGCTGG CACCAGGGCT CGGAAGGTCT CTGAACTCAG AAGGATGGCA TTTTTTGCAA GTTAAAGAAA

ATCAGGATCT GAGTTATTTT GCTAAACTTG GGGGAGGAGG AACAAA
```

Fig. 6. (*Continued.*)

have arrived at the conclusion that the surface of Lp(a) is occupied by apoB, whereas apo(a), though covalently linked to apoB, would be loosely associated with the lipoprotein particles and instead project into the aqueous environment. The facts that apo(a) is heavily glycoslyated (27.9% carbohydrate) and that Lp(a) is relatively more viscous than LDL would support this interpretation. This structural concept has been exploited in devising a method for isolating apo(a) from Lp(a). Ultracentrifugation of Lp(a) in the presence of a reducing agent such as dithiothreitol results in the sedimentation of an essentially lipid-free apo(a) while the Lp(a-) remnant is recovered in the floating fraction (Fless *et al.*, 1985). The apo(a), so isolated, has an apparent molecular weight of 280,000, is distinct in amino acid composition from apoB, and has 71% random structure and a high carbohydrate content, 28% by weight (Fless *et al.*, 1986). Moreover, contrary to apoB, the apo(a) molecule enjoys solubility in water.

Currently nothing is known of the mode of biosynthesis and processing of Lp(a) and on the site of complexation of apoB with apo(a). Preliminary evidence indicates that the intestine is able to produce apo(a) (G. Fless, unpublished observations). It is unclear, however, whether it is secreted in a lipid-free or in a lipid-bound form.

B. cDNA Clone and Apo(a) Gene

No information on this issue is available in the literature. At the time of this writing, clones for apo(a) are being searched in HepG2 cell libraries using monospecific antibodies raised against pure apo(a).

C. Physiological Considerations

Relatively little is known about the physiological role of Lp(a). Since the lipid content of Lp(a) is so much like that of LDL, it is tempting to speculate that the function of these two lipoproteins is analogous, and it is mainly involved in cholesterol transport. However, apo(a) may target Lp(a) for entry into specialized tissues via specific binding sites on the plasma membrane. In this context, there is some controversy whether Lp(a) binds to the apoB,E receptor or to a totally distinct one (Fless and Scanu, 1986). The fractional catabolic rate of Lp(a) is about 30% lower than that of LDL (Krempler *et al.*, 1983). How much of this is contributed by synthesis and catabolism of this particle has not been established. None of the observations reported thus far explains why Lp(a) retains relatively steady plasma levels even after severe dietary manipulations and the action of pharmacological agents such as estrogens, clofibrate, and cholestyramine. The only exception is the anabolic steroid stanazolol, which has been shown to dramatically decrease the levels of plasma Lp(a) (Albers *et al.*, 1984). Overall, these studies appear to suggest that the metabolic control of Lp(a) is different from that of LDL and VLDL.

VII. ApoC-I

A. General Properties, Biosynthesis, and Processing

ApoC-I is a member of the C-peptides which, in plasma, are associated with the triglyceride-rich particles and the HDL class. For a review on the general properties of these peptides, the reader is referred to Scanu *et al.* (1982) and Mahley *et al.* (1984). ApoC-I has been shown to activate LCAT *in vitro* (Soutar *et al.*, 1975); however, whether this activation also takes place *in vivo* or plays a physiological role in lipoprotein metabolism has not been established. ApoC-I has been shown to be synthesized by both the liver and intestine as a preprotein (see below) but the factors regulating synthesis and secretion remain undetermined.

```
                                        27                                  54
CC CGC AGC TCA GCC ACG GCA CAG ATC AGC ACC ACG ACC CCT CCC TCG GGC CTC

                                        81                                 108
GCC ATG AGG CTC TTC CTG TCG CTC CCG GTC CTG GTG GTG GTT CTG TCG ATC GTC
    MET Arg Leu Phe Leu Ser Leu Pro Val Leu Val Val Val Leu Ser Ile Val
                           -20                                         -10

                                       135                                 162
TTG GAA GGC CCA GCC CCA GCC CAG GGG ACC CCA GAC GTC TCC AGT GCC TTG GAT
Leu Glu Gly Pro Ala Pro Ala Gln Gly Thr Pro Asp Val Ser Ser Ala Leu Asp
                                   -1  +1

                                       189                                 216
AAG CTG AAG GAG TTT GGA AAC ACA CTG GAG GAC AAG GCT CGG GAA CTC ATC AGC
Lys Leu Lys Glu Phe Gly Asn Thr Leu Glu Asp Lys Ala Arg Glu Leu Ile Ser
 10                                        20

                                       243                                 270
CGC ATC AAA CAG AGT GAA CTT TCT GCC AAG ATG CGG GAG TGG TTT TCA GAG ACA
Arg Ile Lys Gln Ser Glu Leu Ser Ala Lys MET Arg Glu Trp Phe Ser Glu Thr
         30                                        40

                                       297                                 324
TTT CAG AAA GTG AAG GAG AAA CTC AAG ATT GAC TCA TGA GGA CCT GAA GGG TGA
Phe Gln Lys Val Lys Glu Lys Leu Lys Ile Asp Ser END
                  50

                                       351                                 378
CAT CCA GGA GGG GCC TCT GAA ATT TCC CAC ACC CCA GCG CCT GTG CTG AGG ACT

                                       405                                 432
CCC GCC ATG TGG CCC CAG GTG CCA CCA ATA AAA ATC CTA CCG
```

Fig. 7. Nucleotide sequence of human apoC-I, cDNA. The derived amino acid sequence is also reported. Data from Knott *et al.* (1984b).

B. Properties of the cDNA Clone

Knott *et al.* (1984b) have reported the isolation of cDNA clones (Fig. 7) encoding human apoC-I from an adult liver cDNA library. By Northern blotting using an apoC-I DNA probe, the apoC-I mRNA was found to be represented by two species of about 580 and 560 bases. The apoC-I mRNA was also shown to contain nucleotides each 63 and 40 bp long in the 5′ untranslated region, a termination codon, TGA, and a 3′ untranslated region of 111 bp. From the cDNA sequence, the newly synthesized apolipoprotein was inferred to consist of a polypeptide 83 amino acids long, 26 of them representing the NH_2-terminal signal or prepeptide and the remainder the mature protein. This DNA-derived sequence is identical to that obtained by amino acid sequence analyses. In the liver cDNA library studied the abundance of the apoC-I cDNA clone was found to be about half that of the clones corresponding to apoA-I, apoA-II, and apoE. However, the levels of the apoC-I mRNA were severalfold higher than those of the mRNA of apoA-II and apoE. This observation was attributed to the potential loss of cDNA during the preparation of the library.

C. ApoC-I Gene

No information is available on the gene structure of apoC-I except for its location in chromosome 19 approximately 4 kb from the apoE gene (Tata *et al.*, 1985). No genetic polymorphism has been reported.

VIII. ApoC-II

A. General Properties, Biosynthesis, and Processing

ApoC-II (Scanu *et al.*, 1982; Mahley *et al.*, 1984) is associated with triglyceride-rich particles and HDL. This apolipoprotein has long been recognized as the specific activator of the enzyme lipoprotein lipase, which promotes the hydrolysis of the triglyceride chylomicrons and VLDL. A phospholipase A-I activity has also been described but its physiological relevance has not been clearly established. Sequence studies have shown that apoC-II is a single polypeptide chain, 78 amino acids long with no carbohydrates. Both liver and intestine are able to synthesize this apolipoprotein as a preapolipoprotein (see below).

B. Properties of the cDNA Clone

The studies of the cDNA sequence (Fig. 8) reported thus far (Sharpe *et al.*, 1984; Jackson *et al.*, 1984; Myklebost *et al.*, 1984) have shown that apoC-II has a 5′ untranslated region, yet unidentified, and a 450-bp translated sequence. By

```
                                                  27                                                  54
GAC ACT ATG GGC ACA CGA CTC CTC CCA GCT CTG TTT CTT GTC CTC CTG GTA TTG
        MET Gly Thr Arg Leu Leu Pro Ala Leu Phe Leu Val Leu Leu Val Leu
                  -20                                                -10

                                                  81                                                 108
GGA TTT GAG GTC CAG GGG ACC CAA CAG CCC CAG CAA GAT GAG ATG CCT AGC CCG
Gly Phe Glu Val Gln Gly Thr Gln Gln Pro Gln Gln Asp Glu MET Pro Ser Pro
                      -1  +1                                        10

                                                 135                                                 162
ACC TTC CTC ACC CAG GTG AAG GAA TCT CTC TCC AGT TAC TGG GAG TCA GCA AAG
Thr Phe Leu Thr Gln Val Lys Glu Ser Leu Ser Ser Tyr Trp Glu Ser Ala Lys
                           20                                                  30

                                                 189                                                 216
ACA GCC GCC CAG AAC CTG TAC GAG AAG ACA TAC CTG CCC GCT GTA GAT GAG AAA
Thr Ala Ala Gln Asn Leu Tyr Glu Lys Thr Tyr Leu Pro Ala Val Asp Glu Lys
                                    40

                                                 243                                                 270
CTC AGG GAC TTG TAC AGC AAA AGC ACA GCA GCC ATG AGC ACT TAC ACA GGC ATT
Leu Arg Asp Leu Tyr Ser Lys Ser Thr Ala Ala MET Ser Thr Tyr Thr Gly Ile
     50                                        60

                                                 297                                                 324
TTT ACT GAC CAA GTT CTT TCT GTG CTG AAG GGA GAG GAG TAA CAG CCA GAC CCC
Phe Thr Asp Gln Val Leu Ser Val Leu Lys Gly Glu Glu END
            70

                                                 351                                                 378
CCA TCA GTG GAC AAG GGG AGA GTC CCC TAC TCC CCT GAT CCC CCA GGT TCA GAC


                        α   β
                        ↓   ↓                    405                                                 432
TGA GCT CCC CCT TCC CAG TAG CTC TTG CAT CCT CCT CCC AAC TCT AGC CTG AAT


                                                 459
TCT TTT CAA TAA AAA ATA CAA TTC
```

Fig. 8. Nucleotide sequence of human apoC-II, cDNA. The derived amino acid sequence is also reported. Data from Sharpe *et al.* (1984), Fojo *et al.* (1984), and Jackson *et al.* (1984). α and β indicate insertions of C at both positions (Myklebost *et al.*, 1984).

Northern blot analyses of liver and intestine RNA, the apoC-II mRNA has been found to be 500 bp long; thus it would appear that the 5′ untranslated region is about 50 bp in length. The cDNA sequence also indicates that there is a region 303 bp long coding for 101 amino acids, a termination codon, TAA, and a 3′ untranslated region of 144 bp followed by a poly(A) tail. Of the coded amino acid sequence 22 residues represent the NH_2-terminal extension or prepeptide

and the remaining 79 residues the mature apoC-II. No propeptide has been identified. The DNA-derived sequence agrees with that reported by Hospattankar *et al.* (1984), but differs significantly from that obtained by Jackson *et al.* (1977). The basis for this discrepancy has not been established but could be of a technical nature (Breslow, 1985a). A striking homology has been noted between the NH_2-terminal region comprising residues -2 (Gln), -1 (Gln), +1 (Asp), and +2 (Glu) of apoA-I and that of apoC-II corresponding to residue +5 (Gln), +6 (Gln), +7 (Asp), and +8 (Glu). The Gln-Asp bond is present in both apolipoproteins but only the one in apoA-I is cleaved by the metal-dependent enzyme causing the conversion of proapoA-I to mature apoA-I. The failure of apoC-II to undergo cleavage by this enzyme is unexplained, however, it may relate to the conformation of apoC-II at the lipoprotein surface or to possible hindrance by lipids. Once the metal-dependent enzyme causing the cleavage at the Gln-Asp bond in apoA-I is purified, it would be of interest to test its activity against apoC-II in solution and bound to various lipid matrices.

C. ApoC-II Gene

The apoC-II gene has been recently isolated (Jackson *et al.*, 1984; Breslow, 1985). The data from sequence analyses have shown introns interrupting the codons specifying amino acids −3 and 50. The position of the first intron has not been clearly established but, based on the similarities with other apolipoprotein genes, it may be anticipated that in apoC-II this intron will be interrupting the 5′ noncoding region. A second intron separates the DNA sequence coding for the signal peptide from that coding for the mature protein. The third intron separates the DNA sequence coding for the NH_2-terminal region from that coding for the COOH-terminal portion. No detailed information on the organization of the exons is yet available. Southern blot analyses of human DNA after digestion with the restriction endonuclease *Taq*I have revealed the occurrence of a common polymorphism in the vicinity of the apoC-II gene having in normal individuals an allele frequency of 0.60 (Humphries *et al.*, 1984). The apoC-II gene has been located in chromosome 19 (Jackson *et al.*, 1984; Humphries *et al.*, 1984).

D. ApoC-II Polymorphism

Havel *et al.* (1979) have reported on the polymorphism of apoC-II from the plasma of three hypertriglyceridemic subjects. By the technique of electrofocusing, these patients exhibited two bands, one corresponding to the normal apoC-II isoprotein and the other to one charge unit more acidic. This charge shift was found to be due to a substitution of glutamine for lysine at residue 55. Since this more acidic mutant effectively activated lipoprotein lipase, it did not appear to be involved in the mechanism of hypertriglyceridemia.

IX. ApoC-III

A. General Properties, Biosynthesis, and Processing

ApoC-III (Scanu *et al.*, 1982; Mahley *et al.*, 1984) is a glycoprotein which by amino acid sequence analysis has been reported to represent a single-chain polypeptide 79 amino acids long. The carbohydrate moiety attached to threonine-74 contains 1 mole galactose, 1 mole galactosamine, and 1 or 2 moles sialic acid/mole of apolipoprotein. The nomenclature apoC-III-0, apoC-III-1, and apoC-III-2 refers to the number of sialic acids present in each apoC-III molecule. ApoC-III-1 is the most abundant of three (27%), followed by apoC-III-2 (22%) and apoC-III-0 (14%). Whether these three forms are different in functional terms is unclear. ApoC-III has been shown to inhibit the hepatic uptake of triglyceride-rich remnants presumably by interfering with the interaction of apoE present in these particles with the specific apoE receptor in the liver (Windler *et al.*, 1980; Shelburne *et al.*, 1980). The synthesis of apoC-III can occur in both the liver and the intestine. Experiments in cell-free systems using mRNA from rat liver and intestine have shown that the primary translation product of apoC-III contains a 20-amino acid-long NH_2-terminal extension that can be cotranslationally cleaved by a signal peptidase to yield a product having the same, NH_2-terminus as mature apoC-III (Blaufuss *et al.*, 1984).

B. Properties of the cDNA Clone

The cDNA sequence data (Fig. 9) reported by three laboratories (Sharpe *et al.*, 1984; Karathanasis *et al.*, 1983c, 1985) indicate that the mRNA is at least 507 bp in length including a 297-bp region coding for 99 amino acids, a termination codon, TGA, and a 3′ untranslated region of 187 bp followed by a poly(A) tail. The total size of the mRNA obtained from cDNA sequence data, 550 bp, is compatible with that obtained by Northern blotting analysis of human liver mRNA. The amino acid sequence of apoC-III derived from cDNA analyses has confirmed that the primary translation product of apoC-III has a 20-amino acid-long presegment. The reported cDNA sequence differs from the published amino acid sequence (Brewer *et al.*, 1974) at several sites: residue 32, Glu instead of Ser; residue 33, Ser instead of Gln; residue 37, Gln instead of Ala; and residue 39, Ala instead of Gln. Breslow (1985) has suggested that these differences may have originated from errors in the protein sequence data.

C. Gene Structure

The studies by Karathanasis *et al.* (1983c) in a λ phage system have shown that the apoC-III gene is about 2500 bp from the 3′ end of the apoA-I gene and

```
                                        27                                  54
GAG GCG GGC TGC TCC AGG AAC AGA GGT GCC ATG CAG CCC CGG GTA CTC CIT GTT
                                           MET Gln Pro Arg Val Leu Leu Val
                                          -20

                                        81                                 108
GTT GCC CTC CTG GCG CTC CTG GCC TCT GCC CGA GCT TCA GAG GCC GAG GAT GCC
Val Ala Leu Leu Ala Leu Leu Ala Ser Ala Arg Ala Ser Glu Ala Glu Asp Ala
       -10                                  -1  +1

                                    α
                                    ↓ 135                                  162
TCC CTT CTC AGC TTC ATG CAG GGT TAC ATG AAG CAC GCC ACC AAG ACC GCC AAG
Ser Leu Leu Ser Phe MET Gln Gly Tyr MET Lys His Ala Thr Lys Thr Ala Lys
            10                                  20

                                       189                                 216
GAT GCA CTG AGC AGC GTG CAG GAG TCC CAG GTG GCC CAG CAG GCC AGG GGC TGG
Asp Ala Leu Ser Ser Val Gln Glu Ser Gln Val Ala Gln Gln Ala Arg Gly Trp
                  30       *   *              *       *  40

                                       243                                 270
GTG ACC GAT GGC TTC AGT TCC CTG AAA GAC TAC TGG AGC ACC GTT AAG GAC AAG
Val Thr Asp Gly Phe Ser Ser Leu Lys Asp Tyr Trp Ser Thr Val Lys Asp Lys
                        50                                              60

                                       297                                 324
TTC TCT GAG TTC TGG GAT TTG GAC CCT GAG GTC AGA CCA ACT TCA GCC GTG GCT
Phe Ser Glu Phe Trp Asp Leu Asp Pro Glu Val Arg Pro Thr Ser Ala Val Ala
                                    70

                                                        α β
                                       351              ↓                  378
GCC TGA GAC CTC AAT ACC CCA AGT CCA CCT GCC TAT CCA TCC TGC GAG CTC CTT
Ala END

                                    α
                                    ↓ 405                                  432
GGG TCC TGC AAT CTC CAG GGC TGC CCC TGT AGG TTG CTT AAA AGG GAC AGT ATT

                                       459                                 486
CTC AGT GCT CTC CTA CCC CAC CTC ATG CCT GGC CCC CCT CCA GGC ATG CTG GCC

                                       513
TCC CAA TAA AGC TGG ACA AGA AGC TGC TAT GAG
```

Fig. 9. Nucleotide sequence of human apoC-III, cDNA. The derived amino acid sequence is also reported. Data from Sharpe *et al.* (1984) and Karathanasis *et al.* (1986). An asterisk indicates difference between nucleotide and amino acid sequences.

that these two genes are coded for by opposite DNA strands. The apoC-III gene is about 3000 bp long. Of the three introns, IVS-1 contains approximately 600 bp and occurs in the 5′ untranslated region between bases 13 and 14 upstream of the Met codon that initiates translation. IVS-2 is about 125 bp long and interrupts the codon specifying amino acid +2 in the prepeptide region. IVS-3 is about 1800 bp long and interrupts the codon specifying amino acid 40 of the mature protein. According to these findings, the second intron separates the prepeptide from the mature protein.

D. ApoC-III Polymorphism

The variants described thus far have been associated with abnormalities in plasma lipid levels. They will be discussed in Section XII.

X. ApoE

A. General Properties, Biosynthesis, and Processing

Mature apoE (Mahley *et al.*, 1984; Weisgraber, 1985) is a glycoprotein with a polypeptide chain 299 residues long containing different amounts of carbohydrates. In the normal plasma it is mostly associated with VLDL and in the lighter HDL fractions with HDL_1 and HDL_2. After cholesterol feeding apoE is also found in a lipoprotein called HDL_c, which is larger than HDL_2 and particularly rich in this apolipoprotein. By two-dimensional gel electrophoresis apoE consists of several isoproteins differing in size and/or charge as a consequence of amino acid mutations and sialic acid content (see Section X,D). According to the results obtained in Breslow's laboratory (Fig. 10) (Zannis *et al.*, 1981a, 1982a, 1984), apoE is synthesized and secreted in sialylated form and undergoes desialization in the circulation. For structural reasons yet undefined, apoE is recognized by the LDL (apoB/E) receptor present in hepatic and extrahepatic tissues. It is also a specific ligand for the hepatic apoE receptor involved in the uptake from plasma of chylomicron remnants. As reported by Innerarity *et al.* (1983, 1984) and Weisgraber *et al.* (1983b), in the mature apolipoprotein the receptor binding region is localized between residues 140 and 150 (Fig. 11A). Initially apoE synthesis was described in the liver and the intestine. However, it is now apparent that it may occur in several other tissues such as adrenal glands, ovaries, kidney, brain, and reticuloendothelial cells (Mahley *et al.*, 1984; Weisgraber, 1985; Breslow, 1985, 1986). In a cell-free system the primary translational product of apoE mRNA has been shown to be a preprotein having a presegment attached to the NH_2-terminus of the mature protein (Zannis *et al.*, 1984). This finding has been confirmed by cDNA sequence analyses (see below).

Fig. 10. Alleles of human apoE (Breslow, 1985). (Reproduced, with permission, from the *Annu. Rev. Biochem.* **54,** © 1985 by Annual Reviews Inc.)

B. Properties of the cDNA Clone

The initial cDNA sequence studies by Breslow *et al.* (1982, 1983) have indicated that apoE mRNA is 1163 bp long including a 5′ untranslated region of 67 bp, a region of 951 bp coding for 317 amino acids, a termination codon, TGA, and a 3′ untranslated region of 142 bp. By Northern blot analyses the size of human liver mRNA has been found to be 1150 bp. Initiation of translation is at methionine located in the region containing 18 amino acid residues upstream of the mature protein. This region represents the pre- or signal peptide which is cotranslationally cleaved by a microsomal membrane signal peptidase. The cDNA sequence analysis (Fig. 11A) has revealed that apoE contains eight tandem repeats 22 amino acids long between residues 62 to 237. These repeats are analogous to those observed in apoA-I and apoA-IV and have a similar amphiphilic α-helix character. ApoE exhibits a 51 to 75% homology within each repeat and 72% with the six human apoA-I DNA repeats coding for apoA-I in residues 99 to 230 (Breslow, 1985, 1986). The meaning of these homologies is not yet apparent (see Section XI). More recently McLean *et al.* (1984) reported the complete nucleotide sequence of three full-length cloned cDNAs corresponding to liver apoE mRNA. Two of these DNAs were found to correspond to the normal E3/3 genotype (see below), whereas the third one was an apoE variant. The nucleotide sequence for the normal genotype differed substantially from the cDNA sequence by Breslow *et al.* (1982, 1983) reported in Fig. 11A. The one reported by McLean *et al.* (1984) contained 1157 nucleotides of mRNA sequence and a 5′ terminal nontranslated region of 61 nucleotides, followed in

order by the region corresponding to the 18-amino acid signal peptide, by the mature protein 299 amino acids long, and the 3′-terminal nontranslated region of 142 nucleotides (Fig. 11B). The differences between the two reported sequences by the Breslow and Taylor laboratories are at the level of the coding region (28 nucleotide differences in the third base position of the codon) and in the 3′ nontranslated region (15 nucleotide differences plus an extra segment of 17 nucleotides following nucleotide 966 and a missing nucleotide following nucleotide 1036). The differences appear to be related to technical inaccuracies owing to the fact that the new cDNA sequence obtained by Breslow is similar to that reported by the Taylor group. The variant cDNA clone observed by McLean *et al.* (1984) was characterized by a threonine replacement of the normal alanine at residue 99 and a proline replacement of alanine at residue 152.

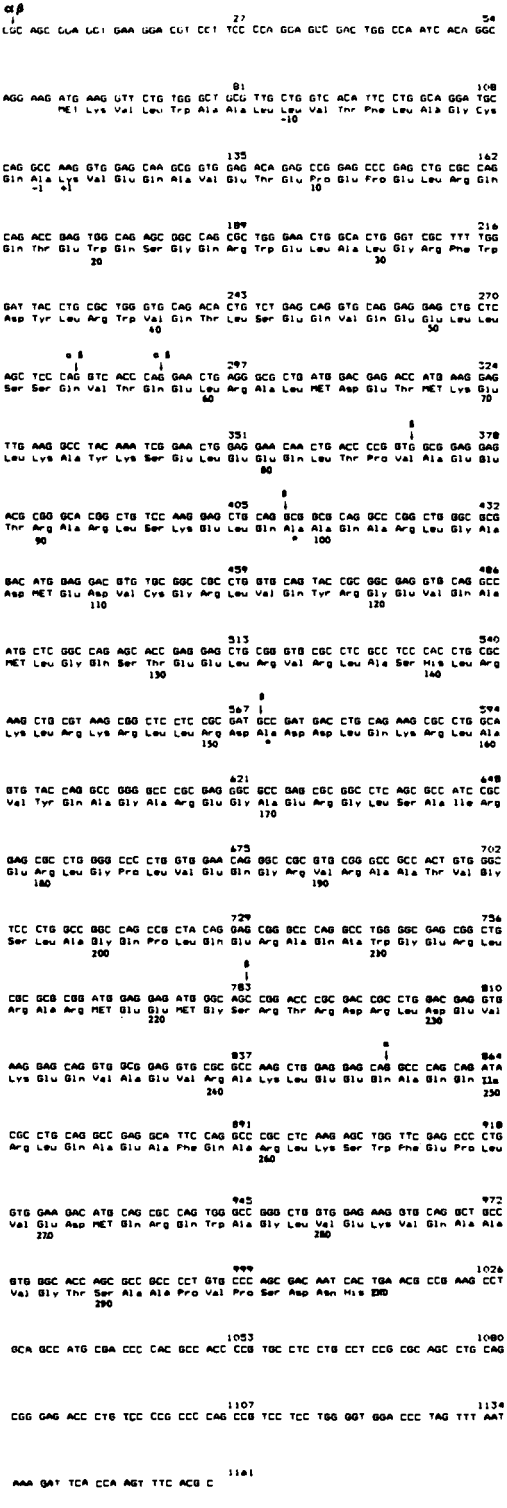
C. ApoE Gene

According to the results of the studies by Das *et al.* (1985), the apoE gene is 3.7 kb long and contains four exons and three introns. The first intron, IVS-1, about 700 bp long, is located in the 5′ untranslated region between bases 23 and 24 upstream of the initiation codon, Met. IVS-2, about 1100 bp long, interrupts the codon specifying amino acid −4 in the prepeptide region. IVS-3, about 600 bp long, interrupts the codon specifying amino acid 61 of the mature protein. The locations of these introns are strikingly similar to those observed in apoA-I, apoC-II, and apoC-III genes. The gene for apoE has been mapped to chromosome 19, where the genes of apoC-I, apoC-II, and apoB,E receptor have also been identified (Breslow, 1985; Lin-Lee *et al.*, 1985).

D. ApoE Polymorphism

The use of one-dimensional isoelectric focusing and two-dimensional gel electrophoresis has permitted a description of the apoE polymorphism. From population studies, six phenotypes have been identified (Zannis and Breslow, 1981; Zannis *et al.*, 1981b). Family studies have helped establish that these phenotypes are the results of a single apoE gene locus having three common alleles (Fig. 12). They have been designated as $\epsilon 4$, $\epsilon 3$, and $\epsilon 2$ and their gene products as E4, E3, and E2. Thus, there are three homozygous phenotypes, E4/4, E3/3, and E2/2, and three heterozygous phenotypes, E4/3, E3/2, and E4/2 (Zannis *et al.*, 1982b). In the normal population the highest frequency is that for the E3/3 phenotype (60%) followed by the E4/3 (20%) and the E3/2 (15%) phenotypes (Table III). The differences among alleles are based on a single amino acid substitution leading to a single charge difference (Fig. 12). E2 differs from E3 by having cysteine instead of arginine in position 158; in turn, E4 differs from E3 by having cysteine replaced by arginine at position 112 (Mahley *et al.*, 1984). The

A


αβ
1 27 54
CGC AGC GCA GCT GAA GGA CGT CCT TCC CCA GCA GCC GAC TGG CCA ATC ACA GGC
81 108
AGG AAG ATG AAG GTT CTG TGG GCT GCG TTG CTG GTC ACA TTC CTG GCA GGA TGC
MET Lys Val Leu Trp Ala Ala Leu Leu Val Thr Phe Leu Ala Gly Cys
-10
135 162
CAG GCC AAG GTG GAG CAA GCG GTG GAG ACA GAG CCG GAG CCC GAG CTG CGC CAG
Gln Ala Lys Val Glu Gln Ala Val Glu Thr Glu Pro Glu Pro Glu Leu Arg Gln
-1 +1 10
189 216
CAG ACC GAG TGG CAG AGC GGC CAG CGC TGG GAA CTG GCA CTG GGT CGC TTT TGG
Gln Thr Glu Trp Gln Ser Gly Gln Arg Trp Glu Leu Ala Leu Gly Arg Phe Trp
20 30
243 270
GAT TAC CTG CGC TGG GTG CAG ACA CTG TCT GAG CAG GTG CAG GAG GAG CTG CTC
Asp Tyr Leu Arg Trp Val Gln Thr Leu Ser Glu Gln Val Gln Glu Glu Leu Leu
40 50
α β α β
1 1 297 324
AGC TCC CAG GTC ACC CAG GAA CTG AGG GCG CTG ATG GAC GAG ACC ATG AAG GAG
Ser Ser Gln Val Thr Gln Glu Leu Arg Ala Leu MET Asp Glu Thr MET Lys Glu
60 70
β
1 351 378
TTG AAG GCC TAC AAA TCG GAA CTG GAG GAA CAA CTG ACC CCG GTG GCG GAG GAG
Leu Lys Ala Tyr Lys Ser Glu Leu Glu Glu Gln Leu Thr Pro Val Ala Glu Glu
80
β
1 405 432
ACG CGG GCA CGG CTG TCC AAG GAG CTG CAG GCG GCG CAG GCC CGG CTG GGC GCG
Thr Arg Ala Arg Leu Ser Lys Glu Leu Gln Ala Ala Gln Ala Arg Leu Gly Ala
90 * 100
459 486
GAC ATG GAG GAC GTG TGC GGC CGC CTG GTG CAG TAC CGC GGC GAG GTG CAG GCC
Asp MET Glu Asp Val Cys Gly Arg Leu Val Gln Tyr Arg Gly Glu Val Gln Ala
110 120
513 540
ATG CTC GGC CAG AGC ACC GAG GAG CTG CGG GTG CGC CTC GCC TCC CAC CTG CGC
MET Leu Gly Gln Ser Thr Glu Glu Leu Arg Val Arg Leu Ala Ser His Leu Arg
130 140
β
1 567 594
AAG CTG CGT AAG CGG CTC CTC CGC GAT GCC GAT GAC CTG CAG AAG CGC CTG GCA
Lys Leu Arg Lys Arg Leu Leu Arg Asp Ala Asp Asp Leu Gln Lys Arg Leu Ala
150 * 160
621 648
GTG TAC CAG GCC GGG GCC CGC GAG GGC GCC GAG CGC GGC CTC AGC GCC ATC CGC
Val Tyr Gln Ala Gly Ala Arg Glu Gly Ala Glu Arg Gly Leu Ser Ala Ile Arg
170
675 702
GAG CGC CTG GGG CCC CTG GTG GAA CAG GGC CGC GTG CGG GCC GCC ACT GTG GGC
Glu Arg Leu Gly Pro Leu Val Glu Gln Gly Arg Val Arg Ala Ala Thr Val Gly
180 190
729 756
TCC CTG GCC GGC CAG CCG CTA CAG GAG CGG GCC CAG GCC TGG GGC GAG CGG CTG
Ser Leu Ala Gly Gln Pro Leu Gln Glu Arg Ala Gln Ala Trp Gly Glu Arg Leu
200 210
β
1 783 810
CGC GCG CGG ATG GAG GAG ATG GGC AGC CGG ACC CGC GAC CGC CTG GAC GAG GTG
Arg Ala Arg MET Glu Glu MET Gly Ser Arg Thr Arg Asp Arg Leu Asp Glu Val
220 230
α
1 837 864
AAG GAG CAG GTG GCG GAG GTG CGC GCC AAG CTG GAG GAG CAG GCC CAG CAG ATA
Lys Glu Gln Val Ala Glu Val Arg Ala Lys Leu Glu Glu Gln Ala Gln Gln Ile
240 250
891 918
CGC CTG CAG GCC GAG GCA TTC CAG GCC CGC CTC AAG AGC TGG TTC GAG CCC CTG
Arg Leu Gln Ala Glu Ala Phe Gln Ala Arg Leu Lys Ser Trp Phe Glu Pro Leu
260
945 972
GTG GAA GAC ATG CAG CGC CAG TGG GCC GGG CTG GTG GAG AAG GTG CAG GCT GCC
Val Glu Asp MET Gln Arg Gln Trp Ala Gly Leu Val Glu Lys Val Gln Ala Ala
270 280
999 1026
GTG GGC ACC AGC GCC GCC CCT GTG CCC AGC GAC AAT CAC TGA ACG CCG AAG CCT
Val Gly Thr Ser Ala Ala Pro Val Pro Ser Asp Asn His END
290
1053 1080
GCA GCC ATG CGA CCC CAC GCC ACC CCG TGC CTC CTG CCT CCG CGC AGC CTG CAG
1107 1134
CGG GAG ACC CTG TCC CCG CCC CAG CCG TCC TCC TGG GGT GGA CCC TAG TTT AAT
1161
AAA GAT TCA CCA AGT TTC ACG C

B

```
                                                                                                                  Met Lys Val Leu Trp Ala Ala Leu Leu Val Thr Phe Leu
CCCCA GCGGAGGTGA AGGACGTCCT TCCCCAGGAG CCGACTGGCC AATCACAGGC AGGAAG ATG AAG GTT CTG TGG GCT GCG TTG CTG GTC ACA TTC CTG
                                -90                                                                -60                -30
                                 *
Ala Gly Cys Gln Ala Lys Val Glu Gln Ala Val Glu Thr Glu Pro Glu Pro Glu Leu Arg Gln Gln Thr Glu Trp Gln Ser Gly Gln Arg
GCA GGA TGC CAG GCC AAG GTG GAG CAA GCG GTG GAG ACA GAG CCG GAG CCC GAG CTG CGC CAG CAG ACC GAG TGG CAG AGC GGC CAG CGC
                           1                                                                  30                  60

Trp Glu Leu Ala Leu Gly Arg Phe Trp Asp Tyr Leu Arg Trp Val Gln Thr Leu Ser Glu Gln Val Gln Glu Glu Leu Leu Ser Ser Gln
TGG GAA CTG GCA CTG GGT CGC TTT TGG GAT TAC CTG CGC TGG GTG CAG ACA CTG TCT GAG CAG GTG CAG GAG GAG CTG CTC AGC TCC CAA
                          90                                                                 120                 150

Val Thr Gln Glu Leu Arg Ala Leu Met Asp Glu Thr Met Lys Glu Leu Lys Ala Tyr Lys Ser Glu Leu Glu Glu Gln Leu Thr Pro Val
GTC ACC CAA GAA CTG AGG GCG CTG ATG GAC GAG ACC ATG AAG GAG TTG AAG GCC TAC AAA TCG GAA CTG GAG GAA CAA CTG ACC CCG GTA
                         180                                                                210                 240

Ala Glu Glu Thr Arg Ala Arg Leu Ser Lys Glu Leu Gln Thr Ala Gln Ala Arg Leu Gly Ala Asp Met Glu Asp Val Cys Gly Arg Leu
GCG GAG GAG ACG CGG GCA CGG CTG TCC AAG GAG CTG CAG ACG GCG CAG GCC CGG CTG GGC GCG GAC ATG GAG GAC GTG TGC GGC CGC CTG
                         270                                                                300                 330

Val Gln Tyr Arg Gly Glu Val Gln Ala Met Leu Gly Gln Ser Thr Glu Glu Leu Arg Val Arg Leu Ala Ser His Leu Arg Lys Leu Arg
GTG CAG TAC CGC GGC GAG GTG CAG GCC ATG CTC GGC CAG AGC ACC GAG GAG CTG CGG GTG CGC CTC GCC TCC CAC CTG CGC AAG CTG CGT
                         360                                                                390                 420

Lys Arg Leu Leu Arg Asp Pro Asp Asp Leu Gln Lys Arg Leu Ala Val Tyr Gln Ala Gly Ala Arg Glu Gly Ala Glu Arg Gly Leu Ser
AAG CGG CTC CTC CGC GAT CCC GAT GAC CTG CAG AAG CGC CTG GCA GTG TAC CAG GCC GGG GCC CGC GAG GGC GCC GAG CGC GGC CTC AGC
                         450                                                                480                 510

Ala Ile Arg Glu Arg Leu Gly Pro Leu Val Glu Gln Gly Arg Val Arg Ala Ala Thr Val Gly Ser Leu Ala Gly Gln Pro Leu Gln Glu
GCC ATC CGC GAG CGC CTG GGG CCC CTG GTG GAA CAG GGC CGC GTG CGG GCC GCC ACT GTG GGC TCC CTG GCC GGC CAG CCG CTA CAG GAG
                         540                                                                570                 600

Arg Ala Gln Ala Trp Gly Glu Arg Leu Arg Ala Arg Met Glu Glu Met Gly Ser Arg Thr Arg Asp Arg Leu Asp Glu Val Lys Glu Gln
CGG GCC CAG GCC TGG GGC GAG CGG CTG CGC GCG CGG ATG GAG GAG ATG GGC AGT CGG ACC CGC GAC CGC CTG GAC GAG GTG AAG GAG CAG
                         630                                                                660                 690

Val Ala Glu Val Arg Ala Lys Leu Glu Glu Gln Ala Gln Gln Ile Arg Leu Gln Ala Glu Ala Phe Gln Ala Arg Leu Lys Ser Trp Phe
GTG GCG GAG GTG CGC GCC AAG CTG GAG GAG CAG GCC CAG CAG ATA CGC CTG CAG GCC GAG GCC TTC CAG GCC CGC CTC AAG AGC TGG TTC
                         720                                                                750                 780

Glu Pro Leu Val Glu Asp Met Gln Arg Gln Trp Ala Gly Leu Val Glu Lys Val Gln Ala Ala Val Gly Thr Ser Ala Ala Pro Val Pro
GAG CCC CTG GTG GAA GAC ATG CAG CGC CAG TGG GCC GGG CTG GTG GAG AAG GTG CAG GCT GCC GTG GGC ACC AGC GCC GCC CCT GTG CCC
                         810                                                                840                 870

Ser Asp Asn His ***
AGC GAC AAT CAC TGA ACGCCGAAGC CTGCAGCCAT GCGACCCCAC GCCACCCCGT GCCTCCTGCC TCCGCGCAGC CTGCAGCGGG AGACCCTGTC CCCGCCCCAG
                         900                                930                                960                 990

CCGTCCTCCT GGGGTGGACC CTAGTTTAAT AAAGATTCAC CAAGTTTCAC GC - Poly(A)
                              1020
```

Fig. 11. (A) Nucleotide sequence of human apoE, cDNA. The amino acid sequence is also reported. α and β indicate differences between the sequences reported by Zannis *et al.* (1984) and McLean *et al.* (1984). An asterisk indicates amino acid substitutions in the variant apoE3 clone reported by McLean *et al.* (1984). (B) Nucleotide sequence of human apoE, cDNA, according to McLean *et al.* (1984).

TABLE III

Prevalence of ApoE Phenotype in Human Subjects[a]

Classification	Prevalence (%)				
	Utermann	Assmann	Breslow	Wardell	Cumming
Phenotype					
E4/4	2.8	2.2	3.0	1.0	1.0
E3/3	59.8	62.2	58.0	51.4	58.3
E2/2	1.0	0.9	1.3	1.4	0.5
E4/3	22.9	19.9	14.0	25.0	24.8
E4/2	1.5	2.9	2.0	1.2	2.8
E3/2	12.0	11.7	22.0	20.0	12.8
No. of subjects	1031	1557	152	426	400
Allele					
ε4	15	14	11	14	15
ε3	77	78	76	74	77
ε2	8	8	13	12	8

[a]From Breslow (1986).

corresponding allele variation in the human apoE gene is shown in Table IV. Besides these alleles, others occurring with a low frequency have been identified. The most significant of these mutations are those occurring in the receptor binding region leading to a functional defect in ligand interaction with the hepatic receptor in the liver (see Fig. 13).

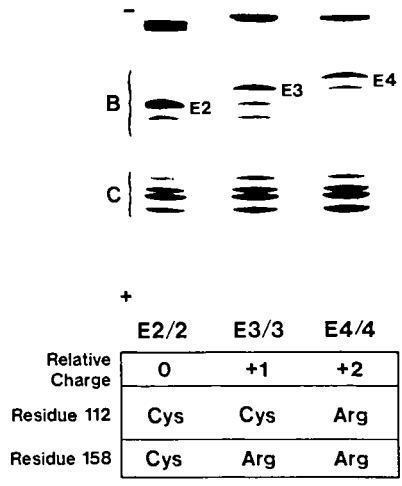


Fig. 12. Isoelectric focusing patterns of E2/E2, E3/E3, and E4/E4. The amino acid substitutions at site A (residue 112) and site B (residue 158) are also presented (Mahley, 1983).

TABLE IV

Allelic Variations in Human ApoE Gene[a]

Allele	Codon at the polymorphic site					Expected base substitution relative to ε3 allele		Amino acid substitution relative to E3/3 phenotype		
	112	127	145	146	158					
ε3	TGC	GGC	CGT	AAG	CGC	None		None		
ε4	CGC	Same	Same	Same	Same	T	C	112	Cys	Arg
ε2	Same	Same	Same	Same	TGC	C	T	158	Arg	Cys
ε2*	Same	Same	TGT	Same	Same	C	T	145	Arg	Cys
ε2**	Same	Same	Same	CAG	Same	A	C	146	Lys	Gln
ε1	Same	GAC	Same	Same	TGC	G	A	127	Gly	Asp
						C	T	158	Arg	Cys

[a]From Karathanasis *et al.* (1985).

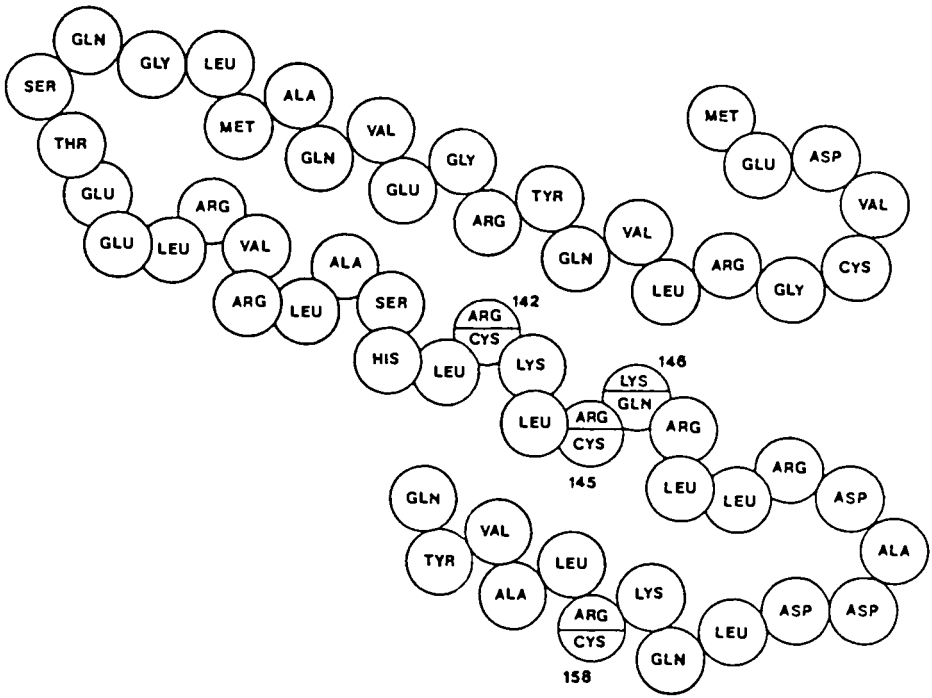


Fig. 13. Region of the apoE sequence involved in acceptor binding. Four amino acid substitution sites are indicated, residues 142, 145, 146, and 158 (Mahley, 1983).

XI. Phylogenetic Relationships among Plasma Apolipoproteins

The studies reviewed in the previous sections have shown that there are many structural homologies among the plasma apolipoproteins studied, namely, apoA-I, apoA-IV, apoC-I, apoC-II, apoC-III, and apoE. Close similarities between the apoA-I and apoA-II genes have also been reported by Lackner *et al.* (1985a). The information on apoB is still too limited to permit determination of the degree of homology with the other proteins. These homologies are expressed in terms of repeating sequences of α-helix amphiphilic segments and by the intron–exon organization. Moreover, we have seen that the apoA-I, apoC-III, and apoA-IV genes are located in chromosome 11 and in close proximity to each other. In addition, the genes of apoC-I, apoC-II, and apoE have been localized in chromosome 19, which is also where the gene of the apoB,E receptor has been reported to be. Finally, the apoA-II gene has been identified in chromosome 1 and the apoB gene in chromosome 2. The recognition that structural homologies exist among many apolipoproteins has led to consideration of a common evolutionary pattern and the possibility that some of the apolipoproteins are members of multigene families. In this context Karathanasis (1985) has shown that in chromosome 11 the genes of apoA-I and apoC-III are only 30 kb apart and that the apoA-IV gene is also closely linked to the apoA-I gene (12 kb distance). By restriction mapping of human chromosomal DNA this author has shown that the apoA-I, apoC-III, and apoA-IV genes are linked and tandemly organized in a 17-kb *Hin*dIII DNA fragment. The apoA-I and apoA-IV genes would both be transcribed in a direction opposite to that of the apoC-III gene. Karathanasis (1985) has taken his findings and previous ones in the literature to suggest that the apoA-I, apoC-III, and apoA-IV genes are derived from a common ancestral precursor. The fact that the apoA-I and apoA-IV genes have 66-bp repeats which are highly homologous to each other (consensus sequence 50%) has led this author to suggest that these two genes arose from intergenic duplication of the apoA-I/A-IV gene precursor that evolved by intragenic amplification of an ancestral 66-bp sequence. The apoC-III gene would also be derived from the same apoC-III/A-I/A-IV ancestor gene, explaining the absence of the 66-bp repeats in the human apoC-III gene either by a divergence mechanism from this ancestral gene or by the loss of the periodic nucleotide sequence during its more recent evolutionary phase. In extending this kind of reasoning to the apoE gene, Karathanasis (1985), based on the extensive homologies of consensus sequences between the apoA-I and those of the apoE and apoA-IV genes, has suggested that the apoE gene has derived from a sequence in evolutionary terms related to the apoA-I/apoA-IV gene precursor even if ultimately these genes are located in different chromosomes, i.e., apoA-I/A-IV genes in chromosome 11 and the apoE gene in chromosome 19.

Boguski *et al.* (1985, 1986) have also carried out extensive studies on the possible relationship between structural homologies among plasma apolipoproteins and their evolution. In particular they have attempted to understand the structural, functional, and evolutionary relationships among the main protein components of rat plasma HDL, apoA-I, apoA-IV, and apoE by taking advantage of the knowledge of their complete nucleotide and amino acid sequence and appropriate computer programs. As a first approach the authors examined the issue of orthologous versus paralogous categories.* In the orthologous category they compared the sequence of human and rat apoA-I, the latter determined in the authors' laboratory. Rat apoA-I was found to differ from human apoA-I by being eight amino acids shorter, by the absence of proline at residue 28, by the absence of glutamine, which in human apoA-I is located at the COOH-terminus, and by the absence of four residues (glutamine, lysine, leucine, histidine) at positions 155 to 159 and of two glycines at position 203–204. In spite of these substitutions, rat apoA-I was found to contain eight repeated amphiphilic sequences each 22 amino acids long and with α-helical conformation analogous to that of human apoA-I (overall sequence homology 64%).

In terms of *paralogous* sequences, Boguski *et al.* (1985, 1986) used the comparison matrix method, which computes matching probability scores for spans of amino acids based on observed frequencies of amino acid replacements in homologous proteins. According to these authors apoA-I, apoA-IV, and apoB are paralogous members of a dispersed gene family which all contain sequences originating from a common ancestral gene. Subsequent divergent evolution would lead to the separation of the three genes. However, those of apoA-I and apoA-IV would diverge at a rate lower than that of the apoE gene. Boguski *et al.* also have suggested that in spite of the overall similarity in sequence organization and shared functions between apoA-I, apoA-IV, and apoE, different rates of evolution among segments within each gene can account for the distinct functions exhibited by each apolipoprotein. In more recent studies these authors expanded their early concepts and have proposed that an ancestral minigene might have generated the multigene family comprising the main plasma apolipoproteins. This would have occurred by the duplication of this ancestral minigene resulting in the formation of two exons. Subsequent steps of the evolution would have been dominated by intraexonic amplification of the repeating units coding for amphiphilic peptides followed by sequence divergence of these repeats and attending functional differences among apolipoproteins. In the case of apoA-I, apoC-II, apoC-III, and apoA-II this would be achieved by duplication

*Orthologous sequences reflect the phylogenetic branching order of the species in which they are found and have identical function. On the other hand, paralogous sequences are the product of a gene duplication that was fixed before speciation. Divergence of such sequences usually results in the evolution of new functions.

and divergence; for apoA-I, apoA-IV, and apoE by multiple unequal crossovers of one of the exons (exon D) followed by duplication and divergence. The information derived from this kind of computer-based analysis provides an explanation of the evolutionary paths leading to shared and distinct functional expressions of the various plasma apolipoproteins derived from a common ancestral gene, yet it fails to provide the detailed mechanism by which these processes have occurred. Because of the current modest information on the apoB gene it is still undetermined how a protein as large as apoB fits into the proposed evolutionary scheme. We may also surmise that the conservation of the repeated amphiphilic sequences in each plasma apolipoprotein may be secondary to a common structural requirement, namely, that of interacting with lipids while retaining solubility in aqueous media.

XII. Apolipoprotein Variants Associated with Pathological States

Gene mutations have been recognized in association with lipoprotein abnormalities and other pathological states. Examples of the better characterized variants are described below.

A. ApoA-I Variants

The variants (Mahley *et al.*, 1984; Breslow, 1985) reported thus far are listed in Table II; they differ in one + or − charge unit and are characterized by either a single amino acid substitution or a deletion of one amino acid. In terms of attending pathology, subjects with apoA-I Marburg characterized by a deletion of lysine 107 (Rall *et al.*, 1984) have mild hypertriglyceridemia and low HDL levels. An interesting mutant is represented by apoA-I Milano, in which arginine is replaced by cysteine at position 173 and in consequence both apoA-I homodimers and apoA-I–apoA-II heterodimers are formed (Weisgraber *et al.*, 1983a). Thus far, all of the subjects with this variant have originated from a single couple. The 33 identified carriers are heterozygotes and conform with an autosomal dominant trait. The biochemical abnormalities are represented by a marked reduction of the plasma levels of HDL cholesterol, apoA-I, and apoA-II, by an elevation of the plasma triglycerides with an increase in their concentrations in LDL and HDL, by the nearly total absence of HDL_2, and by the presence of polydisperse HDL_3. According to Sirtori and Franceschini (1985), subjects with this variant would have a reduced frequency of cardiovascular disease. However, this interesting suggestion needs to be documented further.

Another abnormality of apoA-I characterized by an autosomal recessive form of inheritance is Tangier disease (Mahley *et al.*, 1984; Schaefer, 1984). Homo-

zygotes have an absence of normal plasma HDL and a marked reduction in total apoA-I, which is made up of about equal amounts of proapoA-I and mature apoA-I (Zannis *et al.*, 1982b). The relative abundance of proapoA-I was initially thought to be associated with a defect of the metal-dependent converting enzyme and thus a decrease in the conversion of proapoA-I to mature apoA-I (Breslow, 1985). However, direct experiments have ruled out this early suggestion (Edelstein *et al.*, 1984). The prosegment of proapoA-I Tangier has been shown to have the same amino acid sequence as the normal counterpart (Brewer *et al.*, 1983). Moreover, structural abnormalities of mature apoA-I Tangier have not been clearly established although an anomaly in lipid binding has been reported (Rosseneu *et al.*, 1984). Thus far, the study of the apoA-I Tangier clone has shown no abnormalities at least in the cases examined (Breslow, 1985). It has been observed that if apoA-I is injected intravenously into Tangier subjects it is catabolized more rapidly than by normal subjects (Schaefer, 1984). This can account for both the abnormal proapoA-I : mature apoA-I ratio and also for the low levels of plasma apoA-I in these patients. Unrecognized factors such as increased proteolytic activity by Tangier plasma or the occurrence of membrane receptor abnormalities for apoA-I in these patients could be responsible for the observed abnormalities. Thus, although Tangier disease is commonly listed as an apoA-I variant it is possible that the primary defect does not reside in this apolipoprotein. Still unexplained is the tissue cholesteryl ester deposition responsible for the enlarged yellow tonsils, hepatosplenomegaly, corneal opacity, and perhaps the peripheral neuropathy seen in Tangier patients. It is worth noting that the plasma levels of apoA-II are also low although this protein has been reported to have a normal amino acid composition (Schaefer, 1984).

Another apolipoprotein variant is that associated with a familial deficiency of apoA-I and apoC-III. This disorder is characterized by the total absence in plasma of these two apolipoproteins and by a marked reduction in HDL and apoA-II but normal LDL levels. It was first observed in two sisters, 31 and 33 years old, presenting with xanthomas and premature atherosclerosis (Norum *et al.*, 1982). Their first-degree relatives had plasma levels of HDL, apoA-I, and apoC-III half that of normal. The DNA extracted from the blood lymphocytes of these patients was digested with the restriction enzyme *Eco*RI and then subjected to Southern blotting following hybridization with an apoA-I cDNA probe. A 6.5-kb band was observed by gel electrophoresis (Karathanasis *et al.*, 1983b). In turn, the DNA from normal subjects when examined by the same procedure revealed an additional band 13 kb in size. The first-degree relatives, inclusive of mother and father of the probands, exhibited on the same gels both the normal and the abnormal bands. These findings were taken to indicate that a mutant allele was associated with the apoA-I gene and that the probands were homozygous for this mutant allele. More extensive analyses of the DNA of the homozygotes using several restriction enzymes and apoA-I and apoC-III cDNA probes

have led to the conclusion that the fourth exon of the apoA-I gene is interrupted at approximately the codon specifying residue 80 of the mature protein by sequences corresponding to the apoC-III gene found to be approximately 2.6 kb downstream of the 3′ end of the apoA-I gene (Breslow, 1985, 1986).

Two other disorders characterized by very low levels of plasma HDL (values of HDL cholesterol apoA-I and apoA-II 10–20% of normal) in the homozygous state are Fish-eye disease, a name derived from the severe corneal opacification present in the affected subjects, and HDL deficiency with planar xanthomas in which the affected subjects exhibit a diffuse discoloration of the skin in various parts of the body because of intracellular histocytic deposition of free and esterified cholesterol (Schaefer, 1984). In neither of these two conditions, however, has the molecular defect been established.

B. ApoA-B Variants

Inherited disorders associated with abnormal levels of apoB (Scanu *et al.*, 1975, 1982; Kane, 1983; Breslow, 1985, 1986; Sirtori and Franceschini, 1985) have been recognized although their actual genetic determination has been hampered by our lack of knowledge of the apoB structure and the characteristics of its DNA clone. Since, as discussed in Section V, an important breakthrough in this area has been accomplished, the genetic abnormality of the disorders outlined below is likely to be clarified in the near future.

A. ABETALIPOPROTEINEMIA. The patients affected by this disorder have in their plasma neither apoB nor apoB-containing lipoproteins, i.e., chylomicrons, VLDL, and LDL. This disease is believed to be secondary to a defect in apoB synthesis and to have an autosomal recessive mode of inheritance.

B. HOMOZYGOUS HYPOBETALIPOPROTEINEMIA. Patients with this disorder appear to be phenotypically related to abetalipoproteinemia in terms of a defect in apoB synthesis. Their plasma levels of LDL and apoB are about half of the normal, and the patients may be free of clinical symptoms except for some cases in which neurological manifestation of the type seen in Friedreich ataxia may be present.

C. NORMOTRIGLYCERIDEMIC ABETALIPOPROTEINEMIA. Patients with this rare disorder have normal fat absorption and produce chylomicrons but have either low or absent LDL cholesterol and apoB. According to Malloy *et al.* (1981), who first identified this disorder, these patients are capable of normally producing apoB-48, the intestinal form of apoB, but not apoB-100, the form produced by the liver. Although this clinical abnormality can be viewed as

supporting the concept of two distinct genes for the two forms of apoB, it is also possible that they might be produced by the same gene or represent posttranscriptional variants due to differential gene splicing (Breslow, 1985).

D. FAMILIAL HYPERCHOLESTEROLEMIA. This disorder, which recognizes an autosomal dominant mode of inheritance, is due to a total (homozygote) or partial (heterozygote) defect in the apoB,E receptor. However, in some cases an overproduction of apoB may also be present. Whether this reflects a simple compensatory process or has a genetic determination remains to be established. Of interest in terms of apoB overproduction is the phenotype designated hyperapobetalipoproteinemia (Sniderman *et al.*, 1980). Patients with this disorder have elevated plasma levels of apoB not associated with a comparable increase in LDL cholesterol. This disorder, considered to have an autosomal dominant mode of inheritance, is associated with premature atherosclerosis. Establishing the nature of the genetic and biochemical abnormalities in these subjects is of obvious importance.

C. ApoC Variants

Several kindreds with inherited deficiency of apoC-II have been reported in the literature (Breckenridge *et al.*, 1978, 1982). The homozygous subjects have severe hypertriglyceridemia, absence of apoC-II in either whole plasma or isolated VLDL, and essentially normal lipoprotein lipase. This hypertriglyceridemia can be temporarily reduced by infusion of normal plasma or apoC-II. The heterozygous subjects have reduced absolute amounts of apoC-III and moderate but significant elevation of the very low density lipoproteins. The disease is transmitted as an autosomal recessive trait. The molecular basis for the apoC-II deficiency has not been established. A severe hypertriglyceridemia can also occur in patients with VLDL particles carrying an excess of apoC-II (Stocks *et al.*, 1981). These variant lipoprotein species can activate lipoprotein lipase but do not represent a suitable substrate for the enzymes for reasons which are yet unknown.

In the case of apoC-III, genetic variants have been described associated with hypertriglyceridemia. This may be related to a DNA polymorphism associated with a single base substitution in the 3′ untranslated region of the neighboring apoC-III gene (Karathanasis *et al.*, 1983a). Since variant lipoproteins carrying an excess of apoC-III-2 have been detected in patients with severe hypertriglyceridemia and glucose intolerance (Stocks *et al.*, 1979), it would be important to establish whether the elevation of this apolipoprotein in plasma is associated with a DNA base substitution in the 3′ untranslated region. The genetic abnormality relative to an apoA-I/apoC-III rearrangement has been described in a previous section.

D. ApoE Variants

An apoE polymorphism (Mahley, 1983; Mahley *et al.*, 1984; Breslow, 1985, 1986) appears to be a factor in familial type III hyperlipoproteinemia, a disorder characterized by an elevation of both plasma cholesterol and triglycerides and triglyceride-rich remnant particles and a high incidence of peripheral heart disease. Over 90% of patients with type III have the apoE2/2 phenotype characterized by a defective binding to the specific hepatic apoE receptor. In consequence, the triglyceride-rich remnants carrying this apoE phenotype are cleared more slowly from the plasma. However, homozygosity for the $\epsilon 2$ allele is by itself sufficient for the hypertriglyceridemic expression. Additional factors leading to an overproduction of chylomicron remnants appear to be required. Besides the E2/E2 phenotype, other apoE mutants exhibiting a defective receptor binding activity have been identified. These apoE mutations vary widely in their capacity to interact with the apoE receptor and are associated with varying degrees of hyperlipidemia (Table V). Among these mutants the one designated E2* (Rall *et al.*, 1982) exhibits a substitution of cysteine for arginine in position 145, and another designated E1 is characterized by a $Gly^{127} \rightarrow$ Asp and an $Arg^{158} \rightarrow$ Cys substitution and it is associated with hyperlipidemia and atherosclerosis (Weisgraber *et al.*, 1984). Other alleles with substitutions in the apoE receptor binding region have been classified as E2** ($Lys^{146} \rightarrow$ Gln) (Rall *et al.*, 1983), E3* ($Ala^{99} \rightarrow$ Thr, $Ala^{152} \rightarrow$ Pro), and E3** ($Cys^{112} \rightarrow$ Arg, $Arg^{142} \rightarrow$ Cys) (Breslow, 1985). Additional ones will likely be identified with the systematic

TABLE V

Polymorphism of Human ApoE[a]

Variant	Charge relative to apoE3	Mutation site	Receptor binding activity relative to apoE3 (%)
E1	−2	$Gly^{127} \rightarrow$Asp, $Arg^{158} \rightarrow$Cys	4
E2**	−1	$Lys^{146} \rightarrow$Gln	40
E2*	−1	$Arg^{145} \rightarrow$Cys	45
E2	−1	$Arg^{158} \rightarrow$Cys	<2
E3**	0	$Cys^{112} \rightarrow$Arg, $Arg^{142} \rightarrow$Cys	<20
E3*	0	$Ala^{99} \rightarrow$Thr, $Ala^{152} \rightarrow$Pro	ND[b]
E3	0	—	100
E4	+1	$Cys^{112} \rightarrow$Arg	100
E5	+2	?	

[a]Adapted from Mahley *et al.* (1984).
[b]Not determined.

TABLE VI

Genetic Variant of Human ApoE[a]

Clone	Base pair							Gene product
	9	376	416	455	575	790	865	
pE-368	G	G	G	T	G	C	G	E3
λ apoE #1	C	G	G	C	G	C	G	E4
pHAE-112	C	G	G	T	G	C	A	E3
pHAE-178	C	G	G	T	G	C	A	E3
pHAE-813	C	A	A	T	C	T	G	E3*

[a]Taken from Breslow (1985). (Reproduced, with permission, from the *Annu. Rev. Biochem.* **54,** © 1985 by Annual Reviews Inc.)

application of protein and nucleotide sequence techniques. One genomic and four cDNA clones have been sequenced and the following variants have been identified (Breslow, 1985); cDNA clone pE-368, bp 9; cDNA clones pHAE-112 and pHEA-178, bp 865; and cDNA clone pHAE-813, bp 4, 376, 575, and 790 (see Table VI). This latter gene accounts for allele E3* and the other cDNA clone for variant E3. A genomic clone, λ apoE #1, has one variant site at bp 455 and appears to specify the E4 phenotype. It is of interest that a single point mutation in the region of apoE involved in receptor binding can lead to pathology and in particular to hyperlipidemia and atherosclerosis. In this regard, it has been suggested that the apoE gene locus may be involved in regulating the overall lipid levels in the general population. Evidence has also been provided that the ε2 allele has a stepwise gene dosage effect on lowering the plasma levels of LDL cholesterol and at the same time, in elevating VLDL cholesterol and triglycerides (Breslow, 1985). Moreover, allotype E2 appears to be more frequent in hypertriglyceridemia allotype E4 more frequent in hypercholesterolemia, and both E2 and E4 more frequent in mixed hyperlipidemia. These preliminary findings need corroboration from studies on a large-scale population. It is also important to recognize that other factors such as diet, exercise, age, sex, hormones, and drugs can influence lipid levels and may affect the expression of the apoE gene. The interplay with other genes, for instance, apoB or apoB,E receptor, should also be examined.

XIII. Concluding Remarks

If we look at the progress made during the last 4–5 years, it is apparent that gene cloning and recombinant DNA methods are rapidly generating new insights into the role(s) that apolipoproteins play in lipoprotein structure and metabolism.

Such information is also providing important means for studying the molecular basis of lipoprotein disorders and of their effect on the atherosclerotic process. The relationship between gene polymorphism and hyperlipidemias is beginning to emerge and studies on the factors controlling gene expression are also being investigated. On the structural level we now have the means of producing apolipoproteins of given structural specifications by site-specific mutagenesis and the products so derived should effectively aid in the study of naturally occurring human variants. It is now clear that even single point mutations can affect the functional expression of an apolipoprotein molecule as exemplified by some of the apoE mutations. Very recently the recombinant DNA technology has also provided the long-awaited breakthrough in the area of apoB. For a long time, this apolipoprotein had defied the efforts by many investigators attempting to unravel its structure by classical biochemical methods. The availability of cDNA clones of this apolipoprotein has now paved the way for the study of its primary structure and of its functional properties. The techniques of modern biology are also being applied to the analysis of "minor" apolipoproteins such as apoD and apoH, whose structural and functional properties are essentially unknown. Although not dealt with in this review, important parallel advances are also being made in the areas relating to lipid-modifying enzymes, apolipoprotein receptors, and lipid carrier proteins, all involved in lipid metabolism. Thus, a favorable setting has been reached to begin asking questions about the interrelationships between genetic and environmental factors and about the role they play in either maintaining physiological levels of plasma lipid and lipoprotein or in the pathogenesis of dyslipoproteinemias. Underlying these advances is the awareness that many lipid abnormalities are in fact primary protein disorders whether affecting apolipoproteins, lipid-modifying enzymes, receptors, or carrier proteins. The acquisition of additional knowledge of the factors that control the production and degradation of these proteins should provide better means for understanding the nature of lipoprotein disorders and their management.

Acknowledgments

The author gratefully acknowledges the collaboration of many laboratory colleagues who provided useful comments while the manuscript was being prepared, and in particular R. Byrne, C. Edelstein, G. Fless, and D. Polacek. He also wishes to thank Drs. J. Taylor (The Gladstone Foundation, San Francisco, California), J. Gordon, and M. Boguski (Washington University, St. Louis, Missouri) for granting permission to use materials before their publication. He is also indebted to Rose Scott and Barbara Kass for their patient and invaluable support in the preparation of the manuscript and illustrations. The work carried out in the author's laboratory mentioned in this review was supported by Program Project USPH-HL Grant 18577.

References

Albers, J. J., Taggart, H. McA., Applebaum-Bowden, D., Haffner, S., Chestnut, C. H., and Hazzard, W. R. (1984). *Biochim. Biophys. Acta* **795,** 293–296.

Allison, A. C., and Blumberg, B. S. (1961). *Lancet* **1,** 405–407.

Beisiegel, U., and Utermann, G. (1979). *Eur. J. Biochem.* **93,** 601–608.

Bell-Quint, J., Forte, T., and Graham, P. (1981). *Biochem. Biophys. Res. Commun.* **99,** 700–706.

Berg, K. (1963). *Acta Pathol. Microbiol. Scand.* **59,** 369–382.

Berg, K. (1983). *Prog. Med. Genet.* [N.S.], pp. 35–90.

Blaufuss, M. C., Gordon, J. I., Schonfeld, G., Strauss, A. W., and Alpers, D. H. (1984). *J. Biol. Chem.* **259,** 2452–2456.

Boguski, M. S., Elshourbagy, N., Taylor, J. M., and Gordon, J. I. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 5021–5025.

Boguski, M. S., Elshourbagy, N., Taylor, J. M., and Gordon, J. I. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 992–996.

Boguski, J. S., Birkenmeier, E. H., Elshourbagy, N. A., Taylor, J. M., and Gordon, J. I. (1986). *J. Biol. Chem.*, **261,** 6398–6407.

Boström, K., Wetsten, M., Wiklund, O., Bondjers, G., Lundholm, K., Elias, P., Norfeldt, P. I., and Olofsson, S. O. (1984). *Eur. J. Biochem.* **143,** 101–107.

Breckenridge, W. C., Little, J. A., Steiner, G., Chow, A., and Poapst, M. (1978). *N. Engl. J. Med.* **298,** 1265–1273.

Breckenridge, W. C., Alaupovic, P., Cox, D. W., and Little, J. A. (1982). *Atherosclerosis* **44,** 223–235.

Breslow, J. L. (1985). *Annu. Rev. Biochem.* **54,** 699–727.

Breslow, J. L. (1986). "*In* Biochemistry and Biology of Plasma Lipoproteins" (A. M. Scanu and A. Spector, eds.), pp. 85–143. Marcel Dekker, New York.

Breslow, J. L., McPherson, J., Nussbaum, A. L., Williams, H. W., Lofquist-Kahl, F., Karathanasis, S. K., and Zannis, V. I. (1982). *J. Biol. Chem.* **257,** 14639–14641.

Breslow, J. L., McPherson, J., Nussbaum, A. L., Williams, H. W., Lofquist-Kahl, F., Karathanasis, S. K., and Zannis, V. I. (1983). *J. Biol. Chem.* **258,** 422–427.

Brewer, H. B., Jr., Lux, S. E., Ronan, R., and John, K. M. (1972). *Proc. Natl. Acad. Sci. U.S.A.* **69,** 1304–1308.

Brewer, H. B., Jr., Shulman, R., Herbert, P., Ronan, R., and Wehrly, K. (1974). *J. Biol. Chem.* **249,** 4975–4984.

Brewer, H. B., Jr., Fairwell, T., LaRue, A., Ronan, R., Houser, A., and Bronzert, T. J. (1978). *Biochem. Biophys. Res. Commun.* **80,** 623–630.

Brewer, H. B., Jr., Fairwell, T., Meng, M., Kay, L., and Ronan, R. (1983). *Biochem. Biophys. Res. Commun.* **113,** 934–940.

Bruns, G. A. P., Karathanasis, S. K., and Breslow, J. L. (1984). *Arteriosclerosis (Dallas)* **4,** 97–102.

Cheung, P., and Chan, L. (1983). *Nucleic Acids Res.* **11,** 3703–3715.

Cheung, P., Kao, F. T., Law, M. L., Jones, C., Puck, T. T., and Chan, L. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 508–511.

Das, H. K., McPherson, J., Bruns, G. A. P., Karathanasis, S. K., and Breslow, J. L. (1985). *J. Biol. Chem.* **260,** 6240–6247.

Deeb, S. S., Motulsky, A., and Albers, J. J. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 4983–4986.

DeLamatre, J. G., Hoffmeier, C. A., and Roheim, P. S. (1983). *J. Lipid Res.* **24,** 1578–1586.

Docherty, K., Carroll, R. J., and Steiner, D. F. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 4613–4617.

Edelstein, C., Gordon, J. I., Toscas, K., Sims, H. F., Strauss, A. W., and Scanu, A. M. (1983). *J. Biol. Chem.* **258,** 11430–11433.

Edelstein, C., Gordon, J. I., Vergani, C. R., Catapano, A. L., Pietrini, V., and Scanu, A. M. (1984). *J. Clin. Invest.* **74,** 1098–1103.

Elshourbagy, N. A., Walker, D. W., Boguski, M. S., Gordon, J. I., and Taylor, J. M. (1986). *J. Biol. Chem.* **261,** 1998–2002.

Fidge, W. H. (1980). *Biochim. Biophys. Acta* **619,** 129–141.

Fless, G. M., and Scanu, A. M. (1986). *In* "Biochemistry and Biology of Plasma Lipoproteins" (A. M. Scanu and A. Spector, eds.), pp. 73–83. Marcel Dekker, New York.

Fless, G. M., Rolih, C. A., and Scanu, A. M. (1984). *J. Biol. Chem.* **259,** 11470–11478.

Fless, G. M., ZumMallen, M. E., and Scanu, A. M. (1985). *J. Lipid Res.* **26,** 1224–1229.

Fless, G. M., ZumMallen, M. E., and Scanu, A. M. (1986). *J. Biol. Chem.* (in press).

Fojo, S. S., Law, S. W., and Brewer, H. B., Jr. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 6354–6357.

Galton, D. J., Stocks, J., and Rees, A. (1983). *Clin. Sci.* **64,** 559–563.

Gordon, J. I., Smith, D. P., Alpers, D. H., and Strauss, A. W. (1982). *Biochemistry* **21,** 5424–5430.

Gordon, J. I., Budelier, K. A., Sims, H. F., Edelstein, C., Scanu, A. M., and Strauss, A. W. (1983). *J. Biol. Chem.* **258,** 14054–14059.

Gordon, J. I., Sims, H. F., Edelstein, C., Scanu, A. M., and Strauss, A. W. (1985). *J. Biol. Chem.* **260,** 14824–14831.

Gordon, J. I., Sims, H. F., Strauss, A. W., Edelstein, C., Byrne, R., and Scanu, A. M. (1986). *CRC Crit. Rev. Biochem.* **20,** 37–71.

Havel, R., Kotite, L., and Kane, J. P. (1979). *Biochem. Med.* **21,** 121–128.

Hospattankar, A. V., Fairwell, T., Roman, R., and Brewer, H. B., Jr. (1984). *J. Biol. Chem.* **259,** 318–322.

Humphries, S. E., Berg, K., Gill, L., Cumming, A. M., Robertson, F. W., Stalenhoef, A., Williamson, R., and Borresen, A. L. (1984). *Clin. Genet.* **26,** 389–396.

Innerarity, T. L., Friedlander, E. J., Rall, S. C., Jr., Weisgraber, K. H., and Mahley, R. W. (1983). *J. Biol. Chem.* **258,** 12341–12347.

Innerarity, T. L., Weisgraber, K. W., Arnold, K. S., Rall, S. C., Jr., and Mahley, R. W. (1984). *J. Biol. Chem.* **259,** 7261–7267.

Jackson, C. L., Bruns, G. A., and Breslow, J. L. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 2945–2949.

Jackson, R. L., Baker, H. N., Gilliam, E. B., and Gotto, A. M. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 1942–1945.

Kane, J. P. (1983). *Annu. Rev. Physiol.* **45,** 637–650.

Karathanasis, S. K. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 6374–6378.

Karathanasis, S. K., Zannis, V. I., and Breslow, J. L. (1983a). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 6147–6151.

Karathanasis, S. K., Norum, R. A., Zannis, V. I., and Breslow, J. L. (1983b). *Nature (London)* **301,** 718–720.

Karathanasis, S. K., McPherson, J., Zannis, V. I., and Breslow, J. L. (1983c). *Nature (London)* **304,** 371–373.

Karathanasis, S. K., Zannis, V. I., and Breslow, J. L. (1985). *J. Lipid Res.* **26,** 451–456.

Karathanasis, S. K., Salmon, E., Haddad, I. A., and Zannis, V. I. (1986). *in* "Biochemistry and Biology of Plasma Lipoproteins" (A. M. Scanu and A. Spector, eds.), pp. 475–494. Marcel Dekker, New York.

Knott, T. J., Priestley, L. M., Urdea, M., and Scott, J. (1984a). *Biochem. Biophys. Res. Commun.* **120,** 734–740.

Knott, T. J., Robertson, J. E., Priestly, L. M., Urdea, M., Wallis, S., and Scott, J. (1984b). *Nucleic Acids Res.* **12,** 3909–3915.

Knott, T. J., Rall, S. C., Jr., Innerarity, T. L., Jacobson, S. F., Urdea, M. S., Levy-Wilson, B. *et al.* (1985). *Science* **230,** 37–43.

Krempler, F., Kostner, G. M., Roscher, A., Haslauer, F., Bolzane, K., and Sandhofer, F., (1983). *J. Clin. Invest.* **71,** 1431–1441.

Lackner, K. J., Law, S. W., Brewer, H. B., Sakaguchi, A. Y., and Naylor, S. L. (1984). *Biochem. Biophys. Res. Commun.* **122,** 877–883.

Lackner, K. J., Law, S. W., and Brewer, H. B. (1985a). *Nucleic Acids Res.* **12,** 4597–4608.

Lackner, K. J., Edge, S. B., Gregg, R. E., Hoeg, J. M., and Brewer, H. B. (1985b). *J. Biol. Chem.* **260,** 703–706.

Law, S. W., and Brewer, H. B., Jr. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 66–70.

Law, S. W., Gray, G., Brewer, H. B., Jr., Sakagachi, A. Y., Naylor, S. L. (1984). *Biochem. Biophys. Res. Commun.* **118,** 934–942.

LeBoeuf, R. C., Miller, C., Shively, J. E., Schumaker, V. N., Balla, M. A., and Lusis, A. J. (1984). *FEBS Lett.* **170,** 105–108.

Lin-Lee, Y. C., Kao, F. T., Cheung, P., and Chan, L. (1985). *Biochemistry* **24,** 3751–3756.

Lusis, A. J., West, R., Mehrabian, M., Reuben, M. A., LeBoeuf, R. C., Kaptein, J. S., Johnson, D. F., Schumaker, V. N., Yuhasz, M. P., Schotz, M. C., and Elovson, J. E. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 4597–4601.

McLean, J. W., Elshourbagy, N. A., Chang, D. J., Mahley, R. W., and Taylor, J. M. (1984). *J. Biol. Chem.* **259,** 6498–6504.

Mahley, R. W. (1983). *Klin. Wochenschr.* **61,** 225–232.

Mahley, R. W., Innerarity, T. L., Rall, S. C., and Weisgraber, K. H. (1984). *J. Lipid Res.* **25,** 1277–1294.

Malloy, M. J., Kane, J. P., Hardman, D. A., Hamilton, R. L., and Dolal, K. B. (1981). *J. Clin. Invest.* **67,** 1441–1450.

Menzel, H. J., Kladetzky, L., and Assmann, G. (1982a). *J. Lipid Res.* **23,** 915–922.

Menzel, H. J., Kovary, P. M., and Assmann, G. (1982b). *Hum. Genet.* **62,** 349–352.

Menzel, H. J., Assmann, G., Rall, S. C., Weisgraber, K. H., and Mahley, R. W. (1983). *J. Biol. Chem.* **259,** 3070–3076.

Moore, M. N., Kao, F., Tsao, Y., and Chan, L. (1984). *Biochem. Biophys. Res. Commin.* **123,** 1–7.

Myklebost, O., Williamson, B., Markham, A. F., Myklebost, S. R., Robers, J., Woods, D. E., and Humphries, S. (1984). *J. Bio. Chem.* **259,** 4401–4404.

Nakagawa, S. H., Lau, H. S. H., Kezdy, F. J., and Kaiser, E. T. (1985). *J. Am. Chem. Soc.* **107,** 7087–7091.

Norum, R. A., Jeffrey, M. D., LaKier, J. B., Goldstein, S., Angel, A., Goldberg, R. B., Block, W. D., Noffze, D. K., Dolphin, P. J., Edelglass, J., Bogorad, D. D., and Alaupovic, P. (1982). *N. Engl. J. Med.* **306,** 1513–1519.

Rall, S. C., Jr., Weisgraber, K. H., Innerarity, T. L., and Mahley, R. W. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 4696–4700.

Rall, S. C., Jr., Weisgraber, K. H., Innerarity, T. L., Bersot, T. P., Mahley, R. W., and Blum, C. B. (1983). *J. Clin. Invest.* **72,** 1288–1297.

Rall, S. C., Jr., Weisgraber, H., Mahley, R. W., Ogawa, Y., Fielding, C. J., Utermann, G., Haas, J., Steinmetz, A., Menzel, H. J., and Assmann, G. (1984). *J. Biol. Chem.* **259,** 10063–10070.

Rosseneu, M., Assmann, G., Taveirne, M. J., and Schmitz, G. (1984). *J. Lipid Res.* **25,** 111–120.

Scanu, A. M., Edelstein, C., and Keim, P. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 317–391. Academic Press, New York.

Scanu, A. M., Edelstein, C., and Shen, B. W. (1982). *In* "Lipid–Protein Interactions" (P. C. Jost and O. H. Griffith, eds.), pp. 259–316. Wiley, New York.

Scanu, A. M., Byrne, R., and Edelstein, C. (1984). *J. Lipid Res.* **25,** 1593–1602.
Schaefer, E. J. (1984). *Arteriosclerosis (Dallas)* **4,** 303–322.
Schumaker, V. N., Robinson, M. T., Curtiss, L. K., Butler, R., and Sparks, R. S. (1984). *J. Biol. Chem.* **259,** 6423–6430.
Sharpe, C. R., Sidoli, A., Shelley, C. S., Lucero, M. A., and Shoulders, C. C., *et al.* (1984). *Nucleic Acids. Res.* **12,** 3917–3932.
Shelburne, F., Hanks, J., Meyers, W., and Quarfordt, S. (1980). *J. Clin. Invest.* **65,** 652–658.
Shoulders, C. C., and Baralle, F. E. (1982). *Nucleic Acids Res.* **10,** 4873–4882.
Shoulders, C. C., Kornblihtt, A. R., Munro, B. S., and Baralle, F. E. (1983). *Nucleic Acids Res.* **11,** 2827–2837.
Sirtori, C. R., and Franceschini, G. (1985). *Klin. Wochenschr.* **63,** 481–489.
Snidermann, A. D., Shapiro, S., Marpole, D., Skinner, B., Teng, B., and Kwiterovich, P.O., Jr. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 604–608.
Soutar, A. K., Garner, C. W., Baker, H. N., Sparrow, J. J., and Jackson, R. L. (1975). *Biochemistry* **14,** 3057–3064.
Sparks, C. E., Huatink, O., and Marsh, J. B. (1981). *Can. J. Biochem.* **59,** 693–699.
Steinmetz, A., and Utermann, G. (1985). *J. Biol. Chem.* **260,** 2258–2264.
Stocks, J., Holdsworth, G., and Galton, G. (1979). *Lancet* **2,** 667–671.
Stocks, J., Holdsworth, G., Dodson, P., and Galton, D. J. (1981). *Atherosclerosis* **38,** 1–9.
Swaney, J. B., Reese, H., and Eder, H. A. (1974). *Biochem. Biophys. Res. Commun.* **59,** 513–518.
Tata, F., Henry, I., Markham, A. F., Wallis, S. C., Weil, D., Grzeschik, K. H., Junien, C., Williamson, R., and Humphries, S. E. (1985). *Hum. Genet.* **69,** 345–349.
Utermann, G., Feussner, G., Franceshini, G., Haas, J., and Steinmetz, A. (1982). *J. Biol. Chem.* **257,** 501–507.
Weinberg, R. B., and Spector, M. S. (1985). *J. Biol. Chem.* **260,** 4914–4921.
Weisgraber, K. H. (1986). "*In* Biochemistry and Biology of Plasma Lipoproteins" (A. M. Scanu and A. Spector, eds.), pp. 301–330. Marcel Dekker, New York.
Weisgraber, K. H., Bersot, T. P., and Mahley, R. W. (1978). *Biochem. Biophys. Res. Commun.* **85,** 287–292.
Weisgraber, K. H., Rall, S. C., Jr., Bersot, T. P., Mahley, R. W., Franceschini, G., and Sirtori, C. R. (1983a). *J. Biol. Chem.* **258,** 2508–2513.
Weisgraber, K. H., Innerarity, T. L., Hardes, K. J., Mahley, R. W., Milne, R. W., Marcel, Y. L., and Sparrow, J. T. (1983b). *J. Biol. Chem.* **258,** 12348–12354.
Weisgraber, K. H., Rall, S. C., Jr., Innerarity, T. L., Mahley, R. W., Kuus, W., and Ehnholm, C. (1984). *J. Clin. Invest.* **73,** 1024–1033.
Wettsten, M., Boström, K., Bondjers, G., Jarfeldt, M., Norfeldt, P., Carrella, M., Wiklund, O., Boren, J., and Olofsson, S. (1985). *Eur. J. Biochem.* **149,** 461–466.
Windler, E., Chao, Y., and Havel, R. J. (1980). *J. Biol. Chem.* **255,** 5475–5480.
Wu, A. L., and Windmueller, H. G. (1979). *J. Biol. Chem.* **256,** 3615–3618.
Zannis, V. I., and Breslow, J. L. (1981). *Biochemistry* **21,** 1033–1041.
Zannis, V. I., Jost, P. W., and Breslow, J. L. (1981a). *Am. J. Hum. Genet.* **33,** 11–24.
Zannis, V. I., Breslow, J. L., San Giacomo, T. R., Aden, D. P., and Knowers, B. B. (1981b). *Biochemistry* **20,** 7089–7096.
Zannis, V. I., Kurnit, D. M., and Breslow, J. (1982a). *J. Biol. Chem.* **257,** 536–544.
Zannis, V. I., Breslow, J. L., Utermann, G., Mahley, R. W., and Weisgraber, K. H. (1982b). *J. Lipid Res.* **23,** 911–914.
Zannis, V. I., Lees, A. M., Lees, R. S., and Breslow, J. L. (1982c). *J. Biol. Chem.* **257,** 4978–4986.
Zannis, V. I., McPherson, J., Goldberger, G., Karathanasis, S. K., and Breslow, J. (1984). *J. Biol. Chem.* **259,** 5495–5499.

4 / α_2-Macroglobulin and Related Thiol Ester Plasma Proteins

Lars Sottrup-Jensen

Department of Molecular Biology, University of Aarhus
DK-8000 Aarhus C, Denmark

The Plasma Proteins, Vol. V

I. Introduction

It has recently been established from the results of intense investigations by many groups that the proteinase binding protein α_2-macroglobulin (α_2M) together with the complement proteins C3, C4, and C5 constitute a novel class of structurally and functionally related large plasma proteins. Apart from C5 these proteins contain an internal β-cysteinyl-γ-glutamyl thiol ester, which enables the proteolytically activated forms of α_2M, C3, and C4 to participate in characteristic covalent binding reactions. The thiol ester structure, which in the native proteins can be slowly cleaved by a number of small nitrogen nucleophiles, represents a novel type of postsynthetic modification of proteins.

Traditionally α_2M has been studied within the context of plasma proteinase inhibitors, although by several criteria it is unique (Table I). Whereas most plasma proteinase inhibitors are monomeric proteins of roughly similar size, containing approximately 430–500 residues (54–100 kDa), α_2M is a tetramer whose 180-kDa subunits contain 1451 residues. Furthermore, in contrast to most other proteinase inhibitors, which form 1 : 1 complexes with serine proteinases engaging the active site of the proteinase and the reactive site of the inhibitor, α_2M forms complexes with a wide spectrum of proteinases differing in their substrate specificity and catalytic mechanism. Both 1 : 1 and 2 : 1 proteinase–α_2M complexes can be formed, and the disulfide-bridged dimer (360 kDa) appears to be the functional unit of α_2M. Contrary to "classical" proteinase–inhibitor complexes the bound proteinase is still active, especially toward small synthetic substrates. These features of α_2M have resulted in the "trap" hypoth-

TABLE I

Major Plasma Proteinase Inhibitors[a]

Proteinase inhibitor	Plasma concentration (μM)	Size (kDa)
α_1-Proteinase inhibitor	37–74	53 (394 residues)
Antithrombin III	approx. 5	58 (430 residues)
α_2-Antiplasmin	approx. 1	65 (450 residues)
α_1-Antichymotrypsin	5–10	58 (430 residues)
C1 esterase inactivator	1–3	104 (500 residues)
α_2-thiol proteinase inhibitor (LMW kininogen)	approx. 6	60 (427 residues)
Inter-α-trypsin inhibitor	1–4	180 (1500 residues)
α_2-Macroglobulin	2–4	720 (4 × 1451 residues)

[a]Data collected from Heimburger (1974), Travis and Salvesen (1983), Sasaki *et al.* (1977), Ohkubo *et al.* (1984), and Gounaris *et al.* (1984).

esis for α_2M–proteinase complex formation, where proteolytic cleavage of a particularly exposed peptide stretch near the middle of the 180-kDa subunit (the "bait" region) results in a conformational change of the α_2M tetramer, thereby entrapping the proteinase. The nature of the essentially irreversible proteinase complex formation with α_2M has long remained elusive. The α_2M–proteinase complexes are rapidly cleared from the circulation, indicating an important role of α_2M for controlling proteolytic activity in plasma.

The proteins C3, C4, and C5 are members of the group of 20 plasma proteins now considered to be the functional set of complement, an important part of the immune system. C3 consists of α chains (115 kDa) and β chains (75 kDa); C4 of α chains (95 kDa), β chains (70 kDa), and γ chains (32 kDa); and C5 of α chains (115 kDa) and β chains (75 kDa). Within each protein the individual chains are disulfide bridged. In contrast to the tetrameric α_2M these multiple-chain complement proteins circulate as 190- to 200-kDa proteins, thus representing proteolytically processed "monomers." The physiological activators for C3, C4, and C5 are macromolecular enzyme complexes having a narrow substrate specificity, and in each case the activating event consists of the cleavage of a single arginyl–X bond in the N-terminal part of the respective α chains. C3 is the central protein of complement, operating at a point where the "classical" and

"alternative" pathways of activation converge, eventually resulting in opsonization of cell-bound C3 by macrophages and neutrophils, and in the assembly of the lytic complex, C5b,6,7,8,9. C4 participates in the classical pathway, which is triggered mainly by the aggregation of certain classes of immunoglobulins on antigen binding. In contrast, the alternative pathway, which does not require an immune response, is triggered by binding of C3b to a variety of compounds, particularly polysaccharides from bacterial and yeast cell walls. A schematic representation of the activation pathways of complement is shown in Fig. 1. Both pathways are subjected to a number of intrinsic and extrinsic control mechanisms, among which the "nascent" state of activated C3, C4, and C5 is particularly important.

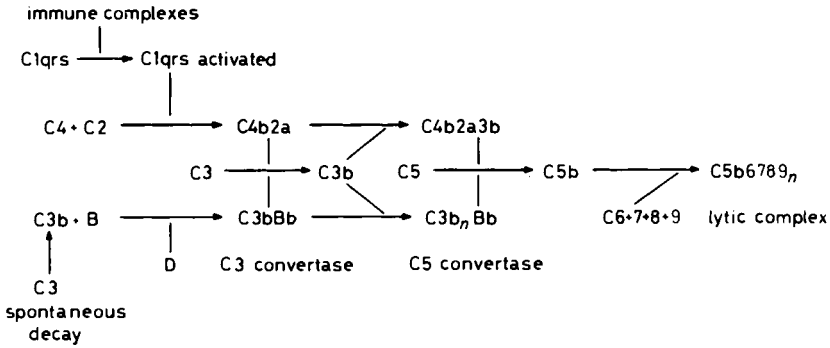


Fig. 1. *Scheme of activation pathways of complement.* Classical pathway: The activation by immune complexes is initiated by binding of the first component (C1) through the C1q subcomponent to the Fc region of the antibody molecule. This results in the sequential activation of the subcomponents C1r and C1s. C1s in turn activates C4 and C2. Nascent C4b becomes covalently bound to the immune complex (to the Fab portion of the antibody) or to cell surface structures and acquires a binding site for C2, which is then activated. The complex between C4b and C2a (C3 convertase) activates C3, which in its nascent state binds covalently to nearby cell surface structures. Part of C3b forms a complex with C4b2a (C5 convertase, C4b2a3b), in which the bound C2a activates C5 to form C5b. Nascent C5b interacts with C6 and C7 and becomes incorporated noncovalently into the sensitized cell membrane, followed by the sequential addition of C8 and C9 to form the terminal lytic complex C5b,6,7,8,9_n. Beyond the activation of C5 no proteolytic processes are believed to take place. Alternative pathway: Nascent C3b, presumably formed in the spontaneous decay of C3, interacts with factor B to form a complex C3bB, which is stabilized by certain polysaccharides and properdin, a plasma protein. In this complex factor B is activated by factor D, a small trypsinlike serine proteinase, to form the complex C3bBb (C3 convertase). This in turn activates C3, more C3b enters the C3bBb complex, and the C5 convertase is formed ($C3b_nBb$). Then assembly of the terminal lytic complex ensues. Subcomponents C1r and C1s are large sophisticated serine proteinases and factor B and C2 are large serine proteinases with an unusual structure. Short accounts of the activation of C1 and the structures of C1q, C1r, C1s, C2, and B have appeared recently (Colomb *et al.*, 1984; Carter *et al.*, 1984; Gagnon, 1984). The assembly of the lytic complex has been reviewed by Bhakdi and Tranum-Jensen (1984), Podack and Tschopp (1984), and Müller-Eberhard (1984).

In contrast to earlier comprehensive reviews on macroglobulins (Starkey and Barrett, 1977; Barrett, 1982; Roberts, 1985; Travis and Salvesen, 1983) and on complement (Porter and Reid, 1979); Müller-Eberhard and Schreiber, 1980; Reid and Porter, 1981), this review will emphasize the close structural and functional relationship between α_2M and related macroglobulins on the one hand and the complement proteins C3, C4, and C5 on the other. Previously these two groups of proteins were thought to be unrelated. This realization is the outcome of the rapid advances in the knowledge about these proteins, which has taken place in recent years, notably since the discovery of the internal reactive thiol esters of α_2M, C3, and C4 in 1980. Recent accounts of this process can also be found in the proceedings from a New York Academy of Science conference on α_2M (Vol. 421, R. D. Feinman, ed., 1983) and in two multiauthor volumes in the Springer Seminar Series on Immunopathology (Vols. 6 and 7, H. Müller-Eberhard, ed., 1983 and 1984).

In this review α_2M, C3, and C4 will be discussed as members of a family of α_2M-related thiol ester plasma proteins. While the existence of two related α-macroglobulins in several animal species has been known for some time, it has become clear only recently that human plasma contains a second thiol ester containing α-macroglobulin, namely, the pregnancy zone protein (PZP). In some rodents, e.g., the mouse and the rat, three distinct thiol ester-containing α-macroglobulins, which are proteinase binding, are found. Thus, the family of α_2M-related thiol ester proteins consists of two subgroups, the proteinase-binding dimeric or tetrameric macroglobulins and the proteins C3 and C4, which in the complement system participate in specific binding reactions of a different kind. Intriguingly, within each of these subgroups an additional protein is known, which does not contain an internal thiol ester, namely, the proteinase-binding ovostatin from hen egg white and complement protein C5. This indicates that it might not be the presence of the thiol ester structure per se, that basically characterizes all these proteins but rather the presence of an underlying common, yet unknown, gross structure, capable of undergoing a particular sort of conformational change upon activation by specific limited proteolysis. An inherent feature of this conformational change appears to be the generation of a short-lived state, the nascent state, where specific binding reactions, covalent or not, can take place, thereby resulting in the assembly of unique macromolecular complexes of vital importance.

II. Brief Historical Perspective

It will be appropriate to summarize those diverse earlier observations, which can now be integrated into a functional concept reflecting the presence of internal thiol esters in α_2M, C3, and C4 activated by a specific conformational change.

These common functional properties of α_2M, C3, and C4 are the following: (1) A slow inactivation by small amines resulting in covalent binding. (2) A fast covalent binding of the proteolytically activated forms to other proteins and to cell surfaces in the nascent state. (3) A unique heat or denaturation fragmentation pattern of the native proteins. (4) An appearance of titratable SH groups during these reactions. (5) A specific conformational change leading to recognition of the proteolytically activated or amine-inactivated forms by specific cell surface receptors.

Gordon *et al.* (1926) were the first to report that a component of complement was unusually sensitive toward ammonia, methylamine, and ethylamine. In contrast to the three components then recognized this new component (the fourth) could be specifically inactivated by low concentrations of amines at a weakly alkaline pH. Pillemer *et al.* (1941) confirmed and extended these observations and concluded that the inactivation was only observed with small amines or hydrazines having a primary amino group. When C4 was later purified it was shown that its hemolytic activity was lost by incubation with low concentrations of hydrazine (Müller-Eberhard and Biro, 1963).

In studies on a plasmin inhibitor later known to be α_2M Pillemer *et al.* (1953) reported that this inhibitor, which was distinct from C4 (Ratnoff *et al.*, 1954), could also be inactivated by hydrazine. Mehl *et al.* (1964) pointed out that α_2M prepared by ammonium sulfate precipitation had lost its proteinase binding activity. This was confirmed by Steinbuch *et al.* (1968), who also noted that inactivation was more effective with hydrazine than with methylamine. From these studies it was not entirely clear whether the inactivation observed was reversible or irreversible.

When purified C3 became available (Müller-Eberhard, 1961) it was shown that it could also be inactivated by hydrazine. In addition, the hemolytic activity was found to be gradually destroyed upon incubation with chaotropes like potassium bromide or upon storage (Müller-Eberhard *et al.*, 1960; Dalmasso and Müller-Eberhard, 1966). Accompanying inactivation of C4 changes in the sedimentation coefficient and the electrophoretic mobility were observed (Müller-Eberhard and Biro, 1963), probably representing the same type of molecular changes, that could be detected in inactivated C3 by immunological techniques (West *et al.*, 1966). Upon proteolytic activation of C3 and C4 distinct changes in electrophoretic mobility were also seen (Pondman and Peetoom, 1964; Müller-Eberhard and Lepow, 1965).

α_2M incubated with hydrazine and methylamine reproducibly migrated slightly faster than native α_2M in agarose gel electrophoresis (Steinbuch *et al.*, 1968), and preparations of α_2M were found to be heterogeneous in acrylamide gel electrophoresis, with the fastest-migrating forms having little or no proteinase binding activity (Saunders *et al.*, 1971). Upon saturation with trypsin the slowest-migrating form representing native α_2M was completely converted to the

fastest-migrating form, which was indistinguishable from that obtained by incubation with methylamine.

Following the demonstration that a fraction of C3 and C4 became strongly associated with erythrocyte membrane constituents or with immune complexes during activation of complement (Dalmasso and Müller-Eberhard, 1964; Willoughby and Mayer, 1965), subsequent studies indicated that activated C3 and C4 expressed a labile binding site of unknown nature, which could react with unspecified acceptors on the erythrocyte surface or in the immune complex (Müller-Eberhard and Lepow, 1965; Müller-Eberhard *et al.*, 1966). If activated C3 and C4 did not react with those acceptors in the short-lived state (the nascent state), inactivated C3 and C4 would accumulate in the fluid phase.

The first clue to the nature of the interaction between activated C3 and cell surfaces was provided by Law and Levine (1977), who showed that a large fraction of cell-bound C3 in fact was covalently bound. Since these covalent bonds could be cleaved by hydroxylamine, they were judged to be esters, probably formed between surface hydroxyl groups and a carboxyl group located in the α' chain of C3 (Law *et al.*, 1979b). The first indication of covalent binding of proteinases to α_2M came from Harpel (cited in Harpel and Rosenberg, 1976), who showed that the light chain of plasmin remained bound to the α_2M subunit or its bait region cleaved fragments, even after extensive denaturation. Similar results were obtained by Steinbuch *et al.* (1976), who also showed that trypsin could be bound covalently to α_2M, and by Granelli-Piperno and Reich (1978). Pochon *et al.* (1978) observed that treatment of α_2M-chymotrypsin with urea failed to dissociate chymotrypsin from the α_2M dimers obtained under these conditions (Jones *et al.*, 1972; Harpel, 1973), and Salvesen and Barrett (1980) demonstrated covalent binding of a variety of proteinases. The binding of activated C4 to immune complexes, presumably to the Fab portion of IgG (Goers and Porter, 1978), was also shown to involve covalent bond formation (Law *et al.*, 1980a,b; Campbell *et al.*, 1980).

A new puzzling observation was made by Harpel and Hayes (1979), who showed that preparations of native α_2M denatured in hot SDS and analyzed by reducing SDS–PAGE contained in addition to the 180-kDa subunit characteristic 120- and 60-kDa fragments. In a subsequent study the possible interference from proteinases was ruled out, and it was concluded that the heat/denaturation fragments resulted from the cleavage of an unusually labile peptide bond in α_2M, not present in a panel of other proteins investigated (Harpel *et al.*, 1979). These results were confirmed and extended by Barrett *et al.* (1979), who like Harpel *et al.* (1979) also showed that the heat fragmentation did not occur with amine-inactivated α_2M. That the methylamine-dependent inactivation of α_2M and the presence of a heat-labile peptide bond were interrelated was apparent from the studies of Swenson and Howard (1979b), which conclusively demonstrated that inactivation with methylamine resulted in the incorporation of 1 mole amine per

mole subunit into a particular Glx residue. This was followed by the precise demonstration that the heat cleavage occurred at the N-terminal side of the methylamine-reactive Glx residue (Howard *et al.*, 1980), thereby converting this residue to a residue of pyroglutamic acid. Subsequent work from the same laboratory established that C3 (Howard, 1980) and C4 (Gorski and Howard, 1980) also displayed a characteristic heat fragmentation pattern and contained a methylamine-reactive Glx residue located in their α chains. A unique reactive site common to α_2M, C3, and C4 was proposed to consist of an internal pyroglutamic acid residue (Howard *et al.*, 1980), which under proper conditions would effect covalent bond formation with added nucleophiles or could undergo cleavage, resulting in polypeptide chain fragmentation.

In a systematic study of the effects of nitrogen nucleophiles, chaotropes, and denaturants on the functional and structural properties of C3 and C4 by Tack, Prahl, and co-workers (Janatova *et al.*, 1979, 1980a,b; Janatova and Tack, 1981) heat fragmentation of the α chains of native C3 and C4 was observed. Moreover, these studies revealed the appearance of a previously unrecognized SH group (maximally 1 mole per mole C3 or C4), in parallel with the loss of hemolytic activity resulting from incubation with hydrazine or hydroxylamine. This SH group also appeared as a result of incubation with chaotropes and with the physiological activator C1s (C4) or the nonphysiological activator trypsin (C3). Sequence studies revealed that the SH group in C3 was contributed by a Cys residue, located only three positions from the methylamine-reactive Glx residue (Tack *et al.*, 1980) in a heptapeptide sequence, identical to that determined earlier in α_2M around its methylamine-reactive Glx residue (Swenson and Howard, 1979b). Additional studies showed that the sequence around the reactive Glx residue of C4 was identical with that of C3 and α_2M (Harrison *et al.*, 1981; Campbell *et al.*, 1981).

Combining the available evidence Tack *et al.* (1980) concluded that the reactive site of C3, C4, and α_2M most likely was constituted by a β-cysteinyl-γ-glutamyl thiol ester, a previously unknown integral structural element of proteins. Additional results on inactivation of C3 and C4 by amines or chaotropes, heat fragmentation, and covalent binding to immune complexes or to cell surfaces immediately confirmed and corroborated this proposal (Law *et al.*, 1980a,b, 1981; Pangburn and Müller-Eberhard, 1980; Reboul *et al.*, 1980; Campbell *et al.*, 1980; R. B. Sim *et al.*, 1981; Gadd and Reid, 1981; Lundwall *et al.*, 1981; von Zabern *et al.*, 1981; Seya and Nagasawa, 1981; Ichihara *et al.*, 1981).

Independently, Sottrup-Jensen *et al.* (1980) showed that α_2M–proteinase complex formation was accompanied by an appearance of SH groups (maximally four SH groups per mole tetramer for 2 mole trypsin or elastase bound). In examining alternative ways of activating the methylamine-reactive Glx residue, thought to be involved also in the formation of covalent α_2M–proteinase com-

plexes, they were intrigued about the possibility that this reaction could be reminiscent of the formation of ε-lysyl-γ-glutamyl cross-links by transglutaminases. Thus, the methylamine-reactive Glx residue identified by Swenson and Howard (1979b) could by analogy with the thiol ester intermediate known in transglutaminases (Folk and Chung, 1973; Folk and Finlayson, 1977; Folk, 1983) be thiol esterified. Since it was also found that incorporation of methylamine into α_2M resulted in the appearance of SH groups in an approximate 1 : 1 stoichiometry, and since proteinases and methylamine acted on the same potential of SH appearance, it was concluded that the methylamine-reactive Glx residue in fact was thiol esterified to a Cys residue, located in a sequence identical to that found in C3 and C4 (Sottrup-Jensen *et al.*, 1981a). The presence of an internal reactive β-cysteinyl-γ-glutamyl thiol ester in α_2M would readily explain the methylamine incorporation, the potential for covalent proteinase binding, and the heat cleavage (Sottrup-Jensen *et al.*, 1980). These results were corroborated by similar results obtained by Salvesen *et al.* (1981) and by Howard (1981). In analogy with activated C3 and C4 (Law *et al.*, 1981), proteolytically activated α_2M was also found to exist in a short-lived nascent state, in which added nucleophiles could be covalently bound to the thiol-esterified Glx residue (Sottrup-Jensen *et al.*, 1981c,d; Salvesen *et al.*, 1981). The relationship between covalent α_2M–proteinase complex formation and binding of small nucleophiles was also studied by Van Leuven *et al.* (1981a,b), Wu *et al.* (1981), and Wang *et al.* (1981). That covalent binding indeed engaged these thiol-esterified Glx residues in α_2M, C3, and C4 was proven by Campbell *et al.* (1981), Sottrup-Jensen and Hansen (1982), and Hostetter *et al.* (1982).

The rapid clearance of α_2M–proteinase complexes from the circulation was first noted by Ohlsson (1971a,b) and was confirmed and extended by studies by Debanne *et al.* (1975) and Kaplan and Nielsen (1979a,b). From these and similar results by Van Leuven *et al.* (1978, 1979) it became evident that several cell types including macrophages and fibroblasts contained receptors for α_2M–proteinase complexes. These cellular receptors would recognize amine-treated and proteinase-treated α_2M equally well (Kaplan *et al.*, 1981; Imber and Pizzo, 1981; Van Leuven *et al.*, 1981a), indicating that the cleavage of the thiol esters in α_2M was intimately associated with a conformational change leading to the exposure of a receptor recognition site.

Among the different receptors for activated C3 and C4 or cleavage products thereof that have been described on different cells, several presumably play a prominent role in the clearance of the activated complement proteins (Gigli and Nelson, 1968; Mantovani *et al.*, 1972; Ehlenberger and Nussenzweig, 1977; Horwitz, 1980; Fearon *et al.*, 1981). In analogy with α_2M the action of amines and chaotropes on C3 and C4 produces a conformationally changed molecule, which resembles the proteolytically activated forms (Janatova *et al.*, 1980b; Janatova and Tack, 1981; Parkes *et al.*, 1981; Isenman *et al.*, 1981; von Zabern

et al., 1981). In contrast to the native proteins these forms readily interact with cellular receptors (Berger *et al.*, 1981; Arnaout *et al.*, 1981; Schreiber *et al.*, 1981) thereby providing for phagocytosis of complexes between C3b or C4b and cells or immune aggregates. It had been suggested earlier, on the basis of limited sequence information, that C5 could be homologous with C3 and C4 (Fernandez and Hugli, 1977). However, this was not firmly demonstrated until longer stretches of sequence from C5 had been determined (Lundwall *et al.*, 1985). Although C5 does not contain an internal thiol ester, activated C5 nevertheless enters a nascent state, in which specific interactions with C6 and membranes take place (Cooper and Müller-Eberhard, 1970; Vogt *et al.*, 1978; DiScipio, 1981a,b; DiScipio *et al.*, 1983).

III. The Family of α_2M-Related Proteins

A. Proteins of the Macroglobulin Subgroup

Following the first description of human α_2M by Schultze *et al.* (1955) this protein has been isolated and characterized by many investigators. Several procedures for the isolation of highly active and native α_2M exist as reviewed by Barrett (1982) and Roberts (1985). It has recently been shown that human plasma contains another α-macroglobulin, structurally and functionally very similar to α_2M, namely, the pregnancy zone protein (PZP) (Sottrup-Jensen *et al.*, 1984d; Sand *et al.*, 1985). In contrast to α_2M this protein is a dimer of 360 kDa. PZP was first described by Smithies (1959) and has been characterized by several groups, e.g., by Bohn (1971), Straube *et al.* (1972), Stimson and Eubank-Scott (1972), von Schoultz and Stigbrand (1974), Bohn and Winckler (1976), Stigbrand *et al.* (1978), Folkersen *et al.* (1978), and Stimson and Farquharson (1978). During pregnancy the plasma level of PZP increases approximately 20- to 100-fold, and levels up to 1000 mg/liter are not uncommon in late pregnancy (Folkersen *et al.*, 1981b; von Schoultz, 1974; Than *et al.*, 1976). Thus PZP is quantitatively the major pregnancy-associated plasma protein. Although less information is available, another large pregnancy-associated protein could also be related to α_2M, namely, the pregnancy-associated plasma protein A (PAPP-A). The level of this protein is roughly one-tenth that of PZP, and it circulates as a tetramer of 800 kDa (Lin *et al.*, 1974; Bischof, 1979; Sutcliffe *et al.*, 1980; Folkersen *et al.*, 1981a; Sinosich *et al.*, 1982).

The presence of two distinct but related αM's in the rat, having similar proteinase-binding properties, has been known for some time (Menninger *et al.*, 1970; Ganrot, 1973a,b; Gordon, 1976; Hudig and Sell, 1979; Nieuwenhuizen *et al.*, 1979; Okubo *et al.*, 1981; Panrucker and Lorscheider, 1982; Schaeufele and Koo, 1982; Nelles and Schnebli, 1982; Van Gool *et al.*, 1982). Rat α_2M, which is an acute-phase reactant, contains 180-kDa subunits, while the subunits of rat

α_1M are cleaved to disulfide-bridged 160- and 40-kDa fragments. These proteins are tetramers of 720 kDa.

Although only one αM has been described in the mouse (Greene *et al.*, 1971) it has recently become clear that two αM's similar to those of the rat are present (Hudson and Koo, 1982; Gonias *et al.*, 1983; Saito and Sinohara, 1985a). In addition, both the rat and the mouse contain a third proteinase-binding αM related to α_2M, namely, the rat α_1-inhibitor III (Gauthier and Ohlsson, 1978; Esnard and Gauthier, 1980; Esnard *et al.*, 1981, 1985) and the mouse murinoglobulin (Saito and Sinohara, 1985a; Yamamoto *et al.*, 1985). The so-called rat "murinoglobulin" (Saito and Sinohara, 1985b) is probably identical with rat α_1-inhibitor III (L. Sottrup-Jensen, unpublished). These proteins, which in analogy with the proteinase-binding dimeric unit of human α_2M probably function as noncovalently associated 360-kDa dimers, are perhaps related to the pregnancy-associated rat and mouse proteins described by Porstmann and Hau (1984) and Waites and Bell (1984). Although rat α_1-inhibitor III has been judged to be related to human inter-α-trypsin inhibitor (Gauthier and Ohlsson, 1978), complexes between inter-α-trypsin inhibitor and trypsin or chymotrypsin have no residual proteolytic activity. In addition, this protein has no capacity for protecting the activity of the bound proteinases against soybean trypsin inhibitor and α_1-proteinase inhibitor, respectively. Finally, N-terminal sequence analysis of human inter-α-trypsin inhibitor has not revealed a sequence which is homologous with those of the members of the macroglobulin family (L. Sottrup-Jensen, unpublished).

Two αM's have also been described in the rabbit (Bloth *et al.*, 1968; Lebreton de Vonne and Mouray, 1968; Debanne *et al.*, 1975) and in the dog (Ohlsson, 1971a,b). Besides the known αM of the pig (Jacquot-Armand and Guinand, 1967, 1976; Baumstark, 1973; Tsuru *et al.*, 1978) another distinct αM is present (Weström, 1979a,b). Although only one αM has been described in the ox and in the horse (Nagasawa *et al.*, 1970a,b; Feldman *et al.*, 1984; Dubin *et al.*, 1984) it is likely that these species and probably all other species will turn out to contain at least two, perhaps three, mutually related αM's.

The plasma of the hen contains an αM strongly related to human α_2M (Nagase *et al.*, 1983). However, an immunologically distinct tetrameric protein, ovostatin, is found in hen egg white (Kitamoto *et al.*, 1983; Nagase *et al.*, 1983; Nagase and Harris, 1983; Feldman and Pizzo, 1984a) that is similar to other proteinase-binding proteins (ovomacroglobulins) found in avian eggs (Miller and Feeney, 1966; Donovan *et al.*, 1969) and in amphibian eggs (Ikai *et al.*, 1983). Unlike α_2M, ovostatin apparently contains no internal thiol esters, but is otherwise closely related. However, the corresponding protein from duck eggs seems to be a thiol ester protein (Nagase *et al.*, 1985). Frog plasma apparently contains two related αM's, a tetrameric and a dimeric protein (Feldman and Pizzo, 1985, 1986).

The relationship between the two αM's found in several animal species and the possible relationship between all vertebrate αM's have been studied by immunological methods (James, 1965; Bütler and Brunner, 1967; Berne *et al.*, 1973; Shortridge *et al.*, 1976; Weström *et al.*, 1983; Carlsson *et al.*, 1985). Although all αM's studied are more or less antigenically related and may be grouped into several subsets, the high discriminatory power of these procedures appears to be less suited for the detection of distant relationships than methods utilizing the known binding properties of αM's, namely, the formation of complexes, in which the bound proteinase is still active, and the incorporation of radiolabeled methylamine. Thus, Starkey and Barrett (1982a) surveyed a number of major vertebrate taxa and found αM's in mammals, birds, reptiles, amphibians, and fish. Even in invertebrates such as the horseshoe crab and the crayfish α_2M-like proteins are present as shown by Quigley and Armstrong (1983, 1985) and Hergenhahn and Söderhall (1985). The αM homolog from plaice has been isolated and characterized (Starkey *et al.*, 1982; Starkey and Barrett, 1982b) and shown to be a noncovalently associated dimeric protein containing proteolytically processed subunits composed of disulfide-bridged 105- and 90-kDa chains.

B. Proteins of the Complement Subgroup

Most studies on C3, C4, and C5 have utilized the human proteins, which have been purified and characterized by many investigators. An example of a large-

TABLE II

Representative Examples of Proteins Related to Human α_2M

Subgroup	Size of subunits (kDa)	Quaternary structure
Macroglobulin		
Human α_2M	180	Tetramer (subunits pairwise disulfide bridged)
Human PZP	180	Dimer/tetramer
Rat α_1M	160 + 40	Tetramer
Rat α_2M	180	Tetramer
Rat α_1-inhibitor III	180	Monomer/dimer?
Plaice αM	105 + 90	Dimer
Ovostatin	180	Tetramer
Complement		
Human C3	75(β)+115(α)	Proteolytically processed "monomers" (disulfide bridged)
Human C4	70(β)+95(α)+32(γ)	
Human C5	75(β)+115(α)	
Cobra venom factor	71(β)+48(α)+28(γ)	

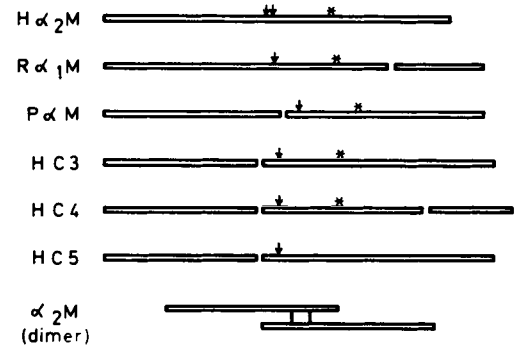


Fig. 2. Schematic representation of the structures and chain organization of examples of α_2M-related proteins. The approximate positions of the activation cleavage sites (↓) and thiol ester sites (*) are shown. The disulfide bridge pattern of human α_2M is known (Sottrup-Jensen *et al.*, 1984c) and the dimeric unit has been shown to contain two interchain bridges (Sottrup-Jensen, 1985; P. E. Jensen and L. Sottrup-Jensen, unpublished). Presumably the subunits of the dimer are arranged in an antiparallel fashion. The individual chains of rat α_1M and plaice αM, C3, C4, and C5 are disulfide-bridged, but their bridge patterns [except for the C3a portion (Huber *et al.*, 1980)] are not known in detail. H, human; R, rat; P, plaice.

scale procedure for the purification of human C3, C4, and C5, which will also yield many other complement components, is given by Hammer *et al.* (1981). Unlike C3 and C5, two isotypes exist for human C4 (C4A and C4B) and for mouse C4 (C4 and Slp) (Passmore and Schreffler, 1970; Roos *et al.*, 1978; O'Neill *et al.*, 1978a,b; Awdeh and Alper, 1980; Mauff *et al.*, 1983a).

As discussed by Gigli and Austen (1971) the complement system is probably present in all vertebrates, although not necessarily containing all of the "classical" nine components of human complement in every case (Jensen *et al.*, 1981). From the rainbow trout Nonaka *et al.* (1981a,b, 1984a,b, 1985a,b) isolated and characterized C5 and two variants of C3 (C3-1 and C3-2). A cobra venom factor, related to *Naja naja* plasma C3 (Alper and Balavitch, 1976), has been purified and characterized (Eggertsen *et al.*, 1981).

Examples of α_2M-related proteins from both subgroups are shown in Table II, and Fig. 2 shows a schematic representation of their structures and chain organization.

C. Available Complete or Partial Sequences

Based on the quantitative N-terminal sequence analysis and peptide mapping of ^{14}C-carboxymethylated tryptic Cys-containing peptides, Swenson and Howard (1979a) concluded that α_2M was composed of four identical 180-kDa subunits. This was supported by the results from the sequence analysis of methionine- or homoserine-containing chymotryptic peptides (Sottrup-Jensen *et al.*,

1979) and from analysis of the new sequences that appear upon complex formation with proteinases (Sottrup-Jensen *et al.*, 1981b; Hall *et al.*, 1981; Mortensen *et al.*, 1981b; Virca *et al.*, 1983). The primary sites of activation cleavage were localized to residues 681–686 (Mortensen *et al.*, 1981b). The sequence results from the early phase were reported by Sottrup-Jensen *et al.* (1979). In this study all major pools of CNBr fragments were subdigested with chymotrypsin in order to produce small peptides, which could readily be handled by paper electrophoretic techniques and by manual sequence determination. The sequence analysis of α_2M proceeded by classical techniques of protein chemistry by the isolation and characterization of the 26 CNBr fragments followed by manual or automated sequence analysis of a variety of peptides from these fragments. The nearly complete sequence was reported in 1983 (Sottrup-Jensen *et al.*, 1983a), and the complete structure, including the localization of the eight glucosamine-based oligosaccharide groups (Dunn and Spiro, 1967a,b) and nearly all of the disulfide bridges, was reported in 1984 (Sottrup-Jensen *et al.*, 1984a,b,c; Kristensen *et al.*, 1984; Stepanik and Sottrup-Jensen, 1984; Welinder *et al.*, 1984). A note of correction (residue 540 is a Cys residue, not a Glu residue) has appeared (Sottrup-Jensen, 1985). The short stretches of sequence from α_2M around the thiol ester site, determined by Swenson and Howard (1979a,b, 1980), were fully confirmed in this work. In a different approach Kan *et al.* (1985) synthesized two mixed oligo DNA probes corresponding to residues 1354–1359 and 1359–1363 of α_2M. These were used to screen a human acute-phase liver cDNA library containing large inserts (Belt *et al.*, 1984). Of the several inserts isolated, a 4.6-kb cDNA clone was shown by sequence analysis to contain the coding sequence for pre-α_2M. Apart from one difference, the sequence deduced for α_2M fully confirmed the sequence determined at the protein level. Short stretches of sequence from rat α_2M (Sottrup-Jensen *et al.*, 1984c; Northemann *et al.*, 1985; Hayashida *et al.*, 1985), mouse αM (Hudson *et al.*, 1980), chicken αM and ovostatin (Nagase *et al.*, 1983), rat α_1M (K. Lonberg-Holm and R. Kutny, unpublished), rat α_1-inhibitor III (L. Sottrup-Jensen, unpublished), and duck ovomacroglobulin (Nagase *et al.*, 1985) have also been determined. Sottrup-Jensen *et al.* (1984d) determined the complete or partial sequences of a random selection of 38 tryptic peptides from human PZP, covering 685 residues (approximately 47%) of the sequence of its subunit.

Extending earlier studies on C3 by Budzko *et al.* (1971), Tack and Prahl (1976) determined a decapeptide sequence of the intact protein, reflecting the two chains of C3. Repeating the sequence determination on the isolated α and β chains Tack *et al.* (1979b) confirmed that the anaphylatoxin of C3, C3a, constituted the first 77 residues of the α chain of C3 as proposed by Budzko *et al.* (1971). The sequence of C3a had been determined previously by Hugli (1975). Based on the additional short C-terminal sequences of the isolated chains of C3 (Tack *et al.*, 1979b), the combined results showed that proteolytic activation of

C3 was effected by the cleavage of a single Arg-Ser bond located as residues 77–78 in the α chain. Studies of Brade *et al.* (1977) and Patel and Minta (1979) showed that C3 is synthesized as a single-chain protein (pro-C3), which prior to secretion is proteolytically processed to the two-chain plasma protein. The order of the chains within pro-C3 was determined as β-α (Goldberger *et al.*, 1981). Following characterization of mRNA coding for mouse pro-C3 by Odink *et al.* (1981), Wiebauer *et al.* (1982) and Domdey *et al.* (1982) isolated and characterized cDNA clones and genomic clones encoding C3. Subsequent work by Lundwall *et al.* (1984) and by Wetsel *et al.* (1984) resulted in the elucidation of the complete coding sequence for mouse prepro-C3. These studies were extended to human prepro-C3, which has also been sequenced at the cDNA level (De Bruijn and Fey, 1985). Short stretches of sequence around the thiol ester site of C3 (human and guinea pig) were determined by Tack *et al.* (1980), Thomas *et al.* (1982), and Thomas and Tack (1983). Short partial sequences of the chains of rainbow trout C3 variants (Nonaka *et al.*, 1984a, 1985a,b) and cobra venom factor (Eggertsen *et al.*, 1981) are also known.

Gigli *et al.* (1977) and Bolotin *et al.* (1977) determined short N-terminal sequences of the isolated α, β, and γ chains of human C4. The sequence of the anaphylatoxin from C4, C4a (Gorski *et al.*, 1979), constituting the first 77 residues of the α chain as supported by the results of Press and Gagnon (1981), was determined by Moon *et al.* (1981). Thus the activation cleavage site of C4 was located at residues 77–78 (Arg-Asn) in the α chain. Studies by Hall and Colten (1977), Goldberger and Colten (1980), Goldberger *et al.* (1980), and Odink *et al.* (1981) showed that C4 is synthesized as a single-chain protein, which is proteolytically processed to the three-chain structure of plasma C4. The order of the chains in pro-C4 was found to be β-α-γ (Goldberger and Colten, 1980). Following the sequence determination around the thiol ester site in C4 (Harrison *et al.*, 1981; Campbell *et al.*, 1981), Chakravarti *et al.* (1983) determined a long segment from the C4d portion. A 16-residue synthetic DNA probe, based on a 5-residue peptide stretch from C4d, was used by Carroll and Porter (1983) to isolate a cDNA fragment. This was then used by Belt *et al.* (1984) to isolate a 5.5-kb cDNA clone, which by sequence analysis was shown to encode prepro-C4A. Partial sequences for the closely related C4B were also reported in this study, along with many shorter stretches of sequence, determined by protein chemistry. Studies by Chakravarti *et al.* (1983) and by Hellman *et al.* (1984) contributed to the localization of sites of genetic variation of C4, especially abundant in the C4d portion. Partial sequences of guinea pig C4 (Whitehead *et al.*, 1983) and mouse C4 and Slp (Parker *et al.*, 1980; Karp *et al.*, 1982b; Ogata *et al.*, 1983; Nonaka *et al.*, 1984b; Levi-Strauss *et al.*, 1985) deduced from the sequence of cDNA clones have also appeared, and recently the complete coding sequence of mouse prepro-C4 was determined (Sepich *et al.*, 1985; Nonaka *et al.*, 1985b).

From sequence studies of the anaphylatoxin from human or porcine C5, C5a (Fernandez and Hugli, 1977, 1978; Gerard and Hugli, 1980), and of the α and α′ chains of C5 (Tack *et al.*, 1979a; DiScipio *et al.*, 1983) the site of proteolytic activation of C5 was determined to be an Arg-Leu bond at position 74–75. Like C3, the two-chain plasma C5 (Nilsson and Mapes, 1973) was found to be synthesized as a single-chain precursor (Ooi and Colten, 1979). In contrast to C3 and C4 no evidence of an internal thiol ester in C5 has been obtained (Law *et al.*, 1980a; Janatova and Tack, 1981; DiScipio, 1981a,b; DiScipio *et al.*, 1983). Lundwall *et al.* (1985) utilized a DNA probe corresponding to residues 19–25 of human C5a and determined the sequence of a cDNA clone, encoding a 438-residue segment of C5 spanning across the C5a portion. Their results showed that like pro-C3 the order of the potential chains in pro-C5 is β-α. Mouse pro-C5 has recently been nearly completely sequenced at the cDNA level (R. A. Wetsel and B. F. Tack, unpublished).

Figure 3 shows a comparison of the known N-terminal sequences of members of the group of α_2M-related proteins.

Human α_2M [a,b]	S V S G K P Q Y M V L V P S L L H T E T T
Human PZP [c]	T E P Q Y M V L V P S L L H T E A V
Rat α_1M [d]	A T G K P X Y V V L V P S E L Y A G V P
Rat α_2M [a]	S A P G K P I Y M V M V P S L L X A
Rat α_1I III [e]	L N G N S K Y M V L V P S Q L Y
Mouse αM [f]	D L A K P Q Y V V L V P I E
Chicken αM [g]	S T V T E P Q Y M V L L P F
Chicken ovostatin [g]	K E P E P Q Y V L M V P A
Duck ovostatin [h]	K E P E P Q Y V L M V P A
Human pro-C3 [i]	S P M Y S I I T P N I L R L E S E
Mouse pro-C3 [j]	I P M Y S I I T P N V L R L E S E
Human pro-C4 [k]	K P R L L L F S P S V V H L G V P
Mouse pro-C4 [l,m]	K P R L L L F S P S V V N L G T P

Fig. 3. Alignment of known N-terminal sequences of members of α_2M-related proteins. (a) Sottrup-Jensen *et al.* (1984c); (b) Kan *et al.* (1985); (c) Sottrup-Jensen *et al.* (1984d); (d) K. Lonberg-Holm and R. Kutny, unpublished; (e) L. Sottrup-Jensen, unpublished; (f) Hudson *et al.* (1980); (g) Nagase *et al.* (1983); (h) Nagase *et al.* (1985); (i) De Bruijn and Fey (1985); (j) Wetsel *et al.* (1984); (k) Belt *et al.* (1984); (l) Sepich *et al.* (1985); (m) Nonaka *et al.* (1985b).

IV. α_2M, PZP, C3, C4, and C5 as a Class of Homologous Proteins

A. Comparison of the Sequences of Human α_2M, C3, and C4

Earlier results had shown that the sequences of the anaphylatoxins C3a, C4a, and C5a, derived from the N-terminal 74–77 residues of the α chains of C3, C4, and C5, were related, indicating a common ancestry of these proteins (Fernandez and Hugli, 1977). In addition, the similar sequences around the thiol ester sites in α_2M, C3, and C4 and their common functional properties strongly indicated that α_2M could be structurally related to the complement proteins. When the complete sequences of human α_2M and mouse pro-C3 were compared, an evolutionary relationship was indeed apparent (Sottrup-Jensen *et al.*, 1985). Thus, eight extended stretches of similar sequence in α_2M and C3 were found, and these stretches showed between 19 and 31% identically placed residues. In all they accounted for 75 and 67%, respectively, of the polypeptide chains of α_2M and pro-C3. These results indicated that α_2M and pro-C3 would contain a number of well-defined different domains of similar structure. Since a number of larger stretches of sequence from human and mouse C4 were also homologous to stretches in α_2M and C3, C4 was judged to be structurally similar with α_2M and C3. Now, when the complete sequences of human α_2M, pro-C3, and pro-C4 are available, a detailed comparison of their sequences might reveal a fundamental domain pattern common to these and other proteins of the family of α_2M-related proteins.

Figure 4 shows an alignment of the complete sequences of human α_2M (Sottrup-Jensen *et al.*, 1984c), human pro-C3 (De Bruijn and Fey, 1985), and human pro-C4 (Belt *et al.*, 1984). In addition, available partial sequences of human PZP (Sottrup-Jensen *et al.*, 1984d; O. Sand and L. Sottrup-Jensen, unpublished) and human C5 (Lundwall *et al.*, 1985) are included. Clearly all five proteins do have many relatively long stretches of similar sequences in common, alternating with stretches of low similarity. In order to give an impression of the size of the similar stretches, which presumably constitute the "cores" of domains of similar tertiary structure in those proteins, positions containing chemically similar residues are pointed out. These residues (D = E = N = Q, T = S, V = M = I = L; Y = F = H = W and K = R) constitute groups of readily exchangeable residues, largely occupying the same positions in homologous proteins. The larger regions of pronounced homology in α_2M, C3, and C4 are emphasized in the schematic representation shown in Fig. 5.

Following a short common N-terminal region of approximately 40 residues, the next approximately 60 residues show little homology and contain stretches in C4, which constitute insertions relative to α_2M and C3. Then three homologous


PZP
α2M
C3
C4
C5

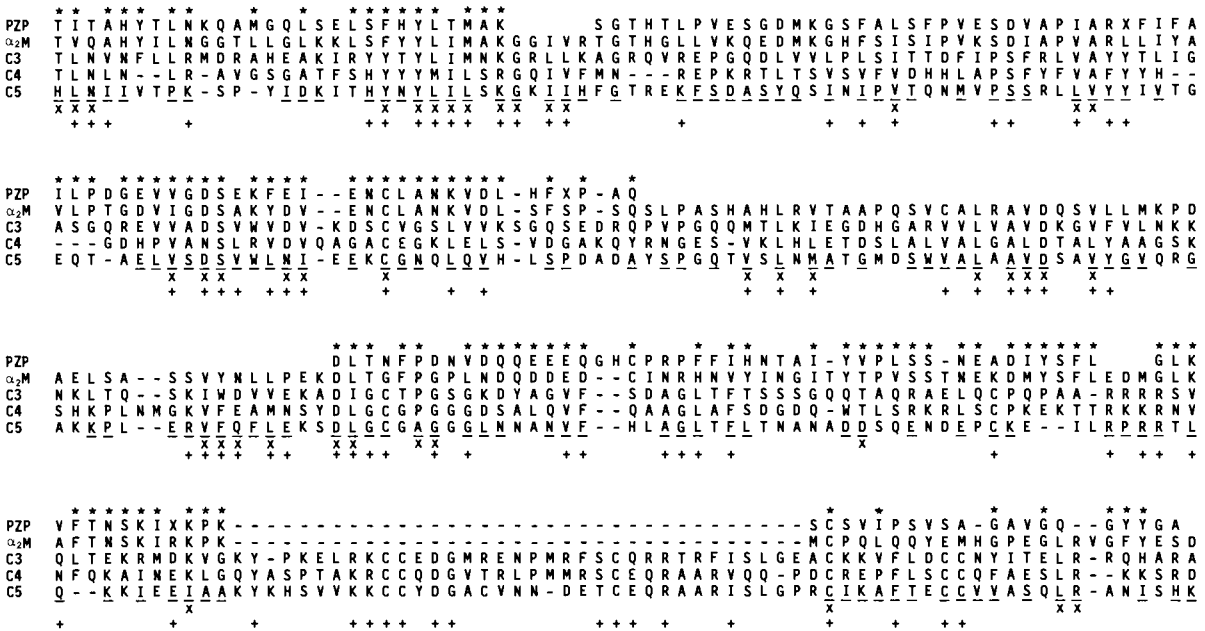
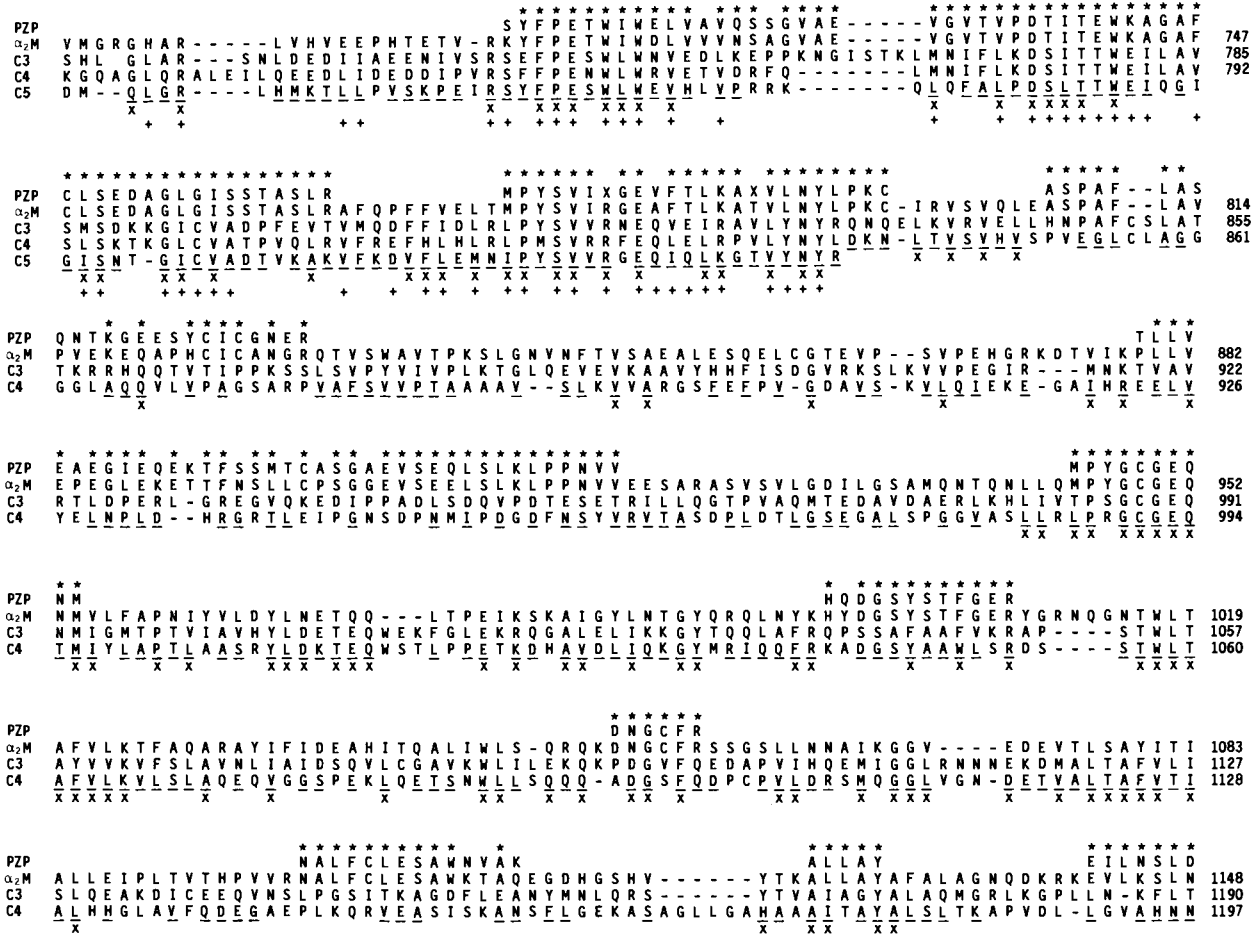


Fig. 4. Alignment of the complete sequences of human α_2M, pro-C3, and pro-C4A. Partial sequences of PZP and pro-C5 are also included. Gaps of varying size have been introduced in order to maximize the homology. Chemically similar residues are defined as: D=N=E=Q; S=T; V=M=I=L; Y=F=H=W; and K=R. The following scoring symbols are used: Identically placed residues in any pair of α_2M/C3/C4 (−). Identically placed residues in α_2M/C3/C4 (×). Identically placed residues in C3/C4/C5 (+). Identically placed residues in C3/C4 (C-terminal extension) (0). Identically placed residues in α_2M/PZP (*). Note that the Appendix to Volume IV (Barker and Putnam, 1984) gives the complete amino acid sequence of human α_2M together with structural features. The Appendix of this volume gives the complete sequences of human pro-C3, C4A, the partial sequence of C5, and other relevant structural data. (*Continued.*)


PZP
α2M
C3
C4
C5

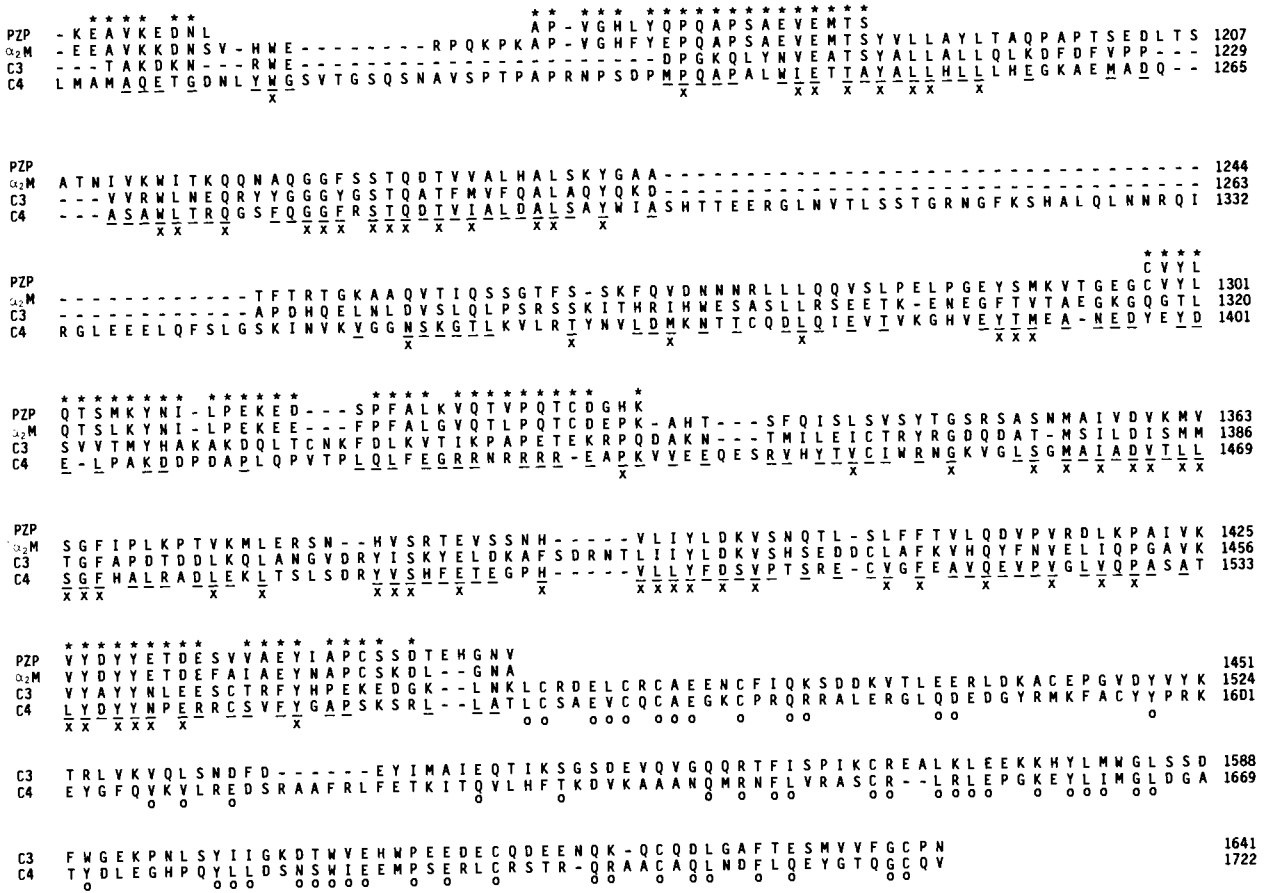


Fig. 4. (*Continued.*)

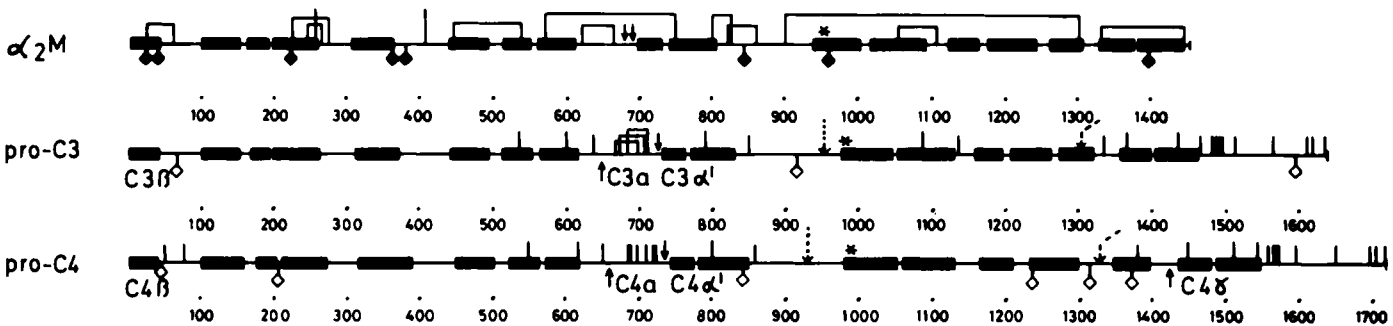


Fig. 5. Schematic representation of the structures of α_2M, C3, and C4, emphasizing the regions of pronounced mutual sequence homology. These regions, which presumably constitute common functional domains, are represented as thick bars. The disulfide bridge pattern (α_2M, C3a) or the localization of half-cysteine residues (C3 and C4) are indicated by vertical lines. The sites of cleavage in pro-C3 and pro-C4, producing the β and α chains and the β, α, and γ chains, respectively, of the mature proteins are shown as arrows below the lines. The activation cleavage sites in all three proteins are shown as arrows above the lines. The localization of the β-cysteinyl-γ-glutamyl thiol ester is shown by an asterisk. Sites of carbohydrate attachment in α_2M are indicated by filled diamonds, and the potential carbohydrate attachment sites in C3 and C4 by open diamonds. Sites of proteolytic cleavage to produce the inactivated forms of C3b and C4b, namely, C3bi and C4bi, respectively, are indicated by broken arrows. The C3d and C4d fragments are produced as a result of secondary cleavages adjacent to the thiol ester domains (dotted arrows). These fragments can readily be separated from the remaining disulfide-bridged sets of fragments, referred to as C3c and C4c, respectively. The N-terminal sequence of the 40 kDa fragment of rat α_1M is: DLSSSDLTTASKIVKWISKQ (L. Sottrup-Jensen, unpublished). This stretch matches well with residues 1199–1218 of human α_2M, indicating that the 40 kDa fragment corresponds to the C-terminal appr. 250 residues of human α_2M, and that rat α_1M is synthesized as a single chain precursor, which is proteolytically processed like proC3 and proC4.

stretches follow, corresponding to residues 96–261 in α_2M, which probably together constitute a larger domain. Preceded by a region of low similarity another homologous domain is located between residues 303 and 355 in α_2M. The next approximately 80 residues are again of low sequence similarity and contain several stretches in which short deletions and insertions occur in either sequence. Then three homologous stretches (corresponding to residues 440–610 in α_2M) follow. These could form a large common domain. Thus, as clearly seen from Fig. 5, the homologous domains of the β chains of C3 and C4 closely match those of the N-terminal 610 residues in α_2M, with only minor displacements due to small insertions in C3 and C4.

Then a stretch of remarkably low sequence similarity between α_2M and C3 or C4 follows. This stretch, which terminates in the activation cleavage sites of all three proteins, contains the sites of proteolytic processing of pro-C3 and pro-C4 to form the β and α chains of the mature proteins. The potential anaphylatoxin structures of C3 and C4 constitute a large fraction of these stretches. While these structures are mutually homologous and homologous to that of C5, α_2M has no such structure, although a weak homology between the sequences of α_2M, C3, and C4 is apparent at their activation cleavage sites (see below). Notably, C3, C4, and C5 contain as part of their anaphylatoxin structure a 34-residue insertion relative to α_2M.

Following the activation cleavage area the three proteins again contain a highly homologous domain, corresponding to residues 719–812 in α_2M. Although a number of scattered identities are found in the following approximately 140 residues, this region presumably is a distinct domain, different for the three proteins. The thiol ester sites of α_2M, C3, and C4 are located in the beginning of an approximately 150-residue stretch of very high sequence similarity (corresponding to residues 942–1091 in α_2M), indicating that this is a domain of highly conserved structure in the three proteins. Following a short stretch of dissimilar sequence, corresponding to residues 1092–1120 in α_2M, which also contains a six-residue insertion in C4, a small homologous domain is again present (corresponding to residues 1122–1160 in α_2M). Another stretch of very similar sequence (corresponding to residues 1174–1241 in α_2M) is preceded by a stretch in which both α_2M and C4 contain larger insertions relative to C3 (residues 1161–1173 in α_2M). Then a stretch of low sequence similarity follows, which in C4 contains a 45-residue insertion relative to α_2M and C3 (at residues 1243–1244 in α_2M). A new stretch of similar sequence is located between residues 1254 and 1301 in α_2M and could be part of a large domain, which also includes the following 140 residues, thus almost extending to the C-terminus of α_2M. This domain contains many small deletions and insertions. Relative to α_2M, C3 (α chain) and C4 (γ chain) contain a C-terminal extension of 159 and 163 residues, respectively. These segments of C3 and C4 have similar sequences.

Although a comparison of the sequences of human α_2M and human C3 gives an average identity of 24%, slightly lower than that of human C3 versus human C4 (29%) (De Bruijn and Fey, 1985), indicating that all three sequences are distantly related, the present analysis shows that about 60% of the sequences of α_2M, C3, and C4 in fact are strongly related. Thus their sequences contain a number of homologous domains of presumably very similar tertiary structure separated by regions of low homology. Some of the latter regions possibly form distinct domains, which could modify an underlying common gross structure of α_2M, C3, and C4, according to the requirements of each protein to specifically interact with different proteins.

B. Partial Sequences of Human PZP and C5

From the partial sequences determined so far from PZP, 68% were identically located in PZP and α_2M (Sottrup-Jensen *et al.*, 1984d). This data and additional partial sequences obtained recently (O. Sand and L. Sottrup-Jensen, unpublished) are shown in Fig. 4. Evidently, large stretches of PZP and α_2M have nearly identical sequences, indicating that the subunits of these proteins are almost identically folded.

The sequence of a cDNA clone coding for 438 residues of human pro-C5 has recently been published (Lundwall *et al.*, 1985). This sequence was found to originate in the C-terminal part of the β chain of C5, continue through the C5a portion, and terminate in the N-terminal part of the α' chain. The sequence of this stretch is also shown in Fig. 4, and is located equivalently with residues 376–795 of α_2M. Apart from the C5a portion the overall sequence homology between C5 and C3 is slightly higher than that between C5 and α_2M (47% versus 33%) based on chemically similar residues. The stretch immediately following the activation cleavage site in C5 (corresponding to residues 703–795 in α_2M) is strongly homologous with those of α_2M, C3, and C4. In fact (R. A. Wetsel and B. F. Tack, unpublished), many stretches in both chains of C5 match well with corresponding stretches in α_2M, C3, and C4. However, the Cys and the Glx residues comprising the thiol ester site are absent in C5 (see below). Taken together, these results suggest that proteolytic activation of C5 triggers basically the same type of conformational change as elicited in α_2M, C3, and C4, presumably involving a structural rearrangement of a number of smaller domains.

C. Secondary Structure Predictions

In an appendix to the sequence determination of human α_2M Welinder *et al.* (1984) analyzed the structure of the α_2M subunit by a variety of procedures for prediction of secondary structure elements in order to possibly obtain information on the domain pattern of α_2M. A composite prediction, based on agreement

of at least two methods, assigned 44% of all residues to β sheets and 8% to α-helices, with the remainder being unordered. Furthermore, a fairly regular pattern of repeating hydrophobic and hydrophilic approximately 10-residue stretches was observed. Thus, in agreement with the results of CD measurements by others (Frenoy *et al.*, 1977; Richman and Verpoorte, 1981; Björk and Fish, 1982; Gonias *et al.*, 1982b; Feldman and Pizzo, 1984b) the α_2M subunit should be classified as an "all β protein." Extending these results to the regions of strong mutual homology between α_2M, C3, and C4 shown in Fig. 5, presumed to be discrete domains of conserved structure, it is likely that most if not all of these domains are constituted by or are part of β barrels, containing 100–200 residues arranged as 5–13 antiparallel β strands (Richardson, 1981). This conclusion is further supported by the CD spectra of C3b (Molenaar *et al.*, 1975; Isenman and Cooper, 1981), which indicate the presence of very little α-helix and about 44% β sheet. Similarly, C5b contains appreciable amounts of β sheet (DiScipio *et al.*, 1983). The α-helix content of native C3 and C5 is largely confined to the anaphylatoxin domains of these proteins (Molenaar *et al.*, 1975; Hugli *et al.*, 1975; Morgan *et al.*, 1974; Huber *et al.*, 1980).

D. Distinctive Features of the Structures of α_2M, PZP, C3, C4, and C5

In contrast to the single chains of the subunit of α_2M and PZP, C3 contains two chains (α and β) and C4 contains three chains (α, β, and γ). As first pointed out by Domdey *et al.* (1982), the potential β and α chains of pro-C3 are connected by the sequence -Arg-Arg-Arg-Arg- (corresponding to residues 648–651 in α_2M), which specifically is removed prior to secretion of the mature protein. Similarly, the potential β and α chains of pro-C4 are connected by the sequence -Arg-Lys-Lys-Arg-, while the potential α and γ chains of C4 are connected by the sequence -Arg-Arg-Arg-Arg- (corresponding to residues 1346–1349 in α_2M). In contrast, the potential β and α chains of pro-C5 are connected by the sequence -Arg-Pro-Arg-Arg-. These sequences are apparently recognized by the processing proteinases, presumably consisting of a proteinase with trypsinlike specificity and a carboxypeptidase B-like enzyme. The complete absence in α_2M of such sequences is consistent with the observation that the subunits of α_2M are not proteolytically processed and exist as single 180-kDa polypeptide chains. Clearly, the sequence -Arg-Lys-Pro-Lys- (residues 661–664 in α_2M), which is located close to the corresponding β–α junctions of C3 and C4, is not recognized by the processing enzymes. In contrast to human α_2M and rat α_2M it is probable that the coding sequences of rat and mouse α_1M will reveal a tetrapeptide sequence similar to those of the complement proteins, since their subunits are proteolytically processed (Table II, Fig. 2). In pro-C4 the maturation cleavages proceed with a relatively low efficiency as revealed by the pres-

ence in plasma of significant amounts of C4 species, which contain uncleaved β and α chains, β and γ chains, or β, α, and γ chains. These incompletely processed pro-C4 molecules are apparently not hemolytically active (Karp *et al.*, 1982a,c; Chan and Atkinson, 1983, 1984; Chan *et al.*, 1983).

Apart from the many common domains of α_2M, C3, and C4 the latter two proteins (and presumably also C5) contain two distinct domains, which are not found in α_2M, namely, the 74–77 residue anaphylatoxin structures and the long C-terminal extensions. As shown by X-ray structure analysis by Huber *et al.* (1980), C3a principally consists of two α-helices (Tyr-15 to Met-27 and Gly-46 to Ser-71, relative numbering) connected by three disulfide bridges, overall resembling a drumstick. The C-terminal extensions of C3 and C4 contain no less than eight identically located half-cystine residues and probably form an intricately folded domain containing a cluster of four disulfide bridges.

It is evident that C3 and C4 are more related than either of these proteins is to α_2M. From the available data this seems to be the case also for C5. On the other hand, PZP is strongly related to α_2M. This in combination with differences in the quaternary structure and physiological role justifies the distinction between the macroglobulin subgroup and the complement subgroup made in Section III. The anaphylatoxin structures and long C-terminal extensions of the complement proteins could represent pieces of distinct genetic origin. Alternatively they could represent material deleted from an ancestral gene. Doolittle (1984) has suggested that the divergence of the genes for the α_2M subgroup and the complement subgroup took place approximately 500 million years ago. Subsequently, divergence of C3, C4, and C5 occurred. Presumably the divergence of α_2M and PZP is a relatively recent event, although the existence of two distinct α-macroglobulins in the frog (Feldman and Pizzo, 1985, 1986) indicates that this may have occurred early.

In contrast to α_2M, where eleven intra- and two interchain disulfide bridges (the latter involving Cys-255 and Cys-408) have been located (Sottrup-Jensen *et al.*, 1984c; Sottrup-Jensen, 1985; P. E. Jensen and L. Sottrup-Jensen, unpublished), only the three bridges in C3a (and by analogy C4a and C5a) have been identified (Huber *et al.*, 1980). Given the many identically located half-cystine residues in C3 and C4, it is likely that their bridge patterns will turn out to be very similar and, importantly, different from that of α_2M. Characteristically, the β chains of C3 and C4, corresponding to the N-terminal 610 residues of α_2M, contain only three (C3) or five (C4) half-cystine residues, unlike the thirteen found in that part of α_2M. From the known pattern of proteolytic degradation of C3 and from partial reduction experiments on C3c (Matsuda *et al.*, 1985), it is likely that one of the two half-cystine residues in the region between the activation cleavage site and the thiol ester site (Fig. 5) is bridged to a half-cystine residue in the β chain, while the other is bridged to a half-cystine residue in the C-terminal part of the α chain. Assuming that the interchain bridges in C4 are identically located, the latter bridge will connect the α and γ chains of C4,

further suggesting the presence of an additional bridge spanning across the α–γ junction. It thus appears that α_2M, C3, and C4 do have a disulfide bridge in common, which spans across 400–600 residues in the linear sequence. This might be important for transmitting the activation cleavage signal to distant parts of the structures. The pronounced differences in the disulfide bridge pattern between α_2M on the one hand and C3 and C4 on the other hand presumably serve to precisely orient the individual domains of α_2M, C3, and C4, thereby specifically creating slightly different versions of the underlying common gross structure.

The sequence location of the eight glucosamine-based carbohydrate groups, comprising approximately 10% by weight of α_2M, is known. The sites of attachment are asparagine residues at positions 32, 47, 224, 373, 387, 846, 968, and 1401 (Sottrup-Jensen *et al.*, 1984c). According to the analyses by Dunn and Spiro (1967a,b), the size and charge of the oligosaccharide chains in α_2M are highly heterogeneous. The smallest carbohydrate groups appear to consist of three mannose and two *N*-acetylglucosamine units, to which variable amounts of galactose, mannose, *N*-acetylglucosamine, *N*-acetylneuraminic acid, and fucose are added. The cDNA sequence encoding human pro-C3 (De Bruijn and Fey, 1985) predicts a single residue in the β chain (Asn-63) and two residues in the α chain (Asn-917 and Asn-1595) as potential carbohydrate attachment sites, compatible with the finding that both chains of C3 contain a few percent carbohydrate (Tack *et al.*, 1979b). Similarly the cDNA sequence encoding pro-C4 (Belt *et al.*, 1984) reveals two residues in the β chain (Asn-41 and Asn-207) and four residues in the α chain (Asn-843, Asn-1230, Asn-1310, and Asn-1372), which are candidate carbohydrate attachment sites. Earlier results had indicated that the α and β chains but not the γ chain of C4 contain 6–8% carbohydrate (Gigli *et al.*, 1977). Recently Chan and Atkinson (1985) found that the α chain of human C4 contains three complex fucose-containing oligosaccharide groups of the biantennary type, while the β chain contains a single high-mannose oligosaccharide group with nine residues of mannose and two residues of *N*-acetylglucosamine. Apart from the thiol ester structure, which represents a novel postsynthetic modification common to α_2M, C3, and C4 (Sections VII and VIII), the α chain of C4 has been reported to contain a residue of tyrosine *O*-sulfate, not found in α_2M, C3, and C5 (Karp, 1983b), and located in the C-terminal part of the α chain (Chan and Atkinson, 1985).

E. Activation Cleavage Regions of α_2M, PZP, C3, C4, and C5

Although α_2M, PZP, C3, C4, and C5 evidently are homologous proteins activated by specific limited proteolysis, an examination of the sequences around their activation cleavage sites reveals significant differences, which undoubtedly relate to the widely differing substrate specificity of the respective activating

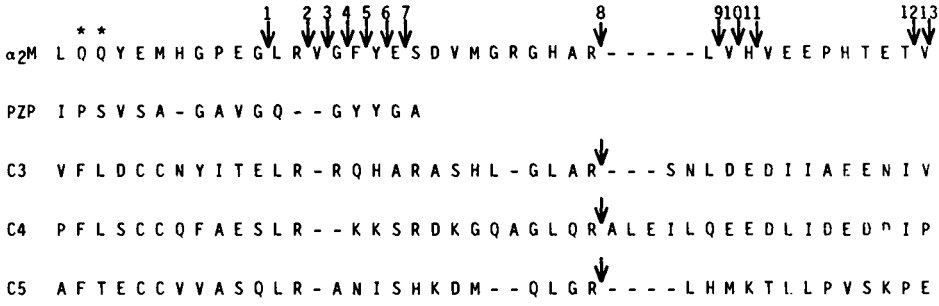


Fig. 6. Comparison of the sequences around the activation cleavage sites of α_2M, PZP, C3, C4, and C5. Cleavage sites are indicated by an arrow. The cleavage sites in PZP are not yet known. Arrows numbered 1 thorugh 13 denote cleavage sites in α_2M for different proteinases: (1) proposed for mammalian collagenases; (2) trypsin, *Streptomyces grieus* trypsin, papain; (3) porcine pancreatic elastase; (4) porcine pancreatic elastase, papain; (5) chymosin, cathepsin G; (6) chymotrypsin; (7) *Staphylococcus aureus* proteinase; (8) trypsin, plasmin, thrombin, thermolysin, subtilisin, *Streptomyces griseus* trypsin; (9) subtilisin, *Streptomyces griseus* proteinase B; (10) human leukocyte elastase; (11) papain; (12) thermolysin; (13) chymosin. The primary sites of cleavage in the bait region of α_2M are those numbered 2 through 7. Putative transglutaminase acceptor sites in α_2M are shown by asterisks. The physiological activators for the complement proteins are C1s for C4, C4b2a or C3bBb for C3, and C4b2a3b or $(C3b)_nBb$ for C5.

proteinases. Thus, the activators of C3, C4, and C5 have a very narrow specificity, while α_2M, PZP, and other macroglobulins can interact with and become activated by many different proteinases differing in their catalytic mechanism and substrate specificity. Figure 6 shows a comparison of the activation cleavage sites for all five proteins. *In vivo,* C4 is activated by the proteinase C1s, C3 by a complex between C2a and C4b or Bb and C3b, and C5 by a complex between C2a, C3b, and C4b or by Bb and aggregated C3b (Fig. 1). All these proteinases basically have trypsinlike specificity as seen from the cleavage of a single arginyl bond in native C3, C4, and C5, and these proteins appear to be the only natural substrates for those enzymes. Presumably, the conformations of the polypeptide chains of C3, C4, and C5 around the cleavage sites differ, forming extended, precisely folded secondary substrate specificity determinants, thus meeting the requirement of each activator. However, the fact that the same proteinase (C2a or Bb) can activate both C3 and C5, dependent on the particular proteins present in the activating complex, is puzzling and points toward subtle changes in the substrate binding area of those proteinases effected by the bound cofactor proteins. Perhaps this is reminiscent of the change in substrate specificity of plasmin, which occurs following complex formation with streptokinase (Kline and Fishman, 1961). Alternatively, it has been suggested, in view of the lack of activation cleavage in C5 by trypsin (Nilsson *et al.*, 1975), that C3b when bound to C5 (Vogt *et al.*, 1978; Isenman *et al.*, 1980; DiScipio, 1981a,b) induces a

conformational change in C5 thereby exposing its activation cleavage site (Vogt *et al.*, 1978).

The primary sites of cleavage in the bait region of α_2M for a variety of proteinases are found in the stretch -Arg^{681}-Val-Gly-Phe-Tyr-Glu-, while fast secondary cleavage preferentially takes place in the stretch -Arg^{696}-Leu-Val-His-. In each case investigated the observed sites of cleavage in α_2M reflect the known primary substrate specificity requirement of the proteinase in question. For those proteinases, where cleavage in the primary cleavage area has not been observed, it is likely that initial cleavage in that region can take place, considering their substrate specificity. These results strongly indicate that the hexapeptide stretch in α_2M (residues 681–686) containing the primary sites of cleavage is uniquely exposed and presumably constitutes a flexible segment, whose conformation readily adapts to the conformation necessary for fast cleavage. In this respect it is of interest that α_2M reacts very fast with proteinases having a relatively broad substrate specificity like trypsin, chymotrypsin, and elastase, while more sophisticated proteinases, such as thrombin and plasmin, which are more dependent on extended substrate binding regions, react slowly (Christensen and Sottrup-Jensen, 1984) (see also Section IX). Although mammalian collagenases are known to form complexes with α_2M (Werb *et al.*, 1974), the site of cleavage is not known. However, as pointed out by Mortensen *et al.* (1981b), the Gly-Leu bond adjacent to the hexapeptide sequence might constitute a specific cleavage site for those collagenases.

Intriguingly, the preferential sites of secondary cleavage closely match the position of the cleavage sites of the complement proteins, and a modest sequence similarity between all sequences is apparent. Under conditions of low proteinase : α_2M ratios, which presumably prevail *in vivo*, it is not known whether cleavage actually takes place at these secondary sites. The presence of two major cleavage areas in the bait region of α_2M raises the question of whether the resultant small peptides can be released from the α_2M–proteinase complex and have any physiological function. Some of these potential peptides clearly resemble the C-terminal active parts of the anaphylatoxins (Caporale *et al.*, 1980; Gerard *et al.*, 1985).

The bait region sequence of PZP, which is also a proteinase-binding protein (Sand *et al.*, 1985), differs markedly from that of α_2M, strongly indicating that PZP and α_2M have different inhibitory spectra (see also Section IX).

Following proteolytic cleavage in the activation region each protein undergoes a conformational change. The structural features for this change, which in the case of α_2M, PZP, C3, and C4 ultimately leads to activation of their internal thiol esters, are not known. It is possible, in analogy with other proteins activated by specific limited proteolysis, that they might reside in the strongly homologous domains just downstream of the activation cleavage sites. In particular, the domain corresponding to residues 740–810 in α_2M has the highest average

hydropathic index of any long stretch in α_2M (and C3 and C4). Thus, it is possible that this domain could constitute a hydrophobic core region, through which the structural constraints released by activation cleavage could be transmitted to other parts of the polypeptide chain, ultimately causing rearrangement of the thiol ester domain(s) and further resulting in activation and exposure of this site (Welinder *et al.*, 1984) (see also Section VIII).

F. A Putative Factor $XIII_a$ Cross-linking Site in α_2M

Apart from fibrinogen and plasma fibronectin, α_2M is a major substrate for factor $XIII_a$ (plasma transglutaminase) (Mosher, 1976; Van Leuven *et al.*, 1981a). The sites of specific incorporation of dansylcadaverine or putrescine into α_2M have been determined as Gln-671 (major site) and Gln-670 (minor site) (Mortensen *et al.*, 1981a). These sites are close to the primary cleavage sites in the bait region of α_2M (Fig. 6) and the accessibility of these residues could possibly be influenced by cleavage in that region. Thus, Van Leuven (1984), using purified factor $XIII_a$, could not demonstrate incorporation of amines into Gln-671 in an α_2M–trypsin complex. Preliminary investigations have indicated that no major cross-linking of α_2M takes place to other proteins during the *in vitro* clotting of blood (Mosher, 1976; Van Leuven *et al.*, 1981a; Sottrup-Jensen *et al.*, 1983a), and the physiological relevance of this site remains to be determined (see also Section X). It is presently not known if PZP, C3, C4, and C5 are substrates for transglutaminases.

G. The Thiol Ester Site

Figure 7 shows an alignment of the sequences around the thiol esters of human α_2M, C3, and C4 and mouse C4 and Slp. In addition, the corresponding stretch of mouse C5 has been included (R. A. Wetsel and B. F. Tack, unpublished). The consensus sequence -Gly-Cys-Gly-Glu-Gln- of human α_2M, PZP, C3, and C4 is not found in mouse C4 and Slp in which an Ala residue replaces the second Gly residue. In Slp the first Gly residue is replaced by a Ser residue. As judged from the generation of SH groups by incubation with methylamine, mouse C4 and Slp do have a functional thiol ester site (Karp *et al.*, 1982c). However, Slp is apparently hemolytically inactive and cannot be cleaved by activated C1s (Ferreira *et al.*, 1978), perhaps due to the presence of several charged residues at the putative activation cleavage site of Slp (-Arg-Lys-Val-Arg-Asp- in Slp versus -Arg-Asn-Asn-His-Asn- in mouse C4 in which the Arg-Asn sequence of mouse C4 is the activation cleavage site) (Nonaka *et al.*, 1984b).

Intriguingly, both components of the thiol ester site, the Cys residue and the Gln residue, are not found in mouse C5, being replaced by a Ser residue and an Ala residue, respectively. This conclusively demonstrates for the first time that

							*				*						
Human α_2M	L	L	Q	M	P	Y	G	C	G	E	Q	N	M	V	L	F	A	P
Human PZP				M	P	Y	G	C	G	E	Q	N	M					
Human C3	L	I	V	T	P	S	G	C	G	E	Q	N	M	I	G	M	T	P
Human C4	L	L	R	L	P	R	G	C	G	E	Q	T	M	I	Y	L	A	P
Mouse C4	L	L	R	L	P	Q	G	C	A	E	Q	T	M	I	Y	L	A	P
Mouse Slp	L	L	R	L	P	R	S	C	A	E	Q	T	M	I	Y	L	A	P
Mouse C5	L	T	H	L	P	K	G	S	A	E	A	E	L	M	S	I	A	P

Fig. 7. Alignment of the sequences around the β-cysteinyl-γ-glutamyl thiol esters of human α_2M, PZP, C3, and C4 and mouse C4 and Slp. The corresponding stretch from mouse C5 is also shown. The Cys and the Gln residues comprising the thiol ester structure are indicated by asterisks. Residues that differ from the pentapeptide sequence -Gly-Cys-Gly-Glu-Gln- found in the human proteins are underlined.

C5, although evidently strongly homologous to α_2M, PZP, C3, and C4, does not contain an internal thiol ester. The polypeptide stretches adjacent to the thiol ester residues are pronouncedly hydrophobic; which in combination with the conserved Pro residues may specify a particular conformation of that site concealed in the native structure (see also Section VIII).

H. Sites of Inactivation Cleavage in C3 and C4

Conformationally changed C3 and C4 (the C3b- and C4b-like conformations, respectively) produced either as a result of proteolytic activation or by incubation with reagents known to inactivate these proteins, e.g., amines and chaotropes, are susceptible to a number of specific cleavages. These physiologically important cleavages, which take place exclusively in the α chains of C3 and C4 (or the α′ chains of C3b and C4b), are dependent (in the fluid phase) on the concerted action of a serine proteinase, factor I (C3b/C4b inactivator), having trypsinlike specificity and two cofactors, factor H for C3b and C4b-binding protein for C4b. The initial cleavage(s) take place at Arg-1281 and Arg-1298 in C3 and at Arg-1318 in C4 and result in inactivated C3b and C4b (C3bi and C4bi, respectively). Subsequent cleavages by factor I or other yet unidentified proteinases take place in the peptide stretch immediately preceding the thiol ester domain and produce the C3d (C4d) and C3c (C4c) fragments (Fig. 5) (Tamura and Nelson, 1967; Lachmann and Müller-Eberhard, 1968; Davis, 1981; Whaley and Ruddy, 1976; Pangburn *et al.*, 1977; Fujita *et al.*, 1978; Nagasawa *et al.*, 1980; Ferreira *et al.*, 1977; von Zabern *et al.*, 1982; E. Sim *et al.*, 1981; Harrison and Lachmann, 1980a,b; Press and Gagnon, 1981; Davis and Harrison, 1982; Davis *et*

al., 1984). When C3b and C4b are bound to the cellular C3b/C4b receptor (CR1) they can also be cleaved by factor I and other proteinases. CR1 thus may serve a cofactor role similar to that of factor H and C4b-binding protein in the fluid phase (Law *et al.*, 1979a; Ruddy and Austen, 1971; Gitlin *et al.*, 1975; Fearon, 1979, 1980; Ross *et al.*, 1982; Medicus *et al.*, 1983; Medof *et al.*, 1982; Medof and Nussenzweig, 1984; Iida and Nussenzweig, 1981). Presumably, these cleavages take place in relatively exposed flexible surface loops, connecting larger functional domains, and are important for the subsequent inactivation and elimination of activated C3 and C4 (see Section X). Other proteinases like trypsin and elastase will also produce similar cleavages. Although trypsin and plasmin have been reported to slightly degrade the β chains of C3 and C4 (Paques, 1980; Nagasawa *et al.*, 1980) no cleavages of physiological significance have been identified in the β chains, which probably form tightly packed structures. A cleavage pattern resembling that of the initial cleavage of C3b and C4b by factor I has also been described in C5 (Nilsson *et al.*, 1975; Yamamoto and Gewurz, 1978; Podack *et al.*, 1978; Wetsel *et al.*, 1980); however, the significance of this remains unclear.

Apart from a minor degradation product from α_2M–trypsin complex (Pan *et al.*, 1980), apparently constituted by residues 961–978 (Sottrup-Jensen *et al.*, 1984c), and that of α_2M–chymotrypsin complex (Harpel, 1973), probably representing a slight secondary proteolysis, the only cleavages observed for α_2M are those associated with the activation cleavages in the bait region. Even at proteinase : α_2M ratios exceeding the binding capacity of α_2M no extensive fragmentation is found, indicating an exceptionally tight packing of the individual domains of the α_2M subunit. However, a recent report (Carlsson *et al.*, 1984) has indicated that proteinases from *Bacteroides gingivalis* (W83, H185) can degrade α_2M (and C3) extensively.

V. Chromosome Assignment, Genetic Variation, and Sites of Synthesis

A. C3

Using somatic cell hybrids between human primary fibroblasts and mouse fibroblastlike cell lines, which carried a limited number of different human chromosomes, the gene for human C3 was localized to chromosome 19 (Whitehead *et al.*, 1982). From the observation that the mouse C3 gene is located outside but linked to the major histocompatibility complex it was localized to chromosome 17 (Penalva da Silva *et al.*, 1978; Natsuume-Sakai *et al.*, 1979b). Most likely there is only one gene for mouse and human C3 (Whitehead *et al.*, 1982; Wiebauer *et al.*, 1982), and the size has been estimated at 24 kb. As the

coding sequence for C3 is only about 5 kb this gene evidently contains a number of introns. The first intron, containing approximately 1000 bases, is located close to the junction between the signal peptide and the β chain portion of C3 (Wiebauer *et al.*, 1982). The C3 gene is polymorphic and the expression is controlled by a single codominant locus (Natsuume-Sakai *et al.*, 1978a). The variants of C3 can be detected by electrophoresis (Penalva da Silva *et al.*, 1978; Alper and Propp, 1968; Rittner and Rittner, 1973; Wieme and DeMulenaare, 1967; Natsuume-Sakai *et al.*, 1978a) or as antigenically distinct forms of C3 (Natsuume-Sakai *et al.*, 1978b, 1979a; Nonaka *et al.*, 1980). Recently a restriction fragment polymorphism was described, which correlated with a protein polymorphism of C3 (Donald and Ball, 1984). The allotypic differences between three mouse proteins (C3-1 alleles) were localized by tryptic fingerprinting to the C3c portion of C3 and were attributed to differences in amino acid sequence (Nonaka *et al.*, 1980). All these C3 variants contained a β chain, which was about 9 kDa smaller than the human β chain. This can perhaps be explained by a difference in carbohydrate content (Fey *et al.*, 1980). The sequence results of Lundwall *et al.* (1984), Wetsel *et al.* (1984), and DeBruijn and Fey (1985) show that the lengths of the β chains of human and mouse C3 differ by only three residues.

Inherited partial or total C3 deficiencies are rare (Alper *et al.*, 1969; Ballow *et al.*, 1975) as reviewed by Alper and Rosen (1984). Individuals with C3 deficiency are subjected to recurrent infections, due to the reduction or the absence of the functions provided by C3 (Alper *et al.*, 1972, 1976; Roord *et al.*, 1983).

Although the primary sites of C3 synthesis are the hepatocytes of the liver (Alper *et al.*, 1969; Brade *et al.*, 1977; Patel and Minta, 1979; Odink *et al.*, 1981; Misumi *et al.*, 1984), several other cell types are also capable of producing C3, e.g., macrophages and monocytes (Cole *et al.*, 1980; Goldberger *et al.*, 1981), fibroblasts (Senger and Hynes, 1978; Whitehead *et al.*, 1981), and peripheral lymphocytes (Sundsmo, 1980). However, monocytes apparently do not secrete functionally active C3 (and C4 and C5 as well) into the culture medium (Einstein *et al.*, 1977; Whaley, 1980). C3, like C4 and C5, is an acute-phase reactant. However, the rise in plasma level is only moderate (Kushner, 1982). In pregnancy the level of C3 is also increased (Propp and Alper, 1968). Except for a possible induction of C3 synthesis by glucocorticoids in a rat hepatoma line (Strunk *et al.*, 1975) no major inflammatory inducers are known.

B. C4

In contrast to the single gene found for C3, several studies have revealed that there are two genes for human C4 (O'Neill *et al.*, 1978a,b; Awdeh *et al.*, 1979; Awdeh and Alper, 1980) located within the class III genes in the major histocompatibility complex on chromosome 6 (Raum *et al.*, 1980; Carroll *et al.*, 1984).

The size of the C4 genes is about 40 kb, and they are separated by about 10 kb of DNA (Carroll and Porter, 1983; Carroll *et al.*, 1984). Thus, the genes for C4 contain about 35 kb of intron sequence. The two human gene products, C4A and C4B, have been shown to correlate with the erythrocyte antigens Rodgers and Chido, respectively, which represent cell-bound C4d fragments (O'Neill *et al.*, 1978a,b; Tilley *et al.*, 1978). Similarly, two forms of C4 are found in the mouse, C4 and its isotype Slp, encoded by closely spaced genes within the major histocompatibility complex on chromosome 17 (Passmore and Schreffler, 1970; Roos *et al.*, 1978; Parker *et al.*, 1979; Meo *et al.*, 1975; Lachmann *et al.*, 1975; Carroll and Capra, 1979). For human C4A, 13 alleles have been found, and for C4B no less than 22 alleles are known (Mauff *et al.*, 1983a), detected either by electrophoresis or by specific antisera (Awdeh and Alper, 1980; O'Neill *et al.*, 1978a). Further typing of the polymorphism of C4 by restriction fragment analysis (Palsdottir *et al.*, 1983; Whitehead *et al.*, 1984) has indicated an additional subdivision of C4 allotypes not revealed by electrophoresis.

Most of the differences between the C4 alleles are found in the α chain part, and particularly in the C4d fragment (Tilley *et al.*, 1978; Mevag *et al.*, 1981; Roos *et al.*, 1982) as further studied by sequence analysis (Lundwall *et al.*, 1981; Chakravarti *et al.*, 1983; Belt *et al.*, 1984; Hellman *et al.*, 1984), but differences in the β chain portion are also seen (Mauff *et al.*, 1983b). From the results of Lundwall *et al.* (1981) and Hellman *et al.* (1984) it is probable that the polymorphism at positions 1054 (Gly/Asp), 1101 (Pro/Leu), 1102 (Cys/Ser), 1105 (Leu/Ile), and 1106 (Asp/His) indeed represent a class difference between C4A (Fig. 4) and C4B as seen from the analysis of distinct peptides covering these positions. The sequence differences at positions 1157 (Asn/Ser), 1182 (Ser/Thr), 1188 (Val/Ala), 1191 (Leu/Arg), 1267 (Ser/Ala), and 1281 (Arg/Val) could represent class differences or, more likely, allelic differences. In spite of all these and other differences not yet detected the sequences of individual C4A or C4B molecules are probably nearly identical. This will presumably also be the case for mouse C4 and Slp (Nonaka *et al.*, 1984b, 1985b; Levi-Strauss *et al.*, 1985; Ogata *et al.*, 1983; Sepich *et al.*, 1985).

Additional sources of heterogeneity in the C4 molecules reside in their incomplete maturation processing (see Section IV), in differences in carbohydrate content, and in differences in the length of the α chains. In some alleles of mouse C4 the α chains are 4 kDa smaller than those in other alleles, due to the absence of some of the carbohydrate groups (Karp *et al.*, 1982a,c). The α chain of Slp is 2 kDa larger than the common α chain of mouse C4, due to the presence of additional carbohydrate groups (Karp *et al.*, 1982b; Karp, 1983a). Following synthesis and secretion both mouse C4 and human C4 gradually are converted to a form in which the α chain is about 5 kDa smaller, due to cleavage in the C-terminal part by unidentified plasma proteinases (Karp *et al.*, 1982c; Chan *et*

al., 1983; Chan and Atkinson, 1984). The variations in the degree of glycosylation are correlated with variations in the hemolytic activity (Karp *et al.*, 1982a).

Many deficiencies in C4 are known, including half-null haplotypes (no C4A or C4B) and double null-haplotypes (Hauptmann *et al.*, 1974; Tappeiner *et al.*, 1978; Ballow *et al.*, 1975; Minta *et al.*, 1981; Awdeh *et al.*, 1981; Kjellman *et al.*, 1982; Mascart-Lemone *et al.*, 1983). In contrast to these deficiencies, which show a recessive type of inheritance, an autosomal dominant type of inheritance of C4 not linked to the major histocompatibility complex has recently been described in a family (Muir *et al.*, 1984). Affected individuals are subjected to recurrent infections and systemic lupus erythematosus. However, C4-deficient guinea pigs have been reported to be normal in all respects (Ellman *et al.*, 1970; Colten, 1982).

Apart from hepatic synthesis (Hall and Colten, 1977; Karp *et al.*, 1982c) C4 and Slp are also synthesized by macrophages and monocytes (Roos *et al.*, 1978; Goldberger and Colten, 1980; Parker *et al.*, 1980; Chan *et al.*, 1983). In some strains of mice the synthesis of Slp is inducible (androgen dependent), while it is constitutive in others (Passmore and Schreffler, 1970; Hansen and Schreffler, 1976; Passmore *et al.*, 1980). Slp is apparently without hemolytic activity in standard assays for C4 and cannot be cleaved by C1s (Ferreira *et al.*, 1978). In contrast, both human C4A and C4B are active. Their hemolytic activity differs and is dependent on the state of proteolytic processing and the amount of carbohydrate present. It appears that C4A and C4B intrinsically have different covalent binding properties (see Section VII).

C. C5

The chromosome assignment of C5 has not yet been made. A few deficiencies in C5 are known (Rosenfeld *et al.*, 1976). Recently, a sexual dimorphism of mouse C5 has been described (Baba *et al.*, 1984). Using isoelectric focusing, females of all strains tested were found to contain one C5 band, while males contained two bands, one of which was identical to that found in females. The expression of the other C5 band could be induced by testosterone, in conformity with earlier reports (Cinader *et al.*, 1964; Urbach and Cinader, 1966; Churchill *et al.*, 1967) that the level of C5 in males is about twice that in females and can be decreased by the administration of estradiol or by castration. In analogy with mouse C4/Slp these results could indicate the presence of two genes for C5 or C5-like molecules.

D. α_2M, PZP, and Other Macroglobulins

Using the somatic cell hybrid technique the gene for human α_2M, of which apparently only one copy exists, has been localized to chromosome 12 (Kan *et*

al., 1985). The question of whether α_2M is polymorphic like C3 and C4 has not been carefully investigated. However, in the cDNA sequence of Kan *et al.* (1985) a Val residue was predicted at position 977, whereas an Ile residue was found by protein sequence determination (Swenson and Howard, 1980; Sottrup-Jensen *et al.*, 1984c) or by sequence analysis of a different cDNA clone (Å. Lundwall, B. F. Tack, and L. Sottrup-Jensen, unpublished). This indicates the presence of allelic forms of the α_2M gene. In the sequence work of Sottrup-Jensen *et al.* (1984c) no evidence for heterogeneity was obtained. However, due to the inherent limitations in protein sequence determination, minor peptide variants could easily have been overlooked. According to Berg and Bearn (1966) human α_2M exhibits a type of X-linked polymorphism, the Xm trait, detected by immunological procedures. This can now be questioned in view of the recent demonstration that PZP, which is usually present in higher amounts in females than in males, is antigenically related to α_2M. The AL-M system described by Leikola *et al.* (1972), which shows an autosomal dominant type of inheritance, could represent true allelic differences, while the electrophoretic polymorphism described by Gallango and Castillo (1974, 1975) cannot be readily evaluated and should be regarded as tentative. Two alleles for rabbit α_2M have been described. Apparently, in the heterozygous rabbit, both allotypic specificities are found in α_2M, indicating that each allelic gene contributes to the same α_2M molecule (Knight and Dray, 1968a,b).

Total deficiencies in α_2M must be exceedingly rare, since Laurell and Jeppson (1975) examined more than 100,000 sera and found none. Only three cases of inherited partial α_2M deficiency are known (Mahour *et al.*, 1978; Bergqvist and Nilsson, 1979; Steenbjerg, 1981). In these cases the defect was inherited as an autosomal dominant trait and did not lead to clinical manifestations.

As with most plasma proteins the major site of synthesis of α_2M is the liver. Thus Kan *et al.* (1985) found a very high representation of α_2M-specific clones in the cDNA library prepared by Belt *et al.* (1984) from human liver. However, in addition to synthesis by hepatocytes, synthesis has also been demonstrated in cultured human fibroblasts (Mosher and Wing, 1976), monocytes, and macrophages (Mosher *et al.*, 1977; Hovi *et al.*, 1977; White *et al.*, 1980), in a variety of melanoma cells (Morgan, 1981), and in other human tumor cells (Bizik *et al.*, 1986). In the mouse α_2M synthesis has been demonstrated in cultured peritoneal macrophages (White *et al.*, 1981). From tissue culture studies synthesis of rat α_1M and α_2M has been found in the liver (Weimer *et al.*, 1965; Benjamin and Weimer, 1966). These studies have been extended using cultures of rat hepatocytes (Andus *et al.*, 1983a,b; Northemann *et al.*, 1983; Koj *et al.*, 1984; Gross *et al.*, 1984; Guillouzo *et al.*, 1984; Bauer *et al.*, 1984; Baumann *et al.*, 1984).

Among the two αM's found in different animal species rat α_2M and rabbit α_1M are definitively acute-phase reactants (Ganrot, 1973a,b; Lebreton de Vonne

and Mouray, 1968). The regulation of their synthesis is not well understood, but according to recent results the hepatic synthesis is under control of a number of factors, of which the synergistic action of glucocorticoids and interleukin-1 and other factors has been recognized (Bauer *et al.*, 1984, 1985; Gross *et al.*, 1984; Guillouzo *et al.*, 1984; Koj *et al.*, 1984; Bauman *et al.*, 1984; Hirata *et al.*, 1985). It has been reported that rat α_1-inhibitor III is a negative acute-phase reactant (Gauthier and Ohlsson, 1978). However, its putative analog in the mouse (pregnancy-associated murine protein-1) seems to be under hormonal control, being induced by estrogens and suppressed by testosterone (Hau *et al.*, 1982).

In contrast to human α_2M, whose level is only slightly increased in pregnancy or during treatment with oral contraceptives (Ganrot and Bjerre, 1967; Mendenhall, 1970; Horne *et al.*, 1970; Chandra, 1972; Chandra *et al.*, 1973), the level of PZP is greatly increased in pregnancy (Bohn, 1971; von Schoultz, 1974; Than *et al.*, 1976; Folkersen *et al.*, 1981a; Stimson, 1972). A significantly increased plasma level of PZP is also seen as a result of oral contraception (DeAlvarez and Afonso, 1967; Beckman *et al.*, 1971, 1973a; Stimson, 1972, 1974; Straube *et al.*, 1973; Horne *et al.*, 1973; Berne, 1973, 1976; Bohn, 1974; Sturdee *et al.*, 1976) and suggests a regulation by estrogens. Likewise high levels of PZP have also been observed in men who received estrogens during treatment of prostatic cancer (Cooper, 1963; Beckman *et al.*, 1973b; Damber *et al.*, 1978; Lanson *et al.*, 1979). The level of PZP is also increased in a number of neoplastic and autoimmune diseases (Berne, 1976; Bohn, 1972; Than *et al.*, 1975b; Stimson, 1975; Bundshuh *et al.*, 1975; Gropp *et al.*, 1977; Anderson *et al.*, 1979; Bauer *et al.*, 1979; Hofmann *et al.*, 1979; Kasukawa *et al.*, 1979), indicating that PZP should be classified as an acute-phase reactant (Bohn, 1972; Horne *et al.*, 1975). The acute-phase protein rat α_2M is also increased in pregnancy (Beaton *et al.*, 1961; Heim, 1962; Weimer and Benjamin, 1965) thereby possibly linking the acute-phase αM's and the pregnancy-associated αM's.

PZP can be detected in plasma a few weeks after conception (Damber *et al.*, 1977), and it has been reported that the initial level of PZP in pregnancy is significantly below the normal range in women who later suffer from spontaneous abortion (Beckman *et al.*, 1974; Berne, 1976; Than *et al.*, 1975a; Damber *et al.*, 1978). The level of PZP returns post partum to the basal within 2 to 8 weeks (Bohn, 1974; Lin *et al.*, 1976; Stimson, 1975). Although PZP thus appears to be regulated by estrogens, no correlation has been found between its plasma level and the level of endogenous estrogens or gonadotropins (Damber *et al.*, 1976, 1977; von Schoultz and Stigbrand, 1982; Westergaard *et al.*, 1982). Estrogen-dependent synthesis of PZP has been demonstrated in cultures of liver cells and normal peripheral leukocytes, judged to originate from the content of monocytes and macrophages (Stimson and Blackstock, 1975; Horne *et al.*, 1978). Neoplastic cell lines derived from histiocytic lymphomas of the monocyte

type synthesize and secrete relatively large amounts of PZP (Lundgren *et al.*, 1979). However, recent results indicate that the primary sites of synthesis are the hepatocytes of the liver (Stimson, *et al.*, 1979; von Schoultz and Stigbrand, 1982).

VI. Shape of the Macroglobulins, C3, C4, and C5

A. Macroglobulins

Using the shadow-casting technique Höglund and Levin (1965) obtained electron micrographs of osmic acid- or formaldehyde-fixed human α_2M indicating a nearly spherical shape about 200 Å in diameter. In a subsequent study Bloth *et al.* (1968) described the characteristic monogramlike shape of human α_2M and rabbit α_1M and α_2M, represented by the Cyrillic letter ж, which has since been associated with α_2M (Fig. 8). They used the negative-contrast method, employing sodium tungstosilicate, potassium phosphotungstic acid, and ammonium molybdate at pH 6.0. The predominant form observed for these three macroglobulins contained three parallel vertical bars connected at their midpoints by a horizontal bar. The length of the molecules was estimated at 170–200 Å, and the width along the horizontal bar was approximately 100 Å (Fig. 8A). The middle vertical bar, which was always shorter than the outer bars (approximately 100 Å), resembled in some specimens two dots, one below and one above the horizontal crosspiece. A certain flexibility of the outer bars was noted. In all preparations examined another type of structure, resembling two beans facing one another or a donut with the dimensions 110 × 125 Å, was also seen. This

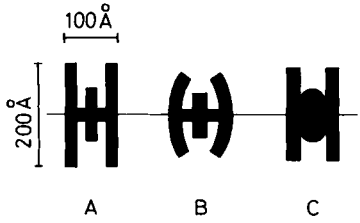


Fig. 8. Representations of α_2M derived from electron micrographs. The α_2M tetramer consists of two dimers, one below and one above the indicated line. These structures can be envisaged as projections of an overall cylinderlike shape in which each dimer contributes to the central bar and contains two short and two long arms. (A) The slender monogramlike structure probably representing native α_2M. (B) The compact form of methylamine-inactivated α_2M. (C) The trypsin–α_2M complex (2 : 1) in which each dimer accommodates one trypsin molecule. The long arms may wrap around the bound proteinase, producing a form resembling that shown in B (Fig. 13, see text for further details).

structure, which was seen when the samples were embedded in a thick layer of contrast medium, was considered to represent a collapsed form. Using a preparation of α_2M, which had been inactivated by exposure to ammonium sulfate, the same monogramlike structure was observed. In studies of rabbit α_1M and rat α_2M by Morelis *et al.* (1969), Gauthier *et al.* (1974), Gauthier and Mouray (1975a), and Bergsma *et al.* (1985) and in studies of human α_2M synthesized by HeLa and 3T3 cell lines by Harris *et al.* (1976), similar forms were seen and, moreover, the structure observed in the rabbit α_1M–trypsin complex was more compact. This corresponded to a contraction of the molecule from about 200 to 150 Å. A similar transition from a slender structure, presumed to represent native human α_2M, to a more compact, well-preserved structure of proteinase-treated α_2M resembling that shown in Fig. 8B was also seen by Barrett *et al.* (1974). Recently, Schramm and Schramm (1982) used computer averaging of selected representations of α_2M molecules and described an open form and a closed form (Figs. 8A and 8B), both assumed to represent inactive α_2M. In contrast to others, they did not observe a closed conformation for the 1 : 2 α_2M : trypsin complex, but rather a structure resembling native α_2M (Fig. 8C).

However, their computer averaging procedure on selected molecules indicated that the two trypsin molecules bound in the complex were located close to one another and near the middle of the structure in agreement with the energy transfer experiments of Pochon *et al.* (1981). In studies on the ovomacroglobulin from the egg of the Cuban crocodile Ikai *et al.* (1983) observed a structure composed of four or five globular nodules, presumed to represent the native ovomacroglobulin. The electron micrographs of the ovomacroglobulin–proteinase complex strongly resembled that of the α_2M–trypsin complex.

Whereas there is little doubt about the overall shape of amine-inactivated or proteinase-treated α_2M, the shape of native α_2M is not agreed upon. Recently Tapon-Bretaudiere *et al.* (1985) reported that native α_2M when examined by the negative-contrast method is represented by donutlike structures 220–250 Å wide, in which four spherules of 85 Å in diameter can be discerned. Other less regular structures were judged to represent different orientations of the tetramers on the supporting grid. In thin uranyl acetate stainings the molecules of α_2M–trypsin complex (1 : 2), similar to those described by others, all had the same dimensions, 155 × 230 Å. The presence of two structures, differing slightly in their details, indicated only two possible orientations when the α_2M–trypsin complexes were lying on their larger side. When the protein was embedded in a thick film of uranyl acetate, structures having the shape of a square with sides of about 155 Å were seen. From the distribution of stain in these structures they were judged to represent a vertical view through hollow cylinders.

In a different approach Branegård *et al.* (1980, 1982) and Österberg and Malmensten (1984) analyzed the structure of human α_2M, α_2M–trypsin com-

plex, and methylamine-treated α_2M by low-angle X-ray scattering in solution. The scattering curves obtained for the different preparations used were compatible with a model of α_2M represented by a hollow cylinder 158–171 Å long and 125–132 Å in diameter containing a central disc about 26 Å in height. The thickness of the walls of this cylinder was estimated at 28 Å. Such a model, although obviously crude, is largely compatible with the different projections of α_2M seen in the electron microscope, either donut-shaped tetrameric forms or the monogramlike structure, if the "walls" are thought to consist of two long and two short arms (Feldman *et al.*, 1985b). Thus α_2M can be envisaged as an assembly of two claw- or basketlike structures, one below the plane through the central disc and one above this plane, with the two cavities pointing in opposite directions. How the monomers are organized in the dimer, which presumably corresponds to one such basket, is not known. Clearly, the peculiar arrangement of the two interchain disulfide bridges within the dimer (involving Cys-255 and Cys-408) must be considered in any attempt at formulating a more detailed model for the structure of α_2M. Each of these basketlike structures presumably contains a binding site for proteinases close to the central disc, which forms the bottom. This model will be further discussed with regard to the mechanism of α_2M–proteinase complex formation (see Section IX).

B. C3, C4, and C5

In contrast to α_2M, electron microscopy has not been successful in giving an idea of the overall shape of C3, C4, and C5, probably due to their "monomeric" structure and lability. Using negative staining Molenaar *et al.* (1975) reported that native C3 was represented by spheres about 230 Å in diameter, as also seen earlier by Suzuki *et al.* (1972). In contrast, C3b appeared as threadlike structures, presumably representing end-to-end linked polymerization products. In a study on the interaction between C4b and C4b-binding protein Dahlbäck *et al.* (1983) observed that negatively stained C4b was represented by a compact structure 90 × 170 Å in size in which four or five domains could be discerned. Electron micrographs of C5 also showed different compact structures without much detail and having the dimensions 104 × 168 A or 140 × 151 Å (DiScipio *et al.*, 1983). In contrast C5b appeared as large aggregates. Given the reported dimensions of the representations of C3, C4, and C5, these were probably composed of aggregated material. Low-angle X-ray scattering data (Österberg *et al.*, 1984) indicate that the size of human C4, e.g., is approximately 21 × 56 × 110 Å, roughly compatible with the size of the individual subunits of α_2M as judged from the electron micrographs and the hollow cylinder model (Fig. 8). Methylamine-inactivated C3 and C4 apparently form dimers, about 235 Å in length (Österberg *et al.*, 1985).

VII. The β-Cysteinyl-γ-glutamyl Thiol Ester Structure in α_2M, C3, and C4

A. General Reaction Scheme of the Internal Thiol Ester in α_2M, C3, and C4

The presence of an internal reactive β-cysteinyl-γ-glutamyl thiol ester structure in α_2M, C3, and C4 readily explains a number of characteristic reactions not found in other proteins and is compatible with much experimental data. The tetrapeptide sequence -Cys-Gly-Glu-Glx- (residues 949–952 in α_2M) is common to human α_2M, C3, and C4 and contains the thiol ester structure. It is represented in Fig. 9 as a 15-member thiolactone ring. The thiol ester can be cleaved in three

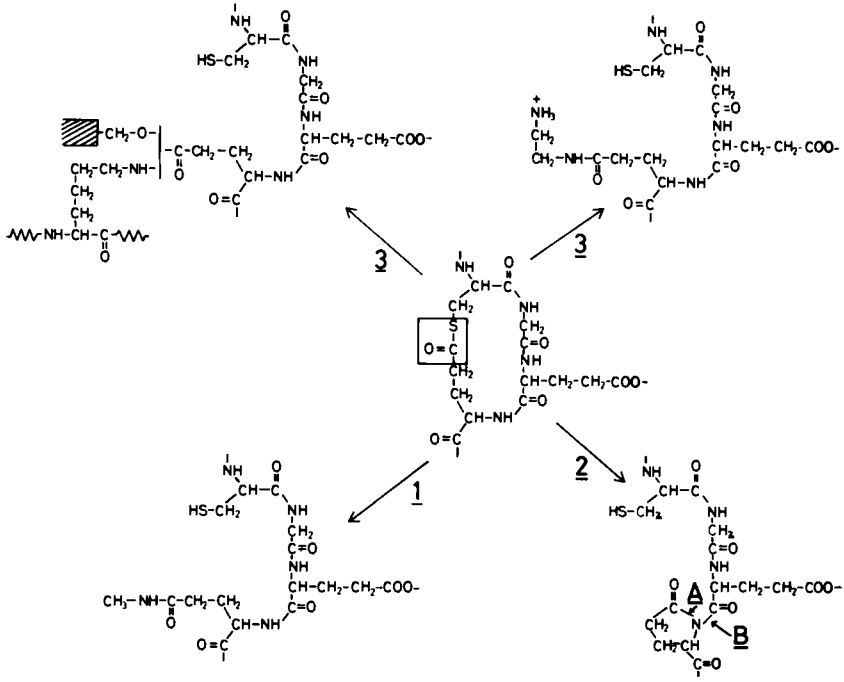


Fig. 9. Reactions of the internal β-cysteinyl-γ-glutamyl thiol esters in α_2M, C3, and C4. The thiol ester in the native proteins can react slowly with small amines, e.g., methylamine (reaction 1). Upon denaturation of the proteins the thiol esters undergo isomerization to form an internal pyroglutamic acid residue (reaction 2). This residue can be cleaved at A or B. Following proteolytic activation of the proteins the thiol esters react rapidly with available nucleophiles either from the solvent or from the surfaces of the targets (proteinases in the case of α_2M, and polysaccharides, membrane constituents, and immune complexes in the case of C3 and C4). This leads to cross-links of the amide type or the ester type (reactions 3). In each of these reactions a free β-cysteinyl SH group appears.

different reactions, 1–3, in each case resulting in the appearance of a readily accessible β-cysteinyl SH group and a γ-substituted Glx residue. Reaction 1 represents the slow cleavage by small nitrogen nucleophiles such as methylamine. In this reaction a residue of γ-glutamylmethylamide is formed. Reaction 2 represents the slow cleavage by the peptide NH group of the thiol-esterified Glx residue, resulting in the formation of a five-member pyrrolidone structure involving that residue (an internal pyroglutamic acid residue). This reaction is observed when the three proteins are denatured, and dependent on the conditions the pyrrolidone ring can be cleaved at two points. Cleavage at A results in hydrolysis, thereby yielding a glutamic acid residue, while cleavage at B results in polypeptide chain fragmentation. The previously thiol-esterified Glx residue now appears as an N-terminal pyroglutamic acid residue in the C-terminal denaturation cleavage fragment. Reaction 3 represents the fast, physiologically important cleavage of the thiol ester, which results from proteolytic activation. In this reaction a substituted amide or an ester involving the thiol-esterified Glx residue is formed, dependent on the presence of suitable nucleophilic groups, e.g., ε-lysyl amino groups on the activating proteinase in the case of α_2M and hydroxyl groups in polysaccharides in the case of C3. Other nucleophiles, e.g., putrescine, which happen to be present during the proteolytic activation and which normally do not react with the thiol ester in the native proteins, can also be covalently bound to the thiol-esterified Glx residue.

B. Evidence for the Thiol Ester Structure in the Native Proteins

Although the presence of a unique reactive site in α_2M, C3, and C4 composed of an internal thiol ester is an attractive hypothesis that is now generally accepted, it should be pointed out that the evidence for that structure is only circumstantial. The ability of the native proteins to react with and covalently incorporate certain small nitrogen nucleophiles into a particular Glx residue evidently requires an activation of that residue, since these reactions proceed readily under physiological conditions. *A priori* an activation of a Glu residue could be brought about by esterification with alcohols, phenols, or thiols, by anhydride formation involving carboxyl or phosphoryl groups, by acylimidazole formation, or by formation of a pseudoactive site as discussed by Swenson and Howard (1979b, 1980) and Howard *et al.* (1980). Based on their available evidence Howard *et al.* (1980) proposed that the reactive site of α_2M was constituted by an internal pyroglutamic acid residue (cf. reaction 2 in Fig. 9). Earlier evidence (Lichtenstein, 1942) indicated that such a structure could react with methylamine albeit under relatively harsh conditions, and internal pyroglutamic acid residues found in polymers of glutamic acid have been shown to spontaneously hydrolyze along

pathway B (Fig. 9) (Battersby and Robinson, 1955; Battersby and Reynolds, 1961).

The demonstration that the amine-reactive Glx residue of α_2M, C3, and C4 is encoded as a Gln residue (Domdey *et al.*, 1982; Belt *et al.*, 1984; Kan *et al.*, 1985) evidently imposes constraints on the possible ways of activation and makes activation by anhydride formation unlikely. Since native α_2M, C3, and C4 contain no free SH groups, the observation by Janatova *et al.* (1980a), Tack *et al.* (1980), Sottrup-Jensen *et al.* (1980), and Janatova and Tack (1981) that a previously unrecognized SH group appears in α_2M, C3, and C4 as a result of incubation with nitrogen nucleophiles immediately suggested that the reactive Glx residue might be activated by thiol esterification. The SH group appearing was contributed by a Cys residue preceding the reactive Glx residue by only three residues in the sequence of α_2M, C3, and C4 (Figs. 4 and 9) (Tack *et al.*, 1980; Sottrup-Jensen *et al.*, 1981a; Howard, 1981; Campbell *et al.*, 1981). Although that Cys residue might have been buried in the native proteins and only exposed as a result of the conformation change, which accompanies the methylamine incorporation, the observed 1 : 1 stoichiometry between methylamine incorporated into the native proteins and β-cysteinyl SH groups appearing in the course of the reaction (Tack *et al.*, 1980; Sottrup-Jensen *et al.*, 1980) makes this interpretation unlikely. The proposal that the reactive site of α_2M, C3, and C4 is constituted by an internal thiol ester is particularly attractive in view of the high reactivity of thiol esters in nucleophilic displacement reactions involving amines (Bruice and Benkovic, 1966).

Additional evidence in favor of a thiol ester of the proposed structure has been provided by Khan and Erickson (1981), who chemically synthesized a variety of 15-member ring structures containing an internal thiol ester based on the sequences (Gly)-Cys-Gly-Glu-Glx-Asn, common to α_2M and C3. The identity of the synthesized model thiol ester peptides was verified by elemental analysis, amino acid analysis, mass spectrometric analysis, and NMR data (300 MHz). The thiolactone structure was found to be in equilibrium with the corresponding isomeric lactam structure containing a free SH group (Khan and Erikson, 1982) (species I and II, respectively, Fig. 10) with a K (lactam/thiolactone) of 10–11, indicating that the ground state of the thiolactone is only about 1.5 kcal/mol higher than that of the lactam. The rate of isomerization (half-time 19–20 min) was found to be 28 times higher than the hydrolysis of the thiol ester producing species III (Fig. 10). These results provide support for isomerization of the thiol ester structure in α_2M, C3, and C4, which is presumed to take place upon denaturation (pathway 2, Fig. 9). As further discussed below this isomerization could also be important in the biosynthesis of the thiol ester and perhaps in the reactions of the nascent state.

Despite extensive efforts (L. Sottrup-Jensen, unpublished) it has not yet been

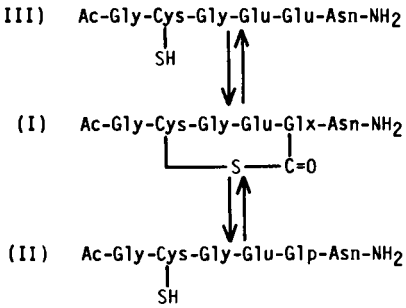


Fig. 10. Isomerization and hydrolysis reactions of model thiol ester peptides (Khan and Erickson, 1981, 1982). "Glp" denotes the internal pyroglutamic acid residue shown in Fig. 9, and the isomerization of I to II is analogous with reaction 2 (Fig. 9). III represents the hydrolysis product of I or II.

possible to obtain a peptide fragment from α_2M in which the thiol ester has not been cleaved, either as a result of denaturation or as a result of proteolytic degradation of the protein (cf. pathways 2 and 3, Fig. 9). Even if such a fragment could be isolated it would possibly isomerize from thiolactone to lactam during the isolation. In order to obtain further evidence for the presence of a thiol ester in native C3 Thomas *et al.* (1983) treated native C3 with tritiated $NaBH_4$. As expected from earlier studies demonstrating thiol ester intermediates in co-enzyme A transferase-catalyzed reactions (Solomon and Jencks, 1969; Sramek and Frerman, 1975), the thiol-esterified Glx residue would be reduced to a residue of α-amino-δ-hydroxyvaleric acid. Intriguingly, no reduction of the thiol-esterified Glx residue could be demonstrated in native C3. In contrast, the highly reactive Glx residue in nascent C3 (obtained by trypsin activation) could be readily reduced to the expected hydroxyamino acid. It is likely that this difference in reactivity of the thiol ester with $NaBH_4$ between the native and the nascent state is a reflection of an increased accessibility and reactivity in the nascent state of C3 (see also Section VIII).

C. Biosynthesis of the Thiol Ester Bond

The mechanism for the *in vivo* formation of the thiol ester structures in α_2M, C3, and C4 is not well understood. However, since the reactive Glx residue in all three proteins is encoded as a Gln residue it is possible that the internal thiol esters can be formed in reactions reminiscent of those of transglutaminases. These enzymes, which are widespread, catalyze the formation of ε-lysyl-γ-glutamyl cross-links in a number of proteins, e.g., fibrin, and the covalent incorporation of amines into Gln residues of a variety of protein substrates (Folk

and Chung, 1973; Folk and Finlayson, 1977; Folk, 1983). During the reactions catalyzed by these enzymes, intermediate thiol esters are formed between an enzyme-bound reactive SH group and a Gln residue of a given protein substrate.

Figure 11 shows three hypothetical ways of forming the thiol esters in α_2M, C3, and C4. In reaction I, which is envisaged as an autocatalytic process, the Cys and the Gln residues are brought together after the folding of approximately two-thirds of the protein. Due to a favorable local environment the SH group can directly react with the carbonyl group of the Gln residue, thereby expelling ammonia and forming the thiol ester structure. As the polypeptide synthesis continues the thiol ester is subsequently sequestered and stabilized in the final protein product. In this regard the thiol ester can be envisaged as a "frozen" intermediate, stable in the absence of nucleophiles. In reaction II the Gln residue is converted to a Glp residue (corresponding to the lactam structure of the model peptides) and ammonia is released. A subsequent isomerization to the thiol ester governed by the folding of the protein then takes place. In reaction III the Gln residue reacts with a transglutaminaselike enzyme, thereby expelling ammonia and forming an intermediary thiol ester, which is then attacked by the Cys-SH group in an intramolecular process to form the internal thiol ester.

Using the heat/denaturation cleavage reaction as a criterion for the presence of a thiol ester in intracellular pro-C4 synthesized by murine peritoneal macrophages, Karp (1983c) reported that the cleavage reaction could be detected 20 min after the onset of synthesis. This reaction could be prevented by pretreatment of the cell lysates with methylamine. Secretion of C4 starts approximately 60 min after synthesis (Fey *et al.*, 1980; Parker *et al.*, 1979). Karp (1983c) further observed that glycosylation preceded the appearance of a functional thiol ester site and obtained evidence showing that nonglycosylated pro-C4 could not undergo the cleavage reaction. This suggests that the degree of glycosylation could be important for the acquisition of a functional thiol ester site. In other studies Karp *et al.* (1982a) have shown that the genetic absence of a particular complex-type oligosaccharide in the C4 α chain is associated with a 75% reduction in hemolytic activity.

Recently Iijima *et al.* (1984) studied the *in vitro* synthesis of rabbit pro-C3 in a rabbit reticulocyte lysate system. Whereas the synthesized product readily reacted with radiolabeled iodoacetamide, no incorporation of methylamine could be demonstrated. However, by including a fraction of a liver homogenate containing cytosol and microsomes, pro-C3 could now incorporate methylamine, suggesting the presence of a functional thiol ester site and strongly indicating the involvement of an enzyme in a postribosomal formation of the thiol ester structure. This factor could be partially purified by DEAE-cellulose chromatography and could be a transglutaminaselike enzyme (reaction III, Fig. 11). Whether this factor accounts for the increase in molecular weight of intracellular pro-C4 found by Karp (1983c) to accompany the appearance of the heat cleavage reaction is

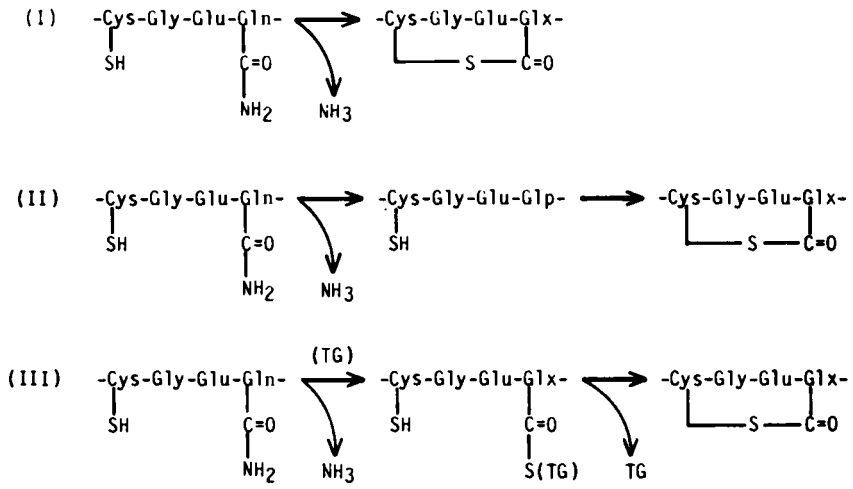


Fig. 11. Possible *in vivo* ways of forming the β-cysteinyl-γ-glutamyl thiol ester structure. Reactions I and II are envisaged as being "spontaneous" processes occurring during the folding of the primary translation products. In contrast, reaction III suggests the involvement of a transglutaminaselike enzyme in the activation of the Gln residue. Glp denotes the internal pyroglutamic acid residue (Fig. 9) and TG denotes a transglutaminase.

not known. In a similar study on the biosynthesis of pro-C3 by rat hepatocytes no heat cleavage was demonstrated (Misumi *et al.*, 1984).

D. Structure and Reactivity of the Thiol Ester in the Native Proteins

Model building has demonstrated that the thiol ester structure in the sequence -Gly-Cys-Gly-Glu-Glx- can be formed without imposing severe constraints on the dihedral angles of its amino acid residues. Thus, Thomas *et al.* (1982) noted that an irregular sharp bond could be made in which the thiol ester group was planar. In this model the glutamic acid residue adjacent to the reactive Glx residue is believed to protrude into the solvent, while the major part of the thiol ester is surrounded by apolar residues, thereby partially shielding the thiol ester from the solvent. Similarly, Howard (1983) noted that model building produced a crowded, but not strained structure, permitted by the presence of the Gly residue that follows the Cys residue, overall resembling a 3_{10} chain reversal structure found in F:S proteins (Adman *et al.*, 1975). In this model the peptide NH group of the reactive Glx residue is positioned close to the carbonyl group of the thiol ester. Davies and Sim (1981) have proposed a different model, in which the Glu residue adjacent to the reactive Glx residue is unprotonized and loops

back to form a hydrogen bond to the carbonyl group, thereby providing an increased electrophilicity of the carbonyl group of the thiol ester. Intramolecular general acid catalysis of this type is known in model systems and can result in greatly increased rates of nucleophilic attack on ester carbonyls (Jencks, 1972; Stewart and Srinavasan, 1978). As a consequence of this model the peptide group connecting the Glu and Glx residues is no longer planar, thereby resulting in an NH group of enhanced nucleophilicity. Considering the predicted secondary structure of α_2M Welinder *et al.* (1984) suggested that the thiol ester loop is located at the turn of a long β hairpin at one end of a proposed β-barrel domain, further protected from solvent by a loop connecting other strands of that barrel. Whatever the precise conformation and localization of the thiol ester in the native proteins is, the NMR spectra of the model peptides (Khan and Erickson, 1982) indicate that their conformation is relatively rigid.

As already noted in the early literature (Gordon *et al.*, 1926; Pillemer *et al.*, 1941; Ratnoff *et al.*, 1954; Taylor and Leon, 1959), the spectrum of amines, which can react relatively rapidly with and inactivate native α_2M, C3, and C4, is restricted to small amines and hydrazines having a primary amino group such as methylamine. The reaction was found to be a typical nucleophilic displacement reaction, with the rate being dependent on the concentration of unprotonated amine (Taylor and Leon, 1959). These results have now been confirmed and extended (Barrett *et al.*, 1979; Salvesen *et al.*, 1981; Pangburn and Müller-Eberhard, 1980). In a recent detailed study the rates for the reaction of a broad spectrum of amines with α_2M were compared with those for the simple model thiol ester, ethyl thiol acetate (Larsson and Björk, 1984). The relative reaction rates of the amines investigated with α_2M did not parallel the relative reaction rates with the small thiol ester. Thus, it was found that secondary and tertiary amines (e.g., dimethylamine and trimethylamine), and primary amines larger than propylamine, which all cleaved ethyl thiol acetate with second-order rate constants comparable to that of methylamine, reacted with α_2M with rate constants considerably lower than that for methylamine. The relative rates [k (α_2M)/k (ethyl thiol acetate)] ranged from 700 (ammonia) and 82 (methylamine) to 16 (propylamine).

Clearly, the thiol ester site in α_2M is inherently more reactive in nucleophilic displacement reactions than a simple thiol ester, provided that the amine can gain access to the site, which presumably is located near the surface but partly shielded by one or more peptide loops. The access to the thiol ester sites is quite discriminatory as seen from the low rate of reaction with dimethylamine versus ethylamine. The second-order rate constants for reaction with methylamine at room temperature and pH 8.0 have been determined as 13.8 (Larsson and Björk, 1984) or 11.6 M^{-1} sec^{-1} (Strickland and Bhattacharya, 1984) for human α_2M, 3.8 M^{-1} sec^{-1} for human C3, and 12.2 M^{-1} sec^{-1} for human C4 (Isenman and Kells, 1982). Thus, whereas the rate constants for α_2M and C4 are similar, that

of C3 is much lower, presumably reflecting a slight difference in the folding of the polypeptide chains at the thiol ester site, which also may be important in determining the range of nucleophiles, which can react with activated C3 (see Section VIII). Whether the enhancement of reaction rate observed is due to an "activated" state of the thiol ester in α_2M is not known, since no data for the reaction of the cyclic model thiol ester peptides with amines have yet appeared. That the cyclic model peptides could have an enhanced reactivity per se is indicated from preliminary results (Khan and Erickson, 1981, 1982) suggesting that the rate of spontaneous hydrolysis of the cyclic model peptides is about two orders of magnitude higher than that of simple thiol esters, for which the pseudo-first-order constants have been estimated at approximately 5×10^{-6} min^{-1} (Morse and Torbell, 1952). Although no accurate estimates are available, it is apparent that the thiol ester in the native proteins is relatively stable. Thus, for C3 Pangburn and Müller-Eberhard (1980) and Pangburn *et al.* (1981) estimated a loss of hemolytic activity at pH 7.3 and 37°C equal to 0.5%/hr for C3. Upon storage or by repeated freezing and thawing the conformation of α_2M, C3, and C4 gradually changes to produce a spectrum of inactive forms, in which the thiol esters eventually can no longer be detected.

The slow inactivation of α_2M, C3, and C4, which frequently yields preparations containing 0.2–0.5 mol SH groups per mole subunit, has been observed by many investigators and has been ascribed to the slow hydrolysis of the thiol ester in the native proteins. However, since aged preparations of α_2M, C3, and C4 usually contain dimers or aggregates of high molecular weight, which resist dissociation by reducing SDS, the spontaneous decay of the thiol ester could to some extent actually be the result of an "activation" process, presumably induced by localized conformational changes. This is supported by the results of Law (1983a,b), who showed that native C3 treated with guanidinium chloride (up to 1.5 *M*) and other denaturants readily would bind glycerol and other nucleophiles covalently. Concomitantly the generation of aggregates was observed. At still higher concentrations of guanidinium chloride, binding of glycerol decreased and the autolytic reaction prevailed. Thus, the localized denaturation-dependent conformational change evidently produced a state showing some similarity with the nascent state of proteolytically activated C3 (see also Section VIII).

The heat/denaturation cleavage reaction of α_2M, C3, and C4 has been studied in detail by Harpel *et al.* (1979), Barrett *et al.* (1979), Howard *et al.* (1980), Janatova *et al.* (1980a), Janatova and Tack (1981), and Sim and Sim (1981). This reaction, also referred to as the autolytic reaction, is a convenient tool for the detection of the native forms of these and related thiol ester-containing proteins (see above). Subjecting each of these proteins to conditions in which their structure is extensively denatured, e.g., by heating in the absence or the presence of strong denaturants such as urea, guanidinium chloride, or SDS and

by incubation in alkaline media, a variable proportion of two new fragments is seen in reducing SDS–PAGE. Characteristically, incubation with weak chaotropes such as KBr produced no such cleavage, consistent with the requirement for an extensive denaturation of the proteins in order to observe the cleavage reaction (Sim and Sim, 1983; Law, 1983a,b). While most studies have indicated that high temperature (70–90°C) and prolonged incubation (up to 5 hr) are necessary for obtaining a high yield of cleavage products, extensive cleavage of α_2M can be obtained in reducing 6 *M* guanidium chloride or in 70% formic acid at room temperature (Sottrup-Jensen *et al.*, 1984a). According to Howard (1981) the products of the autolysis reaction of α_2M can be envisaged as originating from the decay of a common intermediate, representing fully denatured α_2M formed in a relatively fast step. The conformation of the thiol ester structure in the fully denatured state of α_2M probably resembles the conformation of the model thiol ester peptides, in which the ratio of the lactam form to the thiolactone form is 10–11 (Khan and Erickson, 1982). The maximal cleavage observed (60–80%) with α_2M, C3, and C4 (Harpel *et al.*, 1979; Howard, 1981; Sim and Sim, 1981, 1983) indicates that the cleavage of the lactam structure along pathway B (Fig. 9) proceeds roughly three times faster than the cleavage along pathway A, qualitatively in accordance with the high stability of the N-terminal pyroglutamic acid residue formed.

VIII. Conformational Change and the Nascent State

A. Detection and Features of the Conformational Change

Upon incubation with those nucleophiles that are known to inactivate human α_2M, C3, and C4 and by exposure to weakly denaturing conditions these proteins undergo a characteristic change in conformation similar to that seen after activation with proteinases. Since this treatment results in cleavage of the thiol esters there is an apparent coupling between the functional state of thiol esters (native versus cleaved) and the gross conformational state of these proteins.

PAGE systems, which can reveal a conformational change in α_2M, have been described by Steinbuch *et al.* (1968), Saunders *et al.* (1971), Zais and Roberts (1977), Barrett *et al.* (1979), Harpel *et al.* (1979), Nelles *et al.* (1980), and by Van Leuven *et al.* (1981a). Similarly, crossed immunoelectrophoresis can be used (Ganrot and Laurell, 1966; Dott *et al.*, 1985), but is less discriminatory. Conformational changes can also be detected by electron microscopy and by low-angle X-ray scattering (Section VI), or by determination of sedimentation coefficients (Björk and Fish, 1982; Gonias *et al.*, 1982b; Dangott and Cunningham, 1982). For monitoring conformational changes with time determina-

tions of changes in intrinsic fluorescence, dye-mediated fluorescence, circular dichroism, difference UV spectra, and heat capacity have been used (Frenoy *et al.*, 1977; Richman and Verpoorte, 1981; Björk and Fish, 1982; Gonias *et al.*, 1982b; Chlebowski and Williams, 1983; Dangott *et al.*, 1983; Straight and McKee, 1982, 1984; Cummings *et al.*, 1984; Strickland and Bhattacharya, 1984; Strickland *et al.*, 1984; Björk *et al.*, 1985; Larsson *et al.*, 1985; Eccleston and Howard, 1985). Recently the isolation of different active subforms of α_2M by chromatography has been reported (Chlebowski and Williams, 1985) supporting earlier results of Saunders *et al.* (1971), and the "slow" and "fast" forms of α_2M can be separated by high-performance hydrophobic chromatography (Van Leuven *et al.*, 1985).

Similarly, conformational changes in C3 and C4 affecting their sedimentation coefficient, electrophoretic mobility, antigenicity, and spectral properties have been reported, e.g., by Müller-Eberhard *et al.* (1960), Dalmasso and Müller-Eberhard (1966), Müller-Eberhard and Biro (1963), West *et al.* (1966), Pondman and Peetoom (1964), Müller-Eberhard and Lepow (1965), Molenaar *et al.* (1974, 1975), Reboul *et al.* (1979, 1980), von Zabern *et al.* (1981, 1982), Isenman and Cooper (1981), Isenman *et al.* (1981), Isenman and Kells (1982), von Zabern and Gigli (1982), Österberg *et al.* (1984), and Fontaine and Sim (1984). The separation of the hemolytically inactive forms of C3 and C4 from the native forms by chromatography has been reported by Janatova *et al.* (1980b), Janatova and Tack (1981), Parkes *et al.* (1981), and Pangburn *et al.* (1981). For both α_2M and the complement proteins C3 and C4 the final conformational end state, irrespectively of being generated by treatment with amines or proteinases, is effectively recognized by a variety of cellular receptors (Kaplan and Nielsen, 1979a,b; Kaplan *et al.*, 1981; Van Leuven *et al.*, 1979, 1980, 1981a, 1982b; Schreiber *et al.*, 1981; Berger *et al.*, 1981; Arnaout *et al.*, 1981; Imber and Pizzo, 1981; Fuchs *et al.*, 1982; Hanover *et al.*, 1983a). While the activity of C5 is not affected by amines (DiScipio, 1981a,b; DiScipio *et al.*, 1983) as already indicated by the absence of a covalent binding reaction (Janatova and Tack, 1981; Law *et al.*, 1980a) and the lack of the residues constituting the thiol ester site (B. F. Tack and R. A. Wetsel, unpublished) (see Fig. 7), proteolytic activation of C5 nevertheless results in extensive conformational changes as documented by the lack of interaction with complement protein C6 in its native state, changes in its CD spectrum, exposure of hydrophobic sites, and an increased susceptibility to proteolysis (Cooper and Müller-Eberhard, 1970; Yamamoto and Gewurz, 1978, 1980; Podack *et al.*, 1978; Wetsel *et al.*, 1980; Vogt *et al.*, 1978; DiScipio, 1981a; DiScipio *et al.*, 1983; Nilsson *et al.*, 1975).

For C3 (Isenman and Cooper, 1981) spectroscopic techniques sensitive to changes in the backbone conformation (e.g., far-UV CD) or to perturbation of aromatic residues (e.g., intrinsic tryptophan fluorescence) have been of limited value in monitoring the conformational change in C3 following proteolytic activation. In contrast, changes in near-UV CD and fluorescence enhancement of

bound ANS, reporting changes in the local environment of aromatic residues and surface hydrophobicity, respectively, are useful. Thus, Isenman and Cooper (1981) and Isenman *et al.* (1981) demonstrated that the conformational end state for C3b and the C3b-like conformation induced by methylamine or by freezing and thawing were similar as assessed by CD and ANS fluorescence enhancement measurements. The rate of the conformational change following proteolytic conversion from C3 to C3b was very fast and was judged to be limited only by the rate of activation cleavage. This reaction could not be followed by conventional procedures. In contrast, the rate of conformational change following addition of methylamine was slow, the rate being limited by the conformational change itself, not by the initial nucleophilic substitution reaction leading to thiol ester cleavage. Characteristically, the change in the CD and ANS fluorescence signals occurred with a lag phase compared with the loss of hemolytic activity measured in a standard assay. While only about 30% of the final change in ellipticity or ANS fluorescence enhancement had ocurred after 2 hr at 25°C, C3 was essentially completely inactivated at that point. Subsequent addition of trypsin led to a rapid completion of the spectral change. The conformational change induced by methylamine has been analyzed in terms of a model minimally involving three steps as shown in Fig. 12 (reaction I). Following thiol ester cleavage the native protein is converted to an intermediate (step 1). Although this intermediate is formed in a spectroscopically "silent" reaction, its conformation has nevertheless changed, since the stoichiometry between SH groups appearing and methylamine incorporated (Janatova *et al.*, 1980a; Tack *et al.*, 1980) in this reaction is largely correlated with the loss of hemolytic activity observed. The spectroscopically detected, relatively slow unimolecular rearrangements to the end state (steps 2 and 3) account for most of the observed CD and ANS fluorescence changes and result in the acquisition of the C3b-like conformation, characterized by the ability to bind factor B and the susceptibility to cleavage by factors I plus H. The rate constant governing step 2 was determined to be approximately tenfold higher than that for step 3. The factor B binding potential (apparently generated in step 2) appeared significantly faster than the potential for cleavage by factors I plus H, suggesting that their binding sites on C3b are different. However, the rate of cleavage of the C3b-like conformational end state with factors I plus H was found to be smaller than that for C3b obtained by proteolytic activation, suggesting that this state is similar but not identical with the conformational end state of C3b.

As described by Isenman (1983), the cleavage by factors I plus H of either C3b or the C3b-like conformation produced by incubation with amines results in a new conformational change. This change, monitored by a pronounced loss in the prior enhancement of the fluorescence of ANS bound to the C3b-like conformation, indicates the net loss of hydrophobic surface areas of C3b as a result of inactivation cleavage at the points shown in Fig. 5.

In a spectroscopic study of C4 by Isenman and Kells (1982) it was again found

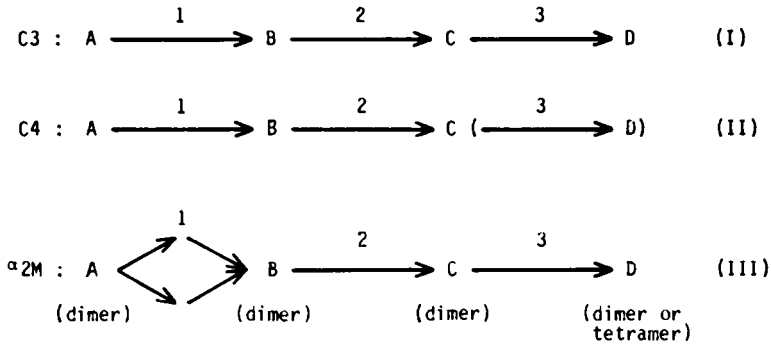


Fig. 12. Suggested schemes of methylamine-dependent conformational changes in C3 (I), C4 (II), and α_2M (III). The species labeled A represent the native conformation; those labeled B and C represent intermediates in the conformational change; and those labeled D represent the end states (the C3b- and C4b-like conformation and inactivated α_2M, respectively). For C3 the following half-times have been determined (50 m*M* methylamine, pH 8.0, 37°C): 2.5, 15, and 112 min for steps 1, 2, and 3, respectively. The loss of hemolytic activity has a half-life of 5.5 min under these conditions. For C4 the half-times are 2.5, 2.5, and 13.9 min (ANS fluorescence) for steps 1, 2, and 3, respectively (20 m*M* methylamine, pH 8.0, 37°C). CD measurements give values of 3.1 and 10.4 min for steps 1 and 2, respectively. The loss of hemolytic activity has a half-time of 2.5 min under these conditions. For α_2M the half-times are 5, 18 (12 at 37°C), and 18 min for steps 1, 2, and 3, respectively (50 m*M* methylamine, pH 8.0, 25°C). The data are taken from Isenman *et al.* (1981), Isenman and Kells (1982), Strickland and Bhattacharya (1984), and Larsson *et al.* (1985).

that the conformational end states generated by cleavage by C1s or by nucleophilic modification with amines were similar. In analogy with C3, the rate of the fast conformational change within the proteolytically activated molecule appeared to be dependent on the rate of activation per se contrasting with the low rate of nucleophile-induced conformational change, which was not limited by the modification event. Unlike C3, the changes in the far-UV CD spectrum observed upon proteolytic cleavage or by amine treatment indicated that some changes in the backbone conformation accompanied these reactions. An additional difference was found in that native C4 produced a considerable enhancement of ANS fluorescence, which decreased upon conversion to C4b or the C4b-like conformation, indicating an overall loss in hydrophobicity accompanying the conformational change(s). As with C3 the loss of hemolytic activity of C4 closely paralleled the extent of nucleophilic substitution with methylamine (Law *et al.*, 1980b; Janatova and Tack, 1981). However, the CD changes showed a shorter lag phase than the ANS fluorescence changes, indicating that they did not report exactly the same transitions. Addition of trypsin to partially conformationally changed amine-treated C4 led to a final conformational change virtually indistinguishable from that of proteolytically activated C4. The CD changes

could not be analyzed in terms of the model shown for C3 (Fig. 12, equation I), and a model involving only two steps was proposed (Fig. 12, equation II). This model, which again presumes that the nucleophilic substitution event is spectroscopically silent (step 1), probably is an oversimplification, since the long lag phase in the ANS fluorescence could not be interpreted well, suggesting at least one additional step. In analogy with that found for C3, the acquisition of the C4b-like properties of methylamine-treated C4 (ability to bind component C2 and potential for cleavage by factor I plus C4b-binding protein) was much slower than the rate of loss of hemolytic activity. The appearance of the potential for factor I plus C4b-binding protein-mediated cleavage roughly correlated with the change in ANS fluorescence, while the C2 binding activity was generated in a faster process.

For human α_2M a number of studies have indicated that the changes in the CD spectra resulting from incubation with amines and proteinases are similar (Björk and Fish, 1982; Gonias *et al.*, 1982b; Straight and McKee, 1982; Dangott *et al.*, 1983). However, the conformational end states are probably not identical, as judged from the lower increase in intrinsic fluorescence upon treatment with methylamine compared with that following treatment with proteinases. This is consistent with scanning calorimetric analyses, which indicate that the thermal stability of the two forms of conformationally changed α_2M differs (Cummings *et al.*, 1984). Like C4 the CD changes indicate small changes in secondary structure. As found for C3 the conformational changes in α_2M lead to an enhancement in the fluorescence of bound ANS (Jacquot-Armand and Krebs, 1973). Larger changes in quaternary structure of the α_2M tetramer have been detected by an increase in sedimentation coefficient, consistent with a decrease in asymmetry and a change toward a more compact structure for both methylamine- and proteinase-treated α_2M (Björk and Fish, 1982; Gonias *et al.*, 1982b; Dangott and Cunningham, 1982). The close correlation between methylamine incorporated and SH groups appearing noted earlier (Sottrup-Jensen *et al.*, 1980) evidently is consistent with a nucleophilic substitution reaction showing pseudo-first-order kinetics as further substantiated by Larsson and Björk (1984) and by Eccleston and Howard (1985). It has been reported that the conformational change(s) leading to inactivation of the proteinase binding potential of α_2M by methylamine also followed a simple pseudo-first-order kinetics (Gonias *et al.*, 1982b; Straight and McKee, 1982; Dangott *et al.*, 1983), indicating a concerted process of thiol ester cleavage and conformational change. This contrasts with the results of Van Leuven *et al.* (1981a, 1982b), which indicated that the methylamine incorporation and the conformational change could be events separated in time. Recent studies by Strickland and Bahttacharya (1984), Larsson *et al.* (1985), and Eccleston and Howard (1985) have provided evidence for a complex mechanism of methylamine-dependent conformational change in α_2M, in which at least three steps can be discerned (Fig. 12, reaction III).

Following initial random nucleophilic attack by methylamine on the thiol esters (step 1, reaction III, Fig. 12), which by analogy with C3 and C4 is spectroscopically silent, a change in intrinsic fluorescence or TNS-mediated fluorescence occurs with a lag phase. The kinetics of this reaction is consistent with a cooperative model, where the conformational change occurs, when both thiol esters in the dimeric unit have been cleaved (step 2). Thus, the thiol ester in each monomeric unit of the dimer appears to maintain the dimeric unit in the native conformation until both bonds have been cleaved. At least one additional slower step (step 3) has been identified by correlating the change in UV absorption with the change in hydrodynamic volume and loss of proteinase binding activity (Larsson *et al.*, 1985). This step, leading to a more extensive conformational change (the "slow" to "fast" form transition) in which the proteinase binding potential is lost, could occur either within the α_2M dimer or cooperatively within the tetramer after completion of the initial changes in both dimers. This model, involving at least two sequential conformational changes, was further supported by the results of pore-limit PAGE, which indicated the presence of two intermediate species in the transition from "slow" to "fast" α_2M (Larsson *et al.*, 1985). These species, which have not been observed in some PAGE systems (Barrett *et al.*, 1979; Van Leuven *et al.*, 1981a), could be equivalent with some of the distinct bands observed earlier in other PAGE systems (Saunders *et al.*, 1971; Nelles *et al.*, 1980) or in isoelectric focusing (Ohlsson and Skude, 1976; Van Leuven *et al.*, 1981a).

Due to the high rate of interaction between α_2M and proteinases like trypsin and chymotrypsin it has not been possible to analyze the conformational changes induced by these enzymes in detail by conventional procedures. Using stopped-flow fluorescence measurements Dangott *et al.* (1983) noted that the conformational change induced by trypsin showed biphasic kinetics, with an overall half-life of 0.5–1 sec. Similarly, Christensen and Sottrup-Jensen (1983) observed in the reaction between α_2M and trypsin (micromolar range of reactants) that the rate of SH appearance contained a very fast and a very slow component. In contrast, the conformational change(s) induced by the relatively slowly reacting thrombin, plasmin, and benzamidine-inhibited trypsin, as monitored by the change in intrinsic fluorescence, fluorescence enhancement of bound TNS, and SH appearance, could adequately be described as a reaction showing second-order kinetics at enzyme : α_2M ratios below 1 : 1 mol/mol (Straight and McKee, 1982; Sottrup-Jensen *et al.*, 1983b; Christensen and Sottrup-Jensen, 1984; Steiner *et al.*, 1985). Under these conditions only one bait region is cleaved and two SH groups appear for one proteinase molecule bound. In this 1 : 1 complex the conformation of the "reacted" dimer is changed, consistent with a concerted or sequential mechanism, in which the initial bait region cleavage is rate determining, and the subsequent conformational change(s) ultimately leading to thiol ester cleavage occur more rapidly. In an additional reaction at higher proteinase:

α_2M ratios the intact dimer of the 'half-reacted'' 1 : 1 α_2M–proteinase complex may interact with a second proteinase molecule resulting in a further conformational change and appearance of the second pair of SH groups, dependent on the cleavage of at least one bait region in that dimer. These features of α_2M will be further discussed in Section IX.

How the initial substitution event in α_2M, C3, and C4 will lead to relatively slow conformational changes within these proteins is presently unknown. While that reaction has been judged to be spectroscopically silent, cleavage of the thiol ester will presumably create a local change in hydrophobicity and hydrogen bonding possibilities at this site, since the thiol ester structure most likely is partially buried in the native proteins, while the SH group appearing is readily accessible for reaction with large aromatic disulfides like DTNB and a variety of alkylating agents. The surface location of the SH group is further indicated by the facile reaction with immobilized mixed disulfides (Tack *et al.*, 1980; Pochon *et al.*, 1983a). It is conceivable that the change in electronic structure at the thiol ester site brought about by nucleophilic substitution to some extent can mimic the result of the physiologically relevant activation, effected by proteolytic cleavage in the activation cleavage sites.

The apparent similarity of the conformational changes effected by amines and those effected by specific limited proteolysis is puzzling in the sense that the triggering events involve structures that are located approximately 250–270 residues apart in the sequence (Fig. 5). At least some of the conformational changes observed are related to the conformational state of the thiol ester (native versus cleaved). Evidently, the thiol ester structure and the domain(s) controlling its conformation and accessibility exert a profound influence on the overall conformation of these proteins. The particular conformational arrangement of the thiol esters in the native proteins seems to be in a delicate tense state, which slowly decays even in the absence of added nucleophiles, and which is particularly sensitive to limited denaturation by freezing and thawing. Whether the thiolactone–lactam equilibrium observed with model peptides plays any role in this process is not known. While it has been suggested that the thiol ester structure could act as a conformational lock, it could equally well be argued that a particular conformation of the proteins locks the thiol ester structure. Intringuingly, Van Leuven *et al.* (1982b) and Björk (1985) reported that alkylating the SH groups appearing in α_2M during nucleophilic substitution with 2,4-dinitrophenyl thiocyanate would effectively prevent the large conformational change detected by PAGE analysis and the inactivation of the proteinase binding properties of α_2M. This effect could be due to a steric interference with polypeptide segments involved in the conformational change(s) and indicates that the cleavage of the thiol esters in human α_2M is not necessarily associated with the large conformational change seen after complex formation with proteinases (the ''slow'' to ''fast'' transition). Similarly, rat α_2M and chicken α_2M did not

undergo the large conformational change upon methylamine treatment, whereas rat α_1M did (Gonias *et al.*, 1983; Feldman and Pizzo, 1984a; Feldman *et al.*, 1984).

Analyzing the reaction between methylamine and bovine α_2M, Strickland *et al.* (1984) and Björk *et al.* (1985) found by spectroscopic methods that no major change in conformation could be detected. Thus the final change seen in human α_2M (Fig. 12, step 3, reaction III) does not proceed very well in bovine α_2M. Upon subsequent addition of trypsin a major conformational change in methylamine-treated bovine α_2M ensued similar to that seen with the native protein, as also detected by a change in sedimentation coefficient (Dangott and Cunningham, 1982). That the nucleophilic substitution nevertheless had led to a change affecting the proteinase binding sites of bovine α_2M was documented by a slight decrease of the binding capacity for trypsin and an increased rate of inactivation of the bound trypsin with soybean trypsin inhibitor (STI). These "anomalous" complexes are presumably similar to those described for methylamine- and 2,4-dinitrophenyl thiocyanate-treated human α_2M (Van Leuven *et al.*, 1982a,b; Björk, 1985).

It thus appears that the methylamine-dependent conformational change(s) in α_2M indeed are mechanistically different from those induced by proteolytic activation. Not only are the methylamine-dependent conformational changes several orders of magnitude slower than those following proteolytic activation, but in the latter case the thiol esters assume a greatly increased reactivity in the nascent state. This is also the case for C3 and C4. While individual steps might be similar, the reaction with amines does not appear to result in conformational changes, which are as extensive as those elicited by proteinases. The minor conformational changes seem to directly dependent on thiol ester cleavage and are correlated with the exposure of receptor recognition sites, concealed in the native macroglobulins (Van Leuven *et al.*, 1981a, 1982a,b; Gonias *et al.*, 1983; Feldman and Pizzo, 1984a,b; Feldman *et al.*, 1984). In contrast, the major conformational changes, which for unknown reasons are not seen with all macroglobulins incubated with methylamine, are responsible for the loss of proteinase binding potential.

The finding that the loss of hemolytic activity of C3 and C4 is directly correlated with cleavage of their thiol esters readily provides an explanation for the inactivation of these proteins by small amines, since their functional activity is crucially dependent on the covalent binding potential provided by the thiol-esterified Glx residues. Only C4 but not the C4b-like conformation of amine-inactivated C4 can interact with activated C1, which is noncovalently associated with an immune complex. In order to rapidly activate C3 the complex formed between C4b (generated by cleavage by C1s) and C2a must be covalently anchored on an immune complex or a target cell. Similarly, for C3 there seems to be absolute requirement for covalent complex formation between C3b and target

cells in order to allow for the fast localized activation of C5 and the assembly of the terminal lytic complex C5b-C9 (von Zabern *et al.*, 1981; Vogt *et al.*, 1978). In addition, the rate of proteolytic cleavage of the C3b- and C4b-like conformations by the physiological activators is much lower than that for the native proteins, suggesting that thiol ester cleavage also affects the conformation or accessibility of their activation cleavage sites (Janatova *et al.*, 1980a; Janatova and Tack, 1981; von Zabern *et al.*, 1981). Any C3b and C4b remaining in the fluid phase will be rapidly inactivated by factors I plus H (or by factor I plus C4b-binding protein) and cleared from the circulation. As discussed by Pangburn and Müller-Eberhard (1980) and Pangburn *et al.* (1981), "spontaneously" inactivated C3 in its C3b-like conformation could be of importance for providing a relatively constant low level of C3b for initiation of the alternative pathway of complement.

Contrary to the complement proteins the role of the thiol esters in the inactivation of α_2M by amines is less obvious. While nucleophilic substitution evidently changes the conformation of the monomeric unit, exposing the receptor recognition site, the conformation of the dimeric functional units of some macroglobulins does not necessarily change drastically, so that proteinase binding is seriously impaired. In this respect it seems fortuitous that human α_2M was the first macroglobulin to be studied. Like amine-inactivated C3 and C4, a contributing element in the inactivation of the proteinase binding activity of α_2M by amines probably resides in the changed conformation of the bait region, rendering it less susceptible to proteolytic cleavage (Wang *et al.*, 1981; Björk *et al.*, 1985; Eccleston and Howard, 1985; L. Sottrup-Jensen, unpublished). As discussed below, it is likely, however, that the covalent binding reaction of nascent α_2M plays a major role in the mechansim of complex formation between α_2M and proteinases.

B. Nature of the Nascent State and the Covalent Binding Reactions

The concept of nascent C3 and C4 was introduced by Müller-Eberhard and Lepow (1965) and Müller-Eberhard *et al.* (1966) in order to explain the observation that unless the activating enzyme complexes for C3 and C4 were firmly associated with cell surface structures and immune complexes their action on C3 and C4 would lead to proteolytically cleaved, but largely intact proteins without hemolytic activity. Furthermore, only a minor fraction of activated C3 and C4 remained bound around the sites of activation, with the rest accumulating in the fluid phase. Thus, it was envisaged that during activation of C3 and C4 these proteins would undergo a conformational change, thereby being transiently activated and expressing a binding site of unknown nature for acceptors in the

vicinity of the site of activation. These acceptors were judged to be immune aggregates on sensitized cells and membrane constituents. The nascent state was found to be very short-lived (Bokisch *et al.*, 1975). Although the association between activated C3 and C4 and immune complexes and cells was initially thought to be due to hydrophobic interactions (Müller-Eberhard, 1975; Capel *et al.*, 1978), the finding by Law and Levine (1977) and Law *et al.* (1979b) that a large fraction of C3b and C4b in fact was covalently bound evidently disclosed an important role of the internal thiol esters in these covalent binding reactions. The concept of a nascent state is also applicable to the covalent binding reactions exhibited by human α_2M and its homologues. While the covalent binding reactions of α_2M, C3, and C4 have now been firmly established, the nature of the highly reactive nascent state is still not known in detail.

The covalent binding reactions of activated C3 and C4 have been examined in detail by Law and Levine (1977) and Law *et al.* (1979b, 1980a,b, 1981, 1984a,b), Law (1983a), Campbell *et al.* (1980, 1981), R. B. Sim *et al.* (1981), Gadd and Reid (1981), Gorski *et al.* (1982), Hostetter *et al.* (1982), and Thomas *et al.* (1983). Analogous studies on covalent α_2M–proteinase complexes have been reported by Harpel (1977), Salvesen and Barrett (1980), Salvesen *et al.* (1981), Sottrup-Jensen *et al.* (1981c,d), Sottrup-Jensen and Hansen (1982), Van Leuven *et al.* (1981b, 1982a,b), Wu *et al.* (1981), Wang *et al.* (1981, 1983, 1984), and Straight and McKee (1984).

The studies showed that activated C3 and C4 could bind covalently to a wide range of immune complexes and cell membrane constituents with both ester bonds and amide bonds. The formation of covalent C3b and C4b dimers has also been seen. *In vitro*, the covalent binding efficiency of C3 and C4 is rather low, usually below 10%. By artificially activating C3 with chaotropes (Law *et al.*, 1984a) the covalent binding efficiency may increase to almost 40%. This contrasts with α_2M and PZP (Sand *et al.*, 1985) where typically more than 50–60% of the proteinase being complexed is covalently bound. In some cases a covalent binding efficiency approaching 100% can be attained. That the site of covalent incorporation of nucleophiles into proteolytically activated α_2M, C3, and C4 indeed was the thiol-esterified Glx residue was clearly demonstrated from sequence studies by Campbell *et al.* (1981), Sottrup-Jensen and Hansen (1982), and Hostetter *et al.* (1982). While it has been reported that the cross-links from activated C4 are formed to residues in the Fd portion of the antibodies, the sequence location of those residues has not yet been determined (Goers and Porter, 1978; Campbell *et al.*, 1981; Gadd and Reid, 1981). In contrast, the sequence locations of the major ϵ-lysyl-γ-glutamyl cross-links in α_2M–trypsin complex are known (Sottrup-Jensen *et al.*, 1983b) (see below).

The covalent binding reaction between activated C3 or C4 and macromolecular targets can be suppressed to a variable extent by a wide range of nucleophiles of greatly differing structure, e.g., mono- and polyhydroxy alcohols,

simple carbohydrates, mono- and diamines, amino acids, and hydroxamic acids. As a result of this competition the nucleophiles themselves become covalently bound to the reactive Glx residues, and in some cases nearly complete incorporation can be achieved. Similarly the extent of covalent α_2M–proteinase complex formation can be greatly reduced by a wide range of competing nucleophiles. In order to become efficiently incorporated, the nucleophiles in question must be present prior to activation, since otherwise no significant incorporation is seen. These observations are compatible with the existence of a proteolytically activated, highly reactive state of these proteins, in which the thiol esters show a vastly increased reactivity toward a wide range of nucleophiles differing in structure. Furthermore, the covalent binding potential rapidly decreases. It has been difficult to obtain accurate estimates of the half-life of this short-lived state, but R. B. Sim *et al.* (1981) indicated that the reactive site in activated C3 decayed with a half-life of about 60 μsec. Sottrup-Jensen *et al.* (1981c) found for α_2M that no significant incorporation of nucleophiles would occur a few seconds after activation. This contrasts with the results of Salvesen *et al.* (1981), which indicated that the decay of the activated state had a half-life of about 112 sec. While it has been generally assumed that for those activated molecules of α_2M, C3, and C4 that fail to react with the targets the thiol-esterified Glx residues are simply hydrolyzed by water to form a Glu residue, no proof has yet been provided. Indeed, the results of Sottrup-Jensen *et al.* (1983b) and Feinman *et al.* (1985) have indicated the possibility that a fraction of the activated thiol ester might undergo a rapid isomerization to the lactam structure, similar to that seen with the model thiol ester peptides (Khan and Erickson, 1982). This could create a relatively long-lived state also capable of covalent binding. Although not investigated in detail, activated α_2M seems to react most readily with N-nucleophiles. The results from several groups show that nascent C3b seems to react preferentially with O-nucleophiles in contrast to nascent C4b, which seems to react most readily with N-nucleophiles. However, evidence has been presented that the isotypes of C4, C4A and C4B, differ in their preference for O- versus N-nucleophiles (Law *et al.*, 1984b; Isenman and Young, 1984). The structural basis for this variability in reactivity is unclear, and inspection of the sequences around the thiol ester sites has not yet provided any clues. It has been suggested that the variation in the hemolytic activity of the allotypes within each C4A and C4B series, apparently effected by a limited number of amino acid substitutions (Belt *et al.*, 1984, 1985) and further modulated by the amount of carbohydrate and degree of proteolytic processing, might enable the total population of C4 species to bind to a wide spectrum of targets (Law *et al.*, 1984b; Dodds *et al.*, 1985).

While the high reactivity of the thiol ester in the nascent state has been generally explained by an "exposure" of the thiol esters as a result of proteolytic or nonenzymatic activation, this does not in itself explain the high reactivity

against a wide range of nucleophiles differing in size and structure. It seems necessary to invoke not only a major change in the accessibility but also a mechanism by which the thiol esters become very reactive. The observed short half-life of the nascent thiol ester structure is compatible with a rate enhancement in this state of several orders of magnitude, perhaps as much as 10^5–10^8. From studies with model thiol esters it is known that the rate of nucleophilic substitution can be increased drastically by imidazole catalysis. Thus, the rate of intramolecular imidazole-catalyzed hydrolysis of *n*-propyl-γ-(4′-imidazolyl)thiol butyrate proceeds at near enzymatic rates (Bruice and Benkovic, 1966). This indicates that the imidazole ring of a histidyl residue positioned in the same steric relationship to the internal thiol ester by the conformational change elicited by activation cleavage could function as a "catalytic" group (Janatova *et al.*, 1980a; Tack *et al.*, 1980). In general, the covalent binding reaction shows some similarity with those catalyzed by transglutaminases (Folk and Chung, 1973; Folk and Finlayson, 1977; Folk, 1983). Perhaps these enzymes could serve as useful models for the reactions of the nascent state of α_2M, C3, and C4 (Parameswaran and Lorand, 1981). It is still speculative that the Glu residue preceding the thiol-esterified Glx residue could serve to increase the electrophilicity of the reactive carbonyl group (Davies and Sim, 1981). Other possible means of increasing the reactivity of the nascent thiol esters could include strain, distortion of bond angles, and hydrophobic effects (Jencks, 1969). Whatever the actual mechanism for increasing the reactivity in the nascent state is, this state can be considered a "pseudoenzymatic" state, characterized by an intermediate that only turns over once and for all.

Whether the covalent binding potential of the nascent state of activated α_2M plays any role in the binding of other nucleophiles besides the proteinase being complexed is not known. When α_2M was complexed with trypsin under conditions, in which all four thiol esters become activated (see Section IX), the nascent complex would covalently bind insulin and TLCK-treated trypsin (Sottrup-Jensen *et al.*, 1981c). Similarly Salvesen and Barrett (1980) reported binding of a variety of proteins to activated α_2M. The finding that hormones such as insulin and human growth hormone (Adham *et al.*, 1968, 1969) and platelet-derived growth factor (Salomon *et al.*, 1982; Huang *et al.*, 1984) will become associated with α_2M during the *in vitro* clotting of blood suggests that this binding might occur under physiological conditions. Recently, Boffa *et al.* (1985) reported that a glycopeptide present in human serum capable of inhibiting the G_1–S transition of dividing rat hepatocytes (Auger *et al.*, 1983) was associated with α_2M. In the case of platelet-derived growth factor, however, the binding was apparently due to disulfide bridge formation (Huang *et al.*, 1984), perhaps as a result of disulfide exchange reactions involving the SH group appearing after thiol ester cleavage.

IX. Formation and Properties of α_2M–Proteinase Complexes

A. Mechanism of α_2M–Proteinase Complex Formation

A tentative structural model for α_2M–proteinase complex formation, which incorporates essential features of the "trap hypothesis" of Barrett and Starkey (1973), the shape of α_2M determined from electron microscopy and X-ray scattering, and the covalent binding potential of the nascent state, is shown in Fig. 13. This model is similar to that presented recently by Feldman *et al.* (1985b) and is compatible with a large body of diverse experimental data. As discussed in Section VI, the disulfide-bridged dimers of α_2M (the "half-molecules") constitute its proteinase-binding functional domains. This is further supported by recent investigations on PZP (Sand *et al.*, 1985), which is a dimeric human α_2M homologue, dimeric frog α_2M (Feldman and Pizzo, 1984b, 1986), and plaice αM (Starkey *et al.*, 1982). In addition, partial reduction of human α_2M results in noncovalently associated dimers (Barrett *et al.*, 1979), which are active in proteinase binding (Gonias and Pizzo, 1983a,b). The exact relationship of those dimers to the disulfide-bridged dimers of native α_2M has not yet been settled, however, since several intrachain bridges have been reduced in addition to the two interchain bridges of the dimer.

In α_2M two such 360-kDa dimers are noncovalently associated to form a structure presumably having the overall shape of a hollow cylinder with the dimensions given in Fig. 8. It is likely that the "walls" of the cylinderlike structure are not massive or contiguous, but rather constituted by a meshlike or lobed structure, in which long and short arms can be seen in electron micrographs. The long arms are rather flexible, being composed of a number of smaller domains, whose relative orientation can change as a result of bait region cleavage or thiol ester cleavage. The complete tetrameric structure probably contains three C_2 symmetry axes. The proteinase-binding unit can thus be visualized as a basketlike structure able to accommodate proteinases of varying sizes. The inner diameter of the basket presumably is around 50–60 Å, thereby possibly limiting the access to proteinases having a maximal diameter of about 40 Å (assuming a spherical shape). Upon complex formation with a proteinase the long arms presumably are in close contact with the proteinase, thereby firmly positioning the proteinase within the lumen of the basket (Feldman *et al.*, 1985b).

As judged from the relatively strong reversible interaction between α_2M and anhydrotrypsin (Tsuru *et al.*, 1978; Sayers and Barrett, 1980; Wu *et al.*, 1981), the bait region sequences are reminiscent of the reactive site of "classical"

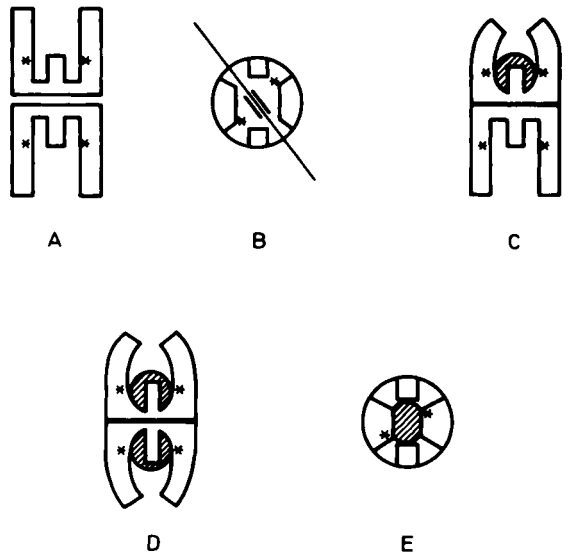


Fig. 13. Model of α_2M in which the bait regions (——) and the reactive thiol esters (*) have been tentatively localized in the dimeric functional unit. (A) The usual monogramlike projection seen in electron micrographs (Fig. 8). (B) A vertical view through the hollow cylinder; the line divides the two monomers in the dimer. The bait regions are thought to be located close to one and another at the bottom of the basketlike dimer. The reactive thiol esters in the dimer could be diametrically located, fairly close to the bait regions, and facing the lumen of the basket. They could possibly be positioned at the edge of a domain in the lower part of the long arms. (C) The 1 : 1 complex formed under conditions of low proteinase activity, where the second dimer is essentially intact. (D) The 2 : 1 complex formed at high levels of proteinase activity. This has led to a considerable compacting of the α_2M tetramer. (E) A vertical view of one dimer, in which the bound proteinase is covalently linked at two points to the diametrically positioned Glx residues (see text for further details).

proteinase inhibitors (Laskowski and Sealock, 1971). However, they evidently function as activation cleavage sites in the interaction of α_2M with active proteinases and are presumably located close to one another at the bottom of the basket. The reactive thiol esters in the dimer could be located in the lower part of the long arms, fairly close to bait regions. Presumably they are located in diametrically opposite positions at the internal surface (see also below).

While maximally two proteinases can be bound to the α_2M tetramer, the estimates of the actual binding ratio vary between one and two for a wide range of proteinases studied. It has now become clear that a major factor controlling the binding ratio is the level of proteolytic activity. Thus, Christensen and Sottrup-Jensen (1984) have identified two distinct routes of complex formation, dependent on the rate of initial bait region cleavage in agreement with the results of Howell *et al.* (1983) (pathways I and II).

$$\alpha_2M + P \xrightarrow{1} \alpha_2M\text{—}P(2SH) \xrightarrow{2} \alpha_2M\text{—}P_2(4SH) \quad \text{(I)}$$

one bait region cleaved; one bait region cleaved

$$\alpha_2M + P \xrightarrow{1} \alpha_2M\text{—}P(4SH) \xrightarrow{2} \alpha_2M\text{—}P_2(4SH) \quad \text{(II)}$$

two bait regions cleaved; two bait regions cleaved

Presumably the conformational changes within the dimeric unit proceed with an overall rate constant, which is independent of the particular proteinase being complexed. In contrast, the rate of bait region cleavage may differ by several orders of magnitude, reflecting the quality of the bait region as a substrate for a given proteinase. Table III shows a comparison of second-order rate constants determined for the association of α_2M with a variety of proteinases.

The cleavage of one bait region is a prerequisite for the formation of a tight, essentially irreversible complex. If this step is rate determining (e.g., with plasmin, thrombin, and benzamidine-inhibited trypsin), a final 1 : 1 complex is formed with three intact bait regions but only two thiol esters cleaved (Christensen and Sottrup-Jensen, 1984; Steiner *et al.*, 1985) (pathway I, step 1). Evidently this suggests that the two subunits within the dimer are strongly interacting, as also found for the methylamine-dependent thiol ester cleavage (Larsson *et al.*, 1985). The actual thiol ester cleavage steps are not known, but evidently both thiol esters of the dimer become activated. Covalent binding of the proteinase takes place from the thiol-esterified Glx residue in both the proteolytically cleaved and uncleaved subunit within the dimer resulting in binding of a fraction of the proteinase through two cross-links (Sottrup-Jensen and Hansen, 1982). Using ^{125}I-labeled proteinases this can readily be seen from the distribution of label among the different fragments obtained from the 360- to 400-kDa dimeric complex upon reduction (Fig. 14) (Salvesen and Barrett, 1980; Sottrup-Jensen *et al.*, 1981d, 1983b; Harpel and Brower, 1983; Straight and McKee, 1984; Wang *et al.*, 1984; Feinman *et al.*, 1985), and it is consistent with the localization of ϵ-lysyl-γ-glutamyl cross-links within the α_2M–trypsin complex (see below). The presence of species of still higher molecular weight containing bound thrombin (Wang *et al.*, 1984; Feinman *et al.*, 1985) or urokinase (Waller *et al.*, 1983; Straight *et al.*, 1985) can perhaps be attributed to covalent binding reactions between unbound enzyme and the thiol esters of intact dimers occurring during denaturation, since the low rate of reaction with thrombin or urokinase will leave some dimers intact.

While the "unreacted" dimer is intact with regard to its bait regions and thiol esters, its conformation is nevertheless somewhat changed as a result of the conformational change within the dimer to react first. That change clearly affects the gross conformation and accessibility of the bait regions in the adjacent dimer, since the subsequent binding of a second proteinase, dependent on the cleavage

TABLE III

Second-Order Rate Constants for Bait Region Cleavage and Association between Proteinases and α_2M

Substance	k_{ass} (M^{-1} sec^{-1})	Estimated half-time of association in plasma (msec)[a,b]
Trypsin	2×10^{7}[c]	
	1.3×10^{7}[d]	19
	1.7×10^{7}[e]	
Leukocyte elastase	4.1×10^{7}[f]	7
Chymotrypsin	1.2×10^{7}[f]	27
Pancreatic elastase	$3.4–4.4 \times 10^{6}$[g]	80
Cathepsin G	3.7×10^{6}[f]	93
Plasmin	5×10^{5}[c]	500
	3×10^{3}[d]	
Plasma kallikrein	2.3×10^{4}[f]	6125
Thrombin	4.9×10^{3}[h]	7×10^{4}
	2.5×10^{3}[i]	
Factor X_a	1×10^{3}[j]	3×10^{5}

[a]Overall rate constant (Travis and Salvesen, 1983).

[b]Estimated from the expression $t_{1/2} = 1/k_{ass}$ (I), where k_{ass} is the second-order rate constant and (I) is the plasma concentration of the inhibitor (Bieth, 1980).

[c]First bait region cleavage step.

[d]Second bait region cleavage step (Christensen and Sottrup-Jensen, 1984).

[e]Overall rate constant (Barrett and Salvesen, 1979).

[f]Overall rate constant (Virca and Travis, 1984).

[g]Overall rate constant (Meyer *et al.*, 1975; Bieth and Meyer, 1984).

[h]Bait region cleavage.

[i]Overall rate constant (Straight and McKee, 1982; Steiner *et al.*, 1985; Björk *et al.*, 1985).

[j]Overall rate constant (Ellis *et al.*, 1982).

of at least one bait region, and ultimately leading to a 2 : 1 proteinase–α_2M complex proceeds more slowly. In this reaction (pathway I, step 2) the remaining two thiol esters become cleaved and the overall process of SH appearance shows negative cooperativity (Christensen and Sottrup-Jensen, 1984). "Half-reacted" 1 : 1 proteinase–α_2M complexes capable of binding one additional proteinase have been reported earlier (Jacquot-Armand and Guinand, 1976; Pochon *et al.*, 1978, 1981) and are only formed under conditions of low proteinase activity. It is likely that α_2M–proteinase complexes formed *in vivo* mainly will be of that type (Christensen and Sottrup-Jensen, 1984; Virca and Travis, 1984).

Under conditions of high proteinase activity, e.g., by using trypsin at micro-

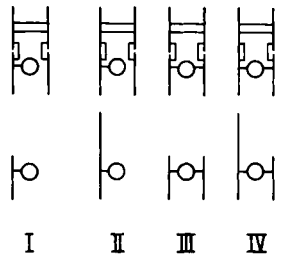


Fig. 14. Schematic representation of different structures in which a labeled proteinase is covalently bound via one or two cross-links to the α_2M dimer. The two interchain disulfide bridges and the bridge spanning the activation cleavage site are shown. In nonreducing SDS–PAGE the dimer containing a bound, labeled proteinase migrates as a 360- to 400-kDa molecular species (upper row). Upon reduction a spectrum of labeled cross-linked species consistent with structures I–IV is found (lower row). Their approximate sizes are 100 (I), 200 (II and III), and 270 kDa (IV).

molar concentrations as done in many *in vitro* experiments, bait region cleavage proceeds rapidly and is not rate determining. Thus, more than one bait region in different dimers may be cleaved before the proteinase is finally bound to one dimer and further cleavages prevented (pathway II, steps 1 and 2). In such complexes all four SH groups in the tetramer appear initially for one proteinase molecule bound, and both 1 : 1 and 2 : 1 proteinase–α_2M complexes form, dependent on the ratio of proteinase to α_2M (Sottrup-Jensen *et al.*, 1980; Christensen and Sottrup-Jensen, 1983; Howell *et al.*, 1983). Under these conditions the stoichiometry between bait regions cleaved and proteinase bound is close to 2 : 1 (Swenson and Howard, 1979a). Christensen and Sottrup-Jensen (1984) estimated that the change from reaction path I to II would be expected to occur between 0.02 and 0.5 μM trypsin at micromolar concentrations of α_2M.

Under conditions of high proteinase activity the conformational change from "slow" to "fast" α_2M is complete at a 1 : 1 ratio of proteinase to α_2M (Van Leuven *et al.*, 1981a; Sottrup-Jensen *et al.*, 1983b; Gonias and Pizzo, 1983c), and the binding of a second proteinase must accordingly occur before the conformational change within the tetramer has gone to completion (Howell *et al.*, 1983; Christensen and Sottrup-Jensen, 1984). While the binding of the second proteinase to α_2M under these conditions has been judged to show some negative cooperativity (Christensen and Sottrup-Jensen, 1983), this has not been found by others (Pochon and Bieth, 1982; Gonias and Pizzo, 1983c; Björk *et al.*, 1984). However, in preformed 1 : 1 complexes (formed at micromolar contrations) of reactants the adjacent dimeric unit can no longer bind proteinases (Saunders *et al.*, 1971; Gonias and Pizzo, 1983c). It has been suggested that the lack of formation of 2 : 1 plasmin–α_2M complexes was due to the large size of this proteinase. Thus, the heavy chain portion was proposed to extend into the second binding site in α_2M (Pochon *et al.*, 1981). However, 2 : 1 complexes do form to

some extent (Straight and McKee, 1982; Christensen and Sottrup-Jensen, 1984), albeit slowly, due to the somewhat changed conformation of the unreacted dimer in the 1 : 1 complex. It is unlikely from the model shown in Fig. 13 that the heavy chain portion of a 1 : 1 plasmin–α_2M complex actually could interfere with the second binding site, since model building indicates that the heavy chain of plasmin must extend from a position in the tertiary structure of the light chain almost diametrically opposite that of the active site due to the presence of the disulfide bridge in plasmin equivalent with the bridge Cys-1–Cys-122 in chymotrypsinogen. Indeed, recent results by Cummings and Castellino (1984) show that the major part of the heavy chain in the plasmin–α_2M complex is solvent accessible and can react with monoclonal antibodies directed against epitopes in the kringle 1–4 domains. In contrast small proteinases like trypsin, chymotrypsin, and elastase are poorly accessible to polyclonal antibodies when complexed to α_2M (Haverback *et al.*, 1962; Rinderknecht *et al.*, 1975; Geokas *et al.*, 1977; Ohlsson and Ohlsson, 1978; Miyata *et al.*, 1981).

The localization of the cross-links within the α_2M–trypsin complex has proven to be technically difficult. However, Sottrup-Jensen *et al.* (1983b) were able to identify five major ϵ-lysyl-γ-glutamyl cross-links formed from the thiol-esterified Glx residues to Lys residues in bovine trypsin. Using the three-dimensional model of trypsin (Huber *et al.*, 1974) these Lys residues were found to be grouped in two surface patches containing Lys-107 and Lys-239; and Lys-145, Lys-188A, and Lys-222, respectively. Strikingly, these patches are located in almost diametrical positions, each having nearly the same distance to the active site Ser-195 (chymotrypsinogen numbering used). This result indicates a concerted sequence of bait region cleavage, fast conformational change, thiol ester cleavage, and covalent bond formation. In particular, the conformational change must result in a fixation of trypsin within the basket, since a more or less random reaction with available nucleophilic side chains on its surface otherwise would result.

These results also suggest that the pair of thiol esters in the dimeric unit of α_2M that minimally becomes activated upon complex formation is located at opposite positions at the lumen of the cavity that presumably closes upon bait region cleavage (Fig. 13). Furthermore, as judged from the distance of each of the reacting patches of lysyl residues to the active site of trypsin, it is likely that the distance between the bait region stretches and the thiol esters is only about 15 Å, i.e., the thiol esters in the dimeric unit are separated by about 30 Å. Clearly this is a minimal estimate, since the domain(s) containing the thiol esters could move a considerable distance upon bait region cleavage. From energy transfer experiments Pochon *et al.* (1983a,b) estimated that the thiol esters within a dimeric unit were located relatively close to one another and close to the proteinase binding site. The two pairs of thiol esters were judged to be separated by 60–80 Å. Thus, it is likely that the restriction of covalent bond formation from

the thiol-esterified Glx residues to these few Lys residues in trypsin to a large extent is dependent on the proper orientation of the side chains of these residues with respect to the thiol esters when the complex forms. Interactions between the active site and the bait region(s) during complex formation and the fast conformational change involving the long arms might be responsible for this.

However, since it is apparent that the covalent binding reaction is not a prerequisite for the formation of relatively stable reversible complexes (Salvesen *et al.*, 1981; Sottrup-Jensen *et al.*, 1981a; Van Leuven *et al.*, 1981a, 1982a,b; Christensen and Sottrup-Jensen, 1983, 1984; Wang *et al.*, 1981, 1983, 1984) it has been suggested that the covalent binding reaction is merely incidental, and that the conformational change per se leading to entrapment of the proteinase adequately describes the mechanism of complex formation (Salvesen *et al.*, 1981; Van Leuven *et al.*, 1981b, 1982a; Travis and Salvesen, 1983; Eccleston and Howard, 1985). Additional support for this suggestion is provided by the proteinase-binding properties of ovostatin, which does not contain thiol esters (Nagase *et al.*, 1983; Nagase and Harris, 1983), and the apparently widely differing extent of covalent binding determined for proteinase–α_2M complexes. Although earlier work on the covalent α_2M–proteinase complex formation focused on the formation of the relatively stable ϵ-lysyl-γ-glutamyl cross-links (Salvesen *et al.*, 1981; Van Leuven *et al.*, 1981b; Sottrup-Jensen, 1981c,d; Wu *et al.*, 1981), the potential of the nascent α_2M–proteinase complex to form cross-links of the ester type (involving Ser, Thr, and Tyr residues) should not be overlooked. Such cross-links could possibly be unstable in the conditions of denaturation employed in the estimation of the fraction of covalent-bound proteinase, e.g., boiling in SDS. This is supported by recent reports showing that the outcome of such experiments depends on the buffers used (Harpel and Brower, 1983; Wang *et al.*, 1984). An additional source of underestimation of the extent of covalent binding resides in the possibility that during denaturation of the complexes some degradation of α_2M can take place as further discussed by Sand *et al.* (1985). From this study it is evident that the extent of covalent binding of many proteinases to PZP is very high (at least 70%), and additional results on α_2M (Wu *et al.*, 1981; Straight and McKee, 1984; Gonias and Pizzo, 1983a,b; Pochon *et al.*, 1983b; Sand *et al.*, 1985) suggest that the covalent binding may proceed with nearly 100% efficiency. Presumably, the full covalent binding potential of α_2M is realized under most conditions *in vivo* and cannot be neglected in formulating a mechanism of α_2M–proteinase complex formation. Further support for an important role of the covalent binding reaction is provided by the results of Wang *et al.* (1981, 1983) and Wu *et al.* (1981). In their studies a large fraction of α_2M-bound trypsin or thrombin, in which the ϵ-amino groups of Lys residues of the proteinases had been blocked or otherwise modified, could be dissociated by other proteinases, albeit slowly. In addition, the extent of bait region cleavage in α_2M by the modified enzymes was much higher than observed

with the native proteinases. Finally, Christensen and Sottrup-Jensen (1983) analyzed the biphasic reaction between STI and α_2M-bound trypsin (Bieth *et al.*, 1981; Wang *et al.*, 1981) and suggested that noncovalently bound trypsin reacted about tenfold faster with STI than covalently bound trypsin did. This interpretation was, however, not shared by Tourbez *et al.* (1984), who observed a similar biphasic reaction with α_2M-bound trypsin and pancreatic trypsin inhibitor (PTI), even when the complexes were formed in the presence of hydroxylamine, which nearly completely abolished covalent complex formation.

B. Inhibitory Spectrum of α_2M and Other Macroglobulins

While it has been generally recognized that α_2M will form complexes with a wide range of proteinases, the rate of complex formation differs greatly for different proteinases as exemplified in Table III. It is apparent that proteinases of relatively narrow substrate specificity like plasmin and thrombin react much less readily than proteinases of relatively broad substrate specificity such as trypsin and chymotrypsin. However, even a proteinase like urokinase will slowly form a complex with α_2M, in which a fraction is covalently bound (Waller *et al.*, 1983; Straight *et al.*, 1985), and mammalian collagenases react readily with α_2M (Abe and Nagai, 1973; Werb *et al.*, 1974). Assuming that a half-life of complex formation of less than 100 msec reflects a physiologically relevant reaction (Travis and Salvesen, 1983), it is likely that the main targets for α_2M are proteinases of relatively broad substrate specificity, perhaps primarily proteinases of cellular origin. However, in evaluating the relative importance of complex formation with α_2M it is indeed puzzling that for most proteinases in plasma there seems to be at least one other inhibitor that reacts as rapidly as α_2M or even more rapidly (Travis and Salvesen, 1983). This could imply a role of α_2M as a "back up" inhibitor, important only under conditions in which the primary inhibitor is being depleted or under deficiency states.

The relative importance of proteinase complex formation with α_2M may be different *in vivo* from expectations based on the rates of complex formation *in vitro*. While *in vitro* studies employing purified α_2M and antithrombin III indicated that only a few percent of thrombin would be complexed to α_2M, thrombin generated in clotting plasma was found to be associated with α_2M to a large extent (Shapiro and Anderson, 1977; Downing *et al.*, 1978). Similarly, Harpel (1977, 1981) found in a plasma system that α_2M would effectively compete with α_2-antiplasmin under conditions where plasmin was generated *in situ* by activation with urokinase, rather than being added directly to plasma. Likewise Fuchs and Pizzo (1983) demonstrated that factor X_a when injected into the mouse circulation rapidly became associated with α_2M, contrary to what would be

expected from *in vitro* data (Ellis *et al.*, 1982). However, in a similar experiment with factor IX_a no complex formation with α_2M was observed (Fuchs *et al.*, 1984). These and other results, e.g., on the distribution of neutrophil elastase and cathepsin G between α_1-proteinase inhibitor and α_2M (Virca and Travis, 1984), demonstrate the difficulties in assigning an unequivocal role of α_2M in inhibition of endogenous proteinases. Given the content in plasma of effective specific proteinase inhibitors, Starkey and Barrett (1977) speculated if a major physiological role of α_2M might reside in the inhibition and clearance of proteinases secreted by invading pathogens and other parasites. Clearly, the bait region sequence of α_2M allows it to be recognized by proteinases of almost any substrate specificity.

In contrast to α_2M only a few studies have been conducted on the related rat and rabbit α_1M and α_2M, particularly with regard to a closer definition of the range of proteinases that will interact with these macroglobulins (see, e.g., Gauthier and Mouray, 1975b, 1976). However, preliminary results on the interaction between PZP and proteinases (Sand *et al.*, 1985) suggest that PZP and α_2M have different, yet overlapping spectra of proteinases as targets as indicated from their different bait region sequences (Sottrup-Jensen *et al.*, 1984d; Sand *et al.*, 1985). Thus, PZP was found to react very slowly with trypsin and *S. aureus* proteinase compared with α_2M. In contrast, both PZP and α_2M reacted rapidly with chymotrypsin and elastase, suggesting that differences in specificity for proteinases may also be found for other pairs of macroglobulins. Given the high efficiency by which α_2M binds proteinases of relatively broad specificity, the physiological role of PZP seems even more puzzling than that of α_2M. It is possible that PZP specifically interacts with certain proteinases of cellular origin and thereby serves to augment an existing potential for proteinase binding under conditions of increased cellular turnover, e.g., in inflammatory processes. The increased level of PZP in such conditions and in pregnancy suggests an intriguing and previously unrecognized correlation in appearance of proteolytic activity (Waites and Bell, 1984; Teixeira and O'Grady, 1985; Sand *et al.*, 1985). Although not investigated in detail, it is possible that the recently described α_2M homologues murinoglobulin and rat α_1-inhibitor III may preferentially act on still other groups of proteinases.

C. Enzymatic Properties of α_2M-Bound Proteinases

Following the demonstration by Haverback *et al.* (1962) and Mehl *et al.* (1964) that proteinases bound to α_2M are enzymatically active, and the later finding of Ganrot (1967) that trypsin bound to α_2M was protected from inhibition by STI, the kinetic properties of the bound enzymes have been investigated by

many groups (see, e.g., reviews by Starkey and Barrett, 1977; Roberts, 1985). In general, the activity of the bound proteinase toward small synthetic substrates is slightly decreased, while the activity toward large substrates such as casein is greatly decreased. From recent experiments (e.g., Christensen and Sottrup-Jensen, 1983; Gonias and Pizzo, 1983b; Cummings and Castellino, 1984) it has been shown that the catalytic efficiency of the bound enzymes toward small synthetic substrates, as expressed by the k_{cat}/K_m value, is decreased two- to fourfold, showing that the active site of the bound enzymes must be readily solvent accessible, i.e., located at or very near the surface of the complex. Similarly, Pochon and Bieth (1982) found that α_2M-bound chymotrypsin could bind to immobilized D-tryptophan methyl ester. While this might be difficult to understand in terms of a burial of the bound enzymes within α_2M, it is compatible with a meshlike shape of α_2M (Fig. 12). Furthermore, the enzymatic activity is identical for both proteinase molecules bound in the 2 : 1 complex (Christensen and Sottrup-Jensen, 1983).

Although the active site of an α_2M-bound proteinase is solvent accessible, it is nevertheless sterically hindered, since an inhibitor like STI reacts only slowly with the bound trypsin forming a ternary complex. The overall second-order rate constants for the association and the dissociation constant for that reaction have been estimated at 5–200 M^{-1} sec^{-1} and 4–100 $\times$ 10^{-6} M, respectively, contrasted with 1.2 $\times$ 10^7 M^{-1} sec^{-1} and 3.6 $\times$ 10^{-12} M for the free enzyme (Bieth *et al.*, 1981; Wang *et al.*, 1981; Christensen and Sottrup-Jensen, 1983). Pochon and Steinbuch (1984) reported that the dissociation constant for hirudin bound to α_2M–thrombin is 1 $\times$ 10^{-7} M, in contrast with a value of about 1 $\times$ 10^{-12} to 1 $\times$ 10^{-14} M for the unbound thrombin. A small inhibitor like PTI reacts more readily with α_2M-bound trypsin than STI, albeit still with a greatly decreased rate (Ganrot, 1967; Tourbez *et al.*, 1984). That steric shielding of the active site of an α_2M-bound proteinase toward substrates of high molecular weight is not complete is further illustrated by the observations that α_2M–thrombin retained some clotting activity (Rinderknecht and Geokas, 1973) and that it could activate factor VIII (Switzer *et al.*, 1983). Likewise, α_2M–plasmin has been reported to retain both fibrinogenolytic and fibrinolytic activity (Harpel and Mosesson, 1973; Veremeenko and Kizim, 1981). While it has been suggested that this activity might be important in modulating physiological processes like blood clotting and fibrinolysis (Harpel and Rosenberg, 1976; Switzer *et al.*, 1983), this seems unlikely, particularly in view of the reported low activities and the rapid clearance from the circulation of α_2M–proteinase complexes. However, under conditions of impaired clearance α_2M–proteinase complexes might exert a local proteolytic activity of possible significance for cellular interactions in the inflammatory state (Abe and Nagai, 1973; Ohlsson and Skude, 1976; Bieth and Klumpp, 1976; Balldin and Ohlsson, 1979; Gaspar *et al.*, 1984; Borth, 1984).

X. Cell Recognition, Receptor-Mediated Endocytosis, and Cellular Effects

A. Macroglobulins

The first indication that α_2M–proteinase complexes are cleared from the circulation was made by Nilehn and Ganrot (1967), who studied the elimination of plasmin generated in patients receiving streptokinase for treatment of deep vein thrombosis. Later Ohlsson *et al.* (1971) and Ohlsson (1971a,b, 1974) reported that the elimination of trypsin–α_2M complexes in the dog was very fast with approximately 85–90% being removed from the circulation in 30 min (half-time 5–8 min). The major organ for clearance was found to be the liver in adult animals, while the bone marrow also contributed significantly in whelps. Kupffer cells and other cells of the reticuloendothelial system were judged to be the most important cells in this regard. Similar results were also reported for the clearance of subtilopeptidase A–α_2M complex in rabbits (Debanne *et al.*, 1973, 1975, 1976) and for the clearance of elastase–α_2M complex in rats (Katayama and Fujita, 1974a,b). Recently, an extensive series of studies on the clearance of macroglobulin–proteinase complexes in the mouse has been conducted by Pizzo and co-workers (Imber and Pizzo, 1981; Gonias and Pizzo, 1981a,b; Fuchs *et al.*, 1982; Fuchs and Pizzo, 1983; Gonias *et al.*, 1982a, 1983; Feldman *et al.*, 1983, 1984, 1985a; Feldman and Pizzo, 1984a). In addition Davidsen *et al.* (1985) and Gliemann *et al.* (1985) have examined the clearance of α_2M–trypsin in rats. In agreement with *in vitro* studies (see below) complexes between human α_2M or its homologues from other species and proteinases from all four classes EC 3.4.21–24 are cleared via specific, saturable membrane-bound receptors. The half-time of elimination is usually 2–5 min, dependent on the particular protein studied, in contrast with many hours for the native macroglobulins. Furthermore, the rate of clearance is largely independent of the particular proteinase being complexed, and with the apparent exception of rat α_2M and one of the mouse αM's the methylamine-treated macroglobulins are cleared equally well via the same pathway. These results strongly indicate that the cleavage of the thiol ester in the native macroglobulins is intimately associated with the exposure of a receptor recognition site on each subunit (Van Leuven *et al.*, 1981a, 1982a; Marynen *et al.*, 1981) (see Section VIII). In view of the pronounced domain structure of α_2M this site is presumably located on a distinct domain and concealed in the native structure. Using a monoclonal antibody Marynen *et al.* (1981, 1982a) characterized a neoantigen on amine or proteinase-treated α_2M not present in the native protein, presumably located at or close to the receptor recognition site as seen from blocking experiments. Prolonged digestion with trypsin resulted in the loss of this antigenic determinant.

Although it is evident from *in vivo* experiments that complexes between pro-

teinases and, e.g., both dog αM's and rabbit αM's clear at roughly similar rates (Ohlsson 1971a,b, 1974; Debanne *et al.*, 1973), and mutual competition for clearance has been observed for some macroglobulins (Gonias *et al.*, 1983), it is not known if this will be the case for all pairs of macroglobulins. Thus, Gonias *et al.* (1983) noted a difference in the clearance between methylamine-treated rat and mouse α_1M's and α_2M's, and J. Gliemann and L. Sottrup-Jensen (unpublished) have observed distinguishing features in the clearance of rat α_1M–proteinase complexes and rat α_2M–proteinase complexes. Proteinases complexed with the third macroglobulin of the rat, α_1-inhibitor III, are also rapidly cleared from the circulation (Gauthier and Ohlsson, 1978; J. Gliemann, K. Lonberg-Holm, and L. Sottrup-Jensen, unpublished). As a result of the rapid elimination, the level of circulating α_2M–proteinase complexes is very low. Using a monoclonal antibody specific for proteinase-treated α_2M Marynen *et al.* (1983) estimated a level of 3.2–4.8 μg/ml for such complexes, corresponding to about 0.12–0.16% of the α_2M present in normal plasma. As expected the level of complexes was found to be much higher in serum (52–86 μg/ml, corresponding to 1.9–3.0%).

From recent analyses of the distribution of ^{125}I-labeled trypsin–α_2M complexes in different tissues of the rat, it has been confirmed that the liver as originally reported by Ohlsson (1971a) is the main organ responsible for clearance. However, compatible with results obtained *in vitro* (Gliemann *et al.*, 1983), the hepatocytes have been identified as the main cells responsible for elimination of α_2M–trypsin complex (Davidsen *et al.*, 1985; Feldman *et al.*, 1985a) and not the Kupffer cells as previously thought. Very similar results have also been reported for the uptake of rat α_1M–subtilisin complex (Bergsma *et al.*, 1985). While the Kupffer cells actively take up and clear αM–proteinase complexes they cannot, by virtue of their low abundancy, contribute significantly to the overall process. However, it is possible that uptake of α_2M–proteinase complexes into Kupffer cells and macrophages or monocytes besides leading to degradation results in the generation of "modulatory signals" (see below). Following clearance the complexes are rapidly degraded to products of low molecular weight presumably by lysosomal proteinases (Ohlsson, 1971b; Katayama and Fujita, 1974b). A number of *in vitro* studies have led to a more detailed description of the binding, endocytosis, and degradation of methylamine- or proteinase-treated α_2M. Apart from hepatocytes (Glieman *et al.*, 1983; Feldman *et al.*, 1985a; Bergsma *et al.*, 1985; Davidsen *et al.*, 1985; Spolarics *et al.*, 1985) and macrophages (Debanne *et al.*, 1975, 1976; Kaplan and Nielsen, 1979a,b; Kaplan and Keogh, 1981; Kaplan *et al.*, 1981; Imber and Pizzo, 1981) several other cell types in culture have been shown to contain receptors for "fast form" α_2M. These include fibroblasts and fibroblastlike cell lines (Mosher *et al.*, 1977; Mosher and Vaheri, 1980; Van Leuven *et al.*, 1977, 1978, 1979, 1980, 1981a; Zardi *et al.*, 1980; Schlessinger *et al.*, 1978; Maxfield *et al.*, 1978, 1981; Willingham *et al.*, 1979; Dickson *et*

al., 1981a,b; Via *et al.*, 1982; Schlegel *et al.*, 1982; Hanover *et al.*, 1983a), adipocytes (Gliemann *et al.*, 1983; Ney *et al.*, 1984), and epididymal epithelial cells (Djakiew *et al.*, 1984, 1985).

The mechanism of endocytosis of α_2M–proteinase complexes seems to be very similar to that first described for human low-density lipoprotein (Anderson *et al.*, 1977; Brown and Goldstein, 1979; Goldstein *et al.*, 1979) and for the internalization of hormones and other proteins into cells (for reviews, see, e.g., Pastan and Willingham, 1981; Steinman *et al.*, 1983; Besterman and Low, 1983). In an extremely simplified version the following events can be recognized: After binding to the cell surface receptors, α_2M–proteinase complexes appear in clathrin-coated pits and are then transferred to receptosomes, which are uncoated endocytic vesicles. In yet unknown processes the Golgi system is thought to be involved in the subsequent transfer of the complexes to lysosomes for degradation. During this process the receptors are separated from the ligands, recycle, and become again incorporated into the cell membrane.

Binding of the α_2M–proteinase complex to its receptor is very tight with apparent K_{diss} values reported to be in the nanomolar range; it requires the presence of calcium ions and can be competitively inhibited by the peptide antibiotic bacitracin (Van Leuven *et al.*, 1981c). A number of amines have been reported to interfere with the clustering of receptor complexes and to inhibit the uptake of α_2M–proteinase complexes (Maxfield *et al.*, 1979; Levitzki *et al.*, 1980; Davies *et al.*, 1980; Van Leuven *et al.*, 1980; Kaplan and Keogh, 1981; Schlegel *et al.*, 1982; Ohlsson *et al.*, 1982), and it has been speculated that cellular transglutaminases somehow could be involved in the uptake process, perhaps in the recycling of the receptors. This is intriguing in view of the presence of a specific transglutaminase acceptor site on α_2M (Mortensen *et al.*, 1981a) and the reported poor uptake into fibroblasts of α_2M–trypsin complex prepared from α_2M modified at this site by dansylcadaverine (Van Leuven *et al.*, 1981a). However, the uptake of a complex between trypsin and α_2M prepared from serum, in which about half of the transglutaminase-reactive Gln-671 residue had been hydrolyzed to a Glu residue, into rat hepatocytes was not suppressed relative to control preparations (Sottrup-Jensen *et al.*, 1984c), in agreement with recent results reported by Van Leuven (1984). Furthermore, Gln-671 has been found to be largely inaccessible to factor $XIII_a$ following complex formation (Van Leuven, 1984). Thus, at present there is no direct evidence for the involvement of transglutaminases in the uptake of α_2M–proteinase complexes. Indeed, as discussed by Tycho *et al.* (1983) many of the results obtained earlier can now be explained by a rise of the pH of the endocytotic vesicles following incubation of cells with amines.

The isolation of functionally active receptors for α_2M–proteinase complexes from fibroblasts has been reported recently. The preparation of Marynen *et al.* (1984b) contained components of 360, 130, and 85 kDa. In contrast, Frey and

Afting (1983) and Hanover *et al.* (1983b) estimated the subunit size of their preparations at 125 and 85 kDa, respectively. The latter species could be identical with the 130- and 83-kDa components seen by Marynen *et al.* (1984b). Apart from the specific receptors for α_2M–proteinase complexes on hepatocytes, Kupffer cells, macrophages, fibroblasts, and adipocytes, the results of immunohistochemical examination indicate that such receptors or other interacting proteins may be found on other cells as well. Thus, Becker and Harpel (1976), using indirect immunofluorescence histochemistry, observed that in addition to hepatocytes immunoreactive α_2M was present on endothelial cells of arteries, veins, and lymphatics. Using basically the same technique Cassiman *et al.* (1980) surveyed a number of human tissues and found immunoreactive α_2M only in connective tissues, leukocytes, and Kupffer cells. In cultured endothelial cells no binding or receptor-mediated endocytosis of α_2M–trypsin complex could be demonstrated (Marynen *et al.*, 1982b). The reason for this discrepancy is presently not clear. As discussed above factor X_a seems to bind exclusively to α_2M *in vivo*. The mechanism of clearance is more complex than that observed for other α_2M–proteinase complexes. Possibly, factor X_a binds initially to thrombin binding sites on the endothelial surface. This binding alters the specificity of factor X_a for the plasma proteinase inhibitors, so that binding to α_2M is favored at the expense of binding to α_1-proteinase inhibitor, thereby resulting in clearance of factor X_a via α_2M (Fuchs and Pizzo, 1983). Nachman and Harpel (1976) further demonstrated α_2M in membrane and granular fractions of platelets, and Ivanyi and Moyes (1980) showed binding of anti (α_2M)-antibodies to a lymphoblastoid cell line. Furthermore, Saksela *et al.* (1981, 1984) found α_2M in normal but not malignant syncytiotrophoblasts and cervical epithelium, compatible with the absence of receptors for α_2M–proteinase complexes in some transformed fibroblast cell lines as shown earlier by Van Leuven *et al.* (1979) and Zardi *et al.* (1980). Studies on the tissue localization of PZP have indicated its presence in the syncytiotrophoblast (Lin and Halbert, 1976; Chemnitz *et al.*, 1982) and on the surface of B lymphocytes and monocytes (Stimson, 1977; Horne *et al.*, 1978a,b, 1979; Thomson *et al.*, 1979). Given the close structural homology between α_2M and PZP their specific tissue distribution remains to be established. However, preliminary investigations suggest that PZP– and α_2M–proteinase complexes are cleared via the same receptors (J. Gliemann and L. Sottrup-Jensen, unpublished).

Although the potential for fast complex formation with proteinases and clearance appears to be the major function of α_2M and PZP, a number of reports have indicated that α_2M or α_2M–proteinase complexes may modulate some cellular functions, which could be relevant with regard to the immune response. Thus, Johnson *et al.* (1982) found that binding of α_2M–proteinase complexes to activated macrophages suppressed the excretion of three distinct proteinases from these cells. Furthermore, the superoxide anion production by activated mac-

rophages was found to be greatly suppressed by methylamine- or trypsin-treated α_2M, but not by native α_2M (Hoffmann *et al.*, 1983). In addition, endocytosis of α_2M–trypsin complexes has also been reported to activate macrophages, resulting in the production of neutral proteinases (Vischer and Berger, 1980). Some of the earlier reported effects of α_2M preparations on immunological reactions have been reviewed by James (1980) and additional observations have appeared. Thus, Cordier and Revillard (1980), Ades *et al.* (1982), and Dickinson *et al.* (1985) reported that α_2M could inhibit the neutrophil-mediated cytotoxicity and Hubbard *et al.* (1981), Miyanaga *et al.* (1982), and Rastogi and Clausen (1985) reported that α_2M (or α_2M–proteinase complexes) could inhibit the mixed lymphocyte reaction and proliferation of lymphocytes. α_2M or associated factors have also been shown to inhibit the growth of various tumor cells (Koo, 1981, 1983). Similar results have also been obtained with PZP (von Schoultz *et al.*, 1973; Stimson, 1976; Björksten *et al.*, 1978; Kasukawa *et al.*, 1979), and it has been speculated that PZP could play a role in the maintenance of the immunologically privileged state of the fetus. Other studies have indicated that α_2M–proteinase complexes could stimulate the locomotion of neutrophils and monocytes (Forrester *et al.*, 1983) and induce dissociation of rosettes formed between B lymphocytes and erythrocytes by cleaving membrane proteins (Mackin *et al.*, 1983). A polyclonal B-cell activator activity has also been attributed to α_2M–proteinase complexes or an associated lymphokine (Teodorescu *et al.*, 1981; Chang *et al.*, 1981, 1983; Ganea *et al.*, 1982).

While it had been suggested earlier that α_2M from patients with cystic fibrosis might be defective in proteinase binding (Shapira *et al.*, 1976, 1977), subsequent studies have not confirmed this. It appears that α_2M prepared from the plasma of these patients is normal in all respects (Parsons and Romeo, 1980; Bridges *et al.*, 1982; Roberts *et al.*, 1982). However, using monoclonal antibodies Eager and Kennett (1984) and Marynen *et al.* (1984a) have been able to demonstrate antigenic differences, which may be related to defective endocytosis of cystic fibrosis α_2M–proteinase complexes.

B. Complement Proteins

While the studies on the interaction between α_2M and cellular receptors almost exclusively have focused on the rapid clearance of α_2M–proteinase complexes, studies on the interaction between the activated complement proteins C3 and C4 and cells have revealed a bewildering diversity and complexity, particularly for C3. For recent comprehensive reviews those of Fearon and Wong (1983) Fearon (1983, 1984), Schreiber (1984), and Arnaout and Colten (1984) should be consulted.

In studies on the phagocytosis of microorganisms Nelson (1953) pointed out that organisms sensitized by reaction with antibodies interacted with erythrocytes

(the immune adherence reaction) in a way that would lead to enhanced phagocytosis by leukocytes. This enhancement was dependent on activation of complement. Subsequent studies with monocytes revealed that the process was dependent on the presence of two types of receptors on the cells, a receptor for the Fc portion of the antibodies and a receptor for the third component of complement (Huber *et al.*, 1968). The efficiency of phagocytosis was dependent on the deposition of C3b by the activating complex of C1, C4b, and C2a on the surface of the particle being ingested (Gigli and Nelson, 1968). The crucial role in opsonization by the covalent binding of activated nascent C3b to surfaces through the thiol-esterified Glx residue has been documented by Schreiber *et al.* (1981) and Hostetter *et al.* (1984), since preformed methylamine-inactivated C3 (in the C3b-like conformation) and otherwise inactivated C3 did not support opsonization. From work of Mantovani *et al.* (1972) and Mantovani (1975) it was indicated that the role of the C3 receptors in phagocytosis was that of attaching the phagocyte to the target cell, while the Fc receptors were mainly responsible for the ingestion of the complexes between C3b and cells. In contrast, Stossel (1973) and Stossel *et al.* (1975) reported that the opsonization could be effected by C3b alone.

Ehlenberger and Nussenzweig (1977) reexamined these phenomena and reported that monocytes in addition to the C3b receptor (CR1) also contained receptors for the C3d fragment (CR2). These receptors were absent on polymorphonuclear leukocytes. A third type of complement receptor (CR3) specific for the C3bi fragment has also been identified (Ross *et al.*, 1973). Cellular receptors for the activation peptides C3a, C4a, and C5a are also known in many cells and play important roles in inflammation, host defense, and immune response (see, e.g., reviews by Weigle *et al.*, 1983; Hugli, 1981, 1984). The different receptors for C3 activation and cleavage products and their cellular distribution are shown in Table IV.

TABLE IV

Tissue Distribution of the Major C3 Receptors[a]

Receptor type	Ligand	Tissue distribution
Anaphylatoxin receptor	C3a, C4a	Neutrophils, basophils, mast cells, eosinophils, macrophages, platelets?
CR1	C3b, C3bi, C3i, C3c, C4b	Erythrocytes, granulocytes, monocytes, macrophages, B lymphocytes, some T lymphocytes, glomerular epithelial cells, mast cells, null cells
CR2	C3bi, C3d	B lymphocytes, monocytes (?), neutrophils
CR3	C3bi	Granulocytes, monocytes, macrophages, null cells

[a]Data taken from Arnaout and Colten (1984) and Schreiber (1984).

The receptor for C3b (CR1) has been characterized by Fearon (1979, 1980) and Dobson *et al.* (1981) and was initially shown to be an integral membrane protein of 205 kDa. However, it is highly polymorphic and allelic species of 160, 190, 220, and 250 kDa have been characterized (Dykman *et al.*, 1983, 1985; Fearon and Wong, 1983; Wong *et al.*, 1983). Recently, a cDNA clone encoding part of CR1 has been characterized (Wong *et al.*, 1985). CR1 will not interact with native C3 (Schreiber *et al.*, 1981; Berger *et al.*, 1981). While C3b generated in the fluid phase binds only weakly to CR1 ($K_{diss} = 5 \times 10^{-7}$ *M*), dimers of C3b are bound with higher affinity ($K_{diss} = 5 \times 10^{-8}$ *M*) (Arnaout *et al.*, 1981) and aggregates of C3b with cells and immune complexes appear to have dissociation constants in the nanomolar range (Fearon and Wong, 1983). In studies on the internalization of soluble C3b-immune complexes by neutrophils and monocytes it was emphasized by Fearon *et al.* (1981) and Abrahamson and Fearon (1982) that the overall process is strongly analogous to that of α_2M–proteinase complexes. Soluble immune complexes containing a few IgG molecules seem to be selectively cleared from the circulation via complement-independent pathways (Mannik *et al.*, 1971; Kijlstra *et al.*, 1981; Leslie, 1985), and clearance via CR1 seems to be important only for certain large immune aggregates (Newman and Johnston, 1979; Horwitz, 1980; Schreiber *et al.*, 1982). In that process a synergistic action with the Fc receptors is seen. Since the erythrocytes bear the bulk of the C3b receptors in the circulation their function in elimination of immune complexes is puzzling. However, Cornacoff *et al.* (1983) have suggested that immune complexes adhering to the erythrocytes are dissociated during passage through the liver and taken over by, e.g., Kupffer cells and other phagocytotic cells for further processing and uptake. Although the details of these reactions are not known, it is possible that a processing similar to the cleavage of fluid-phase C3b to C3bi by factors I plus H, this time with CR1 as a cofactor, takes place (Ross *et al.*, 1982; Pangburn and Müller-Eberhard, 1978; Medicus *et al.*, 1983). This would imply a primary role for CR3 and perhaps CR2 in the uptake of immune complexes. Such a cofactor role of CR1 in the processing of immune complexes has also been suggested in the regulation of complement activation (Fearon, 1979; Iida and Nussenzweig, 1983; Medof *et al.*, 1983; Ross *et al.*, 1982). However, the ability of CR1 to accelerate the decay of the C3 and C5 convertase complexes seems to be inferior to that mediated by the recently discovered decay-accelerating factor (DAF) found on the surface of erythrocytes (Nicholson-Weller *et al.*, 1982; Pangburn *et al.*, 1983a,b; Medof *et al.*, 1985).

CR2 has been isolated from Raji-cell culture medium and characterized by Lambris *et al.* (1981). Although the size of this membrane protein was initially estimated at 72 kDa it now appears that this has been due to proteolysis during preparation, and the size is actually 140 kDa (Bard *et al.*, 1981; Iida *et al.*, 1983; Weiss *et al.*, 1984; Micklem *et al.*, 1984). This protein has recently been identi-

fied as the Epstein–Barr virus receptor of human B lymphocytes (Fingeroth *et al.*, 1984; Frade *et al.*, 1985) and is predominantly expressed on those cells. During activation of C3 a neoantigenic site appears in the C3d portion, which is closely linked to the expression of the CR2 receptor site. Lambris *et al.* (1985) found that some or all of the residues in the sequence -Leu-Tyr-Asn-Val-Glu-Ala- (residues 1204–1209 in pro-C3) constitute the CR2 binding site, with the epitope of the neoantigen being located immediately N-terminal to this sequence. As seen in Fig. 4 this sequence is not particularly well conserved among α_2M, C3, and C4 and, moreover, it is located close to a region where α_2M and C4 have large insertions relative to C3. As further seen from Fig. 5 the receptor recognition site for CR2 is located close to the N-terminal part of a domain that is common to α_2M, C3, and C4, or perhaps in the short stretch connecting this domain with the preceding common domain. Thus, exposure of this recognition site could be effected by a slight rearrangement of these domains after inactivation cleavage. Among other less characterized functions CR2 could be involved in the regulation of lymphocyte activation, such as the proliferative response induced by mitogens or the mixed lymphocyte reaction (Lambris *et al.*, 1982, Meuth *et al.*, 1983).

The existence of CR3 specific for factor I-cleaved C3b (C3bi) was suggested earlier (Ross *et al.*, 1973) and clearly established by Ross and Rabbelino (1979) and Ross and Lambris (1982). This receptor has recently been shown to be identical with a previously known lymphocyte surface antigen MO1 (Beller *et al.*, 1982; Kurzinger *et al.*, 1982; Todd *et al.*, 1982) by Arnaout *et al.* (1983) and Wright *et al.* (1983). CR3 consists of two noncovalently associated chains of 150 (α) and 90 kDa (β). The β subunit is apparently shared with other leukocyte antigens, which are not complement receptors (Kurzinger *et al.*, 1982), indicating that the α subunit contains the C3bi ligand binding site. From studies by Perlman *et al.* (1981) and Schreiber *et al.* (1982) it has been indicated that CR3, present in many different cells, in concert with Fc receptors very efficiently promotes ingestion of C3bi-coated cells and immune complexes, indicating a major role in clearance. This is further strengthened by the identification of individuals having a defective CR3 molecule (Arnaout *et al.*, 1982; Bowen *et al.*, 1982; Fisher *et al.*, 1983; Dana *et al.*, 1984). Although these patients have functional C3b receptors they are subjected to severe, recurrent infections and defective phagocyte functions.

As indicated from *in vitro* studies the covalent deposition of nascent C3 on immune aggregates could proceed with a relatively low efficiency (see Section VIII). If this turns out to be the case *in vivo*, clearance mechanisms for the removal of fluid-phase C3b or its degradation products could exist. These could perhaps involve different receptors in other tissues, and perhaps clearance by hepatocytes might be important in this respect. Recently, Cole *et al.* (1985)

reported that additional C3-binding proteins were present on peripheral leukocytes and various cell lines.

XI. Conclusions and Perspectives

Human α_2M, C3, and C4 are the first well-studied members of a novel class of evolutionarily related large plasma proteins containing internal β-cysteinyl-γ-glutamyl thiol esters. They are sophisticated examples of proteins regulated by specific limited proteolysis and characterized by the ability to undergo a series of specific conformational changes in which biological activities are expressed not only in the final conformational state but also in a short-lived nascent state. In the latter state the activated thiol ester can be considered a "pseudoenzymatic site," turning over once and for all in a process that results in the covalent binding of these proteins to biological targets. While C3 and C4 are proteolytically processed "monomeric" proteins, α_2M is a tetramer containing two dimers, which constitute its functional units. For α_2M the triggering event is the simple encounter with any of a wide spectrum of proteinases, and while the gross conformational change of the dimer per se results in the entrapment of a proteinase by α_2M, the covalent binding potential is presumably important for the fast formation of a tight, irreversible complex at low levels of proteinase. The α_2M–proteinase complex is rapidly cleared from the circulation mainly by the liver as a result of the exposure of previously concealed receptor recognition sites in α_2M.

In contrast, C3 and C4 are activated by complex proteinases of narrow specificity, whose assembly is triggered by immune complexes and certain polysaccharides. The localization of these processes is crucially dependent on the covalent binding of activated C4 and C3 to their targets. In the immune response the complement proteins constitute an important effector system, which principally operates via two routes, one that leads to cell death through the assembly of the terminal complement complex and one that leads to the elimination of the targets. This clearance is mainly effected by C3b and degradation products thereof. Thus, in the encounter of a biological target with α_2M or the complement proteins it effectively becomes destined for rapid clearance from the circulation. In this respect α_2M should perhaps be considered part of the immune system, capable of dealing with potentially injurious proteinases secreted by invading organisms.

It is likely that the plasma of most species will turn out to contain at least two and perhaps three distinct thiol ester-containing proteinase-binding macroglobulins. The macroglobulins evidently form one subgroup and the complement proteins C3 and C4 form another subgroup of the α_2M-related proteins. Howev-

er, within each subgroup a structurally and functionally related protein devoid of thiol esters is known, namely, the proteinase-binding ovostatin from hen egg white and complement protein C5. Although it has not been investigated in detail, activated C5 seems to exist in an analogous nascent state in which it interacts noncovalently with complement protein C6 thereby initiating the assembly of the terminal lytic complex.

The class of α_2M-related proteins can be traced back to the invertebrates and it can be expected that studies of the gene structures of members of this family will contribute significantly to an understanding of the evolutionary history of these proteins.

Although the covalent binding reactions of the nascent state are now known in some detail, the nature of this state and the conformational changes that are characteristic of the class of α_2M-related proteins are not well understood. Likewise, the evidence for the thiol ester structure is only circumstantial. Obviously, besides the use of sophisticated spectroscopic and kinetic procedures, the determination of the three-dimensional structure of any one of these proteins by X-ray crystallography will be of immense value for a detailed understanding of the properties of the whole class of α_2M-related proteins.

While the role of the complement proteins C3, C4, and C5 in the function of the complement system is fairly clear-cut, the role of the macroglobulins in controlling proteolytic activity in plasma is presently unclear, in particular since plasma contains a diversity of specific, fast-acting proteinase inhibitors directed against different groups of proteinases. Initial studies on PZP indicate that individual macroglobulins may have different, yet probably overlapping sets of proteinases as targets. In man and rats one of the macroglobulins is an acute-phase reactant, being greatly elevated not only in inflammation but also in pregnancy, suggesting that this may be the case generally (see Chapter 5). This indicates a role of the macroglobulins in the control of yet unidentified proteinases appearing in plasma, presumably as a result of an increased cellular turnover, common to the inflammatory and pregnant states. Hypothetically the macroglobulins could control elements in the immune response, which depend on proteinases.

Last, the elucidation of the structures of the receptors found in different cells and tissues, which interact with the activated proteins would be expected to lead to an increased understanding of the pathways of clearance and the potential for interaction with other cellular systems.

Acknowledgments

Drs. Brian F. Tack and Salvatore V. Pizzo are thanked for making available unpublished results and for stimulating discussions. Drs. Karl Lonberg-Holm and Rusty Kutny are thanked for un-

published data on the rat macroglobulins. Birthe Hother Vesterager and Aase Sørensen provided expert secretarial assistance in the preparation of this manuscript. This work has been supported by The Harboe Foundation, The Novo Foundation, and The Danish Cancer Society.

References

Abe, S., and Nagai, Y. (1973). *J. Biochem. (Tokyo)* **71,** 897–900.

Abrahamson, D. R., and Fearon, D. T. (1982). *Lab. Invest.* **48,** 162–168.

Ades, E. W., Hinson, A., Chapuis-Cellier, C., and Arnaud, P. (1982). *Scand. J. Immunol.* **15,** 109–113.

Adham, N. F., Wilding, P., Mehl, J., and Haverback, B. J. (1968). *J. Lab. Clin. Med.* **71,** 271–278.

Adham, N. F., Chakmakjian, Z. H., Mehl, J. W., and Bethune, J. E. (1969). *Arch. Biochem. Biophys.* **132,** 175–183.

Adman, E., Watenpaugh, K., and Jensen, L. (1975). *Proc. Natl. Acad. Sci. U.S.A.* **72,** 4854–4858.

Alper, C. A., and Balavitch, D. (1976). *Science 191,* 1275–1276.

Alper, C. A., and Propp, R. P. (1968). *J. Clin. Invest.* **47,** 2181–2191.

Alper, C. A., and Rosen, F. S. (1984). *Springer Semin. Immunopathol.* **7,** 251–261.

Alper, C. A., Johnson, A. M., Birtch, A. G., and Moore, F. D. (1969). *Science 163,* 286–288.

Alper, C. A., Colten, H. R., Rosen, F. S., Robson, A. R., Macnab, G. M., and Gear, J. S. S. (1972). *Lancet* **2,** 1179–1181.

Alper, C. A., Colten, H. R., Gear, J. J. S., Robson, H. R., and Rosen, F. S. (1976). *J. Clin. Invest.* **57,** 522–531.

Anderson, J. M., Stimson, W. H., Gettinby, G., Jhunjkunwala, S. K., and Burt, R. W. (1979). *Eur. J. Cancer* **15,** 709–714.

Anderson, R. G. W., Brown, M. S., and Goldstein, J. L. (1977). *Cell* **10,** 351–364.

Andus, T., Gross, V., Tran-Thi, T.-A., and Heinrich, P. C. (1983a). *FEBS Lett.* **151,** 10–14.

Andus, T., Gross, V., Tran-Thi, T.-A., Schreiber, G., Nagashima, M., and Heinrich, P. C. (1983b). *Eur. J. Biochem.* **133,** 561–571.

Arnaout, M. A., and Colten, H. R. (1984). *Mol. Immunol.* **21,** 1191–1199.

Arnaout, M. A., Melamed, J., Tack, B. F., and Colten, H. R. (1981). *J. Immunol.* **127,** 1348–1354.

Arnaout, M. A., Pitt, J., Cohen, H. J., Melamed, J., Rosen, F. S., and Colten, H. R. (1982). *N. Engl. J. Med.* **306,** 694–699.

Arnaout, M. A., Todd, R. F., III, Dana, N., Melamed, J., Schlossman, S. F., and Colten, H. R. (1983). *J. Clin. Invest.* **72,** 171–179.

Auger, G., Blanot, D., Van Heijenoort, J., Nadal, C., and Gournay, M.-F. (1983). *Eur. J. Biochem.* **133,** 363–369.

Awdeh, Z. L., and Alper, C. A. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 3576–3580.

Awdeh, Z. L., Raum, D., and Alper, C. A. (1979). *Nature (London)* **282,** 205–207.

Awdeh, Z. L., Ochs, H. D., and Alper, C. A. (1981). *J. Clin. Invest.* **67,** 260–263.

Baba, A., Fujita, T., and Tamura, N. (1984). *J. Exp. Med.* **160,** 411–419.

Balldin, G., and Ohlsson, K. (1979). *Surgery (St. Louis)* **85,** 451–456.

Ballow, M., Shira, J. E., Harden, L., Yang, S. Y., and Day, N. K. (1975). *J. Clin. Invest.* **56,** 703–710.

Bard, M., Charrioult, C., and Frade, R. (1981). *FEBS Lett.* **136,** 111–114.

Barker, W. C., and Putnam, F. W. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 4, pp. 361–399. Academic Press, Orlando, Florida.

Barrett, A. J., and Salvesen, G. S. (1979). *In* "The Physiological Inhibitors of Coagulation and Fibrinolysis" (D. Collen, B. Wiman, and M. Verstraete, eds.), pp. 247–254. Elsevier/North-Holland Biomedical Press, Amsterdam.

Barrett, A. J., and Starkey, P. S. (1973). *Biochem. J.* **133,** 709–724.

Barrett, A. J., Starkey, P. M., and Munn, E. A. (1974). *Bayer-Symp.* **5,** 72–77.

Barrett, A. J. (1982). *In* "Methods in Enzymology" (L. Lorand, ed.), Vol. 80, pp. 737–754. Academic Press, New York.

Barrett, A. J., Brown, M. A., and Sayers, C. A. (1979). *Biochem. J.* **181,** 401–418.

Battersby, A. R., and Reynolds, J. J. (1961). *J. Chem. Soc.* pp. 524–531.

Battersby, A. R., and Robinson, J. C. (1955). *J. Chem. Soc.*, pp. 259–267.

Bauer, H. W., Deutschmann, K. E. M., Peter, H. H., and Bohn, H. (1979). *Eur. J. Cancer* **15,** 123–126.

Bauer, J., Birmelin, M., Northoff, G.-H., Northemann, W., Tran-Thi, T.-A., Verberg, H., Decker, K., and Heinrich, P. C. (1984). *FEBS Lett.* **177,** 89–94.

Bauer, J., Weber, W., Tran-Thi, T.-A., Northoff, G.-H., Decker, K., Gerok, W., and Heinrich, P. C. (1985). *FEBS Lett.* **190,** 271–274.

Baumann, H., Jahreis, G. P., Sander, D. N., and Koj, A. (1984). *J. Biol. Chem.* **259,** 7331–7342.

Baumstark, J. S. (1973). *Biochim. Biophys. Acta* **309,** 181–195.

Beaton, G. H., Selby, A. E., Veen, M. J., and Wright, A. M. (1961). *J. Biol. Chem.* **236,** 2005–2008.

Becker, C. G., and Harpel, P. C. (1976). *J. Exp. Med.* **144,** 1–9.

Beckman, L., von Schoultz, B., and Stigbrand, T. (1971). *Acta Obstet. Gynecol. Scand.* **50,** 369–371.

Beckman, L., von Schoultz, B., and Stigbrand, T. (1973a). *Acta Obstet. Gynecol. Scand.* **52,** 157–160.

Beckman, L., von Schoultz, B., and Stigbrand, T. (1973b). *Urol. Res.* **1,** 67–69.

Beckman, G., Beckman, L., Magnusson, S. S., and von Schoultz, B. (1974). *Acta Obstet. Gynecol. Scand.* **53,** 177–180.

Beller, D. I., Springer, T. A., and Schreiber, R. D. (1982). *J. Exp. Med.* **156,** 1000–1009.

Belt, K. T., Carroll, M. C., and Porter, R. R. (1984). *Cell (Cambridge, Mass.)* **36,** 907–914.

Belt, K. T., Yu, C. T., Carroll, M. C., and Porter, R. R. (1985). *Immunogenetics* **21,** 173–180.

Benjamin, D. C., and Weimer, H. E. (1966). *Nature (London)* **209,** 1032–1033.

Berg, K., and Bearn, A. G. (1966). *J. Exp. Med.* **123,** 379–397.

Berger, M., Geither, T. A., Hammer, C. A., and Frank, M. M. (1981). *J. Immunol.* **127,** 1329–1334.

Bergquist, D., and Nilsson, I. M. (1979). *Scand. J. Haematol.* **23,** 433–436.

Bergsma, J., Boelen, M. K., Duursma, A. M., Schutter, W. G., Bouma, J. H. W., and Gruber, M. (1985). *Biochem. J.* **226,** 75–84.

Berne, B. H. (1973). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **32,** 677.

Berne, B. H. (1976). *Protides Biol. Fluids* **24,** 165–179.

Berne, B. H., Dray, S., and Knight, K. L. (1973). *Biochem. Genet.* **8,** 531–535.

Besterman, J. M., and Low, R. B. (1983). *Biochem. J.* **210,** 1–13.

Bhakdi, S., and Tranum-Jensen, J. (1984). *Philos. Trans. R. Soc. London, Ser. B* **306,** 311–324.

Bieth, J., and Klumpp, T. (1976). *Biochim. Biophys. Acta* **439,** 363–367.

Bieth, J. G. (1980). *Bull. Eur. Physiopathol. Respir.* **16,** 183–195.

Bieth, J. G., and Meyer, J.-F. (1984). *J. Biol. Chem.* **259,** 8904–8906.

Bieth, J. G., Tourbez-Perrin, M., and Pochon, F. (1981). *J. Biol. Chem.* **256,** 7954–7957.

Bischof, P. (1979). *Arch. Gynaecol.* **227,** 315–326.

Bizik, J., Vaheri, A., Saksela, O., Kalkkinen, N., Meri, S., and Grotova, M. (1986). *Int. J. Cancer* **37,** 81–88.

Björk, I. (1985). *Biochem. J.* **231,** 451–457.
Björk, I., and Fish, W. W. (1982). *Biochem. J.* **207,** 347–356.
Björk, I., Larsson, L.-J., Lindblom, T., and Raub, E. (1984). *Biochem. J.* **217,** 303–308.
Björk, I., Lindblom, T., and Lindahl, P. (1985). *Biochemistry* **24,** 2653–2660.
Björksten, B., Söderström, T., Damber, M.-G., von Schoultz, B., and Stigbrand, T. (1978). *Scand. J. Immunol.* **8,** 257–262.
Bloth, B., Chesebro, B., and Svehag, S.-E. (1968). *J. Exp. Med.* **127,** 749–756.
Boffa, G. A., Lambin, P., Rius, E., and Nadal, C. (1985). *C.R. Seances Acad. Sci., Ser. 3* **300,** 31–36.
Bohn, H. (1971). *Arch. Gynaekol.* **210,** 440–457.
Bohn, H. (1972). *Arch. Gynaekol.* **213,** 54–72.
Bohn, H. (1974). *Arch. Gynaekol.* **217,** 219–231.
Bohn, H., and Winckler, W. (1976). *Blut* **33,** 377–388.
Bokisch, V. A., Dierich, M. P., and Müller-Eberhard, H. J. (1975). *Proc. Natl. Acad. Sci. U.S.A.* **72,** 1989–1993.
Bolotin, C., Morris, S. C., Tack, B., and Prahl, J. (1977). *Biochemistry* **16,** 2008–2015.
Borth, W. (1984). *Collagen Res.* **4,** 83–95.
Bowen, T. J., Ochs, H. D., Altman, L. C., Price, T. H., Van Epps, D. E., Brautigen, D. L., Rosin, R. E., Perkins, W. D., Babior, B. M., Klebanoff, S. J., and Wedgwood, R. J. (1982). *J. Pediatr.* **101,** 932–940.
Brade, V., Hall, R. E., and Colten, H. R. (1977). *J. Exp. Med.* **146,** 759–765.
Branegård, B., Österberg, R., and Sjöberg, B. (1980). *Int. J. Biol. Macromol.* **2,** 321–323.
Branegård, B., Österberg, R., and Sjöberg, B. (1982). *Eur. J. Biochem.* **122,** 663–666.
Bridges, M. A., Applegarth, D. A., Johannson, J., Wong, L. T. K., and Davidson, G. F. (1982). *Clin. Chim. Acta* **118,** 33–43.
Brown, M. S., and Goldstein, J. L. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 3330–3337.
Bruice, T. C., and Benkovic, S. (1966). "Biorganic Mechanisms." Benjamin, New York.
Budzko, D. B., Bokisch, V. A., and Müller-Eberhard, H.-J. (1971). *Biochemistry* **10,** 1166–1172.
Bundshuh, G., Stober, D., Bayer, H., and Böhmish, S. (1975). *Zentralbl. Gynaekol.* **97,** 49–53.
Bütler, R., and Brunner, E. (1967). *Haematologia* **4,** 375–385.
Campbell, R. D., Dodds, A. W., and Porter, R. R. (1980). *Biochem. J.* **189,** 67–80.
Campbell, R. D., Gagnon, J., and Porter, R. R. (1981). *Biochem. J.* **199,** 359–370.
Capel, P. J. A., Groeneboer, O., Grosveld, G., and Pondman, K. W. (1978). *J. Immunol.* **121,** 2566–2572.
Caporale, L. H., Tippett, P. S., Erickson, B. W., and Hugli, T. E. (1980). *J. Biol. Chem.* **255,** 10758–10763.
Carlsson, J., Herrmann, B. F., Höfling, J. F., and Sundquist, G. K. (1984). *Infect. Immunol.* **43,** 644–648.
Carlsson, L., Folkersen, J., and Stigbrand, T. (1985). *Mol. Immunol.* **22,** 1073–1080.
Carroll, M. C., and Capra, J. D. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 4641–4645.
Carroll, M. C., and Porter, R. R. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 264–267.
Carroll, M. C., Campbell, R. D., Bentley, D. R., and Porter, R. R. (1984). *Nature (London)* **307,** 237–241.
Carter, P. E., Dunbar, B., and Fothergill, J. E. (1984). *Philos. Trans. R. Soc. London, Ser. B* **306,** 293–299.
Cassiman, J.-J., Van Leuven, F., Van Der Schueren, B., and Van Den Berghe, H. (1980). *Cell Tissue Res.* **213,** 301–310.
Chakravarti, D. N., Campbell, R. D., and Gagnon, J. (1983). *FEBS Lett.* **154,** 387–390.
Chan, A. C., and Atkinson, J. P. (1983). *J. Clin. Invest.* **72,** 1639–1649.
Chan, A. C., and Atkinson, J. P. (1984). *J. Immunol.* **132,** 1967–1971.

Chan, A. C., and Atkinson, J. P. (1985). *J. Immunol.* **134,** 1790–1798.
Chan, A. C., Mitchell, K. R., Munns, T. W., Karp, D. R., and Atkinson, J. P. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 268–272.
Chandra, R. K. (1972). *J. Reprod. Fertil.* **28,** 463–464.
Chandra, R. K., Malkani, P. K., and Bhasin, K. (1973). *Obstet. Gynecol.* **41,** 56–59.
Chang, J.-L., Ganea, D., Dray, S., and Teodorescu, M. (1981). *Immunology* **44,** 745–754.
Chang, J.-L., Ganea, D., Dray, S., and Teodorescu, M. (1983). *J. Immunol.* **130,** 267–273.
Chemnitz, J., Hau, J., Svendsen, P., Folkersen, J., Westergaard, J. G., and Christensen, B. C. (1982). *Bibl. Anat.* **22,** 87–92.
Chlebowski, J. F., and Williams, K. (1983). *Biochem. J.* **209,** 725–730.
Chlebowski, J. F., and Williams, K. (1985). *Biochem. J.* **229,** 227–232.
Christensen, U., and Sottrup-Jensen, L. (1983). *Biochim. Biophys. Acta* **747,** 263–275.
Christensen, U., and Sottrup-Jensen, L. (1984). *Biochemistry* **23,** 6619–6626.
Churchill, W. H., Jr., Weintraub, R. M., Borsos, T., and Rapp, H. J. (1967). *J. Exp. Med.* **125,** 657–672.
Cinader, B., Dubishi, S., and Wardlaw, C. (1964). *J. Exp. Med.* **120,** 897–903.
Cole, F. S., Matthews, W. Y., Marino, J. T., Gash, D. Y., and Colten, H. R. (1980). *J. Immunol.* **125,** 1120–1124.
Cole, J. L., Housley, G. A., Dykman, T. R., MacDermott, R. P., and Atkinson, J. P. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 859–863.
Colomb, M. G., Arlaud, G. R., and Villiers, C. L. (1984). *Philos. Trans. R. Soc. London, Ser. B* **306,** 283–292.
Colten, H. R. (1982). *Mol. Immunol.* **19,** 1279–1285.
Cooper, D. W. (1963). *Nature (London)* **200,** 892.
Cooper, N. R., and Müller-Eberhard, H. J. (1970). *J. Exp. Med.* **132,** 775–793.
Cordier, G., and Revillard, J. P. (1980). *Experientia* **36,** 603–605.
Cornacoff, J. B., Hebert, L. A., Smead, W. L., and Vanaman, M. E. (1983). *J. Clin. Invest.* **71,** 236–247.
Cummings, H. S., and Castellino, F. J. (1984). *Biochemistry* **23,** 105–111.
Cummings, H. S., Pizzo, S. V., Strickland, D. K., and Castellino, F. J. (1984). *Biophys. J.* **45,** 721–724.
Dahlbäck, B., Smith, C. A., and Müller-Eberhard, H. J. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 3461–3465.
Dalmasso, A. P., and Müller-Eberhard, H. J. (1964). *Proc. Soc. Exp. Biol. Med.* **117,** 643–650.
Dalmasso, A. P., and Müller-Eberhard, H. J. (1966). *J. Immunol.* **97,** 680–685.
Damber, M. G., von Schoultz, B., Stigbrand, T., and Carlsson, K. (1976). *Acta Obstet. Gynecol. Scand.* **55,** 467–468.
Damber, M. G., von Schoultz, B., Stigbrand, T., Carlström, K., and Furuhjelm, M. (1977). *Acta Obstet. Gynecol. Scand.* **56,** 95–99.
Damber, M. G., von Schoultz, B., Solheim, F., Stigbrand, T., and Carlström, K. (1978). *Obstet. Gynecol.* **51,** 677–681.
Dana, N., Todd, R. F., Pitt, R. F., Springer, T. A., and Arnaout, M. A. (1984). *J. Clin. Invest.* **73,** 153–159.
Dangott, L. J., and Cunningham, L. W. (1982). *Biochem. Biophys. Res. Commun.* **107,** 1243–1251.
Dangott, L. J., Puett, D., and Cunningham, L. W. (1983). *Biochemistry* **22,** 3647–3653.
Davidsen, O., Christensen, E. I., and Gliemann, J. (1985). *Biochim. Biophys. Acta* **846,** 85–92.
Davies, P. J. A., Davies, D. R., Levitzki, A., Maxfield, F. R., Milhaud, P., Willingham, M. C., and Pastan, I. H. (1980). *Nature (London)* **283,** 162–167.
Davies, S. G., and Sim, R. B. (1981). *Biosci. Rep.* **1,** 461–468.
Davis, A. E. (1981). *FEBS Lett.* **134,** 147–150.

Davis, A. E., and Harrison, R. A. (1982). *Biochemistry* **21,** 5745–5749.
Davis, A. E., Harrison, R. A., and Lachman, P. J. (1984). *J. Immunol.* **132,** 1960–1962.
DeAlvarez, R. R., and Afonso, J. S. (1967). *Pa. Med.* **70,** 43–44.
Debanne, M. T., Regoeczi, E., and Dolovich, J. (1973). *Br. J. Exp. Pathol.* **54,** 571–581.
Debanne, M. T., Bell, R., and Dolovich, J. (1975). *Biochim. Biophys. Acta* **411,** 295–304.
DeBanne, M. T., Bell, R., and Dolovich, J. (1976). *Biochim. Biophys. Acta* **428,** 466–475.
De Bruijn, M. H. L., and Fey, G. H. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 708–712.
Dickinson, A. M., Shenton, B. K., Alomran, A. H., Donelly, P. K., and Proctor, S. J. (1985). *Clin. Immunol. Immunopathol.* **36,** 259–265.
Dickson, R. B., Nicolas, J.-C., Willingham, M. C., and Pastan, I. (1981a). *Exp. Cell Res.* **132,** 488–493.
Dickson, R. B., Willingham, M. C., and Pastan, I. (1981b). *J. Biol. Chem.* **256,** 3454–3459.
DiScipio, R. G. (1981a). *Biochem. J.* **199,** 485–496.
DiScipio, R. G. (1981b). *Biochem. J.* **199,** 497–504.
DiScipio, R. G., Smith, C. A., Müller-Eberhard, H. J., and Hugli, T. E. (1983). *J. Biol. Chem.* **258,** 10629–10636.
Djakiew, D., Byers, S. W., and Dym, M. (1984). *Biol. Reprod.* **31,** 1073–1085.
Djakiew, D., Byers, S. W., Lewis, D. M., and Dym, M. (1985). *J. Androl.* **6,** 190–196.
Dobson, N. J., Lambris, J. D., and Ross, G. D. (1981). *J. Immunol.* **126,** 693–698.
Dodds, A. W., Law, S. K., and Porter, R. R. (1985). *EMBO J.* **4,** 2239–2244.
Domdey, H., Wiebauer, K., Kazmaier, M., Müller, V., Odink, K., and Fey, G. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 7619–7623.
Donald, J. A., and Ball, S. P. (1984). *Ann. Hum. Genet.* **48,** 269–273.
Donovan, J. W., Mapes, C. J., Davis, J. G., and Hamburg, R. D. (1969). *Biochemistry* **8,** 4190–4199.
Doolittle, R. F. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 4, pp. 317–360. Academic Press, New York.
Dott, C. S., Howard, A., and Ansell, B. M. (1985). *Clin. Chim. Acta* **146,** 157–166.
Downing, M. R., Bloom, J. W., and Mann, K. G. (1978). *Biochemistry* **17,** 2649–2653.
Dubin, A., Potempa, J., and Silbering, J. (1984). *Biochem. Int.* **8,** 589–596.
Dunn, J. T., and Spiro, R. G. (1967a). *J. Biol. Chem.* **242,** 5519–5555.
Dunn, J. T., and Spiro, R. G. (1967b). *J. Biol. Chem.* **242,** 5556–5563.
Dykman, T. R., Cole, J. L., Iida, K., and Atkinson, J. P. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 1698–1702.
Dykman, T. R., Hatch, J. A., Aqua, M. S., and Atkinson, J. P. (1985). *J. Immunol.* **134,** 1787–1789.
Eager, K. B., and Kennett, R. H. (1984). *Pediatr. Res.* **18,** 999–1004.
Eccleston, E. D., and Howard, J. B. (1985). *J. Biol. Chem.* **259,** 10169–10176.
Eggertsen, G., Lind, P., and Sjöquist, J. (1981). *Mol. Immunol.* **18,** 125–133.
Ehlenberger, A. G., and Nuzzenzweig, V. (1977). *J. Exp. Med.* **145,** 357–371.
Einstein, L. P., Hansen, P. J., Ballow, M., Davis, A. E., Davis, J. S., Alper, C. A., Rosen, F. S., and Colten, H. R. (1977). *J. Clin. Invest.* **60,** 963–969.
Ellis, V., Scully, M., MacGregor, I., and Kakkar, V. (1982). *Biochim. Biophys. Acta* **701,** 24–31.
Ellman, L., Green, I., and Frank, M. M. (1970). *Science* **170,** 74–75.
Esnard, F., and Gauthier, F. (1980). *Biochim. Biophys. Acta* **614,** 553–563.
Esnard, F., Gauthier, F., and Maurizot, J. C. (1981). *Biochimie* **63,** 767–774.
Esnard, F., Gutman, N., El Moujahed, A., and Gauthier, F. (1985). *FEBS Lett.* **182,** 125–129.
Fearon, D. T. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 5867–5871.
Fearon, D. T. (1980). *J. Exp. Med.* **152,** 20–30.
Fearon, D. T. (1983). *Springer Semin. Immunopathol.* **6,** 159–172.

Fearon, D. T. (1984). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **43,** 2553–2557.
Fearon, D. T., and Wong, W. W. (1983). *Annu. Rev. Immunol.* **1,** 243–271.
Fearon, D. T., Kanecko, I., and Tompson, G. G. (1981). *J. Exp. Med.* **153,** 1615–1628.
Feinman, R. D., Yuan, A. I., Windwer, S. R., and Wang, D. (1985). *Biochem. J.* **231,** 417–423.
Feldman, S. R., and Pizzo, S. V. (1984a). *Arch. Biochem. Biophys.* **235,** 267–275.
Feldman, S. R., and Pizzo, S. V. (1984b). *Biochem. Biophys. Res. Commun.* **123,** 771–777.
Feldman, S. R., and Pizzo, S. V. (1985). *Biochemistry* **24,** 2569–2575.
Feldman, S. R., and Pizzo, S. V. (1986). *Biochemistry* **25,** 721–727.
Feldman, S. R., Ney, K. A., Gonias, S. L., and Pizzo, S. V. (1983). *Biochem. Biophys. Res. Commun.* **114,** 757–762.
Feldman, S. R., Gonias, S. L., Ney, K. A., Pratt, C. W., and Pizzo, S. V. (1984). *J. Biol. Chem.* **259,** 4458–4462.
Feldman, S. R., Rosenberg, M. A., Ney, K. A., Michalopoulos, G., and Pizzo, S. V. (1985a). *Biochem. Biophys. Res. Commun.* **128,** 795–802.
Feldman, S. R., Gonias, S. L., and Pizzo, S. V. (1985b). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 5700–5704.
Fernandez, H. N., and Hugli, T. E. (1977). *J. Biol. Chem.* **252,** 1826–1828.
Fernandez, H. N., and Hugli, T. E. (1978). *J. Biol. Chem.* **253,** 6955–6964.
Ferreira, A., Takahashi, M., and Nussenzweig, V. (1977). *J. Exp. Med.* **146,** 1001–1018.
Ferreira, A., Nussenzweig, V., and Gigli, I. (1978). *J. Exp. Med.* **148,** 1186–1197.
Fey, G., Odink, K., and Chapuis, R. M. (1980). *Eur. J. Immunol.* **10,** 75–82.
Fingeroth, J. D., Weis, J. J., Tedder, T. F., Strominger, J. L., Biro, P. A., and Fearon, D. T. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 4510–4514.
Fisher, A., Dechamps-Latscka, B., Gerota, I., Scheinmetzler, C., Virelizier, J. L., Trung, P. H., Lisowska-Grospierre, B., Perez, N., Durandy, A., and Griscelly, C. (1983). *Lancet* **2,** 473–475.
Folk, J. E. (1983). *Adv. Enzymol.* **54,** 1–56.
Folk, J. E., and Chung, S. I. (1973). *Adv. Enzymol.* **38,** 109–191.
Folk, J. E., and Finlayson, J. S. (1977). *Adv. Protein Chem.* **31,** 1–133.
Folkersen, J., Teisner, B., Ahrons, S., and Svehag, S.-E. (1978). *J. Immunol. Methods* **23,** 117–125.
Folkersen, J., Grudzinskas, J. G., Hindersson, P., Teisner, B., and Westergaard, J. G. (1981a). *Am. J. Obstet. Gynecol.* **139,** 910–914.
Folkersen, J., Teisner, B., Grunnet, N., Grudzinskas, J. G., Westergaard, J. G., and Hindersson, P. (1981b). *Clin. Chim. Acta* **110,** 139–145.
Fontaine, M., and Sim, R. B. (1984). *Biochim. Biophys. Acta* **789,** 119–127.
Forrester, J. V., Wilkinson, P. C., and Lackie, J. M. (1983). *Immunology* **50,** 251–259.
Frade, R., Barel, M., Ehlin-Henrikson, B., and Klein, G. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 1490–1493.
Frenoy, J. P., Bourrilon, R., Lippoldt, R., and Edelhoch, H. (1977). *J. Biol. Chem.* **252,** 1129–1133.
Frey, J., and Afting, E.-G. (1983). *Biochem. J.* **214,** 629–631.
Fuchs, H. E., and Pizzo, S. V. (1983). *J. Clin. Invest.* **72,** 2041–2049.
Fuchs, H. E., Shifman, M. A., and Pizzo, S. V. (1982). *Biochim. Biophys. Acta* **716,** 151–157.
Fuchs, H. E., Trapp, H. G., Griffith, M. J., Roberts, H. R., and Pizzo, S. V. (1984). *J. Clin. Invest.* **73,** 1696–1703.
Fujita, T., Gigli, I., and Nussenzweig, V. (1978). *J. Exp. Med.* **148,** 1044–1051.
Gadd, K. J., and Reid, K. B. M. (1981). *Biochem. J.* **195,** 471–480.
Gagnon, J. (1984). *Philos. Trans. R. Soc. London, B Ser.* **306,** 301–309.
Gallango, M. L., and Castillo, O. (1974). *J. Immunogenet.* **1,** 147–151.

Gallango, M. L., and Castillo, O. (1975). *Humangenetik* **26,** 71–74.
Ganea, D., Teodorescu, A., Dray, S., and Teodorescu, M. (1982). *Immunology* **45,** 227–237.
Ganrot, K. (1973a). *Biochim. Biophys. Acta* **295,** 245–251.
Ganrot, K. (1973b). *Biochim. Biophys. Acta* **322,** 62–67.
Ganrot, P. O. (1967). *Acta Chem. Scand.* **21,** 602–608.
Ganrot, P. O., and Bjerre, B. (1967). *Acta Obstet. Gynecol. Scand.* **46,** 126–137.
Ganrot, P. O., and Laurell, C. G. (1966). *Clin. Chim. Acta* **14,** 137–138.
Gaspar, A., Skosey, J. L., Sequeira, W., and Teodorescu, M. (1984). *Clin. Chem. (Winston-Salem, N.C.)* **30,** 1517–1522.
Gauthier, F., and Mouray, H. (1975a). *Int. J. Biochem.* **6,** 95–98.
Gauthier, F., and Mouray, H. (1975b). *Protides Biol. Fluids* **23,** 139–143.
Gauthier, F., and Mouray, H. (1976). *Biochem. J.* **159,** 661–665.
Gauthier, F., and Ohlsson, K. (1978). *Hoppe-Seyler's Z. Physiol. Chem.* **359,** 987–992.
Gauthier, F., Lebreton de Vonne, T., and Mouray, H. (1974). *C. R. Hebd. Seances Acad. Sci.* **168,** 437–439
Geokas, M. C., Brodrick, J. W., Johnson, J. H., and Largman, C. (1977). *J. Biol. Chem.* **252,** 61–67.
Gerard, C., and Hugli, T. E. (1980). *J. Biol. Chem.* **255,** 4710–4715.
Gerard, C., Showell, H. J., Hoeprich, P. D., Hugli, T. E., and Stimler, N. P. (1985). *J. Biol. Chem.* **260,** 2613–2616.
Gigli, I., and Austen, K. F. (1971). *Annu. Rev. Microbiol.* **25,** 309–332.
Gigli, I., and Nelson, R. A. Jr. (1968). *Exp. Cell Res.* **51,** 45–67.
Gigli, I., von Zabern, I., and Porter, R. R. (1977). *Biochem. J.* **165,** 439–446.
Gitlin, J. D., Rosen, F. S., and Lachmann, P. J. (1975). *J. Exp. Med.* **141,** 1221–1226.
Gliemann, J., Larsen, T. R., and Sottrup-Jensen, L. (1983). *Biochim. Biophys. Acta* **756,** 230–237.
Gliemann, J., Davidsen, O., Sottrup-Jensen, L., and Sonne, O. (1985). *FEBS Lett.* **188,** 352–356.
Goers, J. W. F., and Porter, R. R. (1978). *Biochem. J.* **175,** 675–684.
Goldberger, G., and Colten, H. R. (1980). *Nature (London)* **286,** 514–516.
Goldberger, G., Abraham, G. N., Williams, J., and Colten, H. R. (1980). *J. Biol. Chem.* **255,** 7071–7074.
Goldberger, G., Thomas, M. L., Tack, B. F., Williams, J., Colten, H. R., and Abraham, G. N. (1981). *J. Biol. Chem.* **256,** 12617–12619.
Goldstein, J. L., Anderson, R. G. W., and Brown, M. S. (1979). *Nature (London)* **279,** 679–685.
Gonias, S. L., and Pizzo, S. V. (1981a). *Biochim. Biophys. Acta* **678,** 268–274.
Gonias, S. L., and Pizzo, S. V. (1981b). *J. Biol. Chem.* **256,** 12478–12484.
Gonias, S. L., and Pizzo, S. V. (1983a). *Biochemistry* **22,** 536–546.
Gonias, S. L., and Pizzo, S. V. (1983b). *Biochemistry* **22,** 4933–4940.
Gonias, S. L., and Pizzo, S. V. (1983c). *J. Biol. Chem.* **258,** 14682–14685.
Gonias, S. L., Einarson, M., and Pizzo, S. V. (1982a). *J. Clin. Invest.* **70,** 412–423.
Gonias, S. L., Reynolds, J. A., and Pizzo, S. V. (1982b). *Biochim. Biophys. Acta* **705,** 306–314.
Gonias, S. L., Balber, A. E., Hubbard, W. J., and Pizzo, S. L. (1983). *Biochem. J.* **209,** 99–105.
Gordon, A. H. (1976). *Biochem. J.* **159,** 643–650.
Gordon, J., Whitehead, H. R., and Wormall, A. (1926). *Biochem. J.* **20,** 1028–1035.
Gorski, J. P., and Howard, J. B. (1980). *J. Biol. Chem.* **255,** 10025–10028.
Gorski, J. P., Hugli, T. E., and Müller-Eberhard, H. J. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 5299–5302.
Gorski, J. P., Silversmith, R., Fiebeger, S., and Moilanen, T. (1982). *J. Biol. Chem.* **257,** 10948–10954.
Gounaris, A. D., Brown, M. A., and Barrett, A. J. (1984). *Biochem. J.* **221,** 445–452.
Granelli-Piperno, A., and Reich, E. (1978). *J. Exp. Med.* **148,** 223–234.

Greene, N. D., Damain, R. T., and Hubbard, W. J. (1971). *Biochim. Biophys. Acta* **236,** 659–663.
Gropp, C., Lehman, F. G., Bauer, H. W., and Havemann, K. (1977). *Oncology* **34,** 267–272.
Gross, V., Andus, T., Tran-Thi, T.-A., Bauer, J., Decker, K., and Heinrich, P. C. (1984). *Exp. Cell Res.* **151,** 46–54.
Guillouzo, A., Delers, F., Clement, B., Bernard, N., and Engler, R. (1984). *Biochem. Biophys. Res. Commun.* **120,** 311–317.
Hall, P. K., Nelles, L. P., Travis, J., and Roberts, R. C. (1981). *Biochem. Biophys. Res. Commun.* **100,** 8–16.
Hall, R. E., and Colten, H. R. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 1707–1710.
Hammer, C. H., Wirtz, G. H., Reuter, L., Gresham, H. D., and Tack, B. F. (1981). *J. Biol. Chem.* **256,** 3995–4006.
Hanover, J. A., Rudick, J. E., Willingham, M. C., and Pastan, I. (1983a). *Arch. Biochem. Biophys.* **227,** 570–579.
Hanover, J. A., Cheng, S., Willingham, M. C., and Pastan, I. H. (1983b). *J. Biol. Chem.* **258,** 370–377.
Hansen, H. T., and Schreffler, D. C. (1976). *J. Immunol.* **177,** 1507–1513.
Harpel, P. C. (1973). *J. Exp. Med.* **138,** 508–521.
Harpel, P. C. (1977). *J. Exp. Med.* **146,** 1033–1040.
Harpel, P. C. (1981). *J. Clin. Invest.* **68,** 46–55.
Harpel, P. C., and Brower, M. S. (1983). *Ann. N. Y. Acad. Sci.* **421,** 1–9.
Harpel, P. C., and Hayes, M. B. (1979). *In* "The Physiological Inhibitors of Blood Coagulation and Fibrinolysis" (D. Collen, B. Wiman, and M. Verstraete, eds.), pp. 273–280. Elsevier/North-Holland Biomedical Press, Amsterdam.
Harpel, P. C., and Mosesson, M. W. (1973). *J. Clin. Invest.* **52,** 2175–2184.
Harpel, P. C., and Rosenberg, R. D. (1976). *Prog. Hemostasis Thromb.* **3,** 145–189.
Harpel, P. C., Hayes, M. B., and Hugli, T. E. (1979). *J. Biol. Chem.* **254,** 8869–8678.
Harris, J. R., Brown, K. D., and Aiton, J. F. (1976). *Biochim. Biophys. Acta* **427,** 727–747.
Harrison, R. A., and Lachmann, P. J. (1980a). *Mol. Immunol.* **17,** 9–20.
Harrison, R. A., and Lachmann, P. J. (1980b). *Mol. Immunol.* **17,** 219–228.
Harrison, R. A., Thomas, M. L., and Tack, B. F. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 7388–7392.
Hau, J., Svendsen, P., Teisner, B., and Grudzinskas, J. G. (1982). *J. Reprod. Fertil.* **66,** 273–275.
Hauptmann, G., Grosshaus, E., and Heid, E. (1974). *Ann. Dermatol. Syphiligr.* **101,** 479–496.
Haverback, B. J., Dyce, B., Bundy, H. F., Wirtschafter, S. K., and Edmonson, H. A. (1962). *J. Clin. Invest.* **41,** 972–980.
Hayashida, K., Okubo, H., Noguchi, M., Yoshida, H., Kangawa, K., Matsuo, H., and Sasaki, Y. (1985). *J. Biol. Chem.* **260,** 14224–14229.
Heim, W. G. (1962). *Nature (London)* **193,** 491.
Heimburger, N. (1974). *Bayer-Symp.* **5,** 14–22.
Hellman, U., Eggertsen, G., Lundwall, Å., Engström, Å, and Sjöquist, J. (1984). *FEBS Lett.* **170,** 254–258.
Hergenhahn, H.-G., and Söderhall, K. (1985). *Comp. Biochem. Physiol.* **81,** 833–835.
Hirata, Y., Ishibashi, H., Kimura, H., Hayashida, K., Nagono, M., and Okubo, H. (1985). *Inflammation* **9,** 201–209.
Hoffmann, M., Feldman, S. R., and Pizzo, S. V. (1983). *Biochim. Biophys. Acta* **760,** 421–423.
Hofmann, R., Straube, W., and Klausch, B. (1979). *Zentralbl. Gynaekol.* **101,** 457–463.
Höglund, S., and Levin, Ö. (1965). *J. Mol. Biol.* **12,** 866–871.
Horne, C. H. W., Howie, P. W., Weir, R. J., and Goudie, R. B. (1970). *Lancet* **1,** 49–51.
Horne, C. H. W., McLay, A. L. C., Tavadia, H. B., Carmichael, I., Mallingson, A. C., Yeung

Laiwah, A. A. C., Thomas, M. A., and MacSween, R. N. M. (1973). *Clin. Exp. Immunol.* **13,** 603–611.

Horne, C. H. W., Bohn, H., McLay, A. L. C., Wood, E. H., and Thomson, W. D. (1975). *Behring Inst. Mitt.* **58,** 50–53.

Horne, C. H. W., Thomson, A. W., Towler, C. M., MacMillan, F. K., and Gibb, L. M. (1978). *Scand. J. Immunol.* **8,** 75–80.

Horne, C. H. W., Thomson, A. W., Hunter, C. B. J., Van Heyningen, V., Deane, D. L., and Steel, C. M. (1979). *Experientia* **35,** 411–412.

Horwitz, M. A. (1980). *J. Reticuloendothel. Soc.* **28,** 175–265.

Hostetter, M. K., Thomas, M. L., Rosen, F. S., and Tack, B. F. (1982). *Nature (London)* **298,** 72–74.

Hostetter, M. K., Krueger, R. K., and Schmelling, D. J. (1984). *J. Infect. Dis.* **150,** 653–661.

Hovi, T., Mosher, D. F., and Vaheri, A. (1977). *J. Exp. Med.* **145,** 1580–1589.

Howard, J. B. (1980). *J. Biol. Chem.* **255,** 7082–7084.

Howard, J. B. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 2235–2239.

Howard, J. B. (1983). *Ann. N. Y. Acad. Sci.* **421,** 69–80.

Howard, J. B., Vermeulen, M., and Swenson, R. P. (1980). *J. Biol. Chem.* **255,** 3820–3823.

Howell, J. B., Beck, T., Bates, B., and Hunter, M. J. (1983). *Arch. Biochem. Biophys.* **221,** 261–270.

Huang, J. S., Huang, S. S., and Deuel, T. F. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 342–346.

Hubbard, W. J., Hess, A. D., Hsia, S., and Amos, D. B. (1981). *J. Immunol.* **126,** 292–299.

Huber, H., Polley, M. J., Linscott, W. D., Fudenberg, H. H., and Müller-Eberhard, H. J. (1968). *Science* **162,** 1281–1283.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, J., Deisenhofer, J., and Steigemann, W. (1974). *J. Mol. Biol.* **89,** 73–101.

Huber, R., Scholze, H., Paques, E. P., and Deisenhofer, J. (1980). *Hoppe-Seyler's Z. Physiol. Chem.* **361,** 1389–1399.

Hudig, D., and Sell, S. (1979). *Mol. Immunol.* **16,** 547–554.

Hudson, N. W., and Koo, P. H. (1982). *Biochim. Biophys. Acta* **704,** 290–303.

Hudson, N. W., Koo, P. H., and Kehoe, J. M. (1980). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **39,** 1688.

Hugli, T. E. (1975). *J. Biol. Chem.* **250,** 8293–8301.

Hugli, T. E. (1981). *CRC Crit. Rev. Immunol.* **1,** 321–326.

Hugli, T. E. (1984). *Springer Semin. Immunopathol.* **7,** 193–219.

Hugli, T. E., Morgan, W. T., and Müller-Eberhard, H. J. (1975). *J. Biol. Chem.* **250,** 1479–1483.

Ichihara, C., Nagasawa, S., and Koyawa, J. (1981). *J. Biochem. (Tokyo)* **90,** 423–432.

Iida, K., and Nuzzenzweig, V. (1981). *J. Exp. Med.* **153,** 1138–1150.

Iida, K., and Nussenzweig, V. (1983). *J. Immunol.* **130,** 1876–1886.

Iida, K., Nadler, L., and Nussenzweig, V. (1983). *J. Exp. Med.* **158,** 1021–1033.

Iijima, M., Tobe, T., Sakamoto, T., and Tomita, M. (1984). *J. Biochem. (Tokyo)* **96,** 1539–1546.

Ikai, A., Kitamoto, T., and Nishigai, M. (1983). *J. Biochem. (Tokyo)* **92,** 121–127.

Imber, M. J., and Pizzo, S. V. (1981). *J. Biol. Chem.* **256,** 8134–8139.

Isenman, D. E. (1983). *J. Biol. Chem.* **258,** 4238–4244.

Isenman, D. E., and Cooper, N. R. (1981). *Mol. Immunol.* **18,** 331–339.

Isenman, D. E., and Kells, D. I. C. (1982). *Biochemistry* **21,** 1109–1117.

Isenman, D. E., and Young, J. R. (1984). *J. Immunol.* **132,** 3019–3027.

Isenman, D. E., Podack, E. R., and Cooper, N. R. (1980). *J. Immunol.* **124,** 326–331.

Isenman, D. E., Kells, D. I. C., Cooper, N. R., Müller-Eberhard, H. J., and Pangburn, M. K. (1981). *Biochemistry* **20,** 4458–4467.

Ivanyi, J., and Moyes, L. (1980). *Mol. Immunol.* **17,** 1545–1551.

Jacquot-Armand, Y., and Guinand, S. (1967). *Biochim. Biophys. Acta* **133,** 289–300.
Jacquot-Armand, Y., and Guinand, S. (1976). *Biochim. Biophys. Acta* **438,** 239–249.
Jacquot-Armand, Y., and Krebs, G. (1973). *Biochim. Biophys. Acta* **303,** 128–137.
James, K. (1965). *Immunology* **8,** 55–61.
James, K. (1980). *Trends Biochem. Sci.*, Vol. 5, February, pp. 43–47.
Janatova, J., and Tack, B. F. (1981). *Biochemistry* **20,** 2394–2402.
Janatova, J., Prahl, J. W., and Tack, B. F. (1979). *Proc. Int. Congr. Biochem., 11th, 1979,* Abstr. No. 08-1-R81.
Janatova, J., Lorentz, P. E., Schechter, A. N., Prahl, J. W., and Tack, B. F. (1980a). *Biochemistry* **19,** 4471–4478.
Janatova, J., Tack, B. F., and Prahl, J. W. (1980b). *Biochemistry* **19,** 4479–4485.
Jencks, W. P. (1972). *Chem. Rev.* **72,** 705–725.
Jencks, W. P. (1969). "Catalysis in Chemistry and Enzymology." McGraw-Hill, New York.
Jensen, J. A., Festa, E., Smith, D. S., and Cayer, M. (1981). *Science* **214,** 566–569.
Johnson, W. J., Pizzo, S. V., Imber, M. J., and Adams, D. O. (1982). *Science* **218,** 574–576.
Jones, J. M., Creeth, J. M., and Kekwick, R. A. (1972). *Biochem. J.* **127,** 187–197.
Kan, C.-C., Solomon, E., Belt, K. T., Chain, A. E., Hiorns, L. R., and Fey, G. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 2282–2286.
Kaplan, J., and Keogh, E. A. (1981). *Cell (Cambridge, Mass.)* **24,** 925–932.
Kaplan, J., and Nielsen, M. L. (1979a). *J. Biol. Chem.* **254,** 7323–7328.
Kaplan, J., and Nielsen, M. L. (1979b). *J. Biol. Chem.* **254,** 7329–7335.
Kaplan, J., Ray, F. A., and Keogh, E. A. (1981). *J. Biol. Chem.* **256,** 7705–7707.
Karp, D. R. (1983a). *J. Immunol.* **131,** 1405–1410.
Karp, D. R. (1983b). *J. Biol. Chem.* **258,** 12745–12748.
Karp, D. R. (1983c). *J. Biol. Chem.* **258,** 14490–14495.
Karp, D. R., Atkinson, J. P., and Shreffler, D. C. (1982a). *J. Biol. Chem.* **257,** 7330–7335.
Karp, D. R., Parker, K. L., Shreffler, D. C., Slaughter, C., and Capra, J. D. (1982b). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 6347–6349.
Karp, D. R., Shreffler, D. C., and Atkinson, J. P. (1982c). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 6666–6670.
Karp, D. R., Capra, J. D., Atkinson, J. P., and Shreffler, D. C. (1982d). *J. Immunol.* **128,** 2336–2341.
Kasukawa, R., Ohara, M., Yoshida, H., and Yoshida, T. (1979). *Int. Arch. Allergy Appl. Immunol.* **58,** 67–74.
Katayama, K., and Fujita, T. (1974a). *Biochim. Biophys. Acta* **336,** 165–177.
Katayama, K., and Fujita, T. (1974b). *Biochim. Biophys. Acta* **336,** 178–190.
Khan, S. A., and Erickson, B. W. (1981). *J. Am. Chem. Soc.* **103,** 7374–7376.
Khan, S. A., and Erickson, B. W. (1982). *J. Biol. Chem.* **257,** 11864–11867.
Kijlstra, A., Van Es, L. A., and Daha, M. (1981). *Immunology* **43,** 345–352.
Kitamoto, T., Nakashima, M., and Ikai, A. (1983). *J. Biochem. (Tokyo)* **92,** 1679–1682.
Kjellman, M., Laurell, A.-B., Löw, B., and Sjöholm, A. G. (1982). *Clin. Genet.* **22,** 331–339.
Kline, D. L., and Fishman, J. B. (1961). *J. Biol. Chem.* **236,** 2807–2812.
Knight, K. L., and Dray, S. (1968a). *Biochemistry* **7,** 1165–1171.
Knight, K. L., and Dray, S. (1968b). *Biochemistry* **7,** 3830–3835.
Koj, A., Gouldie, J., Regoeczi, E., Sauder, D., and Sweeney, G. D. (1984). *Biochem. J.* **224,** 505–514.
Koo, P. H. (1981). *Cancer Res.* **42,** 1788–1797.
Koo, P. H. (1983). *Cancer Lett.* **18,** 169–177.
Kristensen, T., Wierzbicki, D. M., and Sottrup-Jensen, L. (1984). *J. Biol. Chem.* **259,** 8313–8317.
Kurzinger, K., Ho, M. K., and Springer, T. A. (1982). *Nature (London)* **296,** 669–670.

Kushner, I. (1982). *Ann. N. Y. Acad. Sci.* **389,** 39–48.
Lachmann, P. J., and Müller-Eberhard, H. J. (1968). *J. Immunol.* **100,** 691–698.
Lachmann, P. S., Grennam, D., Martin, A., and Demant, P. C. (1975). *Nature (London)* **258,** 242, 243.
Lambris, J. D., Dobson, N. J., and Ross, G. D. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 1828–1832.
Lambris, J. D., Cohen, P. L., Dobson, N. J., Wheeler, P. W., Papamichail, M., and Ross, G. D. (1982). *Clin. Res.* **30,** 514A.
Lambris, J. D., Ganu, V. S., Hirani, S., and Müller-Eberhard, H. J. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 4235–4239.
Lanson, Y., Lanson, M., Gutman, N., and Mure, J. P. (1979). *J. Urol. Nephrol.* **85,** 399–407.
Larsson, L.-J., and Björk, I. (1984). *Biochemistry* **23,** 2802–2807.
Larsson, L.-J., Olson, S. T., and Björk, I. (1985). *Biochemistry* **24,** 1585–1593.
Laskowski, M., Jr., and Sealock, R. W. (1971). *In* "The Enzymes" (P. D. Boyer, ed.), 3rd ed., Vol. 3, pp. 376–474. Academic Press, New York.
Laurell, C.-B., and Jeppsson, J.-O. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 229–264. Academic Press, New York.
Law, S. K. (1983a). *Ann. N. Y. Acad. Sci.* **421,** 246–258.
Law, S. K. (1983b). *Biochem. J.* **211,** 381–389.
Law, S. K., and Levine, R. P. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 2701–2705.
Law, S. K., Fearon, D. T., and Levine, R. P. (1979a). *J. Immunol.* **122,** 759–765.
Law, S. K., Lichtenberg, N. A., and Levine, R. P. (1979b). *J. Immunol.* **123,** 1388–1394.
Law, S. K., Lichtenberg, N. A., Holcombe, F. H., and Levine, R. P. (1980a). *J. Immunol.* **125,** 634–639.
Law, S. K., Lichtenberg, N. A., and Levine, R. P. (1980b). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 7194–7198.
Law, S. K., Minich, T. M., and Levine, R. P. (1981). *Biochemistry* **20,** 7457–7463.
Law, S. K., Minich, T. M., and Levine, R. P. (1984a). *Biochemistry* **23,** 3267–3272.
Law, S. K., Dodds, A. W., and Porter, R. R. (1984b). *EMBO J.* **3,** 1819–1823.
Lebreton de Vonne, T., and Mouray, M. H. (1968). *C.R. Hebd. Seances Acad. Sci., Ser. D* **266,** 1076–1079.
Leikola, J., Fudenberg, H. H., Fasukawa, R., and Milgrom, F. (1972). *Am. J. Hum. Genet.* **24,** 134–144.
Leslie, R. G. Q. (1985). *Mol. Immunol.* **22,** 513–519.
Levi-Strauss, M., Tosi, M., Steinmetz, M., Klein, J., and Meo, T. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 1746–1750.
Levitzki, A., Willingham, M., and Pastan, I. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 2706–2710.
Lichtenstein, N. (1942). *J. Am. Chem. Soc.* **64,** 1021–1028.
Lin, T.-M., and Halbert, S. P. (1976). *Science* **193,** 1249–1252.
Lin, T.-M., Halbert, S. P., Kiefer, D., and Spellacy, W. N. (1974). *Int. Arch. Allergy Appl. Immunol.* **47,** 35–53.
Lin, T.-M., Halbert, S. P., and Kiefer, D. (1976). *J. Clin. Invest.* **57,** 466–472.
Lundgren, E., Damber, M. G., Roos, G., von Schoultz, B., Stigbrand, T., Nilsson, K., and Alexander, T. T. (1979). *Int. J. Cancer* **24,** 45–48.
Lundwall, Å., Hellman, U., Eggertsen, G., and Sjöquist, J. (1981). *Mol. Immunol.* **19,** 1655–1665.
Lundwall, Å., Wetsel, R. A., Domdey, H., Tack, B. F., and Fey, G. H. (1984). *J. Biol. Chem.* **259,** 13851–13856.
Lundwall, Å. B., Wetsel, R. H., Kristensen, T., Whitehead, A. S., Woods, D. E., Ogden, R. C., Colten, H. R., and Tack, B. F. (1985). *J. Biol. Chem.* **260,** 2108–2112.
Mackin, W. M., Mayer, E. P., Dray, S., and Reiter, H. (1983). *Immunology* **49,** 519–528.

Mahour, G. H., Song, M. K., Adham, N. F., and Rinderknecht, H. (1978). *Pediatrics* **61,** 894–897.
Mannik, M., Arend, W. P., Hall, A. P., and Gillilaud, B. C. (1971). *J. Exp. Med.* **133,** 713–739.
Mantovani, B. (1975). *J. Immunol.* **115,** 15–20.
Mantovani, B., Rabinovitch, M., and Nussenzweig, V. (1972). *J. Exp. Med.* **135,** 780–792.
Marynen, P., Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1981). *J. Immunol.* **127,** 1782–1786.
Marynen, P., Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1982a). *FEBS Lett.* **137,** 241–244.
Marynen, P., Van Der Schueren, B., Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1982b). *Haemostasis* **11,** 210–214.
Marynen, P., Van Leuven, F., and Cassiman, J.-J. (1983). *Ann. N. Y. Acad. Sci.* **421,** 401–403.
Marynen, P., Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1984a). *Biochim. Biophys. Acta* **797,** 187–193.
Marynen, P., Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1984b). *J. Biol. Chem.* **259,** 7075–7079.
Mascart-Lemone, F., Hauptmann, G., Goetz, J., Duchateau, J., Delespesse, G., Vray, B., and Dab, I. (1983). *Am. J. Med.* **75,** 295–304.
Matsuda, T., Seya, T., and Nagasawa, S. (1985). *Biochem. Biophys. Res. Commun.* **127,** 264–269.
Mauff, G., Alper, C. A., Awdeh, Z., Batchelor, J. R., Bertrams, T., Bruun-Petersen, G., Dawkins, R. L., Demant, P., Edwards, J. Grosse-Wild, H., Hauptmann, G., Klona, P., Lamm, L., Mullenhauer, E., Nerl, C., Olaisen, B., O'Neill, G. O., Rittner, C., Roos, M. H., Shanes, V., Teisberg, P., and Wells, L. (1983a). *Immunobiology* **164,** 184–191.
Mauff, G., Stener, M., Weck, M., and Bender, K. (1983b). *Hum. Genet.* **64,** 186–188.
Maxfield, F. R., Schlessinger, J., Shechter, Y., Pastan, I., and Willingham, M. C. (1978). *Cell (Cambridge, Mass.)* **14,** 805–810.
Maxfield, F. R., Willingham, M. C., Davies, P. J. A., and Pastan, I. (1979). *Nature (London)* **277,** 661–663.
Maxfield, F. R., Willingham, M. C., Haigler, H. T., Dragsten, P., and Pastan, I. H. (1981). *Biochemistry* **20,** 5353–5358.
Medicus, R. G., Melamed, J., and Arnaout, M. A. (1983). *Eur. J. Immunol.* **13,** 465–470.
Medof, M. E., and Nussenzweig, V. (1984). *J. Exp. Med.* **159,** 1669–1685.
Medof, M. E., Iida, K., Mold, C., and Nussenzweig, V. (1982). *J. Exp. Med.* **156,** 1739–1754.
Medof, M. E., Lam, T., Prince, G. M., and Mold, C. (1983). *J. Immunol.* **130,** 1336–1340.
Medof, M. E., Kinoshita, T., and Nussenzweig, V. (1985). *J. Exp. Med.* **160,** 1558–1578.
Mehl, J. W., O'Connel, W., and Degroot, J. (1964). *Science* **145,** 821–822.
Mendenhall, H. W. (1970). *Am. J. Obstet. Gynecol.* **107,** 388–399.
Menninger, F. F., Jr., Ester, H. J., and Bogden, A. E. (1970). *Clin. Chim. Acta* **27,** 385–394.
Meo, T., Krasteff, T., and Schreffler, D. C. (1975). *Proc. Natl. Acad. Sci. U.S.A.* **72,** 4536–4540.
Meuth, J. L., Morgan, E. L., DiScipio, R. G., and Hugli, T. E. (1983). *J. Immunol.* **130,** 2605–2611.
Mevag, B., Olaisen, B., and Teisberg, P. (1981). *Scand. J. Immunol.* **14,** 303–307.
Meyer, J.-F., Bieth, J., and Metais, P. (1975). *Clin. Chim. Acta* **62,** 45–53.
Micklem, K. J., Sim, R. B., and Sim, E. (1984). *Biochem. J.* **224,** 75–86.
Miller, H. T., and Feeney, R. E. (1966). *Biochemistry* **5,** 952–958.
Minta, J. O., Urowitz, M. B., Gladman, D. D., Irizawa, T., and Biggar, W. D. (1981). *Clin. Exp. Immunol.* **45,** 72–80.
Misumi, Y., Takami, N., and Ikehara, Y. (1984). *FEBS Lett.* **175,** 63–67.
Miyanaga, O., Okuba, H., Kudo, J., Ikuta, T., and Hirata, Y. (1982). *Immunology* **47,** 351–356.
Miyata, K., Nakamura, M., and Tomoda, K. (1981). *J. Biochem. (Tokyo)* **89,** 1231–1237.
Molenaar, J. L., Müller, M. A. C., Engefriet, C. P., and Pondman, K. W. (1974). *J. Immunol.* **112,** 1444–1451.

Molenaar, J. L., Helder, A. W., Müller, M. A. C., Goris-Mulder, M., Jonker, L. S., Brouwer, M., and Pondman, K. W. (1975). *Immunochemistry* **12,** 359–364.

Moon, K. E., Gorski, J. P., and Hugli, T. E. (1981). *J. Biol. Chem.* **256,** 8685–8692.

Morelis, P., Ambrosioni, J.-C., Got, R., and Fontanges, R. (1969). *C.R. Hebd. Seances Acad. Sci., Ser. D* **269,** 1453–1454.

Morgan, A. C., Jr. (1981). *JNCI, J. Natl. Cancer Inst.* **72,** 557–562.

Morgan, W. T., Vallota, E. H., and Müller-Eberhard, H. J. (1974). *Biochem. Biophys. Res. Commun.* **57,** 572–577.

Morse, B., and Torbell, D. (1952). *J. Am. Chem. Soc.* **74,** 416–425.

Mortensen, S. B., Sottrup-Jensen, L., Hansen, H. F., Rider, D., Petersen, T. E., and Magnusson, S. (1981a). *FEBS Lett.* **129,** 314–317.

Mortensen, S. B., Sottrup-Jensen, L., Hansen, H. F., Petersen, T. E., and Magnusson, S. (1981b). *FEBS Lett.* **135,** 295–300.

Mosher, D. F. (1976). *J. Biol. Chem.* **251,** 1639–1645.

Mosher, D. F., and Vaheri, A. (1980). *Biochim. Biophys. Acta* **627,** 113–122.

Mosher, D. F., and Wing, D. A. (1976). *J. Exp. Med.* **143,** 462–467.

Mosher, D. F., Saksela, O., and Vaheri, A. (1977). *J. Clin. Invest.* **60,** 1036–1045.

Muir, W. A., Hedrick, S., Alper, C. A., Ratnoff, O. D., Schachter, B., and Wisnieski, J. J. (1984). *J. Clin. Invest.* **74,** 1509–1514.

Müller-Eberhard, H. J. (1961). *Acta Soc. Med. Ups.* **66,** 152–170.

Müller-Eberhard, H. J. (1975). *Annu. Rev. Biochem.* **44,** 697–724.

Müller-Eberhard, H. J. (1984). *Springer Semin. Immunopathol.* **7,** 93–141.

Müller-Eberhard, H. J., and Biro, C. E. (1963). *J. Exp. Med.* **118,** 447–466.

Müller-Eberhard, H. J., and Lepow, I. H. (1965). *J. Exp. Med.* **121,** 819–833.

Müller-Eberhard, H. J., and Schreiber, R. D. (1980). *Adv. Immunol.* **29,** 1–53.

Müller-Eberhard, H. J., Nilsson, U. R., and Aronson, T. (1960). *J. Exp. Med.* **111,** 201–215.

Müller-Eberhard, H. J., Dalmasso, A. P., and Calcott, M. A. (1966). *J. Exp. Med.* **124,** 33–54.

Nachman, R. L., and Harpel, P. C. (1976). *J. Biol. Chem.* **25,** 4514–4521.

Nagasawa, S., Sugihara, H., Hoon Han, B., and Suzuki, T. (1970a). *J. Biochem. (Tokyo)* **67,** 809–819.

Nagasawa, S., Hoon Han, B., Sugihara, H., and Suzuki, T. (1970b). *J. Biochem. (Tokyo)* **67,** 821–832.

Nagasawa, S., Ichihara, C., and Stroud, R. M. (1980). *J. Immunol.* **125,** 578–582.

Nagase, H., and Harris, E. D., Jr. (1983). *J. Biol. Chem.* **258,** 7490–7498.

Nagase, H., Harris, E. D., Jr., Woessner, J. F., and Brew, K. (1983). *J. Biol. Chem.* **258,** 7481–7489.

Nagase, H., Brew, K., and Harris, E. D., Jr. (1985). *Prog. Clin. Biol. Res.* **180,** 283–285.

Natsuume-Sakai, S., Hayakawa, J. I., and Takahashi, M. (1978a). *J. Immunol.* **121,** 491–498.

Natsuume-Sakai, S., Amano, S., Hayakawa, J. I., and Takahashi, M. (1978b). *J. Immunol.* **121,** 2025–2029.

Natsuume-Sakai, S., Moriwaka, K., Amano, S., Hayakawa, J. I., Kaidoh, T., and Takahashi, M. (1979a). *J. Immunol.* **123,** 216–221.

Natsuume-Sakai, S., Hayakawa, J. I., Amano, S., and Takahashi, M. (1979b). *J. Immunol.* **123,** 947–948.

Nelles, L. P., and Schnebli, H. P. (1982). *Hoppe-Seyler's Z. Physiol. Chem.* **363,** 677–682.

Nelles, L. P., Hall, P. K., and Roberts, R. C. (1980). *Biochim. Biophys. Acta* **623,** 46–56.

Nelson, R. A., Jr. (1953). *Science* **118,** 733–737.

Newman, S. L., and Johnston, R. B., Jr. (1979). *J. Immunol.* **123,** 1839–1846.

Ney, K. A., Gidwitz, S., and Pizzo, S. V. (1984). *Biochemistry* **23,** 3395–3403.

Nicholson-Weller, A., Burge, J., Fearon, D. T., Weller, P. F., and Austen, K. F. (1982). *J. Immunol.* **129,** 184–189.

Nieuwenhuisen, W., Emeis, J. J., and Hemmink, J. (1979). *Biochim. Biophys. Acta* **580,** 129–139.
Nilehn, J.-E., and Ganrot, P. O. (1967). *Scand. J. Clin. Lab. Invest.* **20,** 113–121.
Nilsson, U. R., and Mapes, J. (1973). *J. Immunol.* **111,** 293.
Nilsson, U. R., Mandle, R. J., Jr., and McConnel-Mapes, J. A. (1975). *J. Immunol.* **114,** 815–824.
Nonaka, M., Natsuume-Sakai, S., Kaidoh, T., Nonaka, M., and Takahashi, M. (1980). *J. Immunol.* **125,** 2025–2030.
Nonaka, M., Yamaguchi, N., Natsuume-Sakai, S., and Takahashi, M. (1981a). *J. Immunol.* **126,** 1489–1494.
Nonaka, M., Natsuume-Sakai, S., and Takahashi, M. (1981b). *J. Immunol.* **126,** 1495–1898.
Nonaka, M., Iwaki, M., Nakai, C., Nozaki, M., Kaidoh, T., Nonaka, M., Natsuume-Sakai, S., and Takahashi, M. (1984a). *J. Biol. Chem.* **254,** 6327–6333.
Nonaka, M., Takahashi, M., Natsuume-Sakai, S., Nonaka, M., Tanaka, S., Shimizu, A., and Honja, T. (1984b). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 6822–6826.
Nonaka, M., Nonaka, M., Irie, M., Tanabe, K., Kaidoh, T., Natsuume-Sakai, S., and Takahashi, M. (1985a). *J. Biol. Chem.* **260,** 809–815.
Nonaka, M., Nakayama, K., Yeul, Y. D., and Takahashi, M. (1985b). *J. Biol. Chem.* **260,** 10936–10943.
Northemann, W., Andus, T., Gross, V., and Heinrich, P. C. (1983). *Eur. J. Biochem.* **137,** 257–262.
Northemann, W., Heisig, M., Kunz, D., and Heinrich, P. C. (1985). *J. Biol. Chem.* **260,** 6200–6205.
Odink, K., Fey, G., Wiebauer, K., and Diggelman, H. (1981). *J. Biol. Chem.* **256,** 1453–1458.
Ogata, R. T., Shreffler, D. C., Sepich, D. S., and Lilly, S. P. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5061–6065.
Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H., and Sasaki, M. (1984). *Biochemistry* **23,** 5691–5697.
Ohlsson, K. (1971a). *Acta Physiol. Scand.* **81,** 269–272.
Ohlsson, K. (1971b). *Biochim. Biophys. Acta* **236,** 84–91.
Ohlsson, K. (1974). *Bayer-Symp.* **5,** 96–105.
Ohlsson, K., and Olsson, A.-S. (1978). *Hoppe-Seyler's Z. Physiol. Chem.* **359,** 1531–1539.
Ohlsson, K., and Skude, G. (1976). *Clin. Chim. Acta* **66,** 1–7.
Ohlsson, K., Ganrot, P. O., and Laurell, C.-B. (1971). *Acta Chir. Scand.* **137,** 113–121.
Ohlsson, K., Polling, Å., and Stenberg, P. (1982). *Hoppe-Seyler's Z. Physiol. Chem.* **363,** 213–220.
Okubo, H., Miyanaga, O., Nagano, M., Ishibashi, H., Kudo, J., Ikuta, T., and Shibata, K. (1981). *Biochim. Biophys. Acta* **668,** 257–267.
O'Neill, G. J., Yang, S. Y., and Dupont, B. (1978a). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 5165–5169.
O'Neill, G. J., Yang, S. Y., Tegoli, J., Berger, R., and Dupont, B. (1978b). *Nature (London)* **273,** 668–670.
Ooi, Y. M., and Colten, H. R. (1979). *J. Immunol.* **123,** 2494–2498.
Österberg, R., and Malmensten, B. (1984). *Eur. J. Biochem.* **143,** 541–544.
Österberg, R., Eggertsen, G., Lundwall, Å., and Sjöquist, J. (1984). *Int. J. Biol. Macromol.* **6,** 195–198.
Österberg, R., Nilsson, U. R., and Eggertsen, G. (1985). *J. Biol. Chem.* **260,** 12970–12973.
Palsdottir, A., Cross, S. J., Edwards, J. H., and Carroll, M. C. (1983). *Nature (London)* **306,** 615–616.
Pan, Y. T., Mukherjee, A. K., Horowitz, P. M., and Elbein, A. D. (1980). *Biochem. Biophys. Res. Commun.* **92,** 703–709.
Pangburn, M. K., and Müller-Eberhard, H. J. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 2416–2420.

Pangburn, M. K., and Müller-Eberhard, H. J. (1980). *J. Exp. Med.* **152,** 1102–1114.
Pangburn, M. K., Schreiber, R. D., and Müller-Eberhard, H. J. (1977). *J. Exp. Med.* **146,** 257–270.
Pangburn, M. K., Schreiber, R. D., and Müller-Eberhard, H. J. (1981). *J. Exp. Med.* **154,** 856–867.
Pangburn, M. K., Schreiber, R. D., and Müller-Eberhard, H. J. (1983a). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5430–5434.
Pangburn, M. K., Schreiber, R. D., Trombold, J. S., and Müller-Eberhard, H. J. (1983b). *J. Exp. Med.* **157,** 1971–1980.
Panrucker, D. E., and Lorscheider, F. L. (1982). *Biochim. Biophys. Acta* **705,** 174–183.
Paques, E. P. (1980). *Hoppe-Seyler's Z. Physiol. Chem.* **361,** 445–456.
Parameswaran, K. N., and Lorand, L. (1981). *Biochemistry* **20,** 3703–3711.
Parker, K. L., Roos, M. H., and Shreffler, D. C. (1979). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 5853–5857.
Parker, K. L., Shreffler, D. C., and Capra, J. D. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 4275–4278.
Parkes, C., DiScipio, R. G., Kerr, M. H., and Porter, R. R. (1981). *Biochem. J.* **193,** 963–970.
Parsons, M., and Romeo, G. (1980). *Clin. Chim. Acta* **100,** 215–224.
Passmore, H. C., and Shreffler, D. C. (1970). *Biochem. Genet.* **4,** 351–365.
Passmore, H. C., Kubo, J. K., Singh, S. K., and Klein, J. (1980). *Immunogenetics* **11,** 397–405.
Pastan, I. H., and Willingham, M. C. (1981). *Annu. Rev. Physiol.* **43,** 239–250.
Patel, F., and Minta, J. D. (1979). *J. Immunol.* **122,** 1582–1586.
Penalva da Silva, F., Hoecker, G. F., Day, N. K., Vienne, K., and Rubinstein, P. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 963–967.
Perlman, H., Perlman, P., Schreiber, R. D., and Müller-Eberhard, H. J. (1981). *J. Exp. Med.* **153,** 1592–1603.
Pillemer, L., Seifter, J., and Ecker, E. E. (1941). *J. Immunol.* **40,** 89–95.
Pillemer, L., Ratnoff, O. D., Blum, L., and Lepow, I. H. (1953). *J. Exp. Med.* **97,** 573–589.
Pochon, F., and Bieth, J. G. (1982). *J. Biol. Chem.* **257,** 6683–6685.
Pochon, F., and Steinbuch, M. (1984). *FEBS Lett.* **177,** 109–111.
Pochon, F., Amand, B., Lavalette, D., and Bieth, J. (1978). *J. Biol. Chem.* **253,** 7496–7499.
Pochon, F., Favaudon, V., Tourbez-Perrin, M., and Bieth, J. (1981). *J. Biol. Chem.* **256,** 547–550.
Pochon, F., Favaudon, V., and Bieth, J. (1983a). *Biochem. Biophys. Res. Common.* **111,** 964–969.
Pochon, R., Steinbuch, M., Lambin, P., and Kichenin, V. (1983b). *FEBS Lett.* **161,** 51–54.
Podack, E. R., and Tschopp, J. (1984). *Mol. Immunol.* **21,** 589–603.
Podack, E. R., Kolb, W. P., and Müller-Eberhard, H. J. (1978). *J. Immunol.* **120,** 1841–1848.
Pondman, K. W., and Peetoom, F. (1964). *Immunochemistry* **1,** 65–90.
Porter, R. R., and Reid, K. B. M. (1979). *Adv. Protein Chem.* **33,** 1–71.
Porstmann, T., and Hau, J. (1984). *Lab. Anim.* **18,** 344–348.
Press, E. M., and Gagnon, J. (1981). *Biochem. J.* **199,** 351–357.
Propp, R. P., and Alper, C. A. (1968). *Science* **162,** 672–673.
Quigley, J. P., and Armstrong, P. B. (1983). *Ann. N. Y. Acad. Sci.* **421,** 119–124.
Quigley, J. P., and Armstrong, P. B. (1985). *J. Biol. Chem.* **260,** 12715–12719.
Rastogi, S. C., and Clausen, J. (1985). *Immunobiology* **169,** 37–44.
Ratnoff, O. D., Lepow, I. H., and Pillemer, L. (1954). *Bull. Johns Hopkins Hosp.* **94,** 169–179.
Raum, D., Donaldson, V. H., Rosen, F. S., and Alper, C. A. (1980). *Curr. Top. Hematol.* **3,** 111–174.
Reboul, A., Thielens, N., Villiers, M. B., and Colomb, M. G. (1979). *FEBS Lett.* **103,** 156–161.
Reboul, A., Thielens, N., Villiers, M. B., and Colomb, M. G. (1980). *FEBS Lett.* **115,** 118–122.
Reid, K. B. M., and Porter, R. R. (1981). *Annu. Rev. Biochem.* **50,** 433–464.
Richardson, J. S. (1981). *Adv. Protein Chem.* **34,** 167–339.
Richman, J. B. Y., and Verpoorte, J. H. (1981). *Can. J. Biochem.* **59,** 519–523.

Rinderknecht, H., and Geokas, M. C. (1973). *Biochim. Biophys. Acta* **295,** 233–244.
Rinderknecht, H., Carmack, C., and Geokas, M. C. (1975). *Immunochemistry* **12,** 1–8.
Rittner, C. H., and Rittner, B. (1973). *Vox Sang.* **25,** 21–27.
Roberts, R. C. (1985). *Rev. Hematol.* (in press).
Roberts, R. C., Nelles, L. P., Hall, P. K., Salvesen, G. S., and Mischler, E. (1982). *Pediatr. Res.* **16,** 416–423.
Roord, J. J., Daha, M., Kuis, W., Verbrugh, H. A., Verhoef, Y., Zegers, B. J. M., and Stoop, J. W. (1983). *Pediatrics* **71,** 81–87.
Roos, M. H., Atkinson, J. P., and Shreffler, D. C. (1978). *J. Immunol.* **121,** 1106–1115.
Roos, M. H., Mollenhauer, E., Demant, P., and Rittner, C. (1982). *Nature (London)* **298,** 854–856.
Rosenfeld, S. I., Kelley, M. E., and Leddy, J. P. (1976). *J. Clin. Invest.* **57,** 1626–1634.
Ross, G. D., and Lambris, J. D. (1982). *J. Exp. Med.* **155,** 96–110.
Ross, G. D., and Rabellino, E. M. (1979). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **38,** 1467.
Ross, G. D., Polley, M. J., Rabellino, E. M., and Grey, H. M. (1973). *J. Exp. Med.* **138,** 798–811.
Ross, G. D., Lambris, J. D., Cain, J. A., and Newman, S. L. (1982). *J. Immunol.* **129,** 2051–2060.
Ruddy, S., and Austen, K. F. (1971). *J. Immunol.* **107,** 742–750.
Saito, A., and Sinohara, H. (1985a). *J. Biol. Chem.* **260,** 775–781.
Saito, A., and Sinohara, H. (1985b). *J. Biochem. (Tokyo)* **98,** 501–516.
Saksela, O., Wahlström, T., Lehtovirta, P., Seppäla, M., and Vaheri, A. (1981). *Cancer Res.* **41,** 2507–2513.
Saksela, O., Wahlström, T., Meyer, B., and Vaheri, A. (1984). *Cancer Res.* **44,** 2942–2946.
Salomon, D. S., Bano, M., Smith, K. B., and Kidwell, W. R. (1982). *J. Biol. Chem.* **257,** 14093–14101.
Salvesen, G. S., and Barrett, A. J. (1980). *Biochem. J.* **187,** 695–701.
Salvesen, G. S., Sayers, C. A., and Barrett, A. J. (1981). *Biochem. J.* **195,** 453–461.
Sand, O., Folkersen, J., Westergaard, J. G., and Sottrup-Jensen, L. (1985). *J. Biol. Chem.* **260,** 15723–15735.
Sasaki, M., Minakata, K., Yamamoto, H., Niwa, M., Kato, T., and Ito, N. (1977). *Biochem. Biophys. Res. Commun.* **76,** 917–924.
Saunders, R., Dyce, B. J., Vannier, W. E., and Haverback, B. J. (1971). *J. Clin. Invest.* **50,** 2376–2383.
Sayers, C. A., and Barrett, A. J. (1980). *Biochem. J.* **189,** 255–261.
Schaeufele, J. T., and Koo, P. H. (1982). *Biochem. Biophys. Res. Commun.* **108,** 1–7.
Schlegel, R., Dickson, R. B., Willingham, M. C., and Pastan, I. H. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 2291–2295.
Schlessinger, J., Shechter, Y., Willingham, M. G., and Pastan, I. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 2659–2663.
Schramm, H. J., and Schramm, W. (1982). *Hoppe-Seyler's Z. Physiol. Chem.* **363,** 803–812.
Schreiber, R. D. (1984). *Springer Semin. Immunopathol.* **7,** 221–249.
Schreiber, R. D., Pangburn, M. K., and Müller-Eberhard, H. J. (1981). *Bio-sci. Rep.* **1,** 873–880.
Schreiber, R. D., Pangburn, M. K., Bjornson, A. B., Brothers, M. A., and Müller-Eberhard, H. J. (1982). *Clin. Immunol. Immunopathol.* **23,** 335–357.
Schultze, H. E., Göllner, I., Heide, K., Schönenberger, M., and Schwick, H. G. (1955). *Z. Naturforsch. B: Anorg. Chem. Org. Chem., Biochem., Biophys., Biol.* **10B,** 463–473.
Senger, D. R., and Hynes, R. O. (1978). *Cell* **15,** 375–384.
Sepich, D. S., Noonan, D. J., and Ogata, R. T. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 5895–5899.
Seya, T., and Nagasawa, S. (1981). *J. Biochem. (Tokyo)* **89,** 659–664.
Shapira, E., Rao, G. J. S., Wessel, H. U., and Nadler, H. L. (1976). *Pediatr. Res.* **10,** 812–817.
Shapira, E., Martin, C. L., and Nadler, H. L. (1977). *J. Biol. Chem.* **252,** 7923–7929.
Shapiro, S. S., and Anderson, D. B. (1977). *In* "Chemistry and Biology of Thrombin" (R. L.

Lundblad, J. W. Fenton, II, and K. G. Mann, eds.), pp. 361–374. Ann Arbor Press, Ann Arbor, Michigan.

Shortridge, K. F., Belyavin, G., and Bidwell, D. E. (1976). *Comp. Biochem. Physiol.* **54,** 319–321.

Sim, E., Wood, A. B., Hsiung, L.-M., and Sim, R. B. (1981). *FEBS Lett.* **132,** 55–60.

Sim, R. B., and Sim, E. (1981). *Biochem. J.* **193,** 129–141.

Sim, R. B., and Sim, E. (1983). *Ann. N. Y. Acad. Sci.* **421,** 259–276.

Sim, R. B., Twose, T. M., Paterson, D. S., and Sim, E. (1981). *Biochem. J.* **193,** 115–127.

Sinosich, M. J., Davey, M. W., Ghosh, P., and Grudzinskas, J. G. (1982). *Biochem. Int.* **5,** 777–786.

Smithies, O. (1959). *Adv. Protein Chem.* **14,** 65–133.

Solomon, F., and Jencks, W. P. (1969). *J. Biol. Chem.* **244,** 1079–1081.

Sottrup-Jensen, L. (1985). *J. Biol. Chem.* **260,** 6500.

Sottrup-Jensen, L., and Hansen, H. F. (1982). *Biochem. Biophys. Res. Commun.* **107,** 93–100.

Sottrup-Jensen, L., Stepanik, T. M., Jones, C. M., Petersen, T. E., and Magnusson, S. (1979). *In* "The Physiological Inhibitors of Blood Coagulation and Fibrinolysis" (D. Collen, B. Wiman, and M. Verstraete, eds.), pp. 255–271. Elsevier/North-Holland Biomedical Press, Amsterdam.

Sottrup-Jensen, L., Petersen, T. E., and Magnusson, S. (1980). *FEBS Lett.* **121,** 275–279.

Sottrup-Jensen, L., Hansen, H. F., Mortensen, S. B., Petersen, T. E., and Magnusson, S. (1981a). *FEBS Lett.* **123,** 145–148.

Sottrup-Jensen, L., Lønblad, P. B., Stepanik, T. M., Petersen, T. E., Magnusson, S., and Jörnvall, H. (1981b). *FEBS Lett.* **127,** 167–173.

Sottrup-Jensen, L., Petersen, T. E., and Magnusson, S. (1981c). *FEBS Lett.* **128,** 123–126.

Sottrup-Jensen, L., Petersen, T. E., and Magnusson, S. (1981d). *FEBS Lett.* **128,** 127–132.

Sottrup-Jensen, L., Stepanik, T. M., Wierzbicki, D. M., Jones, C. M., Lønblad, P. B., Kristensen, T., Mortensen, S. B., Petersen, T. E., and Magnusson, S. (1983a). *Ann. N. Y. Acad. Sci.* **421,** 41–60.

Sottrup-Jensen, L., Hansen, H. F., and Christensen, U. (1983b). *Ann. N. Y. Acad. Sci.* **421,** 188–201.

Sottrup-Jensen, L., Stepanik, T. M., Jones, C. M., Lønblad, P. B., Kristensen, T., and Wierzbicki, D. M. (1984a). *J. Biol. Chem.* **259,** 8293–8303.

Sottrup-Jensen, L., Lønblad, P. B., Jones, C. M., and Stepanik, T. M. (1984b). *J. Biol. Chem.* **259,** 8304–8309.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Wierzbicki, D. M., Jones, C. M., Lønblad, P. B., Magnusson, S., and Petersen, T. E. (1984c). *J. Biol. Chem.* **259,** 8318–8327.

Sottrup-Jensen, L., Folkersen, J., Kristensen, T., and Tack, B. F. (1984d). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 7353–7357.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Lønblad, P. B., Jones, C. M., Wierzbicki, D. M., Magnusson, S., Domdey, H., Wetsel, R. A., Lundwall, Å., Tack, B. F., and Fey, G. H. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 9–13.

Spolarics, Z., Mandl, J., Machovich, R., Lambin, P., Garzo, T., Antoni, F., and Horvath, I. (1985). *Biochim. Biophys. Acta* **845,** 389–395.

Sramek, S. J., and Frerman, F. E. (1975). *Arch. Biochem. Biophys.* **171,** 27–35.

Starkey, P. M., and Barrett, A. J. (1977). *In* "Proteinases in Mammalian Cells and Tissues" (A. J. Barrett, ed.), pp. 663–696. Elsevier/North-Holland Biomedical Press, Amsterdam.

Starkey, P. M., and Barrett, A. J. (1982a). *Biochem. J.* **205,** 91–95.

Starkey, P. M., and Barrett, A. J. (1982b). *Biochem. J.* **205,** 105–113.

Starkey, P. M., Fletcher, T. C., and Barrett, A. J. (1982). *Biochem. J.* **205,** 97–104.

Steenbjerg, S. (1981). *Thromb. Res.* **22,** 491–495.

Steinbuch, M., Pejaudier, L., Quentin, M., and Martin, V. (1968). *Biochim. Biophys. Acta* **154,** 228–231.

Steinbuch, M., Audran, R., Lambin, P., and Fine, J. M. (1976). *Protides Biol. Fluids* **23,** 133–138.

Steiner, J. P., Bhattacharya, P., and Strickland, D. K. (1985). *Biochemistry* **24,** 2993–3001.
Steinman, R. M., Mellman, I. S., Muller, W. A., and Cohn, Z. A. (1983). *J. Cell Biol.* **96,** 1–27.
Stepanik, T. M., and Sottrup-Jensen, L. (1984). *J. Blol. Chem.* **259,** 8310–8312.
Stewart, R., and Srinavasan, R. (1978). *Acc. Chem. Res.* **11,** 271–275.
Stigbrand, T., Damber, M.-G., and von Schoultz, B. (1978). *Acta Chem. Scand., Ser. B* **B32,** 717–719.
Stimson, W. H. (1972). *Clin. Biochem.* **5,** 3–12.
Stimson, W. H. (1974). *J. Endocrinol.* **61,** 30–31.
Stimson, W. H. (1975). *Lancet* **1,** 777–779.
Stimson, W. H. (1976). *Clin. Exp. Immunol.* **25,** 199–206.
Stimson, W. H. (1977). *Clin. Exp. Immunol.* **28,** 445–452.
Stimson, W. H., and Blackstock, T. L. (1975). *Experientia* **31,** 371–373.
Stimson, W. H., and Eubank-Scott, L. (1972). *FEBS Lett.* **23,** 298–302.
Stimson, W. H., and Farquharson, D. M. (1978). *Int. J. Biochem.* **9,** 839–843.
Stimson, W. H., Farquharson, D. M., Shepherd, A., and Anderson, D. M. (1979). *J. Clin. Lab. Immunol.* **2,** 235–238.
Stossel, T. P. (1973). *J. Cell Biol.* **58,** 346–356.
Stossel, T. P., Field, R. J., Gitlin, J. D., Alper, C. A., and Rosen, F. S. (1975). *J. Exp. Med.* **141,** 1329–1347.
Straight, D. L., and McKee, P. A. (1982). *Biochemistry* **21,** 4550–4556.
Straight, D. L., and McKee, P. A. (1984). *J. Biol. Chem.* **259,** 1272–1284.
Straight, D. L., Hassett, M. A., and McKee, P. A. (1985). *Biochemistry* **24,** 3902–3907.
Straube, W., Klausch, B., Hoffmann, R., and Brock, J. (1972). *Arch. Gynaekol.* **212,** 230–245.
Straube, W., Goretzlehner, G., Hoffmann, R., and Klausch, B. (1973). *Zentralbl. Gynaekol.* **95,** 1281–1286.
Strickland, D. K., and Bhattacharya, P. (1984). *Biochemistry* **23,** 3115–3124.
Strickland, D. K., Steiner, J. P., Feldman, S. R., and Pizzo, S. V. (1984). *Biochemistry* **23,** 6679–6685.
Strunk, R. C., Tashjian, A. H., and Colten, H. R. (1975). *J. Immunol.* **114,** 331–337.
Sturdee, D. W., Burnett, D., Moore, B., and Bradwell, A. R. (1976). *Clin. Chim. Acta* **72,** 233–239.
Sundsmo, J. S. (1980). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **39,** 1200.
Sutcliffe, R., Kukulska-Langlands, B. M., Coggins, J. R., Hunter, J. B., and Gore, C. H. (1980). *Biochem. J.* **191,** 799–809.
Suzuki, I., Okada, H., Kawachi, S., Takahashi, M., and Nahioka, K. (1972). *Jpn. J. Exp. Med.* **42,** 309–312.
Swenson, R. P., and Howard, J. B. (1979a). *J. Biol. Chem.* **254,** 4452–4456.
Swenson, R. P., and Howard, J. B. (1979b). *Proc. Natl. Acad. Sci. U.S.A.* **76,** 4313–4316.
Swenson, R. P., and Howard, J. B. (1980). *J. Biol. Chem.* **255,** 8087–8091.
Switzer, M. E. P., Gordon, H. J., and McKee, P. A. (1983). *Biochemistry* **22,** 1437–1444.
Tack, B. F., and Prahl, J. W. (1976). *Biochemistry* **15,** 4513–4521.
Tack, B. F., Morris, S. C., and Prahl, J. W. (1979a). *Biochemistry* **18,** 1490–1496.
Tack, B. F., Morris, S. C., and Prahl, J. W. (1979b). *Biochemistry* **18,** 1497–1503.
Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M. L., and Prahl, J. W. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 5764–5768.
Tamura, N., and Nelson, R. A., Jr. (1967). *J. Immunol.* **99,** 582–589.
Tapon-Bretaudiere, J., Bros, A., Couture-Tosi, E., and Delain, E. (1985). *EMBO J.* **4,** 85–89.
Tappeiner, G., Scholtz, S., Linert, J., Albert, E. D., and Wolff, K. (1978). *Colloq.—Inst. Natl. Sante Rech. Med.* **80,** 399–404.
Taylor, A. B., and Leon, M. A. (1959). *J. Immunol.* **83,** 284–290.

Teixeira, N. A., and O'Grady, J. (1985). *Biochem. Soc. Trans.* **13,** 373–374.
Teodorescu, M., Chang, J.-L., and Skosey, J. L. (1981). *Int. Arch. Allergy Appl. Immunol.* **66,** 1–12.
Than, G. N., Csabo, I. F., Karg, N. J., Szabo, D. G., and Novak, P. F. (1975a). *IRCS Med. Sci.: Libr. Compend.* **3,** 94.
Than, G. N., Csabo, I. F., Szabo, D. G., Karg, N. J., and Novak, P. F. (1975b). *Arch. Gynaekol.* **218,** 125–130.
Than, G. N., Csabo, I. F., Szabo, D. G., Karg, N. J., and Novak, P. F. (1976). *Vox Sang.* **30,** 134–138.
Thomas, M. L., and Tack, B. F. (1983). *Biochemistry* **22,** 942–947.
Thomas, M. L., Janatova, J., Gray, W. R., and Tack, B. F. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 1054–1058.
Thomas, M. L., Davidson, F. F., and Tack, B. F. (1983). *J. Biol. Chem.* **258,** 13580–13586.
Thomson, A. W., Hunter, C. B. J., Cruickshank, N., and Horne, C. H. W. (1979). *Int. Arch Allergy Appl. Immunol.* **58,** 251–259.
Tilley, C. A., Romans, D. G., and Crookson, M. C. (1978). *Nature (London)* **276,** 713–715.
Todd, R. F., III, Van Agthoven, A., Schlossman, S. F., and Terhorst, C. (1982). *Hybridoma* **1,** 329–337.
Tourbez, M., Dimicoli, J.-L., Pochon, F., and Bieth, J. G. (1984). *Biochim. Biophys. Acta* **789,** 74–79.
Travis, J., and Salvesen, G. S. (1983). *Annu. Rev. Biochem.* **52,** 655–709.
Tsuru, D., Kado, K., Fujiwara, K., Tomimatsu, M., and Ogita, K. (1978). *J. Biochem. (Tokyo)* **83,** 1345–1353.
Tycho, B., DiPaola, M., Yamashiro, D. J., Fluss, S., and Maxfield, F. R. (1983). *Ann. N. Y. Acad. Sci.* **421,** 424–433.
Urbach, G., and Cinader, B. (1966). *Proc. Soc. Exp. Biol. Med.* **122,** 779–782.
Van Gool, J., Ladiges, N. C. J. J., and Boers, W. (1982). *Inflammation* **6,** 127–135.
Van Leuven, F. (1984). *Mol. Cell. Immunol.* **58,** 121–128.
Van Leuven, F., Verbrüggen, R., Cassiman, J.-J., and Van Den Berghe, H. (1977). *Exp. Cell Res.* **109,** 468–471.
Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1978). *Exp. Cell Res.* **117,** 273–282.
Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1979). *J. Biol. Chem.* **254,** 5155–5160.
Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1980). *Cell* **20,** 37–43.
Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1981a). *J. Biol. Chem.* **256,** 9016–9022.
Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1981b). *J. Biol. Chem.* **256,** 9023–9027.
Van Leuven, F., Marynen, P., Cassiman, J.-J., and Van Den Berghe, H. (1981c). *FEBS Lett.* **134,** 83–87.
Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1982a). *Biochem. J.* **201,** 119–128.
Van Leuven, F., Marynen, P., Cassiman, J.-J., and Van Den Berghe, H. (1982). *Biochem. J.* **203,** 405–411.
Van Leuven, F., Cassiman, J.-J., and Van Den Berghe, H. (1985). *Sci. Tools* **32,** 41–43.
Veremeenko, K. N., and Kizim, A. I. (1981). *Thromb. Res.* **23,** 317–322.
Via, D. P., Willingham, M. C., Pastan, I., Gotto, A. M., Jr., and Smith, L. C. (1982). *Exp. Cell Res.* **141,** 15–22.
Virca, G. D., and Travis, J. (1984). *J. Biol. Chem.* **259,** 8870–8874.
Virca, G. D., Salvesen, G. S., and Travis, J. (1983). *Hoppe-Seyler's Z. Physiol. Chem.* **364,** 1297–1302.
Vischer, T. L., and Berger, D. (1980). *J. Reticuloendothel. Soc.* **28,** 427–435.

Vogt, W., Schmidt, G., Van Buttlar, B., and Dieminger, L. (1978). *Immunology* **34,** 29–40.
von Schoultz, B. (1974). *Am. J. Obstet. Gynecol.* **119,** 792–797.
von Schoultz, B., and Stigbrand, T. (1974). *Biochim. Biophys. Acta* **359,** 303–310.
von Schoultz, B., and Stigbrand, T. (1982). *In* "Pregnancy Proteins" (J. G. Grudzinskas, B. Teisner, and M. Seppälä, eds.), pp. 167–175. Academic Press, New York.
von Schoultz, B., Stigbrand, T., and Tärnvik, A. (1973). *FEBS Lett.* **38,** 23–26.
von Zabern, I., and Gigli, I. (1982). *J. Immunol.* **128,** 1439–1442.
von Zabern, I., Nolte, R., and Vogt, W. (1981). *Scand. J. Immunol.* **13,** 413–431.
von Zabern, I., Bloom, E. L., Chu, V., and Gigli, I. (1982). *J. Immunol.* **128,** 1433–1438.
Waites, G. T., and Bell, S. C. (1984). *J. Reprod. Fertil.* **70,** 581–589.
Waller, E. K., Schleuning, W. D., and Reich, E. (1983). *Biochem. J.* **215,** 123–131.
Wang, D., Wu, K., and Feinman, R. D. (1981). *Arch. Biochem. Biophys.* **211,** 500–506.
Wang, D., Wu, K., and Feinman, R. D. (1983). *Arch. Biochem. Biophys.* **222,** 117–122.
Wang, D., Yuan, A. I., and Feinman, R. D. (1984). *Biochemistry* **23,** 2807–2811.
Weigle, W. O., Goodman, M. G., Morgan, E. L., and Hugli, T. E. (1983). *Springer Semin. Immunopathol.* **6,** 173–194.
Weimer, H. E., and Benjamin, D. C. (1965). *Am. J. Physiol.* **209,** 736–744.
Weimer, H. E., Benjamin, D. C., and Darcy, D. A. (1965). *Nature (London)* **208,** 1221–1222.
Weiss, J. J., Tedder, T. F., and Fearon, D. T. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 881–885.
Welinder, K. G., Mikkelsen, L., and Sottrup-Jensen, L. (1984). *J. Biol. Chem.* **259,** 8328–8331.
Werb, Z., Burleigh, M. C., Barrett, A. J., and Starkey, P. M. (1974). *Biochem. J.* **139,** 359–368.
West, C. D., Davis, N. C., Forristal, J., Herbst, J., and Spitzer, R. (1966). *J. Immunol.* **96,** 650–658.
Westergaard, J. G., Bach, A., Teisner, B., Hau, J., and Grudzinskas, J. G. (1982). *J. Reprod. Fertil.* **66,** 695–698.
Weström, B. R. (1979a). *Hoppe-Seyler's Z. Physiol. Chem.* **360,** 1861–1867.
Weström, B. R. (1979b). *Hoppe-Seyler's Z. Physiol. Chem.* **360,** 1869–1878.
Weström, B. R., Karlsson, B. W., and Ohlsson, K. (1983). *Hoppe-Seyler's Z. Physiol. Chem.* **364,** 375–381.
Wetsel, R. A., Ware, C. F., and Kolb, W. P. (1980). *J. Immunol.* **124,** 1545.
Wetsel, R. A., Lundwall, Å., Davidson, F., Gibson, T., Tack, B. F., and Fey, G. H. (1984). *J. Biol. Chem.* **259,** 13857–13862.
Whaley, K. (1980). *J. Exp. Med.* **151,** 501–516.
Whaley, K., and Ruddy, S. (1976). *Science* **193,** 1011–1013.
White, R., Janoff, A., and Godfrey, H. P. (1980). *Lung* **158,** 9–14.
White, R., Habicht, G. S., Godfrey, H. P., Janoff, A., Barton, E., and Fox, C. (1981). *J. Lab. Clin. Med.* **97,** 718–729.
Whitehead, A. S., Sim, R. B., and Bodmer, W. F. (1981). *Eur. J. Immunol.* **11,** 140–146.
Whitehead, A. S., Solomon, E., Chambers, S., Bodmer, W. F., Povey, S., and Fey, G. H. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 5021–5025.
Whitehead, A. S., Goldberger, G., Woods, D. E., Markham, A. F., and Colten, H. R. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5387–5391.
Whitehead, A. S., Woods, D. E., Fleischmek, E., Chin, J. E., Yunis, E. J., Katz, A. J., Gerald, P. S., Alper, C. A., and Colten, H. R. (1984). *N. Engl. J. Med.* **310,** 88–91.
Wiebauer, K., Domdey, H., Diggelman, H., and Fey, G. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 7077–7081.
Wieme, R. J., and DeMulenaare, B. (1967). *Nature (London)* **214,** 1042–1043.
Willingham, M. C., Maxfield, F. R., and Pastan, I. H. (1979). *J. Cell Biol.* **82,** 614–625.
Willoughby, W. F., and Mayer, M. M. (1965). *Science* **150,** 907–908.
Wong, W. W., Wilson, J., and Fearon, D. (1983). *J. Clin. Invest.* **72,** 685–693.

Wong, W. W., Klickstein, L. B., Smith, J. A., Weis, J. H., and Fearon, D. T. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 7711–7715.

Wright, S. D., Rao, P. E., Van Voorhis, W. C., Craignyle, L. S., Iida, K., Talle, M. H., Westberg, E. F., Goldstein, G., and Silverstein, S. C. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5699–5703.

Wu, K., Wang, D., and Feinman, R. D. (1981). *J. Biol. Chem.* **256,** 10409–10414.

Yamamoto, K.-C., and Gewurz, H. (1978). *J. Immunol.* **120,** 2008–2014.

Yamamoto, K.-C., and Gewurz, H. (1980). *J. Immunol.* **125,** 1745–1750.

Yamamoto, K., Tsujino, Y., Saito, A., and Sinohara, H. (1985). *Biochem. Int.* **10,** 463–469.

Zais, D. P., and Roberts, R. C. (1977). *Clin. Chem. (Winston-Salem, N.C.)* **23,** 590–592.

Zardi, L., Carnemolla, B., Gagnasso, D., and Santi, L. (1980). *Eur. J. Cancer* **16,** 35–42.

5 Synthesis, Processing, and Secretion of Plasma Proteins by the Liver and Other Organs and Their Regulation

Gerhard Schreiber

The Russell Grimwade School of Biochemistry
University of Melbourne
Parkville, Victoria 3052, Australia

The Plasma Proteins, Vol. V

I. Introduction: Coordinate Regulation of Plasma Protein Synthesis and Mechanism of Synthesis and Secretion

Plasma proteins are best defined operationally. They are proteins which can be isolated in relatively large amounts from blood plasma. However, their occurrence in the body is not confined to the vascular compartment. To the contrary, for most plasma proteins the total extravascular amount exceeds the intravascular pool. With the exception of the immunoglobulins (Chapter 2) and some apolipoproteins (Chapter 3), the plasma proteins found in the blood plasma are synthesized in the liver (Miller *et al.*, 1951; Miller and Bale, 1954; Schreiber *et al.*, 1966; for review, see Schreiber and Urban, 1978). A single liver cell can synthesize several different plasma proteins simultaneously, i.e., individual hepatocytes are not specialized in the synthesis of one plasma protein species (Foucrier *et al.*, 1979; Courtoy *et al.*, 1981; Kraemer *et al.*, 1981). However, the liver is not the exclusive site of synthesis for plasma proteins. Quantitatively, the liver/blood plasma system is the most important system in the body involving plasma proteins, but similar logistic functional needs, i.e., providing and maintaining extracellular environments of appropriate composition, exist in other areas of the body. Examples are the choroid plexus/cerebrospinal fluid system, participating in establishing an appropriate extracellular environment in the central nervous system, and the yolk sac membranes/yolk sac/fetal system, mediating in chemical communication and transport between mother and fetus. It is remarkable that in so distant and, superficially regarded, apparently unrelated tissues such as liver, choroid plexus, and yolk sac, similar plasma protein genes are expressed in response to similar logistic functional needs.

The isolation, physicochemical properties, and various metabolic aspects of plasma proteins have been described in several excellent monographs (Allison, 1974, 1976; Putnam, 1975a,b,c; Hitzig, 1977; Blombäck and Hanson, 1979; Glaumann *et al.*, 1983). The emphasis in this chapter will be on the regulation of the rates of synthesis of plasma proteins and the link between function and

regulation. A short recapitulation of important data on structure and function of proteins will be included, where appropriate.

Because of the ease of isolation, abundance and accessibility of material, and the high rates of synthesis, plasma proteins have for many years been an attractive system for the study of protein synthesis in mammalian systems. No special features were found in the mechanism of synthesis and secretion which would distinguish the production of plasma proteins in the liver from that of other proteins to be exported from cells. Plasma proteins are synthesized by polyribosomes bound to the rough endoplasmic reticulum. Plasma proteins analyzed until now (cf., e.g., Table I) possess a presegment rich in hydrophobic amino acids. An "internal" presegment, similar to that described for ovalbumin (Lingappa *et al.*, 1979; Braell and Lodish, 1982; Meek *et al.*, 1982), has not been observed yet for plasma proteins. The data summarized in Table I show alanine, or, in three cases, an amino acid of very similar structure (glycine), as the carboxy-terminal amino acid of the presegment in 18 out of the 22 cases listed (indicated by a box in Table I). Only for four other polypeptide chains, albumins and the α and β chains of fibrinogen, is an amino acid carrying a hydroxyl group found at the carboxy terminus of the presegment. Perhaps the specificity of the signal peptidase is somewhat limited or, possibly, two different signal peptidases might be involved in the processing of plasma proteins, one having the typical elastaselike specificity for alanine and the other reacting with hydroxyl amino acids. The occurrence of several distinct signal peptidases was reported recently for *Escherichia coli* (Tokunaga *et al.*, 1984). From the rough endoplasmic reticulum the proteins to be secreted are transported in vesicles to the Golgi apparatus and from there in so-called secretory vesicles to the cell membrane. During their transit from rough endoplasmic reticulum to the cell membrane precursor plasma proteins are modified both during and after translation. The presegment is removed early and, therefore, plasma protein precursors still containing the presegment can be obtained usually only by translation of mRNA in cell-free protein synthesizing systems which do not contain processing proteinases (e.g., for albumin, see Strauss *et al.*, 1977a,b; Yu and Redman, 1977). However, later intermediates in the intracellular maturation of plasma proteins have been isolated from liver. These intermediates contained a short prosegment at the N-terminus (Urban *et al.*, 1974, 1976; Russell and Geller, 1975; Edwards *et al.*, 1976a,b,c; Rosen and Geller, 1977; Millership *et al.*, 1980) or, in the case of glycoproteins, a precursor carbohydrate moiety (Schreiber *et al.*, 1979, 1981; Nagashima *et al.*, 1980, 1981; Urban *et al.*, 1982a,b). Differences in secretion kinetics and intracellular pool sizes suggest that plasma protein precursors are transported through the liver cells individually or, perhaps, in groups of related proteins (Morgan and Peters, 1971b; Schreiber, 1979; Schreiber *et al.*, 1979; Morgan, 1983; Ledford and Davis, 1983; Lodish *et al.*, 1983; Fries *et al.*, 1984). In the case of transferrin, there might even be separate intracellular handling of

TABLE I

Amino Acid Sequences of the Presegments of Plasma Proteins around the Recognition Site for the Signal Peptidase[a]

Plasma protein	Source	Amino acid position (1 = N-terminal amino acid in the mature protein)						Reference
		−3	−2	−1	1	2	3	
Transthyretin	Rat	SER	GLU	ALA	GLY	PRO	GLY	(a)
α_1-Acid glycoprotein	Rat	LEU	GLU	ALA	GLN	ASN	PRO	(b)
α_1-Antitrypsin	Baboon	SER	LEU	ALA	GLU	ASP	PRO	(c)
Retinol-binding protein	Human	ALA	ALA	ALA	GLU	ARG	ASP	(d)
α_{2u}-Globulin	Rat	GLY	HIS	ALA	GLU	GLU	ALA	(e)
Transferrin	Rat	CYS	LEU	ALA	VAL	PRO	ASP	(f)
β_2-Microglobulin	Human	LEU	GLU	ALA	ILE	GLN	ARG	(g)
Fibrinogen γ chain	Rat	GLY	LEU	ALA	GLN	TYR	THR	(h)
Fibrinogen γ chain	Human	CYS	VAL	ALA	TYR	VAL	ALA	(i)
Apolipoprotein E	Rat	CYS	LEU	ALA	GLU	GLY	GLU	(j)
Apolipoprotein E	Human	CYS	GLN	ALA	LYS	VAL	GLU	(k)
Apolipoprotein A-I	Rat	CYS	GLN	ALA	TRP	GLU	PHE	(l)
Apolipoprotein A-I	Human	SER	GLN	ALA	ARG[v]	HIS	PHE	(m)
Apolipoprotein C-III	Human	ALA	ARG	ALA	SER	GLY	ALA	(n)
Apolipoprotein A-IV	Rat	THR	GLN	ALA	GLU	VAL	THR	(o)
Apolipoprotein C-I	Human	ALA	GLN	GLY	THR	PRO	ASP	(p)
Apolipoprotein C-II	Human	VAL	GLN	GLY	THR	GLN	GLN	(n,q)
Apolipoprotein A-II	Human	LEU	GLU	GLY	ALA	LEU	VAL	(n)
Fibrinogen α chain	Human	ALA	TRP	THR	ALA	ASP	SER	(r)
Fibrinogen β chain	Human	VAL	LYS	SER	GLN	GLY	VAL	(s)
Albumin	Rat	ALA	PHE	SER	ARG[v]	GLY	VAL	(t)
Albumin	Human	SER	ALA	TYR	SER[v]	ARG	GLY	(u)

[a]From Dickson (1985), modified with permission.
[b]Ricca and Taylor (1981).
[c]Kurachi *et al.* (1981).
[d]Colantuoni *et al.* (1983).
[e]Unterman *et al.* (1981).
[f]Aldred *et al.* (1984).
[g]Suggs *et al.* (1981).
[h]Crabtree and Kant (1982a).
[i]Chung *et al.* (1983b).
[j]McLean *et al.* (1983).
[k]Zannis *et al.* (1984).
[l]Poncin *et al.* (1984).
[m]Cheung and Chan (1983).
[n]Sharpe *et al.* (1984).
[o]Gordon *et al.* (1982); Boguski *et al.* (1984).
[p]Knott *et al.* (1984).
[q]Myklebost *et al.* (1984).
[r]Rixon *et al.* (1983); Kant *et al.* (1983).
[s]Chung *et al.* (1983a).
[t]Sargent *et al.* (1981).
[u]Dugaiczyk *et al.* (1982).
[v]First amino acid of the prosegment.

the same protein dependent on whether it is taken up from outside to import iron or whether it is to be exported from the cell. Various aspects of the secretion of plasma proteins by the liver have recently been summarized by several authors (see Glaumann *et al.*, 1983).

In healthy individuals, the concentration of most plasma proteins in the blood is kept within a relatively constant range. This seems to indicate the existence of a regulatory system, with the particular protein, or a derivative thereof, as the signal and the synthesizing or degrading system for the protein as the means effecting regulation of the concentration. In early work on the regulation of plasma protein synthesis in the liver, the system was perturbed by lowering the concentration of the signal compound, in most cases albumin. This was achieved by bleeding or plasmapheresis, i.e., replacement of a portion of blood by a suspension of erythrocytes in physiological saline or by a solution of dextran. An increase in the rate of synthesis of albumin in the liver was observed in such experiments by a number of authors (Wasserman *et al.*, 1956; Rothschild *et al.*, 1961, 1969a; Moore, 1965; Tracht *et al.*, 1967). Alterations in the rate of synthesis of albumin in the liver were also found when the level of albumin in plasma was decreased by inducing nephritis either by injecting an anti-rat kidney antiserum from rabbits immunized with rat kidney homogenate or by injecting the aminonucleoside of puromycin (Drabkin and Marsh, 1955; Marsh and Drabkin, 1958, 1960; Braun *et al.*, 1962a,b; Marsh *et al.*, 1966; Katz *et al.*, 1967, 1968). The plasma level of albumin decreases also after partial hepatectomy (Chanutin *et al.*, 1938; Bengmark *et al.*, 1968). An increase in the rate of incorporation of radioactive amino acid into albumin after partial hepatectomy has been observed (Guidotti *et al.*, 1959; Majumdar *et al.*, 1967). However, using methods correcting for changes in precursor amino acid pool sizes, no significant changes in the rate of synthesis of albumin per gram liver tissue were found after partial hepatectomy (Mutschler and Gordon, 1966; Schreiber *et al.*, 1971).

Albumin is by far the most abundant among the proteins in blood plasma. However, no single specific function is associated with albumin. It serves as a carrier for various important substances, such as fatty acids, bilirubin, and many drugs. Albumin is the main contributor to the maintenance of oncotic pressure, but its complete absence from the blood due to a genetic deficiency is tolerated in both humans (Bennhold *et al.*, 1954; Boman *et al.*, 1976; Dammacco *et al.*, 1980) and rats (Nagase *et al.*, 1979) without affecting health. A function as a "metabolic adapter," whose rate of synthesis can be turned down when rates of synthesis increase strongly for other proteins, thus keeping the overall rate of synthesis of protein constant, has been suggested for the synthesis of albumin in liver (Schreiber *et al.*, 1982; Dickson *et al.*, 1982; Schreiber and Howlett, 1983).

Both age of animals and protein content of their diets influence the rate of albumin synthesis. For adult humans with a daily intake of 70 gm protein, Hoffenberg *et al.* (1966) determined a rate of albumin synthesis of 151 mg/-

day/kg body weight (corresponding to 7.1 mg/day/gm liver weight) and Kelman *et al.* (1972a) measured a rate of synthesis of albumin of 245 ± 98 mg/day/kg body weight (corresponding to 11 mg/day/gm liver weight). Adult rats synthesized 8.4 mg albumin/day/gm liver weight (Peters and Peters, 1972). Fasting or protein depletion in the diet reduced the rate of albumin synthesis (Hoffenberg *et al.*, 1966; Kirsch *et al.*, 1968; Haider and Tarver, 1969; Morgan and Peters, 1971a; Kelman *et al.*, 1972a; Peters and Peters, 1972). For a diet containing 64% protein ("high-protein diet") both a reduction (Peters and Peters, 1972) and an increase in albumin synthesis (Haider and Tarver, 1969) have been found for rats. The rate of albumin synthesis responded also to changes in amino acid supply in investigations with the isolated perfused liver (Gordon, 1966; Rothschild *et al.*, 1968, 1969a; John and Miller, 1969; Kirsch *et al.*, 1969; Hoffenberg *et al.*, 1971; Kelman *et al.*, 1972b). Conditions affecting the rate of albumin synthesis have been discussed elsewhere by numerous authors (Waterlow, 1969; Peters, 1970; Rothschild *et al.*, 1972a,b, 1973, 1975; Schreiber and Urban, 1978).

Maintaining an appropriate "internal milieu" in the body is the function common to virtually all plasma proteins. Perturbing the "internal milieu" might therefore be a powerful challenge of the regulatory systems involved in maintaining the normal composition of the blood plasma and of tissue fluid. During conditions associated with loss of tissue fluids (wounds and burns) or damage to tissue with ensuing degradation of damaged tissue (inflammation) the amount and composition of extracellular fluid are affected. A series of physiological and biochemical reactions, summarized as the general acute-phase response to trauma and inflammation, is initiated in the defense mechanism of the body. A typical feature in this defense mechanism is a rearrangement of the pattern of concentration of proteins in blood plasma. Inducing an acute experimental inflammation is therefore a very effective means of challenging the regulatory mechanisms involved in maintaining appropriate plasma protein concentrations. This method has been used more frequently to investigate the regulation of plasma protein synthesis in various studies in recent years (see Koj, 1974; Kushner *et al.*, 1981; Schreiber and Howlett, 1983).

II. Response of the Plasma Protein Synthesizing System in the Liver to Trauma and Inflammation

The response of the body to trauma and inflammation in higher animals is characterized by a typical sequence of events. Locally, vasodilation (after an early transient vasoconstriction of arterioles for a few seconds) and an increased permeability of the microvascular system produces three typical signs of inflammation: rubor, tumor, and calor (redness, swelling, and heat) (Celsus, 1713; see

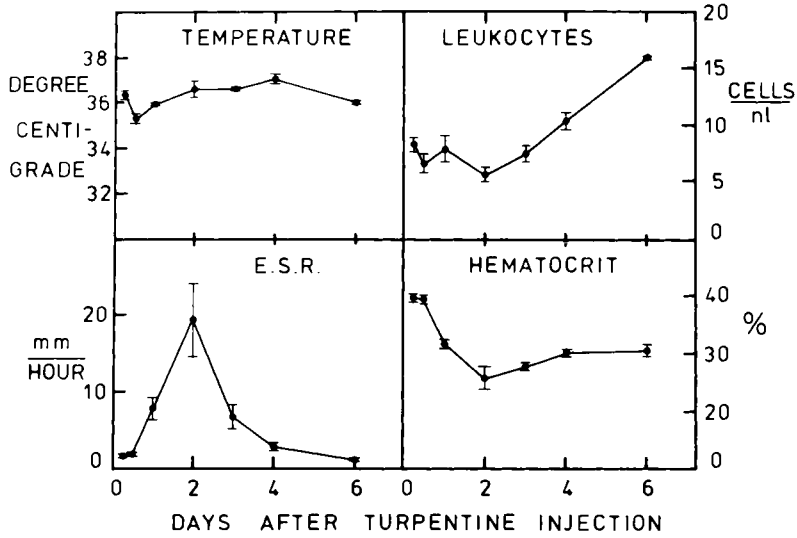


Fig. 1. General parameters of an acute inflammation, produced by subcutaneous injection of 0.5 ml turpentine per 100 gm body weight into the back of male Buffalo rats, weighing between 250 and 300 gm. From Schreiber and Howlett (1983), with permission.

also Virchow, 1860; Büchner, 1956; Robbins and Cotran, 1979). In the general response to acute trauma and inflammation an early phase is characterized by increases in body temperature and in the sedimentation rate of erythrocytes. This is followed by a later phase characterized by an increase in the concentration of leukocytes in the blood. In rats the change in the temperature of the body is small compared with the more pronounced rises of temperature seen in humans and rabbits. The general parameters of an acute inflammation produced in rats by subcutaneous injection of turpentine are summarized in Fig. 1.

The cause for the increase in the sedimentation rate of erythrocytes during acute inflammation is a change in the concentrations of plasma proteins, with an increase in the level of fibrinogen being the most important contribution. The kinetics of the change in concentration of individual plasma proteins in plasma from rats suffering from an acute experimental inflammation is summarized in Figs. 2 and 3.

A. Physiology of the General Response to Trauma and Acute Inflammation

The destruction of an area of tissue and the following proteolytic removal of the damaged tissue create various logistic problems for the "healthy parts" of

the body. When proteinases are released from lysosomes and phagocytic cells to catalyze the breakdown of protein in the area affected by trauma (Melloni *et al.*, 1981; Gordon and Cross, 1981), the healthy regions of the organism and, in particular, the information and transport system based on proteins in the bloodstream have to be protected from inappropriate proteolysis. Thus, it is not surprising that many of the acute-phase proteins are proteinase inhibitors or help to establish a demarcation border between damaged and intact tissue. The acute-

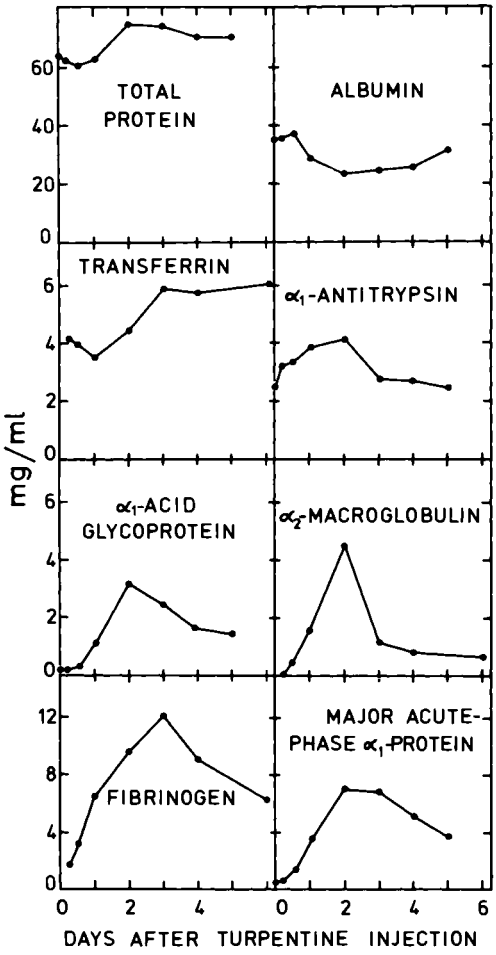


Fig. 2. Concentrations of various proteins in the serum (total protein, albumin, α_1-antitrypsin, α_1-acid glycoprotein, major acute-phase α_1-protein) or plasma (transferrin, α_2-macroglobulin, fibrinogen) of rats suffering for various periods of time from an acute inflammation produced as described in the legend to Fig. 1. From Schreiber and Howlett (1983), with permission.

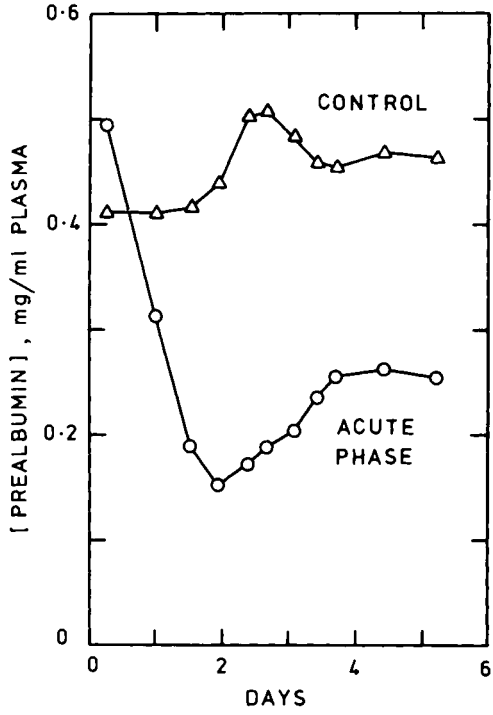


Fig. 3. Concentration of prealbumin (transthyretin) in the plasma of healthy rats and rats suffering for various periods of time from an acute inflammation. The inflammation was induced by turpentine injection as described in the legend of Fig. 1. From Dickson *et al.* (1982), with permission.

phase proteins are summarized in Table II according to their function (or, in some cases, hypothetical function).

One acute-phase proteinase inhibitor might have an interesting dual function. The major acute-phase α_1-protein of the rat, also called thiostatin (Cole *et al.*, 1985a), not only inhibits proteinases with cysteine in their active center (Esnard and Gauthier, 1983), such as cathepsins or papain, but also contains the sequences for T-kinin (Okamoto and Greenbaum, 1983a,b; Greenbaum, 1984) and bradykinin in its polypeptide chain (Cole *et al.*, 1985a,b). Possibly, it is the substrate for the proteolytic generation of vasoactive oligopeptides (Okamoto and Greenbaum, 1983c; Barlas *et al.*, 1985) which are then involved in the control of the blood flow in and around the area of tissue damage (Haddy *et al.*, 1970), of the vascular permeability (Kline *et al.*, 1973), and in the generation of pain (see Armstrong, 1970; Lim, 1970). Pain, "dolor," is the fourth of the typical signs of inflammation as summarized by Celsus (1713).

TABLE II

Acute-Phase Proteins[a]

Protein	Concentration in plasma (gm/liter)	Factor by which the plasma concentrations change during acute inflammation
Positive Acute-Phase Proteins		
1. Proteinase inhibitors		
α_1-Antitrypsin[b]	2–4	2–4
α_1-Antichymotrypsin[c]	0.3–0.6	2–5
α_2-Macroglobulin (rat)[d]	0.014 (rat)	>100 (rat)
Major acute-phase α_1-protein of the rat (= thiostatin = T-kininogen)[e]	0.5 (rat)	20 (rat)
2. Proteins with transport or binding function		
Hemopexin[f]	0.5–1.15	2 (human); up to 10 (rat)
Haptoglobin[g]	1–2.2	2–4
Ceruloplasmin[h]	0.15–0.6	<2
Transferrin[i]	2–4	1.5[y]
3. Proteins involved in "structural protection"		
Fibrinogen[j]	2–4.5	2–4 (human); 7 (rat)
4. Proteins influencing the activity of other cells		
Complement system[k] in particular component C3	0.55–1.2	<2
α_1-Acid glycoprotein[l]	0.55–1.4	2–4 (human); 20 (rat)
C-reactive protein[m]	0.8 (mg/liter)	Up to 3000 (human); 2–4 (rat)
Serum amyloid A-related protein[n] (human, mice)	0.1 (mg/liter)	>100
Serum amyloid P-component-related protein[o] (mice)	10–60 (mg/liter)	3–25
5. Intracellular acute-phase proteins		
Metallothionein[p] (rodent, human)		2.6[z]
Actin[q] (mouse)		5[aa]
Negative Acute-Phase Proteins		
Albumin[r]	35–55	0.6–0.7
Transthyretin[s]	0.1–0.4	0.3–0.4
Retinol-binding protein[t]	0.03–0.06	
α_2-HS glycoprotein[u]	0.4–0.85	
Transcortin (rat)[v]	0.03–0.036	
α_{2u}-Globulin (rat)[w]		0.25[z]
Apolipoprotein A-IV (rat)[x]		0.25–0.3[z]

[a]In human, unless indicated otherwise.

[b]Heimburger *et al.* (1964); Heide and Schwick (1973); Clamp (1975); Laurell and Jeppsson (1975); baboon: Kurachi *et al.* (1981); rat: Urban *et al.* (1982b); Roll and Glew (1981); human: Kurachi *et al.* (1982).

TABLE II (*Continued*)

[c]Heimburger and Haupt (1965); Clamp (1975); Laurell and Jeppsson (1975).
[d]Gauthier and Mouray (1976); Gordon (1976); Nieuwenhuizen *et al.* (1979); Okubo *et al.* (1981); Northemann *et al.* (1985); Schreiber *et al.* (1986).
[e]Darcy (1964, 1966); Gordon and Louis (1969); Jeejeebhoy *et al.* (1977); John and Miller (1969); Urban *et al.* (1979, 1982a).
[f]Putnam (1975b).
[g]Gerbeck *et al.* (1967); Black *et al.* (1970); Yang *et al.* (1983).
[h]Jamieson (1965); Rydén (1972); Poulik and Weiss (1975); Kingston *et al.* (1977); Cooper and Ward (1979); Frieden (1979); Noyer *et al.* (1980); Dwulet and Putnam (1981a,b).
[i]Chasteen (1977); Aisen and Brown (1977); Aisen (1981); Schreiber *et al.* (1979, 1982); Morgan (1981, 1983); MacGillivray *et al.* (1982).
[j]Mester and Szabados (1968); Doolittle (1975).
[k]Müller-Eberhard and Nilsson (1960); Müller-Eberhard *et al.* (1960); Alper (1974); Bokisch *et al.* (1975); Edwards *et al.* (1977); Hodgson *et al.* (1977); Laurell (1979); Müller-Eberhard (1975a).
[l]Schmid *et al.* (1973).
[m]Oliveira *et al.* (1979); Kushner *et al.* (1981).
[n]Levin *et al.* (1973); Rosenthal and Franklin (1975); Rosenthal *et al.* (1976); Sipe *et al.* (1976); Anders *et al.* (1977); Bausserman *et al.* (1980); Kushner *et al.* (1981); Gorevic and Franklin (1981).
[o]Pepys *et al.* (1978, 1979a); Kushner *et al.* (1981).
[p]Mouse: Durnam *et al.* (1980); human: Karin and Richards (1982); rat: Mercer and Hudson (1982); Schreiber *et al.* (1986).
[q]Morrow *et al.* (1981).
[r]Peters (1975).
[s]Kanda *et al.* (1974); Mita *et al.* (1984); Dickson (1985); Dickson *et al.* (1985b); Sundelin *et al.* (1985b).
[t]Peterson (1971); Moody (1982); Sundelin *et al.* (1985a).
[u]Schultze *et al.* (1962); Lebreton *et al.* (1979).
[v]Chader and Westphal (1968).
[w]Dickson (1985).
[x]Tu *et al.* (1985).
[y]Increase is delayed by two to three days compared with other acute-phase proteins.
[z]mRNA level in liver compared with that in the liver of healthy rats.
[aa]Change in rate of protein synthesis.

Other acute-phase proteins have a salvaging or scavenging function. Thus, hemopexin and haptoglobin can bind heme or hemoglobin released in the area of destroyed tissue.

Another plasma protein involved in keeping a very "precious" compound, iron, within the bounds of the body is transferrin. The increase observed for the concentration of transferrin in the bloodstream during the acute-phase response is delayed compared with that of other acute-phase proteins. The concentration of transferrin in blood plasma starts to rise only when other acute-phase plasma proteins have already reached their maximum in the change of concentration. The change in the concentration of transferrin in plasma during the acute-phase response might be a secondary phenomenon brought about by the decrease in plasma iron concentration during inflammation. Infections, tumors, and turpentine abscesses cause a decreased release of iron from stores in the body, in particular from the cells of the reticuloendothelial system, resulting in low serum

iron levels (see Heilmeyer and Begemann, 1955; Tietz, 1976). The injection of partially purified leukocytic endogenous mediator (see Section II,D,1) leads also to a lowering of the concentration of iron and zinc ions in plasma as well as to the production of fever (Kampschmidt *et al.*, 1973b). It is well known that the transferrin concentration in plasma increases in iron-deficiency anemia (Laurell, 1947). This increase in the concentration of transferrin in the blood plasma is due to an increase in the rate of synthesis of transferrin (Awai and Brown, 1963; Masuya and Kozuru, 1963; Cromwell, 1964; Morton *et al.*, 1976). A nutritional iron deficiency also increases the rate of expression of the transferrin gene (McKnight *et al.*, 1980a,b).

The function of some acute-phase proteins is not yet completely understood. An example in this group is α_1-acid glycoprotein. It is the plasma protein with the highest content of carbohydrate. This large carbohydrate portion might be involved in the mechanism of its function. α_1-Acid glycoprotein has been reported to influence the activity of various other cells in the bloodstream (see Section II,B,4,b).

In contrast to the well-characterized antiproteolytic or scavenging action of some acute-phase proteins, the interpretation of the functional significance of the decrease in plasma concentration, observed for several other plasma proteins, is largely based on speculation. One explanation offered for the understanding of the changes in synthesis rates of these so-called negative acute-phase proteins is that of metabolic adaptation (Schreiber *et al.*, 1982; Dickson *et al.*, 1982; Schreiber and Howlett, 1983). A decrease in the demand for aminoacyl-tRNA and ATP, caused by the decrease in synthesis rates of negative acute-phase proteins, can compensate for the increase in the demand caused by the acceleration in the rates of synthesis of positive acute-phase proteins. Albumin, with its high rate of synthesis, large body pool, slow turnover, and without a function absolutely indispensable for healthy survival (Bennhold *et al.*, 1954; Boman *et al.*, 1976; Dammacco *et al.*, 1980), would be particularly appropriate for such a role of a metabolic adapter. A more detailed quantitative discussion of this aspect can be found elsewhere (Schreiber and Howlett, 1983).

B. Positive Acute-Phase Proteins

1. Proteinase Inhibitors

A. α_1-ANTITRYPSIN. α_1-Antitrypsin consists of one polypeptide chain, has a molecular weight of 54,000, and contains 12.4% carbohydrate (data for the protein in humans). α_1-Antitrypsin, ovalbumin, angiotensinogen, and antithrombin III form a superfamily of proteins related by similarities in amino acid sequence (Carrell *et al.*, 1979, 1980; Hunt and Dayhoff, 1980; Kurachi *et al.*,

1982; Leicht *et al.*, 1982; Doolittle, 1983, 1984; Tanaka *et al.*, 1984). The untranslated 5′-flanking regions of the genes for α_1-antitrypsin, α_1-acid glycoprotein, and haptoglobin share a common consensus sequence which could be of regulatory significance (see Section II,B,4,b). α_1-Antitrypsin concentrations increase only very moderately after induction of an acute inflammation. Genetic variants are of clinical significance (see Section VI). α_1-Antitrypsin has a fairly broad specificity concerning the inhibition of proteinases (Heimburger, 1972). Trypsin, chymotrypsin, plasmin, and pancreatic kallikrein are all inhibited by α_1-antitrypsin. Of particular importance is the inhibitory effect of α_1-antitrypsin for elastase.

Complementary cDNA clones from baboons (Kurachi *et al.*, 1981) and human genomic DNA clones (Kurachi *et al.*, 1982) for α_1-antitrypsin have been isolated. The human α_1-antitrypsin gene was about 5 kb in size and contained three intervening sequences in the 3′ end of the coding region. Recombinant human α_1-antitrypsin cDNA was expressed in both *Escherichia coli* (Courtney *et al.*, 1984) and in the yeast *Saccharomyces cerevisiae* (Cabezón *et al.*, 1984). This led to the development of α_1-antitrypsin cDNA variants coding for α_1-antiproteinase inhibitors which, in addition, were oxidant-resistant (Rosenberg *et al.*, 1984; Travis *et al.*, 1985; Courtney *et al.*, 1985), which could be important, possibly, for applications in the treatment of emphysema and thrombosis.

A very interesting observation is the recent discovery of a regulatory mutation for the expression of the α_1-antitrypsin gene in wild mice (Berger and Baumann, 1985) causing abundant expression of α_1-antitrypsin in the kidney. Normally, the α_1-antitrypsin gene is expressed only in the liver.

Several recent reviews on the structure, function, and clinical importance of α_1-antitrypsin are available (Jeppsson, 1978; Travis and Salvesen, 1983; Janoff, 1985).

B. α_1-ANTICHYMOTRYPSIN. Antichymotrypsin also consists of one polypeptide chain. Human antichymotrypsin has a molecular weight of 68,000 and a carbohydrate content of 26.8%. Its levels in the plasma increase up to about fivefold after injury and remain elevated for a protracted period (Daniels *et al.*, 1974). Antichymotrypsin inhibits chymotrypsin (Heimburger, 1972), cathepsin C, and chymase (for review, see Travis and Salvesen, 1983).

C. α_2-MACROGLOBULIN. α_2-Macroglobulin is a tetramer with a molecular weight of 760,000 and contains 8.2% carbohydrate. In the rat its plasma concentration increases 320-fold during acute experimental inflammation (Schreiber *et al.*, 1982). The increase of α_2-macroglobulin levels in rat plasma during the acute-phase response depends on the presence of corticosteroid hormones (Weimer and Benjamin, 1965). In humans, α_2-macroglobulin plasma concentrations do not increase during acute tissue damage and infections (Schumacher and

Schlumberger, 1963), but increases of the plasma level are seen during pregnancy, administration of estrogens, and in patients suffering from nephrotic syndrome (Ganrot and Bjerre, 1967; Housley, 1968; Horne *et al.*, 1971). α_2-Macroglobulin is an unspecific inhibitor of proteinases. The inhibited proteinases are transiently bound to a specific section of the polypeptide chain of the inhibitor, the so-called "bait" region (Mortensen *et al.*, 1981; see also Chapter 4 of this volume).

The implications of the fact that α_2-macroglobulin is an acute-phase protein only in the rat are not clear. Certain parasites can release proteinases which cleave immunoglobulins produced by the host, thus preventing the buildup of an immunological defense. If α_2-macroglobulin inhibits such parasite-produced proteinases it might improve the resistance of rats to parasite infections. The cDNA for rat α_2-macroglobulin has been cloned and a section of the nucleotide sequence has been determined (Northemann *et al.*, 1985).

D. MAJOR ACUTE-PHASE α_1-PROTEIN OF THE RAT (= THIOSTATIN = T-KININOGEN). The major acute-phase α_1-protein of the rat consists of one polypeptide chain and has a molecular weight of 56,000 and a carbohydrate content of 19%. An increase in the concentration of an α_1-globulin in plasma during the acute-phase response of the rat was first described by Darcy (1957, 1964, 1966), Gordon and Darcy (1967), and Gordon and Louis (1969). Also the protein described by Jeejeebhoy *et al.* (1977) and the α_1-acid glycoprotein of John and Miller (1969) are probably identical with the major acute-phase α_1-protein. The protein was analyzed in detail by Urban *et al.* (1979). It is synthesized in the liver via a precursor protein (Urban *et al.*, 1982a). It inhibits cysteine proteinases released from lysosomes during inflammation (Esnard and Gauthier, 1983), such as cathepsins. Not surprisingly, it also inhibits papain, whose amino acid sequence is partially homologous to that of cathepsin H and B (Takio *et al.*, 1983). Therefore, the name "thiostatin" has been suggested for the protein to indicate the inhibitory effect on cysteine proteinases (Cole *et al.*, 1985a; Esnard, 1985). A similar proteinase inhibitor is also found in human serum (Pagano *et al.*, 1984b).

Molecular cloning and the amino acid sequence derived from the nucleotide sequence analysis (Fig. 4A) of the cDNA for major acute-phase α_1-protein of the rat showed a homology of 70% with bovine and human kininogens (Fig. 4B). In particular, major acute-phase α_1-protein contained the sequence for T-bradykinin, i.e., Ile-Ser-Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg, between amino acid positions 358 and 368 and the sequence for bradykinin between positions 360 and 368 (box in Fig. 4A and longest box in second to last row of Fig. 4B). These observations and the data of Okamoto and Greenbaum (1983c) and Barlas *et al.* (1985) suggest that, in addition to the inhibitory effect on cysteine proteinases, major acute-phase α_1-protein can be a substrate for the proteolytic

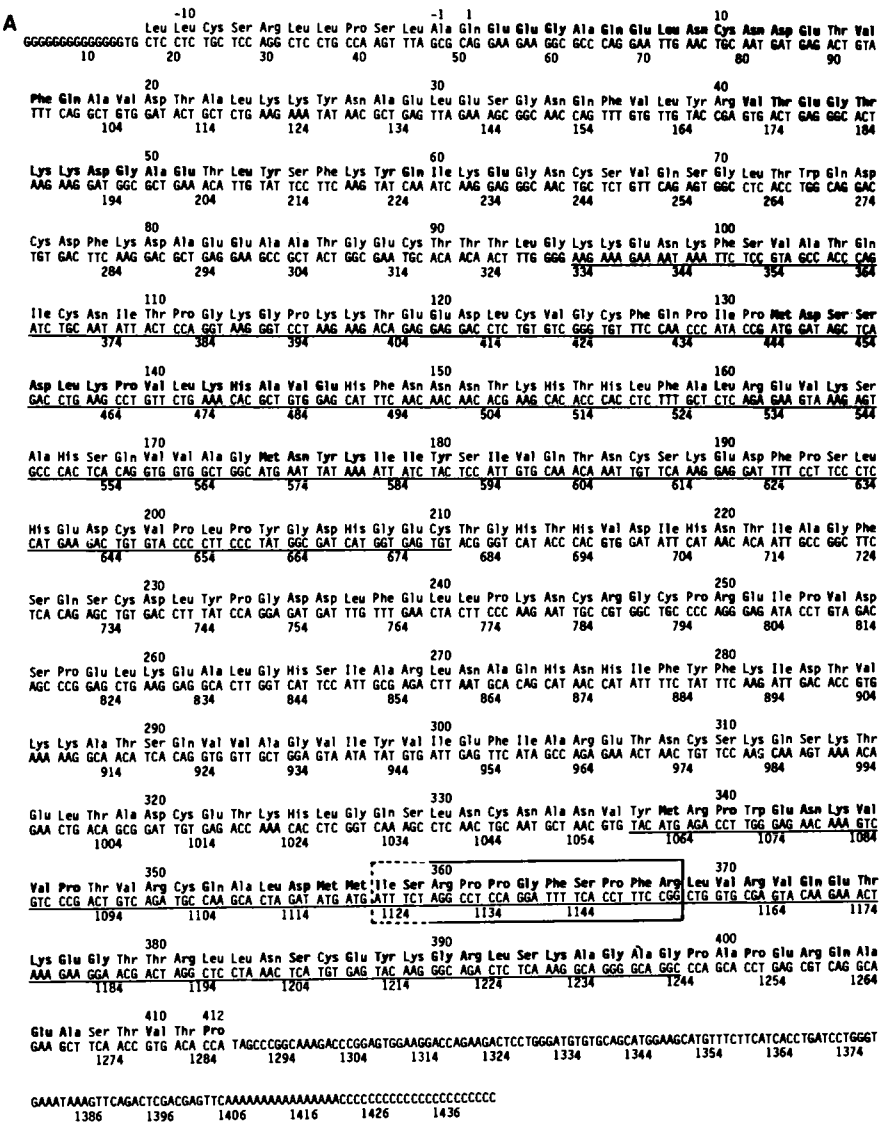


Fig. 4.A. Nucleotide sequence of major acute-phase α_1-protein cDNA and amino acid sequence of major acute-phase α_1-protein. The nucleotide sequence derived from independent sequencing of both strands is indicated by underlining. Amino acid sequences derived from both nucleotide sequencing and direct analysis of major acute-phase α_1-protein fragments are indicated by bold letters. Numbering of amino acids begins at the N-terminus of the mature protein. The sequence of bradykinin is indicated by a solid box and that of T-kinin by a dotted extension of the box. Adapted from Cole *et al.* (1985b), with permission.

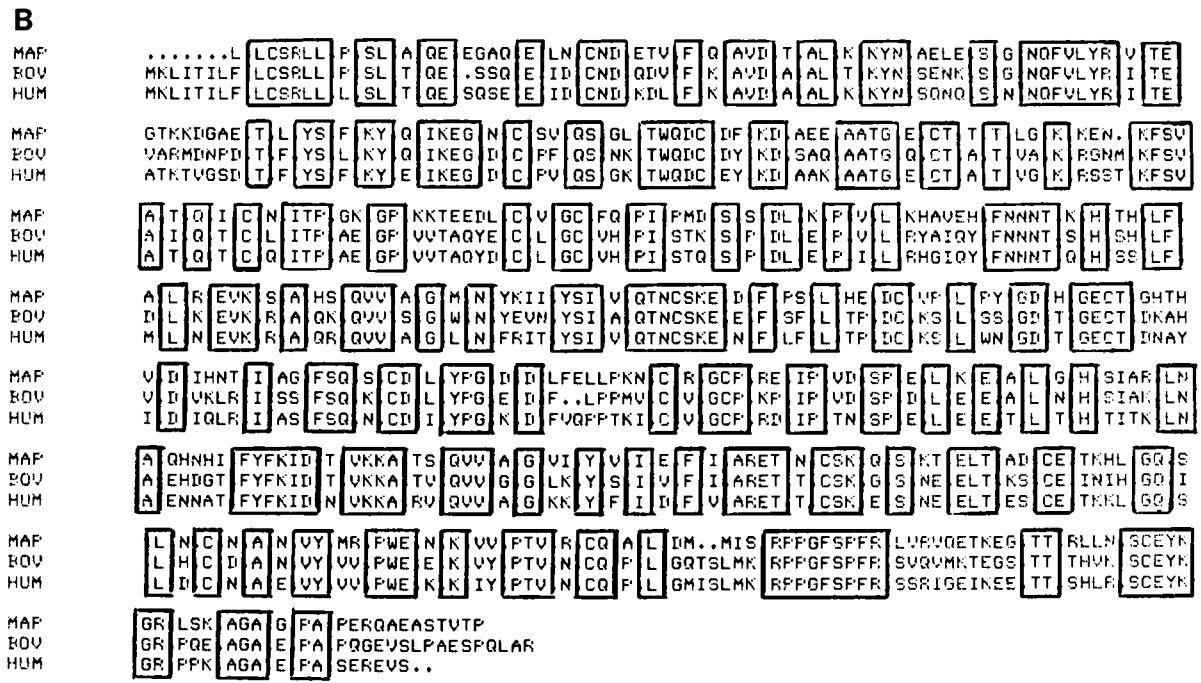


Fig. 4.B. Comparison of the amino acid sequences for bovine kininogen, human kininogen, and major acute-phase α_1-protein of the rat. Based on sequence data from Nawa *et al.* (1983), Ohkubo *et al.* (1984), Anderson *et al.* (1984), and Cole *et al.* (1985a,b). Conserved sequences are indicated by boxes. The fully conserved sequence for the nonapeptide bradykinin is located in the right half of the second to last row. In the rat, possibly, Ile-Ser-bradykinin is released from the major acute-phase α_1-protein (= thiostatin = T-kininogen) by the action of proteinases (Okamoto and Greenbaum, 1983c; Barlas *et al.*, 1985).

generation of vasodilating and pain-inducing oligopeptides. This leads to the "calor" (heat), the "rubor" (redness), the "dolor" (pain), and the "tumor" (swelling), i.e., the four cardinal local symptoms of an acute inflammation (Celsus, 1713; see also Virchow, 1860; Büchner, 1956; Robbins and Cotran, 1979; Hurley, 1983). Both high- and low-molecular-weight kininogens in both humans (for amino acid sequences, see Lottspeich *et al.*, 1984, and the Appendix to this volume) and bovines (for primary structures of proteins and mRNAs, see Nawa *et al.*, 1983) have been found recently to be thiol proteinase inhibitors (Müller-Esterl *et al.*, 1985; Sueyoshi *et al.*, 1985), but they are not acute-phase proteins (Pagano *et al.*, 1984a).

A sequence arginylbradykinin and the liberation of bradykinin by glandular kallikreins from high-molecular-weight kininogen have been described by Kato *et al.* (1985).

A total of 12 independent genomic clones for the major acute-phase α_1-protein of the rat has been isolated from a rat genomic DNA library of 2×10^6 independent recombinants (Fung *et al.*, 1986). The size of the gene for the major acute-phase α_1-protein of the rat is about 20 kb and contains ten introns. A single gene exists for both high-molecular-weight and low-molecular-weight kininogens in bovines (Kitamura *et al.*, 1983). A comparative nucleotide sequence analysis of the flanking region for the acute-phase kininogen (i.e., major acute-phase α_1-protein of the rat = T-kininogen) and the non-acute-phase kininogens (human and bovine kininogens) will be of interest in the further investigation of the mechanism of the general acute-phase response (see also Section II,B,4,b).

2. *Proteins with Transport or Binding Function*

A. HEMOPEXIN. Hemopexin, a monomer, has a molecular weight of 57,000 in the human and a carbohydrate content of 22.6%. Hemopexin increases up to 10-fold during the acute-phase response in the rat and less strongly in humans during various diseases (Cleve *et al.*, 1968) and inflammation (Kushner *et al.*, 1972). It forms a complex with heme released by *in vivo* hemolysis. The hemopexin–heme complex is taken up by endocytosis in the liver, where heme is metabolized further. Thus, intravascular hemolysis can lead to a decrease of the plasma concentration of hemopexin, masking the effect of an increased rate of synthesis of hemopexin due to an acute-phase reaction (see Section VI,C). The properties and the metabolism of hemopexin have been reviewed elsewhere by other authors (Müller-Eberhard and Liem, 1974; Putnam, 1975b; Keyser, 1983). The primary structure of human hemopexin was determined directly (Takahashi *et al.*, 1985; see the Appendix to this volume) and was also deduced from the nucleotide sequence in cloned cDNA (Altruda *et al.*, 1985).

B. HAPTOGLOBIN. Haptoglobin, a tetramer with the structure $\alpha_2\beta_2$ and a molecular weight in the human of 103,772 with a carbohydrate content of 16%, increases only moderately during the acute-phase response. Mature haptoglobin is synthesized by proteolytic processing of a larger precursor protein containing two different subunits in one continuous polypeptide chain (Haugen *et al.*, 1981). Haptoglobin has been separated by starch gel electrophoresis into the three phenotypes Hp1-1, Hp2-1, and Hp2-2 (Smithies, 1965; Schultze and Heremans, 1966). The difference between the phenotypes is due to a variation in the composition of the α chains (Sutton, 1970). Haptoglobin has been reviewed in greater detail by other authors (e.g., Sutton, 1970; Giblett, 1974; Putnam, 1975a). Complementary DNA clones have been isolated for both human (Yang *et al.*, 1983) and rat prohaptoglobin (Goldstein and Heath, 1984) and characterized, including the sequence coding for the region joining the α and the β subunit (see Chapter 1). The untranslated 5′ flanking region of the genes for α_1-antitrypsin, α_1-acid glycoprotein, and haptoglobin share a homologous section (Dente *et al.*, 1985) (see Section II,B,4,b).

C. CERULOPLASMIN. Ceruloplasmin, a single-chain molecule with a molecular weight of 135,000 and a carbohydrate content of 7% in the human, increases only very slightly during the acute-phase response. Ceruloplasmin has been the subject of numerous reviews (see references in Table II and Gutteridge, 1978). Its complete amino acid sequence was elucidated recently (Takahashi *et al.*, 1984). In addition to its function in the transport of copper (Hsieh and Frieden, 1975), ceruloplasmin can act as a ferrooxidase (Osaki *et al.*, 1966), oxidizing ferrous to ferric ions which are then bound to apotransferrin. Thus, ceruloplasmin might be involved indirectly in the transport of iron. Ceruloplasmin and coagulation factors V and VIII constitute a family of structurally related proteins (Church *et al.*, 1984). A model for the structure of human ceruloplasmin has been developed, based on internal triplication, hydrophilic/hydrophobic character, and secondary structure of domains (Ortel *et al.*, 1984).

D. TRANSFERRIN. Transferrin consists of one polypeptide chain, contains 6% carbohydrate (human), and has a molecular weight of 79,600 in the human and 76,500 in the rat. It is synthesized via a precursor protein with an N-terminal presegment, but without a prosegment (Schreiber *et al.*, 1979; Aldred *et al.*, 1984). It has been the subject of numerous reviews (see references in Table II). Probably, it is not involved directly in the early acute-phase response, and the delayed increase in its concentration in the plasma after inducing an inflammation ocurs more likely in response to lowered serum iron levels (see Section II,A). Genetic atransferrinemia is not compatible with survival to adulthood (Heilmeyer *et al.*, 1961).

The complete amino acid sequence of human transferrin has been established

(MacGillivray *et al.*, 1982). Molecular cloning has been reported for the cDNA for part of the transferrin mRNA from both human (Uzan *et al.*, 1984; Yang *et al.*, 1984) and rat liver (Aldred *et al.*, 1984). Using the human transferrin cDNA as a probe for hybridization, the chromosomal location of the transferrin gene was found to be 3q21 → 3qter (Huerre *et al.*, 1984; see also Chapter 1). A genomic library of rat DNA of 2×10^6 independent recombinants was found to contain 25 independent transferrin clones. The size of the transferrin gene in the rat is at least 16 kb and the gene contains more than seven introns (Fung and Schreiber, 1986).

The role of the change in the rate of transcription of the transferrin gene during the acute-phase response has been discussed in Section II,A. Nonhepatic sites of transferrin synthesis of functional importance are the yolk sac (see Section III,C), the choroid plexus (see Section V), and the Sertoli cells in the testis (Wright *et al.*, 1981).

3. Proteins Involved in "Structural Protection": Fibrinogen

Fibrinogen, a hexamer of the structure $(A\alpha)_2(B\beta)_2\gamma_2$ with a molecular weight in the human of 340,000 and a carbohydrate content of 3.7%, is one of the acute-phase proteins moderately increasing in concentration. An increased demand for fibrinogen occurs whenever tissue damage has led to an opening of blood vessels. After severing blood vessels, a closed circulatory system is quickly re-established by a complex mechanism involving both contraction of muscles in the vessel wall (one of the effects of bradykinin) and formation of blood clots. Larger losses of fibrinogen can occur during fibrinous exudation, leading to deposition of fibrin in body cavities such as the pericardial or pleural cavity (e.g., during rheumatic pericarditis or pneumococcal pneumonia).

The complicated arrangement of the six subunits of three different types in the fibrinogen molecule poses the riddle of how such an intricate structure is formed in the cell. An early concept (Doolittle, 1973) suggested that all three of the fibrinogen subunits were coded by a single mRNA species. Similarly to the formation of the tertiary structure of insulin from proinsulin, the fibrinogen subunits were thought to be generated from one large polypeptide precursor by proteolytic processing. However, in an analysis of the sedimentation pattern of polyribosomes involved in fibrinogen synthesis during centrifugation through a sucrose gradient, the size of polyribosomes synthesizing fibrinogen subunits did not suggest the presence of a single large fibrinogen mRNA species (Bouma *et al.*, 1975). Other studies (Yu *et al.*, 1980; Crabtree and Kant, 1981; Nickerson and Fuller, 1981) demonstrated that each of the fibrinogen subunits possesses its own mRNA species with a section coding for a separate presegment. The isolated mRNA could be translated into fibrinogen subunits in an *in vitro* system

(Crabtree and Kant, 1981) and individual cloning of the cDNA for the various chains of fibrinogen was possible (Chung *et al.*, 1981). There is a single copy of each gene for the Aα, Bβ, and γ chain of fibrinogen in humans (Kant *et al.*, 1985). A coordinate accumulation of the messenger RNA for the Aα, Bβ, and γ chains of fibrinogen has been observed in rat liver under conditions producing an increase in the rate of synthesis of fibrinogen (Crabtree and Kant, 1982b). The assembly of fibrinogen commences while nascent incomplete Bβ chains are still attached to polyribosomes. Previously formed Aα and γ chains, drawn from their intracellular pools, combine independently with the growing Bβ chains (Yu *et al.*, 1984). A homology in sections of the 5′ flanking regions of the genes for the fibrinogen chains was considered to be a potential structural basis for regulation leading to the coordinate expression of the genes for the three subunits (Fowlkes *et al.*, 1984).

Numerous reviews of the literature on various aspects of the chemistry and biochemistry of fibrinogen are available (see Doolittle, 1973, 1975, 1984; Fuller, 1983).

4. *Proteins Influencing the Activity of Other Cells*

A. THE COMPLEMENT SYSTEM. The complement system is thought to "complement" the function of the antibody system in the defense of the organism against foreign cells and their products. It consists of 11 different proteins. Several of these are of monomeric structure and others are composed of several subunits. Their molecular weights range from 75,000 to 410,000. Some of the components of the complement system, in particular component C3, are acute-phase proteins moderately increasing in human plasma during acute inflammation. Activation of the system occurs in a complex sequence of events including numerous specific proteolytic cleavages of its components. The activated system lyses foreign cells, such as invading microorganisms, attracts phagocytic leukocytes to the site of antigens, stimulates ingestion and degradation of antigens by phagocytes, and causes mast cells to release histamine, one of the local mediators of inflammation. The complement system is the subject of several detailed reviews (see references in Table II and Müller-Eberhard and Liem, 1974; Müller-Eberhard, 1975b; Porter and Reid, 1979; Kerr, 1981; Reid and Porter, 1981). The nucleotide sequences of the cDNA coding for the α chain of murine complement component C3 (Wetsel *et al.*, 1984) and the cDNA and genomic DNA for the murine β chain (Lundwall *et al.*, 1984) have both been reported. Also, the primary structure of human complement component C3 has been derived recently from the coding sequence of its cDNA (de Bruijn and Fey, 1985) (see Appendix). Complement components C3 and C4 and α_2-macroglobulin have a common evolutionary origin (Sottrup-Jensen *et al.*, 1985) (see Chapter 4).

B. α_1-ACID GLYCOPROTEIN. α_1-Acid glycoprotein is a monomer of 39,500 Da and contains 45% carbohydrate (for review, see Schmid, 1975). α_1-Acid glycoprotein levels increase slightly in human plasma and strongly in rat plasma during the acute-phase response. The function of α_1-acid glycoprotein is not understood yet. It has been reported to inhibit the interaction of malaria parasites with the surface of red blood cells (Friedman, 1983). *In vitro*, α_1-acid glycoprotein also inhibits the aggregation of platelets (Snyder and Coodley, 1976), the transformation of lymphocytes, and the phagocytic activity of macrophages (Chiu *et al.*, 1977). The carbohydrate moiety seems to be important in the interaction of α_1-acid glycoprotein with the lymphoid cell surface (Bennett and Schmid, 1980) leading to nonspecific immune suppression (Cheresh *et al.*, 1984). Carcinoembryonic antigen and α_1-acid glycoprotein possess immunological similarity (Ochi *et al.*, 1982a,b). α_1-Acid glycoprotein also has some similarity in amino acid sequence with epidermal growth factor (Toh *et al.*, 1985). The synthesis of α_1-acid glycoprotein has been studied in primary cultures of rat hepatocytes which displayed changes in rates of synthesis similar to those observed for the liver *in vivo* during the acute-phase response (Howlett *et al.*, 1981; Andus *et al.*, 1983). The cDNA has been cloned for both rat (Ricca and Taylor, 1981; Ricca *et al.*, 1981; Birch *et al.*, 1983) and human α_1-acid glycoprotein (Board *et al.*, 1985). An α_1-acid glycoprotein-like protein was synthesized by *E. coli* containing a recombinant plasmid (Birch *et al.*, 1983). There are at least two genes coding for α_1-acid glycoprotein in the human genome (Dente *et al.*, 1985). The α_1-acid glycoprotein gene is composed of six exons and five introns. A consensus sequence exists in the untranslated 5′ flanking region for α_1-acid glycoprotein (Dente *et al.*, 1985), α_1-antitrypsin (Baumann *et al.*, 1983), and haptoglobin (Bensi *et al.*, 1985).

C. C-REACTIVE PROTEIN. C-reactive protein (for review, see Kushner *et al.*, 1981) was detected initially in the serum of patients suffering from pneumococci infections. It precipitated with the C-polysaccharide from the pneumococcal cell walls in the presence of calcium ions (Tillett and Francis, 1930). C-reactive protein is composed of five identical subunits of 21,000 Da each, which are noncovalently bound. It does not contain carbohydrate and cannot be found in the rat. Partial amino acid sequence homology suggests that it is a member of a superfamily of proteins together with serum amyloid P-component-related protein (Osmand *et al.*, 1977; Oliveira *et al.*, 1979). The level of C-reactive protein in the plasma of healthy individuals is extremely low and, therefore, more than 1000-fold increases in plasma concentration can be observed during the acute-phase response. C-reactive protein binds to receptors on mononuclear phagocytes and lymphocytes. It stimulates phagocytosis and inhibits the aggregation of platelets. During the course of the acute-phase response the secretion of

C-reactive protein becomes more efficient (Macintyre *et al.*, 1985). Using cDNA prepared for human C-reactive protein as a hybridization probe, its gene could be located on chromosome 1 (Whitehead *et al.*, 1983).

D. SERUM AMYLOID A-RELATED PROTEIN. Serum amyloid A-related protein (Levin *et al.*, 1973; for reviews, see Kushner *et al.*, 1981; Gorevic and Franklin, 1981) is probably the precursor protein for amyloid A protein which is found in the form of extracellular deposits at various sites of the body during certain diseases (see Section VI,E). Ethionine, which inhibits the synthesis of protein specifically in the liver by competing with *S*-adenosylmethionine, produces a decrease of both the synthesis of serum amyloid A-related protein in the liver and the appearance of amyloid in tissue (Kisilevesky *et al.*, 1979). Serum amyloid A-related protein consists of a large complex with a molecular weight of between 100,000 and 200,000. It is composed of subunits of 12,000 to 14,000 Da.

The subunit of amyloid A protein has been completely sequenced for humans (Levin *et al.*, 1972). The sequence of its 76 amino acids is identical to that of the amino-terminal portion of serum amyloid A-related protein but it lacks about 30 to 40 more amino acids at the carboxyl terminus if compared with serum amyloid A-related protein. Its molecular weight is about 8400.

Serum amyloid A-related protein is found in humans, mice, and guinea pigs, but not in rats. It can be obtained from the high-density lipoprotein fraction of the serum proteins (Benditt and Eriksen, 1977). Producing an acute-phase response by application of endotoxin or casein led to an increase in the production of serum amyloid A-related protein in mice (Benson *et al.*, 1977). Etiocholanolone had a similar effect in humans (McAdam *et al.*, 1978). The increase in the plasma levels of serum amyloid A-related protein seemed to depend on a substance derived from macrophages and could be prevented by the application of protein synthesis inhibitors (Sipe, 1978; Sipe *et al.*, 1979). Immunohistochemically, the serum amyloid A-related protein was detected mainly in hepatocytes (Sipe *et al.*, 1978).

A genetically deficient mouse strain, C3H/HEJ, lacks serum amyloid A-related protein. Serum amyloid A-related protein has been described to suppress the antibody response of mouse spleen cells *in vitro* (Benson *et al.*, 1975; Benson and Aldo-Benson, 1979). The biosynthesis and the processing of the precursor protein for human serum amyloid A-related protein has been studied and structural variants were defined using complementary DNA (Sipe *et al.*, 1985).

E. SERUM AMYLOID P-COMPONENT-RELATED PROTEIN. Serum amyloid P-component-related protein is partially homologous to C-reactive protein (see Section II,B,4,c), but does not cross-react immunologically with serum amyloid A protein. It binds to polysaccharides and amyloid fibrils (Pepys *et al.*, 1979b)

and is a constituent of the glomerular basement membrane (Dyck *et al.*, 1980). It leads to the agglutination of complement-coated erythrocytes (Hutchcraft *et al.*, 1981). An increase in its concentration during the acute-phase response has been reported only in mice (see Kushner *et al.*, 1981). In humans it is not an acute-phase reactant (Pepys *et al.*, 1978). Probably, it gives rise to the "P-component," which amounts to about 5–10% of the dry weight of all amyloid.

5. Intracellular Acute-Phase Proteins

Several intracellular proteins are found in increased amounts in hepatocytes during the acute-phase response. Examples are metallothionein (Schreiber *et al.*, 1986) and actin (Morrow *et al.*, 1981). Actin might participate in the intracellular transport system for the export of plasma proteins. The function of metallothionein is related to the transport of copper, but is not fully understood (for review, see Kägi and Nordberg, 1979; Brady, 1982). The induction of its mRNA in the mouse by bacterial endotoxin is independent of metals and glucocorticoid hormones (Durnam *et al.*, 1984). The role of various specific sections (consensus sequences) of the promoter regions of human and mouse metallothionein genes could be identified (Durnam *et al.*, 1984; Karin *et al.*, 1984; Stuart *et al.*, 1984).

C. Negative Acute-Phase Proteins

Plasma proteins whose concentrations decrease during the acute-phase response are called negative acute-phase proteins (lower part of Table II). An interpretation for the role of albumin during the acute-phase response (metabolic adapter) has been offered in Section II,A. The functional implications of the decrease in the concentration of the other negative acute-phase proteins listed in Table II are not yet understood. It is possible that they simply share their regulatory mechanism with that of albumin. It is also not known whether the structural homologies observed between transthyretin (prealbumin), gastrointestinal hormones, and other proteins (Jörnvall *et al.*, 1981) have any regulatory implications.

In the case of retinol-binding protein, a 10-fold increase in the rate of excretion in the urine has been observed after surgical trauma (Ramsden *et al.*, 1978). Such an increase in excretion could be the reason for the low level of retinol-binding protein in plasma observed during the acute-phase response. Binding to transthyretin (Trägårdh *et al.*, 1980) is believed to prevent the loss of retinol-binding protein into the urine. More retinol-binding protein will be lost from the blood if not enough transthyretin is available for binding. The measurement of the levels of mRNA for retinol-binding protein in liver extracts is required to

decide whether or not the rate of synthesis of retinol-binding protein changes during the acute-phase response.

D. Factors Involved in Initiating or Mediating the Acute-Phase Response of the Plasma Protein Synthesizing System in the Liver

The concept that a humoral agent, produced by or related to white blood cells, induces the general response of the organism to trauma, acute diseases, and experimental inflammation is relatively old. In most early studies the increase in body temperature and the concentration of leukocytes in the blood were the parameters used to monitor the effect of such an agent. Thus, Billroth (1875) produced fever by injecting filtrates of pus. Menkin (1936) tried to identify organic substances other than histamine from inflammatory exudates, produced by injection of turpentine, which could increase vascular permeability. A factor influencing leukocytes directly, leukotoxin (Menkin, 1937), was distinguished from a leukocytosis-promoting factor (Nettleship, 1938; Menkin, 1940) and a fever-inducing factor, pyrexin (Menkin, 1945). Bennett and Beeson (1953a,b) extracted and tried to characterize a fever-inducing substance from granulocytes obtained from blood, peritoneal exudates, or the sites of dermal lesions produced by Schwartzman and Arthus reactions. Atkins and Wood (1955) demonstrated a transferable pyrogenic substance which appeared in the circulation of rabbits during the course of experimental fever induced by injecting typhoid vaccine. The earlier literature has been summarized by Menkin (1956), Lawrence and Landy (1969), and Atkins and Bodel (1972). More recent summarizing discussions of the field were published by Dinarello and Wolff (1978), Dinarello (1979, 1981), Larsen and Henson (1983), Morimoto *et al.* (1984), Kampschmidt (1984), and Blatteis (1984). An endogenous pyrogen activity has been reported to appear in human plasma even after as small an irritation as exercise (Cannon and Kluger, 1983).

Studies on the agents initiating the acute-phase response of plasma protein synthesis in the liver are complicated by the fact that these agents occur and are active in only minimal amounts and possess a half-time of only minutes in the circulation (Lorber *et al.*, 1971; Kampschmidt and Upchurch, 1980). Isolation to homogeneity, proof of purity, and elucidation of the primary structure of the proteins are therefore very difficult to achieve. In this regard, the development of a radioimmunoassay (Dinarello *et al.*, 1977), an *in vitro* bioassay (Ritchie and Fuller, 1981), and the recently reported molecular cloning of the cDNA for interleukin-1 (Lomedico *et al.*, 1984; March *et al.*, 1985) constitute considerable progress.

It is possible that a family of related mediators of the general acute-phase response exists (Dinarello *et al.*, 1974; Farrar and Hilfiker, 1982; Nurmi and Largen, 1982; Oppenheim *et al.*, 1982; Darlington and Lachman, 1984; for

review, see Kampschmidt, 1984). Assessment of the full array of features involved in the general acute-phase response for a mediator preparation containing only one molecular species would be desirable. In most studies the effects of a mediator preparation on the metabolism of only a few plasma proteins, or on the changes in plasma concentration of iron and copper, or on body temperature are reported. It seems that proteins, or protein preparations, designated endogenous pyrogen, leukocytic endogenous mediator, or interleukin-1 can all induce the synthesis of acute-phase proteins in the liver, but it has also been claimed that another protein, hepatocyte-stimulating factor, and not interleukin-1 induces the synthesis of fibrinogen in hepatocytes (Ritchie and Fuller, 1983; Wolowski and Fuller, 1985).

1. Leukocytic Endogenous Mediator

The term leukocytic endogenous mediator (LEM) was introduced by Pekarek *et al.* (1972a). Release of leukocytic endogenous mediator into the bloodstream was thought to lead to a decrease of plasma iron and zinc during the acute-phase response (Kampschmidt and Upchurch, 1962, 1968; Pekarek and Biesel, 1971). Leukocytic endogenous mediator was believed to be related to the so-called endogenous pyrogen (Kampschmidt and Upchurch, 1969, 1970a,b), a small lipoprotein with a molecular weight of about 13,000, possessing SH groups essential for its function (Kozak *et al.*, 1968). Increased levels, after application of mediator preparations, were reported for the following plasma proteins: glycoproteins (Jamieson *et al.*, 1983), fibrinogen (Homburger, 1945; Kampschmidt and Upchurch, 1974; Merriman *et al.*, 1975, 1977, 1978; Wannemacher *et al.*, 1975; Bornstein and Walsh, 1978; Gordon and Limaos, 1979; Rupp and Fuller, 1979a,b; Weidner *et al.*, 1979), α_1-acid glycoprotein (= seromucoid) (Wannemacher *et al.*, 1975), α_1- and α_2-macrofetoprotein (=α_2-macroglobulin) (Eddington *et al.*, 1971, 1972; Kampschmidt *et al.*, 1973b; Thompson *et al.*, 1976; Gordon and Limaos, 1979; Wannemacher *et al.*, 1975), ceruloplasmin (Pekarek *et al.*, 1972b; Wannemacher *et al.*, 1975; Bornstein and Walsh, 1978), hemopexin (Merriman *et al.*, 1978), C-reactive protein (Merriman *et al.*, 1975; Bornstein and Walsh, 1978; Morley and Kushner, 1982), haptoglobin (Kampschmidt and Upchurch, 1974; Wannemacher *et al.*, 1975; Palmer, 1976; Bornstein and Walsh, 1978; Gordon and Limaos, 1979; Hooper *et al.*, 1981), and serum amyloid A-related protein (Sipe *et al.*, 1979, 1982; McAdam and Dinarello, 1980; Sztein *et al.*, 1981; McAdam *et al.*, 1982; Phong *et al.*, 1982; Luger *et al.*, 1983).

2. Interleukin-1

The so-called lymphocyte-activating factor (Gery *et al.*, 1971, 1972; Gery and Waksman, 1972), later called interleukin-1 (Mizel and Farrar, 1979), was found

to be very similar to leukocytic endogenous mediator (Rosenwasser *et al.*, 1979; Murphy *et al.*, 1980) and behaved similarly to leukocytic endogenous mediator during purification (Kampschmidt *et al.*, 1982). Its molecular weight is between 11,000 and 15,000 (Lachman, 1982; Nurmi and Largen, 1982; Oppenheim *et al.*, 1982). However, during gel filtration, activity peaks of 30,000 to 40,000 Da, and even larger than 70,000 Da, were also obtained (Nurmi and Largen, 1982). Interleukin-1 had many of the effects previously demonstrated for leukocytic endogenous mediator (Merriman *et al.*, 1977; Bornstein, 1982; McAdam *et al.*, 1982). In particular, it stimulated the synthesis of serum amyloid A-related protein (Sztein *et al.*, 1981; McAdam *et al.*, 1982; Sipe *et al.*, 1982; Staruch and Wood, 1982), α_1-acid glycoprotein, and fibrinogen (Darlington and Lachman, 1984). It induces stable E rosette formation and the synthesis of the T-cell growth factor interleukin-2 by peripheral T lymphocytes (Oppenheim *et al.*, 1982). The pyrogenic and the mitogenic actions of interleukin-1 seem to be connected (Duff and Durum, 1983) and there also seems to be a relationship between interleukin-1 and a homogeneous interferon-inducing factor of 22,000 Da (Van Damme *et al.*, 1985). In the brain, interleukin-1 has been reported to stimulate the proliferation of astroglial cells after injury (Giulian and Lachman, 1985). In the muscle, interleukin is involved in the control of protein degradation, together with prostaglandins and Ca^{2+} ions (Goldberg *et al.*, 1984). Full-length cDNA clones have been reported for interleukin-1 (Lomedico *et al.*, 1984; March *et al.*, 1985).

3. Hepatocyte-Stimulating Factor

Recently, a so-called hepatocyte-stimulating factor has been partially purified and shown to induce the synthesis of fibrinogen in hepatocytes in culture (Ritchie and Fuller, 1983; Wolowski and Fuller, 1985). Interleukin-1 had no stimulating effect on the synthesis of fibrinogen in this system. On the other hand, hepatocyte-stimulating factor did not stimulate the growth of fibroblasts, which would have been characteristic for interleukin-1.

4. Hormones

Trauma and inflammation induce a change in the hormonal status of the body (for reviews, see Beisel, 1981; Schade *et al.*, 1982). Effects of hormones on the synthesis of plasma proteins have been shown in both the isolated liver perfusion system (Griffin and Miller, 1973, 1974; for review, see Miller and Griffin, 1975; Miller, 1976) and cultures of hepatoma cells (Fuller *et al.*, 1979; Feinberg *et al.*, 1983) or of hepatocytes (Jeejeebhoy *et al.*, 1977; Grieninger *et al.*, 1978, 1983; Chen and Feigelson, 1978; Liang and Grieninger, 1981). However, most of the effects of leukocytic endogenous mediator occur also in adrenalectomized or hypophysectomized rats (Thompson *et al.*, 1976; Kampschmidt *et al.*, 1973a;

Wannemacher *et al.*, 1975). As outlined above (Section II,B,5), the site of the regulation of mouse metallothionein-I mRNA synthesis by bacterial endotoxin differs from that by metals and glucocorticoid hormones (Durnam *et al.*, 1984). The presence of corticosteroid hormones is an absolute prerequisite for the induction of α_2-macroglobulin in the rat (Weimer and Benjamin, 1965; Heim and Ellenson, 1967; Miller and Griffin, 1975). Glucocorticoid hormones have also been shown to play a role in the induction of α_1-acid glycoprotein (Jeejeebhoy *et al.*, 1977; Vannice *et al.*, 1984; Reinke and Feigelson, 1985) and fibrinogen (Grieninger *et al.*, 1978, 1983; Rupp and Fuller, 1979a; Hertzberg *et al.*, 1981). The synthesis of α_{2u}-globulin is under the control of several hormones (Kurtz and Feigelson, 1977; Kurtz, 1981; Antakly *et al.*, 1982), including thyroid hormones (Kurtz *et al.*, 1976) as well as glucocorticoid and sexual hormones (Kurtz *et al.*, 1978).

III. Perinatal Development of the Plasma Protein Synthesizing System

A. Concentrations of Plasma Proteins in Rats during Development

The pattern of concentrations of proteins in the blood plasma of young mammals is quite different from that found in adults. The concentrations of six different proteins measured in the blood plasma of very young rats are plotted against time after birth in Fig. 5. The concentration of α_2-macroglobulin is much higher than normally observed in healthy adult rats (Weimer *et al.*, 1967; Van Gool and Ladiges, 1969; Hudig and Sell, 1978), whereas albumin, transferrin, transthyretin (=prealbumin), and α_1-acid glycoprotein all have lower levels after birth than those observed in adult animals. The concentrations of albumin and transferrin rise steadily after birth, overshooting adult levels after about 14 to 15 days. The time course for the plasma concentrations of transthyretin and α_1-acid glycoprotein is interesting. These two proteins start to rise in concentration sharply at the time of weaning, when the levels of thyroxine and corticosterone increase and the diet changes drastically (see Henning, 1981).

B. The Acute-Phase Response of Plasma Protein Synthesis in the Developing Rat Liver

Because of the high plasma level of α_2-macroglobulin after birth, the question arises as to whether the plasma protein synthesizing system in neonatal rat liver is in the acute-phase mode. It is conceivable that the trauma of birth creates the first acute-phase response in life and that the pattern of plasma concentration of

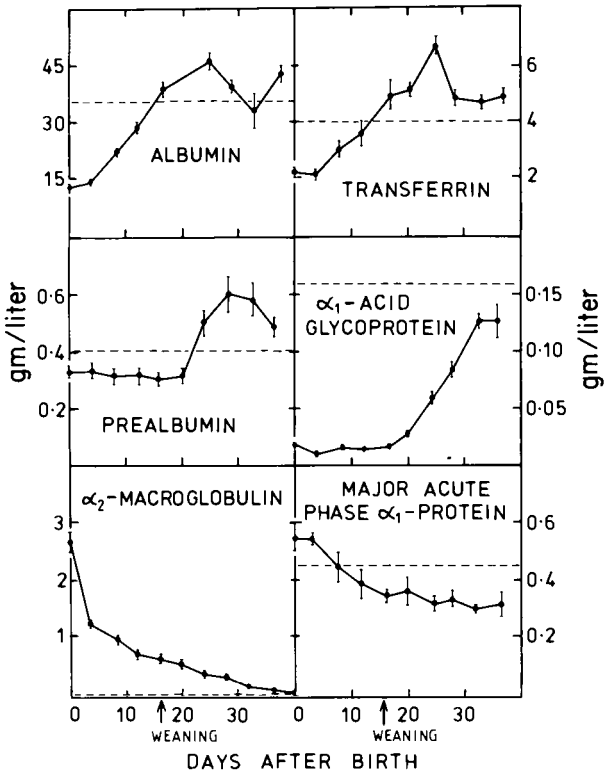


Fig. 5. Changes in the concentration of plasma proteins during development. Each time point represents six rats. Error bars indicate ±1 standard error interval. The concentrations measured in healthy adult rats are indicated by a broken line. From Thomas and Schreiber (1985), with permission.

proteins observed after birth (Fig. 5) is a "perinatal acute-phase response." To answer this question an acute experimental inflammation was induced in four-day-old rats and the levels of proteins in blood plasma were measured (Fig. 6). A typical acute-phase response pattern similar to that found in adult rats was obtained, indicating that an "immature" acute-phase response system in the liver of neonates is not the explanation for the particular pattern of concentration of proteins in plasma observed after birth.

C. Expression of Plasma Protein Genes in the Yolk Sac and Placenta

Certain precautions are required if the synthesis of plasma proteins in tissues other than liver is to be investigated. The incorporation of radioactive amino

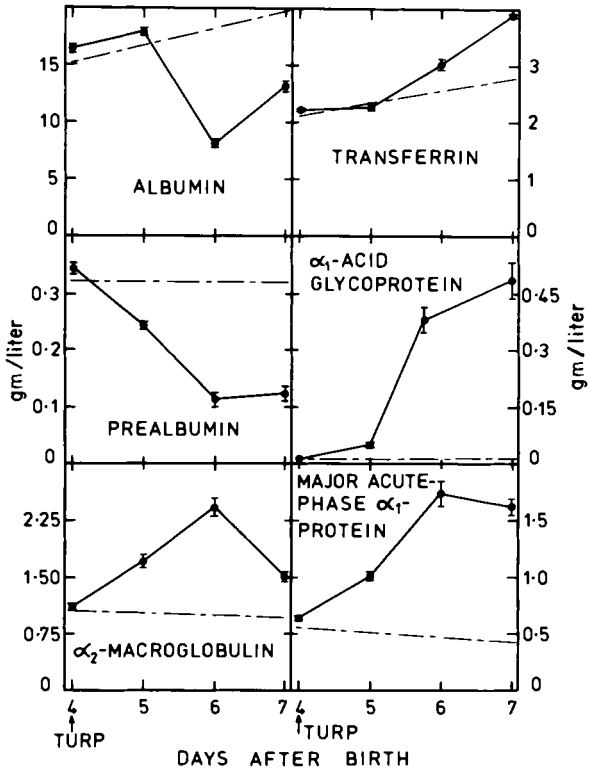


Fig. 6. Changes in the concentration of plasma proteins in neonatal rats during acute experimental inflammation. Each time point represents six rats. Mean values ±1 standard error are given. For comparison, the concentrations which would have been measured under healthy conditions (such as those in Fig. 5) are indicated by a broken line. From Thomas and Schreiber (1985), with permission.

acids into proteins is one of the methods to study the site of synthesis *in vivo* of plasma proteins in the body. This method is complicated by the fact that, after injection of radioactive amino acids, radioactively labeled plasma protein is synthesized and secreted by the liver and distributed throughout the body. Therefore, tissues must be removed within the so-called "secretion time" (Schreiber *et al.*, 1971) or "minimum transit time" (Peters and Peters, 1972), i.e., the time between injection of radioactive amino acid and first appearance of labeled protein in the bloodstream. However, when specific cDNA probes are available, it is possible to determine intracellular levels of specific mRNAs by hybridization. In this way, the expression of plasma protein genes in various nonhepatic tissues can be investigated without interference by plasma proteins made and secreted by the liver. The study of nonhepatic plasma protein synthesis is of particular interest for tissues or organs which might be involved in the control of

the extracellular "milieu" in areas of the body which are separated from the general circulation by structures that do not allow free permeation of proteins.

As discussed above, transport and communication between tissues and cells of the body are important functions of the liver/blood plasma system. The question arises of how plasma proteins are involved in the communication between fetus and mother. There is no open connection between fetal and maternal circulatory systems. The passage of compounds from mother to fetus and from fetus to mother has to occur across one or several layers of cells. In the rat, this passage can occur at two sites, the chorioallantoic placenta and the yolk sac (vitellinic placenta) (Rowett, 1960; Ramsey, 1975). Obviously, it is important for proper growth of the fetus that iron is transported from the mother in sufficient amounts to the growing organism. It may be expected that the usual protein carrier for iron, transferrin, will be involved in this transport. It is an intriguing question whether this transferrin is derived from plasma transferrin provided by the maternal liver and whether and how this maternal transferrin interacts perhaps with a transcellular iron transport protein and/or transferrin originating in the fetal liver.

TABLE III

Expression of Plasma Protein Genes by Fetal Membranes[a]

Name of protein	Level of mRNA in tissue (percentage of value found in livers from 65-day-old male rats[b]) Yolk sac	Placenta
Albumin	<0.03	<0.03
Transferrin	4	2
Ceruloplasmin	8	<0.1
Transthyretin	<0.5	9
Retinol-binding protein	1	6
Major acute-phase α_1-protein	0.5	0.2
Fibrinogen	<0.2	<0.2
α_1-Acid glycoprotein	11	12
α_2-Macroglobulin	3600	2200
Apolipoprotein A-I	24	82
Apolipoprotein A-IV	6	98
Apolipoprotein E	14	8

[a]Total RNA was prepared from homogenates from yolk sac and placenta from 19 days pregnant inbred Buffalo rats and hybridized on nitrocellulose filters with ^{32}P-labeled cDNA probes for various plasma proteins. From Thomas and Schreiber (1986), with permission.

[b]About 200 gm body weight.

Dancis *et al.* (1957) described the incorporation of 2-[^{14}C]glycine by human placenta into proteins immunologically related to plasma globulins, whereas albumin did not become radioactive. Yeoh and Morgan (1974) reported the incorporation of [^{14}C]leucine into both transferrin and albumin in the rat yolk sac, but a lack of such incorporation in placenta. Miura *et al.* (1979) isolated α-fetoprotein messenger RNA from mouse yolk sac. Muglia and Locker (1984) detected mRNA for α-fetoprotein, and to a much lesser extent for albumin, in the yolk sac of rats using specific cDNA probes for the respective proteins. Using specific cDNA probes for a number of plasma proteins, the expression of plasma protein genes in both placenta and yolk sac was studied in the rat. The results are summarized in Table III (Thomas and Schreiber, 1986). Apparently, the yolk sac is an important site of plasma protein synthesis. A considerable amount of α_2-macroglobulin mRNA is found in extracts from rat placenta.

IV. Molecular Mechanisms of the Regulation of Plasma Protein Synthesis in the Liver

The changes in the pattern of concentrations of plasma proteins observed during the acute-phase response (Figs. 2 and 3) can result from changes in the distribution, the rate of degradation, or the rate of synthesis of plasma proteins.

Although certain changes occur in the equilibration between extra- and intravascular space for plasma proteins during the acute-phase response (Schreiber *et al.*, 1982), the increase or decrease in concentration of proteins in the blood plasma during acute inflammation cannot be explained by a change of distribution. It has been shown for albumin, the major acute-phase α_1-protein of the rat, transferrin, α_1-acid glycoprotein (Schreiber *et al.*, 1982), and transthyretin (=prealbumin) (Dickson *et al.*, 1982) that the total body pools show similar changes after inducing an acute inflammation to the concentrations of plasma proteins in the bloodstream (Figs. 7A and 7B).

Theoretically, it would be possible that an increase in the concentration in plasma of a particular protein is caused by a decrease in the rate of its degradation, and, vice versa, a decrease of the plasma concentration of a protein might be caused by an increase in the rate of degradation without rates of synthesis being changed. However, no changes were observed when the rate of degradation of transferrin (Schreiber *et al.*, 1982) and transthyretin (Dickson *et al.*, 1982) were measured during experimental inflammation (Figs. 8A and 8B).

In contrast to the lack of an effect of acute experimental inflammation on the rate of degradation of plasma proteins, distinct changes are observed in the pattern of rates of incorporation of amino acids into plasma proteins during acute inflammation (Fig. 9). The changes in the rates of incorporation of amino acids into proteins occur in the same direction as the changes of concentrations in the

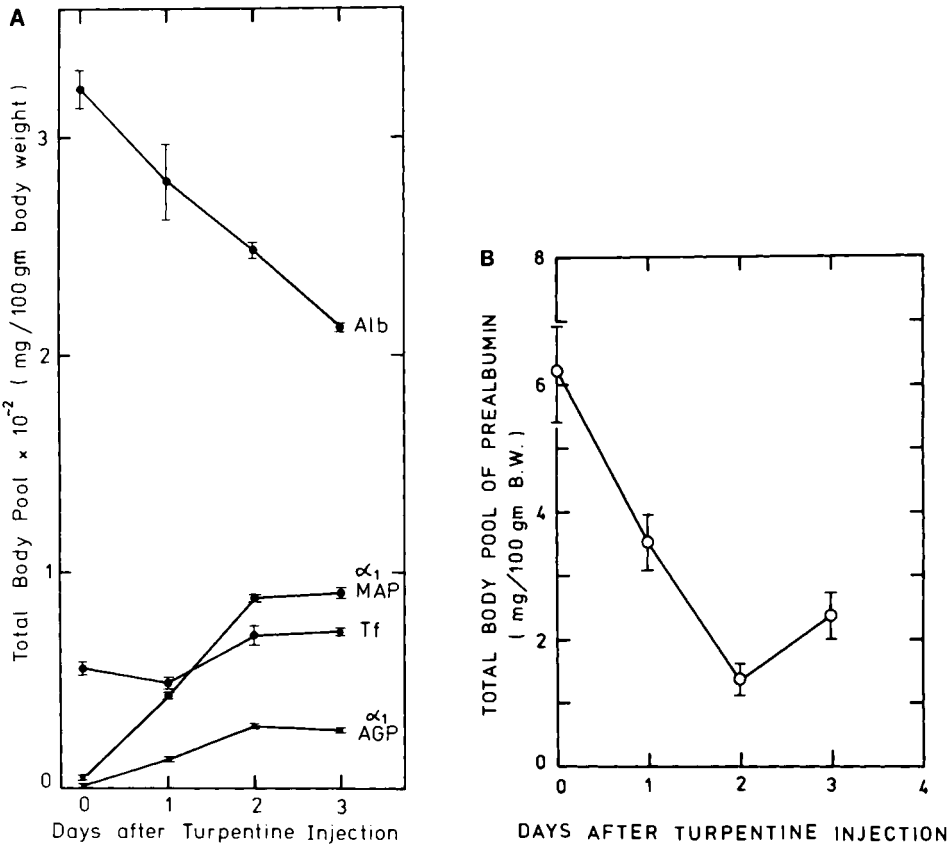


Fig. 7. (A) Total body pools of albumin (Alb), major acute-phase α_1-protein (α_1MAP), transferrin (Tf), and α_1-acid glycoprotein (α_1AGP) during acute inflammation produced by subcutaneous injection of 0.5 ml of mineral turpentine per 100 gm body weight into the back of male Buffalo rats. From Schreiber *et al.* (1982), with permission. (B) Effect of acute inflammation on the total body pool of transthyretin (prealbumin) in rats. From Dickson *et al.* (1982), with permission.

bloodstream, however, the maxima of the changes in incorporation rates precede the maxima of the changes in plasma concentrations by about 24 hr. For albumin, transferrin, major acute-phase α_1-protein, and α_1-acid glycoprotein, it was shown that these proteins are synthesized from a common amino acid precursor pool in the liver (Schreiber *et al.*, 1982). Therefore, a method could be developed for the calculation of the synthesis rates of the four proteins and the changes in the rates of synthesis during the acute-phase response based on measuring incorporation of radioactive amino acids into protein. The obtained values are summarized in Table IV.

A change in the rate of synthesis of a particular protein can be brought about

by a change in the rate of translation or transcription of its mRNA. As outlined above (Section III,C), intracellular levels of specific mRNAs can be measured conveniently by hybridization if appropriate cDNA probes are available. The influence of an acute experimental inflammation on the levels of mRNAs for various plasma proteins in the liver is summarized in Figs. 10A and 10B. The changes in intracellular mRNA levels for plasma proteins during acute inflamma-

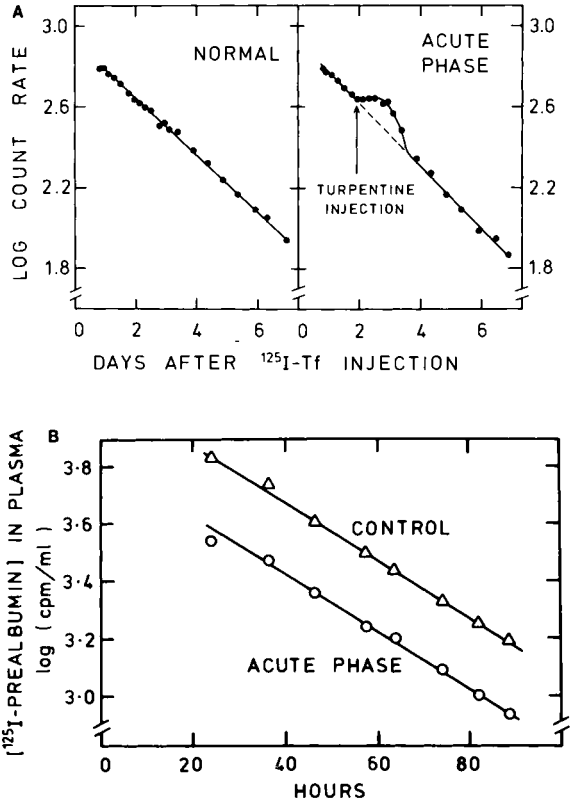


Fig. 8. (A) Disappearance of ^{125}I-labeled transferrin (Tf) from a healthy male Buffalo rat (normal, 260 gm body weight), and from a rat (acute phase, 255 gm body weight) injected subcutaneously with 1.3 ml of turpentine into four different sites 2 days after the injection of ^{125}I-labeled transferrin. Radiation was detected by two NaI detectors on opposite sides of the animal. The deviation from the straight line in the semilogarithmic plot for the acute-phase rat is caused by the transient accumulation of fluid containing plasma proteins in the inflamed area. This increases the efficiency with which ^{125}I radiation is detected, since the half-thickness for ^{125}I radiation in the rat tissue was only 2 cm. With permission from Schreiber *et al.* (1982). (B) Rate of removal of ^{125}I-labeled transthyretin (^{125}I-labeled prealbumin) from the bloodstream of healthy rats and rats suffering from an acute inflammation. With permission from Dickson *et al.* (1982).

tion reflect closely the changes in incorporation rates of radioactive amino acids into the proteins: only slight deviations from average are obtained if the rate of incorporation of amino acid into a protein divided by the mRNA level is plotted against the time after inducing inflammation (Fig. 11). Apparently, the mRNA for a particular protein is translated with the same speed independently of whether the absolute rates of synthesis increase or decrease compared with normal values.

Changes in mRNA levels might be brought about by changes in the rate of transcription or in messenger RNA stability. At present it is not possible to ascertain the contribution of each of these two possible mechanisms to the

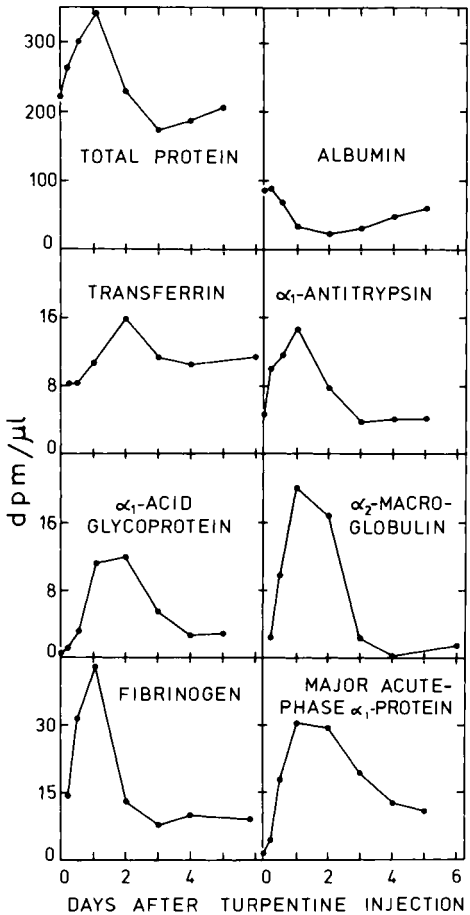


Fig. 9. Incorporation of L-[1-^{14}C]leucine into plasma proteins in the bloodstream during acute experimental inflammation. Adapted from Schreiber *et al.* (1982), with permission.

TABLE IV

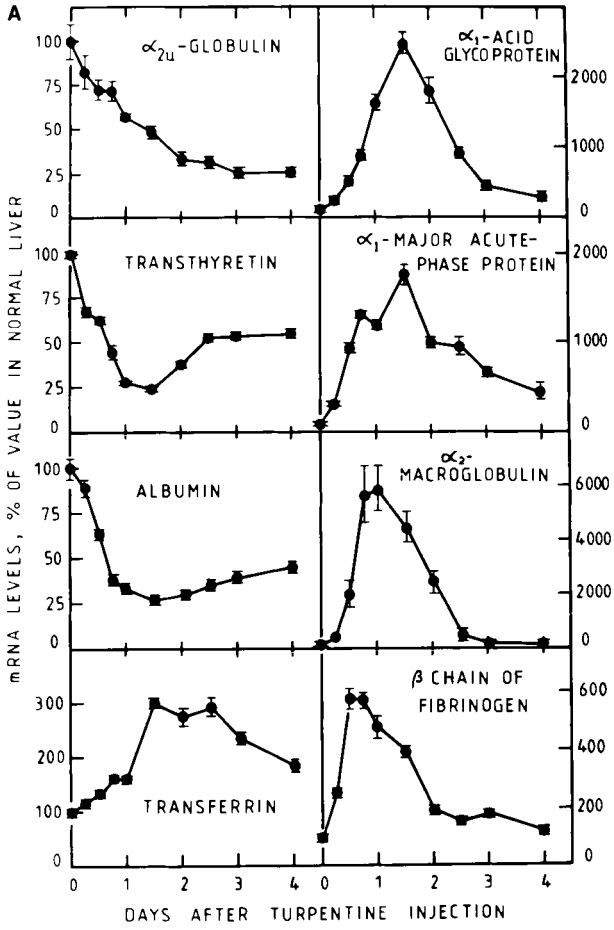
Synthesis Rates of Plasma Proteins in Rats[a]

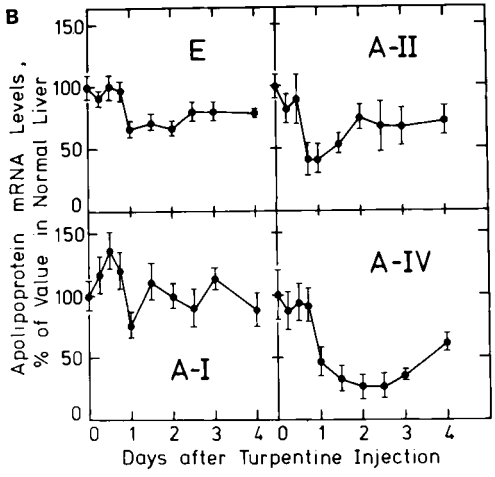
	Albumin	Transferrin	Major acute-phase α_1-protein	α_1-Acid glycoprotein
Rate of incorporation in % of that into total plasma protein	41 (H) 9.2 (I)	5.1 (H) 3.1 (I)	0.53 (H) 9.0 (I)	0.24 (H) 3.2 (I)
Factor by which the rate of synthesis changes 24 hr after inducing inflammation	0.37	1.0	42	22
Rate of synthesis in healthy rats in mg/100 gm body weight/day	91	19	2.3	1.0
Rate of synthesis 24 hr after inducing inflammation in mg/100 gm body weight/day	32	19	97	22

[a]Healthy rats on a diet containing 20% protein (H) and rats 24 hr after inducing an inflammation plus fasting (I). Each value is the mean from four animals. From Schreiber *et al.* (1982), with permission (modified).

changes in mRNA levels observed in the liver during the acute-phase response. For proteins induced by hormones in the oviduct, it has been shown recently that actively expressed genes are found to be enriched in the nuclear matrix-associated DNA (Ciejek *et al.*, 1983; Buttyan *et al.*, 1983; Robinson *et al.*, 1983). A similar relationship between gene expression and association of the gene with a nuclear matrix can also exist for plasma protein genes, as indicated in studies with the genes for albumin and α_1-acid glycoprotein (Fig. 12). Tissue or cell specificity of the expression of the genes for albumin and α-fetoprotein was shown to be related to a differential sensitivity of the chromatin for the two genes toward digestion with deoxyribonuclease I (Nahon *et al.*, 1984).

With more information becoming available on the primary structure of the genomic DNA for plasma proteins (Chapter 1, and Appendix) and on protein factors initiating the acute-phase response, as well as the development of transfection and transgenic animal systems (see Section II,D), a better understanding of the mechanism leading to changes in the rates of transcription of plasma protein genes during the acute-phase response is to be expected in the near future. The molecular cloning of the cDNA for many plasma proteins has provided the means to determine the concentrations of specific mRNAs by hybridization in liquid phase or on solid supports. These techniques combined with the application of inhibitors specific for RNA polymerase II (see Kedinger *et al.*, 1970; Wieland, 1972), the enzyme involved in the synthesis of mRNA in eukaryotic cells, should allow insight into the stability of mRNA, the second possible site for pretranslational regulation of protein synthesis.


A
α_{2u}-GLOBULIN
α_1-ACID GLYCOPROTEIN
TRANSTHYRETIN
α_1-MAJOR ACUTE-PHASE PROTEIN
ALBUMIN
α_2-MACROGLOBULIN
TRANSFERRIN
β CHAIN OF FIBRINOGEN
mRNA LEVELS, % OF VALUE IN NORMAL LIVER
DAYS AFTER TURPENTINE INJECTION


B
E
A-II
A-I
A-IV
Apolipoprotein mRNA Levels, % of Value in Normal Liver
Days after Turpentine Injection

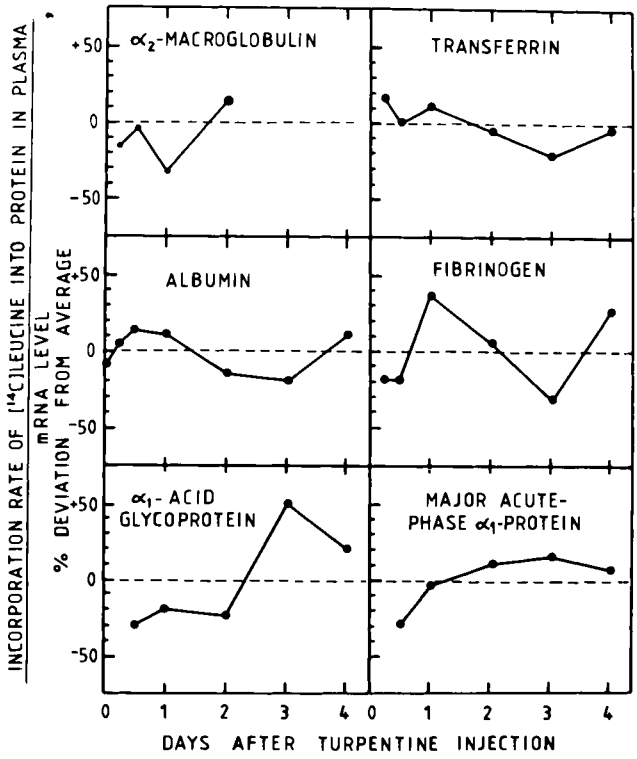


Fig. 11. Relationship between rates of incorporation of radioactive leucine into plasma proteins and levels of corresponding messenger ribonucleic acids in the liver. Rates of incorporation of L-[1-^{14}C]leucine into plasma proteins, obtained from Fig. 9, were divided by the levels of mRNA given in Fig. 10A. The average of all ratios obtained in this way for a particular protein was calculated and considered to be 100% for comparison with the ratios for individual time points. From Schreiber *et al.* (1986), with permission.

V. Plasma Protein Synthesis in the Choroid Plexus

The brain is separated from the body mechanically by being suspended in the cerebrospinal fluid (reducing the weight of the human brain from more than 1 kg to about 50 gm) and chemically by the blood/brain barrier and the blood/choroid plexus/cerebrospinal fluid barrier. The cerebrospinal fluid is secreted by the

Fig. 10. Levels of mRNA in the liver of rats suffering for different lengths of time from an acute inflammation induced by subcutaneous injection of turpentine. Each point is the mean ±1 standard error for eight rats. (A) mRNA levels for plasma proteins; (B) mRNA levels for apolipoproteins. From Schreiber *et al.* (1986) and Tu *et al.* (1986), with permission.

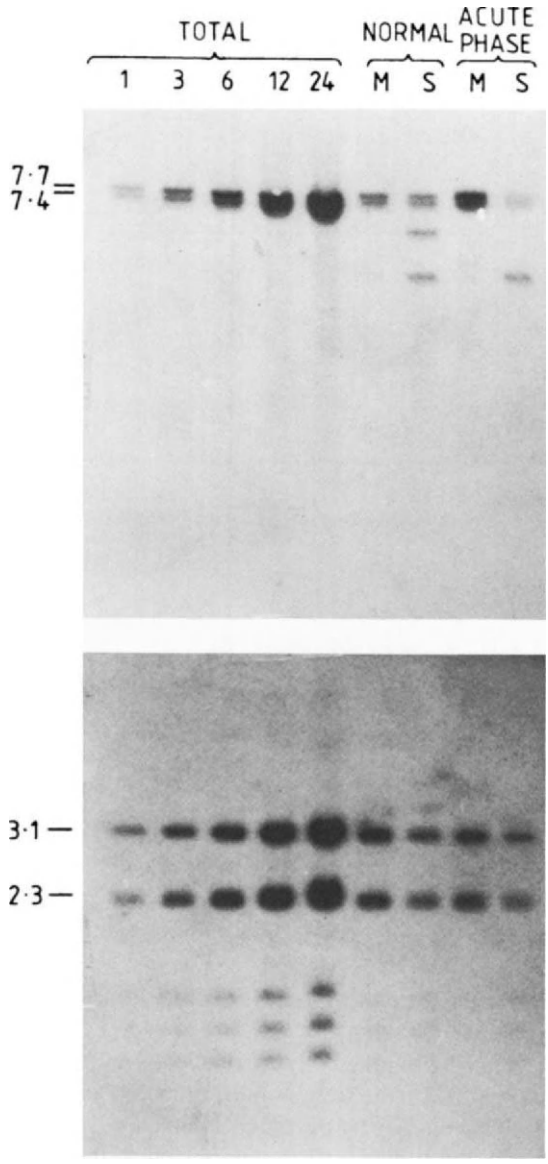


Fig. 12. Association of the genes for α_1-acid glycoprotein (upper panel) and albumin (lower panel) with the nuclear matrix in the liver of healthy rats and of rats with an acute experimental inflammation. Two micrograms of matrix DNA (M) and supernatant DNA (S) isolated from nuclei from both normal and acute-phase liver, together with various amounts of *Eco*RI-cleaved total DNA (1,3,6,12, and 24 μg), was separated by electrophoresis through 0.8% agarose, transferred to Zeta-probe membrane (Bio-Rad), and hybridized with a ^{32}P-labeled α_1-acid glycoprotein cDNA probe

TABLE V

Protein Concentrations in Blood Plasma and Cerebrospinal Fluid (CSF) in Humans[a]

Protein	MW ($\times 10^{-3}$)	Concentration		Ratio CSF/plasma ($\times 10^4$)
		CSF (mg/liter)	Plasma (mg/liter)	
β_2-Microglobulin	12	1.1	2	5,500
Transthyretin	55	17.3	250	690
Transferrin	77	14	2,600	54
Albumin	66	155	40,000	39
α_1-Acid glycoprotein	40	3.5	980	36
α_2-HS protein	49	1.7	600	28
Ceruloplasmin	151	0.9	370	24
α_1-Antitrypsin	45	7	3,000	23
α_2-Macroglobulin	800	4.6	3,000	15
IgG	160	15	10,000	15
Haptoglobin	100 200 400	2.24	4,800	5
Plasminogen	140	0.25	700	3.6
Fibrinogen	350	0.6	2,600	2.3

[a]Compiled from Geigy Scientific Tables (Diem, 1960; Lentner, 1981).

choroid plexus in the lateral and third and fourth ventricles (for review, see Cserr, 1971). It participates in providing the humoral environment of the cells of the central nervous system. The protein concentrations in plasma and cerebrospinal fluid are summarized in Table V. Although the total concentration of protein in cerebrospinal fluid is very low, the proportion of transthyretin (=prealbumin) to albumin is much higher in cerebrospinal fluid than in plasma (Table V; see also Weisner and Kauerz, 1983; Weisner and Röthig, 1983).

The proteins of the cerebrospinal fluid were believed to originate from the blood plasma by transcellular transfer through the choroid plexus (Møllgård *et al.*, 1979; Dziegielewska *et al.*, 1981). A functional leak (Brightman, 1975) or the selective intracellular digestion (Hurley *et al.*, 1981) of proteins during their

(upper panel) or an albumin cDNA probe (lower panel). The difference in the relative intensities of the bands in the autoradiograph indicates that the two α_1-acid glycoprotein gene fragments (7.7 and 7.4 kb) are preferentially associated with the nuclear matrix (M) fraction to a far greater extent in the acute phase than in the normal liver. This matrix association of the DNA for the α_1-acid glycoprotein and albumin genes is in direct correlation with the level of expression of these genes in normal and acute-phase liver. From Birch and Schreiber (1986), with permission.

passage from the bloodstream to the ventricles through the choroid plexus were believed to produce a relative enrichment in the cerebrospinal fluid for certain plasma proteins, such as transthyretin and transferrin. However, synthesis *de novo* of transthyretin (prealbumin) in the choroid plexus was also discussed as an explanation for the occurrence of transthyretin (prealbumin) in both choroid plexus tissue and the cerebrospinal fluid (Agnew *et al.*, 1980; Aleshire *et al.*, 1983; Weisner and Kauerz, 1983). Analysis of the composition of mRNA in choroid plexus by hybridization to specific cDNA probes showed very high levels of mRNA for transthyretin (prealbumin) and transferrin in choroid plexus (Figs. 13, 14, and 15). No mRNA could be detected in choroid plexus for the Bβ subunit of fibrinogen, for major acute-phase α_1-protein, for albumin, for retinol-binding protein, and for α_1-acid glycoprotein. Other regions of the brain were found to contain low levels of transferrin mRNA but no transthyretin mRNA (Fig. 14). Within the choroid plexus, transthyretin mRNA seems to be located only in the choroid epithelial cells. The stroma of the choroid villi or the ependyma of the ventricles do not contain transthyretin mRNA (Figs. 16A and 16B). The transthyretin mRNA in choroid plexus is very similar if not identical in structure with that in liver (Fig. 17). The relative levels of mRNA for transferrin and transthyretin in choroid plexus, brain without choroid plexus, and liver are summarized in Table VI.

The extremely high levels of transthyretin mRNA (prealbumin mRNA) in choroid plexus tissue suggest very active synthesis of transthyretin (prealbumin) in the choroid plexus. Therefore, protein synthesis was investigated directly in choroid plexus by incubating choroid plexus pieces *in vitro* with radioactive leucine in a medium optimized for protein synthesis (Schreiber and Schreiber,

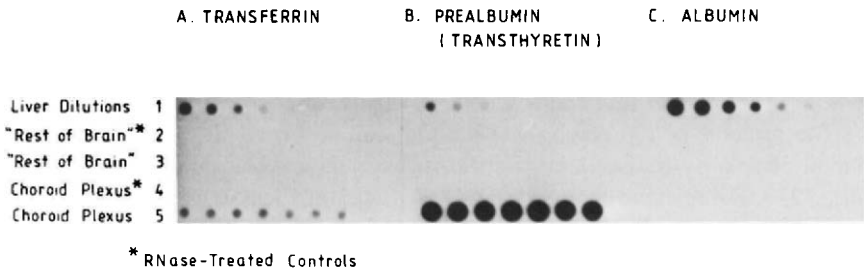


Fig. 13. Determination of mRNA for transferrin (A), prealbumin (B), and albumin (C) by hybridization to specific cDNA. Row 1, cytoplasmic extracts corresponding to 2500, 1000, 500, 250, 100, 50, 25, and 10 μg liver, wet weight, per spot. The extracts used in rows 2 to 5 were prepared from the tissue of seven individual animals and processed separately. Rows 2 and 3, cytoplasmic extracts corresponding to 2500 μg of brain tissue, wet weight, per spot, excluding choroid plexus. Rows 4 and 5, cytoplasmic extracts corresponding to 500 μg choroid plexus per spot. Rows 2 and 4, cytoplasmic extracts incubated with ribonuclease prior to processing for hybridization. From Dickson *et al.* (1985a), with permission.

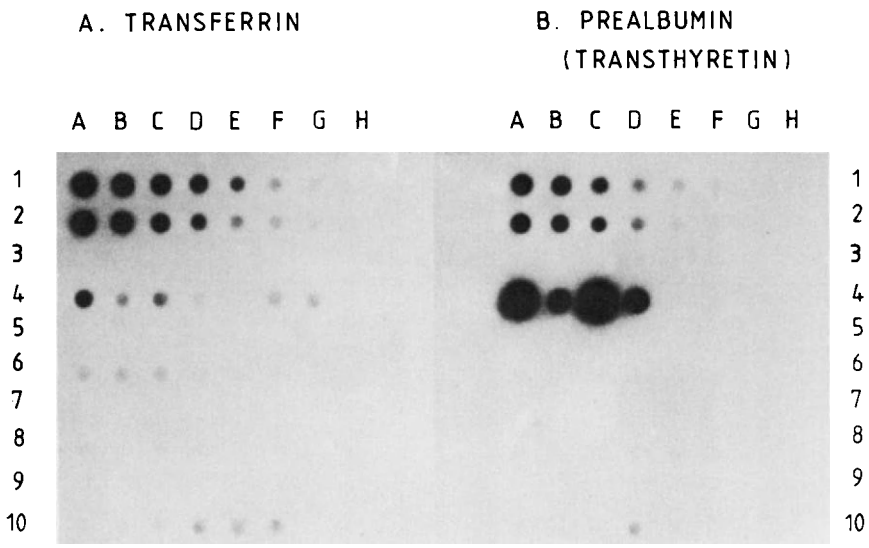


Fig. 14. Distribution of the mRNA for transferrin and prealbumin in the central nervous system analyzed by hybridization to the corresponding specific cDNAs on nitrocellulose filters.

Key to spotting patterns: Rows 1 and 2 are serial dilutions of liver extracts from two rats, amounts of tissue as in Fig. 13. Row 4, tissue pooled from three rats: 4A and B, extract derived from 250 and 25 μg of choroid plexus, respectively, from lateral and third ventricles; 4C and D, extract derived from 250 and 25 μg choroid plexus, respectively, from fourth ventricle; 4E, pia mater from ventral brain surface (extract from 500 μg tissue); 4F and G, cervical spinal cord tissue free of pia mater, F without central canal ependyma (derived from 2500 μg tissue) and G with central canal ependyma (derived from 2500 μg tissue). Rows 6, 8, and 10 correspond to 2500 μg tissue per spot, tissue dissected free of pia mater and ventricular ependyma from three separate rats; 6A,B,C, thalamus; 6D,E,F, hypothalamus; 8A,B,C, pituitary; 8D,E,F, cerebral cortex; 10A,B,C, cerebellum; 10D, E,F, medulla. Spots 3A–C,E,F,G,5A–F, 7A–F, and 9A–F contain material similar to the spots immediately below them, but extracts were incubated with ribonuclease. 3H and 4H are ribonuclease-treated controls for the liver samples in 1A and 2A, respectively. Positions not identified above were not used. From Dickson *et al.* (1985a), with permission.

1973). The results of such studies are shown in Figs. 18 and 19. Very active incorporation of radioactive leucine into transthyretin (prealbumin) was obtained. About 20% of total radioactive protein in choroid plexus was transthyretin. The synthesized transthyretin was secreted with a typical delay ("secretion time") into the medium. About 50% of the radioactive protein in the medium was transthyretin.

In contrast to the liver, transthyretin (prealbumin) mRNA levels did not change in choroid plexus during an acute-phase response induced by injection of turpentine or talcum or by superficial burns, indicating an independent regulation of the expression of the transthyretin (prealbumin) genes in the two tissues (Table VII).

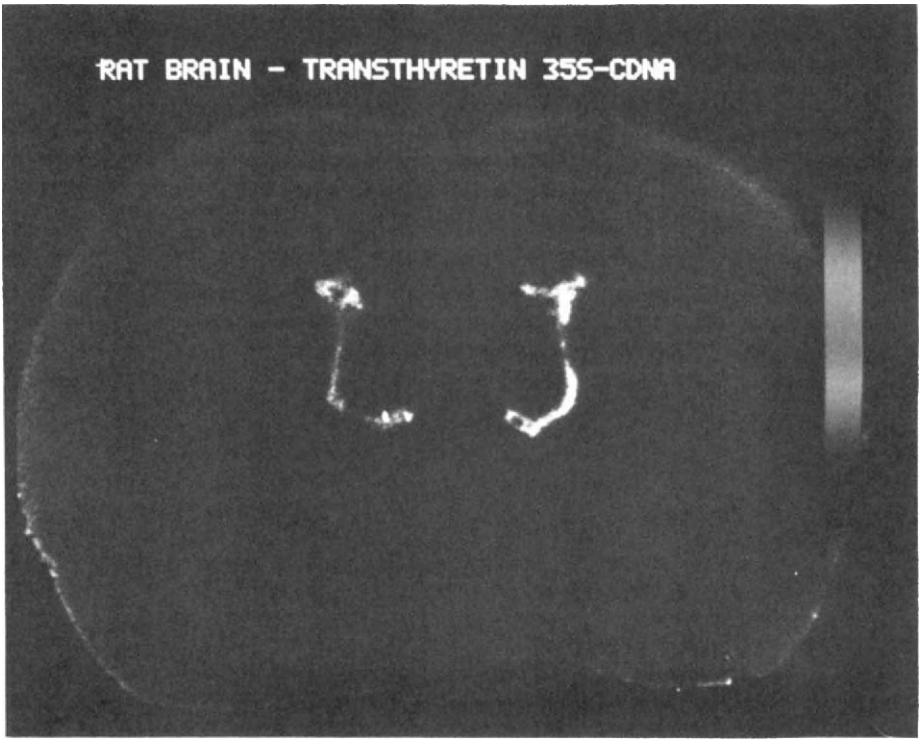


Fig. 15. Computerized densitometry of a coronal section through rat brain hybridized *in situ* with transthyretin cDNA labeled with ^{35}S. The bar to the right is a calibration bar. The choroid plexus tissue gives a bright, positive signal, whereas the rest of the brain shows no radioactivity.

The functional significance of the high rate of synthesis of transthyretin in the choroid plexus is not yet fully understood. Transthyretin is the main thyroid hormone-transporting protein in the bloodstream of rats (Davis *et al.*, 1970; Sutherland and Brandon, 1976). It might also transport thyroid hormones in the cerebrospinal fluid flowing from the lateral and third and fourth ventricles to other areas of the brain and the spinal cord. In the human, thyroxine-binding globulin is the most important thyroid hormone carrier (Woeber and Ingbar, 1968; for reviews, see Ingbar and Woeber, 1981; Hoffenberg and Ramsden, 1983). However, it is possible that the relative importance of thyroid hormone-carrying proteins differs in blood and cerebrospinal fluid. Human thyroxine-binding globulin and transthyretin (prealbumin) from rabbits, rats, and humans are partly homologous in sequence (Sundelin *et al.*, 1985b).

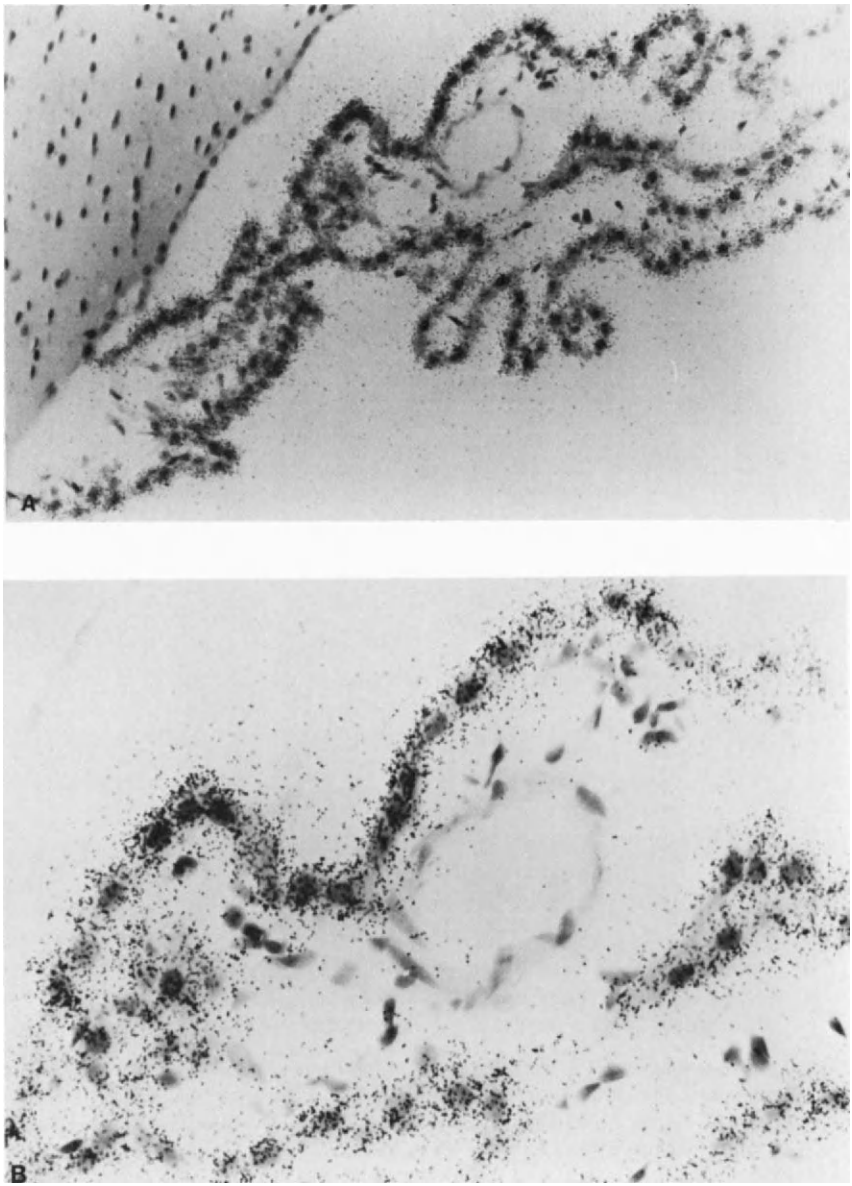


Fig. 16. *In situ* hybridization followed by autoradiography of a section of rat choroid plexus in which mRNA was hybridized *in situ* to transthyretin cDNA labeled with ^{35}S. (B) is a higher magnification of a central region of (A). Silver grains are seen only above the epithelial cells. Ependyma and stroma of the choroid villi, in particular vascular endothelial cells, are free of silver grains.

How is blood transthyretin related then to transthyretin in the cerebrospinal fluid? The "free thyroid hormone" hypothesis claims that free, but not protein-bound, thyroid hormone determines the characteristics of thyroid hormone action on cells (for review, see Ingbar and Woeber, 1981). The main target organs for thyroid hormones, indicated, for example, by stimulation of uptake or stimulation of other metabolic activities, are the liver, the kidney, and the adrenal glands. When ^{125}I-labeled transthyretin was injected into the bloodstream, a pattern was obtained for the binding of transthyretin to various tissues, which

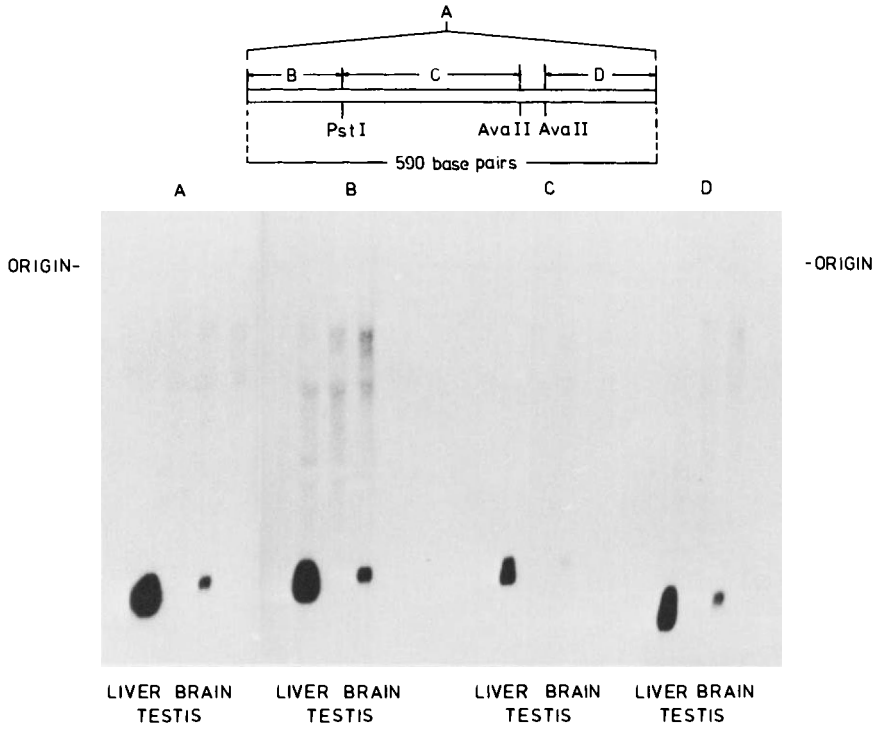


Fig. 17. Northern blot analysis of transthyretin mRNA (= prealbumin mRNA) from liver and brain. Polyadenylated RNA from liver, brain, and testes, 7 μg per track, was subjected to electrophoresis in 1.4% agarose gel containing formaldehyde. Panel A shows the result obtained when ^{32}P-labeled transthyretin cDNA containing the whole coding region plus some flanking regions was used for hybridization. Panels B, C, and D show the results obtained when subfragments of transthyretin cDNA, representing different sections of transthyretin cDNA, as indicated at the top of the figure, were used for hybridization. The horizontal bar in the upper part of the figure represents the insert of transthyretin cDNA in bacteriophage λgt11Amp3 used for cloning; the sites of cleavage by restriction enzymes are also shown. The mRNAs, separated by electrophoresis, which hybridized to any cDNA probe containing a part of the coding sequence for transthyretin were found to be identical in size. From Dickson *et al.* (1985b), with permission.

TABLE VI

Relative Levels of mRNA for Transferrin and Transthyretin (=Prealbumin) in Choroid Plexus, Brain Tissue Free of Choroid Plexus ("Rest of Brain"), and Liver as Determined by Hybridization to Specific cDNA Probes[a]

Tissue	mRNA for	
	Transferrin	Transthyretin
Choroid plexus	100	12,000
"Rest of brain"	0.01	0
Liver	100	100

[a]Ribonuclease-treated samples gave values similar to background. A transferrin cDNA of 1540 base pairs (Aldred *et al.*, 1984) and a transthyretin cDNA of 600 base pairs (Dickson *et al.*, 1985b) were used for preparation of ^{32}P-labeled cDNA probes. From Dickson *et al.* (1985a), with permission.

was similar to the pattern of thyroid hormone responsiveness of tissues (Fig. 20). Interestingly, the fourth highest amount of binding of transthyretin was seen in choroid plexus. The question of specific binding of transthyretin to cell surface sites of target or uptake organs for thyroid hormones seems to deserve further investigation. Crude membrane fractions from rat liver were found indeed to exhibit specific binding of transthyretin (P. W. Dickson and G. Schreiber, unpublished, 1985).

For transferrin, receptors have been found on the endothelium of brain capillaries (Jefferies *et al.*, 1984). Special conditions, beyond the scope of this review, exist for the transport and synthesis of plasma proteins in the developing brain.

VI. Human Acute-Phase Proteins and Their Physiopathological Role and Clinical Significance. Modulation of the Acute-Phase Response

A. General Remarks. Role of Malnutrition and Liver Damage

The acute-phase response consists of a number of well-integrated reactions initiated by one or several related signal proteins released from monocytes. It occurs in principally similar ways whenever monocytes or monocyte-derived

cells are activated, e.g., during destruction and degradation of tissue (Table VII). Monitoring the acute-phase response by one of the various methods available (measuring of the sedimentation rates of erythrocytes, electrophoretic separation and quantitation of serum proteins, immunochemical determination of specific proteins in blood plasma, etc.) will therefore provide an indication for the presence of a disease of an organic nature as opposed to psychosomatic disturbances (Werner and Cohnen, 1967; Werner, 1969; Cooper and Ward, 1979). The effect on plasma protein concentrations has been demonstrated, for example, for injuries (Owen, 1967; Clarke *et al.*, 1971), burned patients (Zeineh and Kukral, 1970), surgical trauma (Aronsen *et al.*, 1971), myocardial infarction (Johannsson *et al.*, 1972), carcinomatous diseases (Cooper and Stone, 1979; Baskies *et al.*, 1979), infection during leukemia (Mackie *et al.*, 1979), inoculation hepatitis

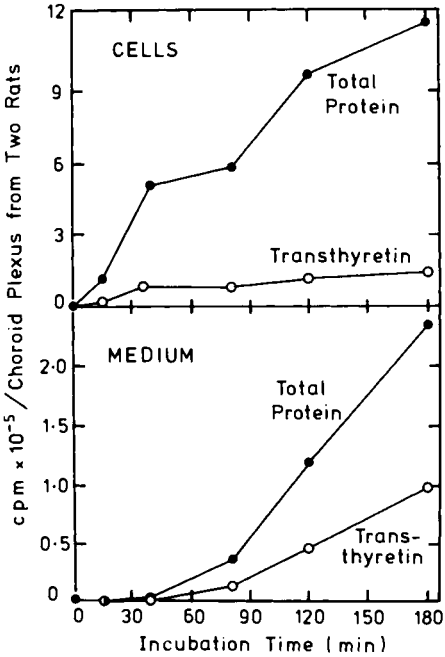


Fig. 18. Incorporation of L-[U-^{14}C]leucine into proteins secreted by choroid plexus incubated *in vitro*. Upper figure: [^{14}C]leucine incorporated into transthyretin and total protein in choroid plexus cells. Lower figure: [^{14}C]leucine incorporated into transthyretin and total protein in the incubation medium. Abscissa, incubation time; ordinate, [^{14}C]leucine incorporated into protein. Incubation conditions were as described by Schreiber and Schreiber (1972). Choroid plexus from two rats was used per time point. In control incubations 20 μ*M* cycloheximide inhibited incorporation into total protein by 97% (curve not shown). Transthyretin was measured after isolation by immunoprecipitation. From Dickson *et al.* (1986a), with permission.

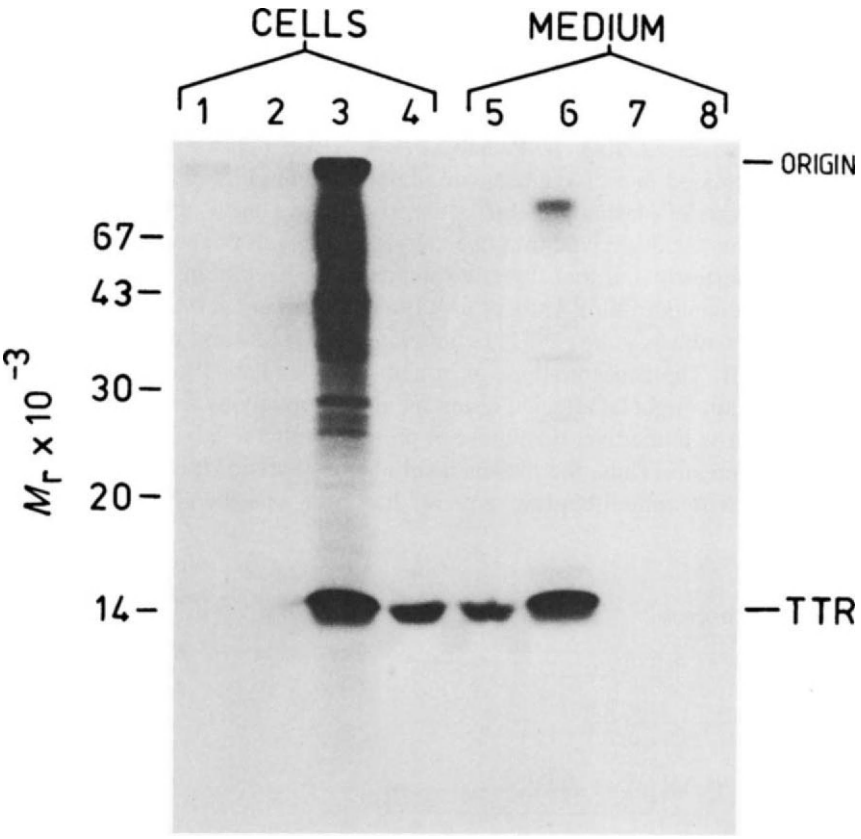


Fig. 19. Synthesis and secretion of proteins by isolated choroid plexus incubated for 180 min with L-[U^{14}C]leucine. Electrophoresis of proteins in 15% polyacrylamide gel containing 0.1% sodium dodecyl sulfate was followed by fluorography. Lanes 1 and 8, cell homogenate and medium, respectively, of choroid plexus incubated with 20 μM cycloheximide. Lanes 2 and 7, protein isolated by immunoprecipitation from cell homogenate and medium, respectively, with γ-conglutinin antiserum after previous addition of 24 μg γ-conglutinin. Lanes 3 and 6, proteins isolated from tissue homogenate and medium, respectively. Lanes 4 and 5, immunoprecipitation of tissue homogenate and medium, respectively, with transthyretin antiserum after adding 75 μl rat serum as carrier. The positions of authentic transthyretin (TTR) from rat serum and molecular weight standards are indicated. From Dickson *et al.* (1986a), with permission.

(Kindmark and Laurell, 1972), rheumatic disorders (Denko and Gabriel, 1979), gingivitis (Norman *et al.*, 1979), and Crohn's disease and ulcerative colitis (Weeke and Jarnum, 1971).

Despite the generally uniform nature of the acute-phase response, certain variations will be observed when one or more of the components of the system

are subjected to additional interfering influences. Since the acute-phase proteins of the blood plasma are synthesized and secreted by the liver, an impairment of the function of the liver will produce a modified acute-phase response. The underlying base for an observed hypoalbuminemia, for example, could be a chronic liver disease (Zilva and Pannall, 1979). Alcohol has also been shown to lead to a decreased rate of synthesis of albumin (Rothschild *et al.*, 1971). The rate of synthesis of plasma proteins in the liver is, of course, dependent on the supply of amino acids derived from the diet. Plasma levels of several proteins, in particular transferrin and transthyretin (prealbumin), have been observed to decrease during malnutrition (Antia *et al.*, 1968; McFarlane *et al.*, 1969; Gabr *et al.*, 1971; Ingenbleek *et al.*, 1972; Grant *et al.*, 1973; Masawe and Rwabwogo-Atenyi, 1973). The concentrations of proteins with a relatively fast turnover in the bloodstream might be affected faster and more strongly by a decrease in their rate of synthesis in the liver than those of proteins with a relatively low turnover in the bloodstream. Thus, the plasma level of transthyretin (prealbumin), alone or complexed to retinol-binding protein, has been considered to be a more

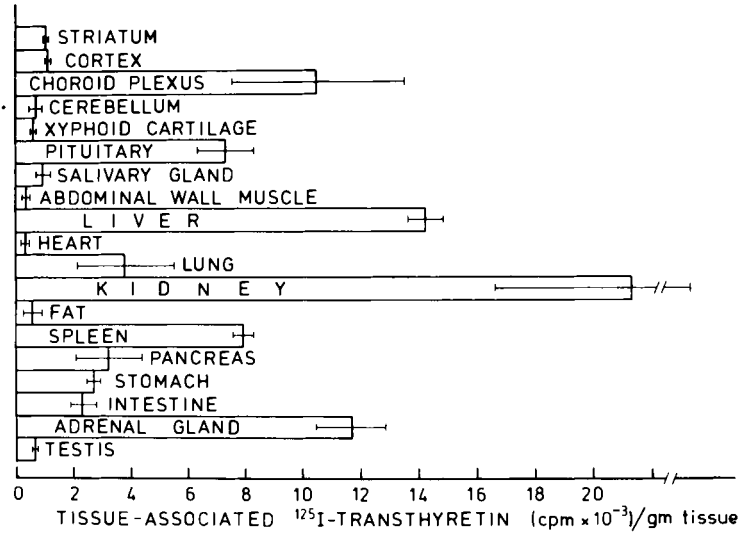


Fig. 20. Binding of transthyretin by various tissues. Eight rats were anaesthetized with 1 ml per animal of intraperitoneally injected equithesin. The abdominal cavity was opened and either ^{125}I-labeled transthyretin or ^{125}I-labeled albumin (2.5×10^6 cpm in 0.5 ml 0.9% NaCl per rat) was injected into the caval vein. After 10 min, 4 ml of blood was withdrawn into 1 ml of 3.5% trisodium citrate. The tissues indicated in the figure were removed immediately thereafter and homogenized. The amount of ^{125}I-labeled albumin in each sample was used to calculate the blood content in tissues. Tissue-bound ^{125}I-labeled transthyretin was obtained after correcting for the ^{125}I-labeled transthyretin originating from blood in the tissue sample.

TABLE VII

Levels of mRNAs for Transthyretin, Albumin, α_1-Acid Glycoprotein, and Major Acute-Phase α_1-Protein in Liver and Choroid Plexus during the Acute-Phase Response to Inflammation[a]

Protein and condition	Liver		Choroid plexus	
	cpm/g	%	cpm/mg	%
Transthyretin				
Healthy	295 ± 45	100	8,926 ± 605	100
Burn	245 ± 19	83	8,448 ± 903	95
Talc	170 ± 6	58	9,203 ± 742	103
Turpentine	80 ± 20	27	7,660 ± 277	86
Albumin				
Healthy	2,921 ± 119	100	Not detected	
Burn	2,275 ± 150	78		
Talc	1,628 ± 146	56		
Turpentine	1,662 ± 128	57		
α_1-Acid Glycoprotein				
Healthy	681 ± 93	100	Not detected	
Burn	5,853 ± 249	859		
Talc	15,266 ± 775	2,241		
Turpentine	21,770 ± 851	3,196		
Major Acute-Phase α_1-Protein				
Healthy	72 ± 17	100	Not detected	
Burn	1,432 ± 60	1,998		
Talc	1,759 ± 122	2,454		
Turpentine	2,364 ± 81	3,297		

[a]For five rats per group mRNA was measured in tissue extracts by dot hybridization of specific [^{32}P]cDNA to mRNAs. Backgrounds were subtracted before calculation. Means ± standard errors are given. Inflammation was induced by subcutaneous injection of mineral turpentine or intraperitoneal injection of a suspension of talcum in physiological saline or by producing limited superficial burns 24 hr before sacrifice. From Dickson *et al.* (1986a,b), with permission.

sensitive indicator for the assessment of malnutrition than the plasma concentration of albumin (Ingenbleek *et al.*, 1972, 1975a,b; Smith *et al.*, 1973; Gofferje, 1978; Schwandt *et al.*, 1979; Shetty *et al.*, 1979; Ingenbleek, 1982). An interesting recent addition to the spectrum of acute-phase proteins are some of the apolipoproteins. Apolipoproteins are also synthesized in tissues other than the liver, e.g., in the intestine. Amount of food and composition of diet influence the concentration of apolipoproteins in plasma. A discussion of the clinical importance of plasma lipoprotein levels is beyond the scope of this chapter (for reviews, see Eisenberg and Levy, 1975; Blum *et al.*, 1977; Gordon *et al.*, 1977; Osborne and Brewer, 1977; Smith *et al.*, 1978; Owen and McIntyre, 1982; Schwandt, 1982). Apolipoproteins are the subject of Chapter 3.

B. Nephrotic Syndrome, Protein-Losing Enteropathy, and Genetic Analbuminemia

Another clinical condition leading to hypoproteinemia, which could mask or modulate certain features of the pattern of plasma protein concentrations during the acute-phase response, is the nephrotic syndrome. Albumin, α_1-acid glycoprotein, transferrin, and α_1-antitrypsin may all be excreted in the urine, whereas proteins of larger size tend to be retained in the bloodstream (Schultze and Heremans, 1966). Albumin losses during the nephrotic syndrome can be compensated partly by an increased synthesis of albumin in the liver (Marsh and Drabkin, 1958). A large increase in the loss of plasma proteins into the gastrointestinal tract can occur during stomach and intestinal disorders ("protein-losing enteropathy"), with ensuing hypoproteinemia (Rothschild *et al.*, 1969b; Waldmann *et al.*, 1969; Yssing *et al.*, 1969; Beathard, 1982). Finally, analbuminemia can be caused by a genetic deficiency in the expression of the albumin gene. However, this does not lead to any severe consequences for health (Dammacco *et al.*, 1980). An analbuminemic strain of rats was analyzed in greater detail. A seven-base-pair deletion in an intron of the albumin gene in those rats (Esumi *et al.*, 1983) led to the production of an albumin precursor mRNA which was not processed and transported into the cytoplasm of the liver cells (Esumi *et al.*, 1982). The concentration of total protein in the plasma of the analbuminemic rats was about the same as that in healthy rats (Esumi *et al.*, 1979). An increase in the concentration of other plasma proteins, such as transferrin and immunoglobulin G, compensated for the absence of albumin. Injected ^{3}H-labeled albumin was cleared from the bloodstream of analbuminemic rats with a half-life of 8.0 days compared with a value of about 3.5 days measured for healthy rats.

C. Disseminated Intravascular Coagulation and Hemolysis

Disseminated intravascular coagulation can be the cause of an acute decrease in the plasma concentration of fibrinogen despite a simultaneous increase in the rate of fibrinogen synthesis in the liver due to an acute-phase response.

Similarly, hemopexin and haptoglobin both increase in plasma concentration during the acute-phase response (see Table II and Section II,B,2,a,b). However, the two proteins bind heme and hemoglobin, respectively, during intravascular hemolysis (Fink *et al.*, 1967; Daniels, 1975). If tissue damage or inflammation is complicated by intravascular hemolysis, an acute-phase-type increase in the plasma level of haptoglobin or hemopexin might be prevented by formation of haptoglobin–hemoglobin and hemopexin–heme complexes and their removal from the bloodstream by reticuloendothelial and liver cells, respectively.

D. Consequences of Insufficient Plasma Concentrations or of Molecular Modifications of Antiproteinases

Various mutations in the gene for α_1-antitrypsin can lead to an impairment of the synthesis and secretion of α_1-antitrypsin by the liver, resulting in severe lowering of the plasma concentration (for reviews, see Norum *et al.*, 1977; Morse, 1978; Jeppsson, 1978). Clinical consequences of the deficiency of α_1-antitrypsin in plasma are chronic obstructive bronchitis and emphysema (Guenter *et al.*, 1968; Meiers *et al.*, 1968; Stein *et al.*, 1971; Stevens *et al.*, 1971; Janoff, 1985). The pathogenetic mechanism of the disease may be insufficient inhibition of proteinases released from leukocytes during minor inflammation in the lung with ensuing destruction of lung tissue (Eriksson, 1965; Kueppers and Bearn, 1966). A second clinical feature in some patients suffering from α_1-antitrypsin deficiency is the development of hepatic cirrhosis (Sharp *et al.*, 1969). Abnormal α_1-antitrypsin accumulates intracellularly in the liver (Gordon *et al.*, 1972; Lieberman *et al.*, 1972). This accumulating α_1-antitrypsin has an abnormal carbohydrate moiety (Bell and Carrell, 1973; Jeppsson *et al.*, 1975).

An interesting variant of α_1-antitrypsin is α_1-antitrypsin Pittsburgh (Lewis *et al.*, 1978). A change in one amino acid (Owen *et al.*, 1983) converts the proteolytic specificity of α_1-antitrypsin from the normal anti-elastase specificity to antithrombin specificity. The result is a thrombin-inhibiting protein with the regulatory characteristics of an acute-phase protein. Therefore, the clinical picture is characterized by severe, repetitive hemolytic attacks in situations where an acute-phase response occurs.

E. Extracellular Deposition of Plasma Protein-Derived Fibrous Proteins: Amyloidosis and Alzheimer's Disease

In various protracted diseases, such as rheumatoid arthritis, multiple myelomas, osteomyelitis, and others, an extracellular deposition of fibrous protein is found in one or more sites of the body. The deposited fibrous protein is characterized histochemically by staining with Congo red, birefringence, and crystal violet metachromasia. The latter is due to mucopolysaccharides associated with the protein fibrils (Cooper, 1974). The fibrous protein, called amyloid, is probably derived from plasma proteins as precursors (Glenner, 1980). The protein is predominantly in the β-pleated sheet conformation (Eanes and Glenner, 1968; Termine *et al.*, 1972). The accumulated protein masses can eventually interfere with the function of parenchymatous organs, usually the kidney and the liver (see Gorevic and Franklin, 1981; Cohen, 1983). The major cause of death in generalized amyloidosis is renal failure.

A correlation between the protein deposition and aging is well documented (Wright *et al.*, 1969). In particular, without exception such deposition was found in the choroid plexus of older patients (Zalka, 1928). Also the uptake of amino acids and peptides by the choroid plexus has been shown to change in aging rats (Huang, 1984).

In Alzheimer's disease, the so-called "senile plaques," which are associated with the neurofibrillary tangles (see Schoene, 1979), consist of a central extracellular core of fibrous protein surrounded by an accumulation of cellular processes and enlarged presynaptic axon terminals (Gonatas *et al.*, 1967). Alzheimer's disease is the most common cause of adult-onset dementia (for review, see Price *et al.*, 1985). The senile plaques are found in small numbers in the brains of apparently healthy aged monkeys and humans, but they are abundant in patients suffering from Alzheimer's disease, particularly in the cortex, the hippocampus, and the amygdala. There is a clear correlation between the abundance of plaques and the severity of dementia (Tomlinson *et al.*, 1970).

Based on the precursor plasma protein, one can distinguish five types of amyloid proteins. In primary and myeloma-associated amyloidosis the deposited protein has similarities with the light chains of immunoglobulins. A second type of amyloid protein is derived from prohormones, e.g., for insulin, glucagon, and thyrocalcitonin (Sletten *et al.*, 1976).

The other three amyloid proteins are related to proteins discussed in this chapter. In chronic infections or inflammatory diseases and the so-called Familial Mediterranean Fever, the deposited amyloid is probably produced from serum amyloid A-related protein by proteolytic removal of the carboxy-terminal portion of the serum amyloid A protein, since the first 76 amino acids from the N-terminus are identical for the two proteins (cf. also Section II,B,4,d and Gorevic and Franklin, 1981).

Amyloid P-component (see Section II,B,4,e), related in structure to C-reactive protein, but inducible during the acute-phase response only in mice, is found in an amount of about 5 to 10% of all amyloid deposits in both humans and mice.

Variants of transthyretin (prealbumin) have been found in the amyloid fibril protein isolated from patients with familial amyloidotic polyneuropathy (Costa *et al.*, 1978), originally observed in Portugal (Saraiva *et al.*, 1984) but also in patients of Swedish (Dwulet and Benson, 1984) and Jewish origin (Pras *et al.*, 1983). Transthyretin (prealbumin) has also been reported to be the major constituent in the neuritic plaques associated with the neurofibrillary tangles and lesions in the brains of patients suffering from Alzheimer's disease (Shirahama *et al.*, 1982).

The demonstration that the synthesis of transthyretin amounts to 50% of total protein secreted and 20% of total protein synthesized by the choroid plexus (see Section V) suggests that the choroid plexus plays a role in providing transthyretin

(prealbumin) for the extracellular space in the brain. The investigation of protein synthesis in the choroid plexus during conditions leading to the deposition of amyloid protein seems, therefore, to be desirable.

VII. Concluding Remarks

The common functional denominator of plasma proteins is to provide and maintain a transport and communication system creating an extracellular humoral environment of appropriate composition in the body. It is fitting that, with the exception of the immunoglobulins and certain apolipoproteins, most of the plasma proteins of the bloodstream are synthesized in the liver, the general control organ for the concentration of many other compounds in the blood. Three main functional groups can be distinguished among plasma proteins. The first is that of the transport proteins. It includes albumin, transferrin, ceruloplasmin, hemopexin, haptoglobin, transthyretin, retinol-binding protein, transcortin, transcobalamin, and others. The function of the second main group is to maintain and protect the circulatory system. This group contains the proteins involved in blood coagulation, which seal off disrupted sections of blood vessels, and the antiproteinases, which neutralize the effect of proteinases liberated, for example, in damaged tissue undergoing proteolytic degradation. The third group of plasma proteins made in the liver contains proteins which are precursor substrates for the generation of "signal" compounds of lower molecular weight, such as angiotensin and kinins. Angiotensin and kinins are formed by partial specific proteolysis from precursor proteins, angiotensinogen, and kininogens.

The destruction of a part of the body, for example, by inflicting a wound or inducing an inflammation, poses a severe challenge to the system maintaining homeostasis. The body responds with a well-coordinated set of changes, the so-called general acute-phase reaction. Part of this is a rearrangement in the pattern of concentrations of proteins in the bloodstream. The mechanism leading to the change in the pattern of concentrations of plasma proteins is a change in their rates of synthesis in the liver.

The stability of plasma proteins seems to remain unaffected by the acute-phase response. Thus, inducing an experimental inflammation is a convenient and reproducible way to provide alterations in the synthesis rates of plasma proteins for studies of the control of gene expression. The immediate cause for the change in plasma protein synthesis rates in the liver in response to trauma or inflammation elsewhere in the body is a change in the levels of mRNAs. Whether this is a consequence of altered gene transcription or whether and how far the altered stability of mRNAs contributes is an open question.

Some areas in the body have their own extracellular environment of special composition. One of these areas is the brain. It is separated from the body by the

blood/brain barrier and the blood/choroid plexus/cerebrospinal fluid barrier. An extracellular environment with a composition closely similar to that of blood plasma would be incompatible with the proper functioning of the central nervous system. For example, it is mandatory that the neurons of the central nervous system are protected from the large changes in concentration of amino acids such as glutamate, aspartate, and glycine, or of the hormones adrenaline and noradrenaline, which can occur in the bloodstream.

Another area of the body with special requirements for the composition of the extracellular milieu is the growing fetus. Supply of substrates and control by hormones have to be adjusted to the quite different needs in the maternal and the fetal organisms. It is understandable then that both brain and fetus are separated from the main body by a system of one or more cell layers, although several logistic problems such as the transport of sparely water-soluble compounds by carrier proteins are clearly similar in cerebrospinal fluid and in maternal and fetal blood plasma. Thus, similar protein species are found in the three extracellular fluids.

Recently, a large number of specific cDNA probes for various plasma proteins have become available. These probes allow the very sensitive analysis of the tissue distribution of specific plasma protein mRNAs and the study of gene expression. Both choroid plexus and yolk sac, tissues interfacing different extracellular compartments, express plasma protein genes very actively and, in some cases, even more strongly than the liver. The functional implications of this synthesis of plasma proteins in the interfacing tissues are intriguing. Receptor-mediated contact between plasma proteins in the bloodstream and cells of the interfacing tissue would introduce both specificity and an opportunity for control in the chemical communication between the extracellular systems. It is a fascinating problem for future research to elucidate how the same plasma protein genes are expressed at such different sites as liver, choroid plexus, and yolk sac, where similar logistic problems have to be solved by the organism.

Acknowledgments

The preparation of this chapter would not have been possible without the expert help of Miss Elsie Gill for word processing and layout of text and tables. I thank Drs. P. Dickson and G. Howlett for critically reading the manuscript. I am very grateful to Mrs. Angela Aldred for proofreading of the manuscript and to Miss Danielle Bannister, Miss Helen Birch, and Miss Felice de Jong for assistance in preparing the reference list. Dr. R. Anderson advised and helped with the preparation of the photographs presented as Figs. 16A and 16B. *In situ* hybridization histochemistry (Figs. 15, 16A, and 16B) was carried out by F. Stauder and R. P. Hudson and computerized densitometry (Fig. 15) by Dr. F. Mendelsohn. The experimental work on which this chapter is based was supported by grants from the Australian Research Grants Scheme.and the National Health and Medical Research Council of Australia.

References

Agnew, W. F., Alvarez, R. B., Yuen, T. G. H., and Crews, A. K. (1980). *Cell Tissue Res.* **208,** 261–281.

Aisen, P. (1981). *In* "Iron in Biochemistrv and Medicine II" (A. Jacobs and M. Worwood, eds.), pp. 87–129. Academic Press, New York.

Aisen, P., and Brown, E. B. (1977). *Semin. Hematol.* **14,** 31–53.

Aldred, A. R., Howlett, G. J., and Schreiber, G. (1984). *Biochem. Biophys. Res. Commun.* **122,** 960–965.

Aleshire, S. L., Bradley, C. A., Richardson, L. D., and Parl, F. F. (1983). *J. Histochem. Cytochem.* **31,** 608–612.

Allison, A. C., ed. (1974). "Structure and Function of Plasma Proteins," Vol. 1. Plenum, New York.

Allison, A. C., ed. (1976). "Structure and Function of Plasma Proteins," Vol. 2. Plenum, New York.

Alper, C. A. (1974). *In* "Structure and Function of Plasma Proteins" (A. C. Allison, ed.), Vol. 1, pp. 195–222. Plenum, New York.

Altruda, F., Poli, V., Restagno, G., Argos, P., Cortese, R., and Silengo, L. (1985). *Nucleic Acids Res.* **13,** 3841–3859.

Anders, R. F., Natvig, J. B., Sletten, K., Husby, G., and Nordstoga, K. (1977). *J. Immunol.* **118,** 229–234.

Anderson, K. P., Martin, A. D., and Heath, E. C. (1984). *Arch. Biochem. Biophys.* **233,** 624–635.

Andus, T., Gross, V., Tran-Thi, T.-A., Schreiber, G., Nagashima, M., and Heinrich, P. C. (1983). *Eur. J. Biochem.* **133,** 561–571.

Antakly, T., Lynch, K. R., Nakhasi, H. L., and Feigelson, P. (1982). *Am. J. Anat.* **165,** 211–224.

Antia, A. U., McFarlane, H., and Soothill, J. F. (1968). *Arch. Dis. Child.* **43,** 459–462.

Armstrong, D. (1970). *Handb. Exp. Pharmakol.* [N.S.] **25,** 434–481.

Aronsen, K.-F., Ekelund, G., Kindmark, C.-O., and Laurell, C.-B. (1971). *Scand. J. Clin. Lab. Invest.* **29,** Suppl. 124, 127–136.

Atkins, E., and Bodel, P. (1972). *N. Engl. J. Med.* **286,** 27–34.

Atkins, E., and Wood, W. B., Jr. (1955). *J. Exp. Med.* **102,** 499–516.

Awai, M., and Brown, E. B. (1963). *J. Lab. Clin. Med.* **61,** 363–396.

Barlas, A., Okamoto, H., and Greenbaum, L. M. (1985). *Biochem. Biophys. Res. Commun.* **129,** 280–286.

Baskies, A. M., Chrétien, P. B., Wolf, G. T., and Weiss, J. F. (1979). *Surg. Forum.* **30,** 516–518.

Baumann, H., Firestone, G. L., Burgess, T. L., Gross, K. W., Yamamoto, K. R., and Held, W. A. (1983). *J. Biol. Chem.* **258,** 563–570.

Bausserman, L. L., Herbert, P. N., and McAdam, K. P. W. J. (1980). *J. Exp. Med.* **152,** 641–656.

Beathard, G. F. (1982). *In* "Serum Protein Abnormalities: Diagnostic and Clinical Aspects" (S. E. Ritzmann and J. C. Daniels, eds.), pp. 173–211. Liss, New York.

Beisel, W. R. (1981). *In* "Infection: The Physiologic and Metabolic Responses of the Host" (M. C. Powanda and P. G. Canonico, eds.), pp. 147–172. Elsevier/North-Holland Biomedical Press, Amsterdam.

Bell, O. F., and Carrell, R. W. (1973). *Nature (London)* **243,** 410–411.

Benditt, E. P., and Eriksen, N. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 4025–4028.

Bengmark, S., Olsson, R., and Rehnström, B. (1968). *Scand. J. Gastroenterol.* **3,** 525–528.

Bennett, I. L., Jr., and Beeson, P. B. (1953a). *J. Exp. Med.* **98,** 477–492.

Bennett, I. L., Jr., and Beeson, P. B. (1953b). *J. Exp. Med.* **98,** 493–508.

Bennett, M., and Schmid, K. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 6109–6113.

Bennhold, H., Peters, H., and Roth, E. (1954). *Verh. Dtsch. Ges. Inn. Med.* **60,** 630–634.

Bensi, G., Raugei, G., Klefenz, H., and Cortese, R. (1985). *EMBO J.* **4,** 119–126.
Benson, M. D., and Aldo-Benson, M. A. (1979). *J. Immunol.* **122,** 2077–2082.
Benson, M. D., Aldo-Benson, M. A., Shirahama, T., Borel, Y., and Cohen, A. S. (1975). *J. Exp. Med.* **142,** 236–241.
Benson, M. D., Scheinberg, M. A., Shirahama, T., Cathcart, E. S., and Skinner, M. (1977). *J. Clin. Invest.* **59,** 412–417.
Berger, F. G., and Baumann, H. (1985). *J. Biol. Chem.* **260,** 1160–1165.
Billroth, T. (1875). "Die allgemeine chirurgische Pathologie und Therapie," 7th ed., pp. 101–102. Reimer, Berlin.
Birch, H. E., Nagashima, M., Simpson, R. J., and Schreiber, G. (1983). *Biochem. Int.* **6,** 653–661.
Birch, H. E., and Schreiber, G. (1986). *Biochem. Biophys. Res. Commun.* **137,** 633–639.
Black, J. A., Chan, G. F. Q., Hew, C. L., and Dixon, G. H. (1970). *Can. J. Biochem.* **48,** 123–132.
Blatteis, C. M. (1984). *In* "Thermal Physiology" (J. R. S. Hales, ed.), pp. 539–546. Raven Press, New York.
Blombäck, B., and Hanson, L. A. (1979). "Plasma Proteins." Wiley, New York.
Blum, C. B., Levy, R. I., Eisenberg, S., Hall, M. III, Goebel, R. H., and Berman, M. (1977). *J. Clin. Invest.* **60,** 795–807.
Board, P. G., Jones, I., and Bentley, A. (1985). *Proc. Aust. Biochem. Soc.* **17,** 64.
Boguski, M. S., Elshourbagy, N., Taylor, J. M., and Gordon, J. I. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 5021–5025.
Bokisch, V. A., Dierich, M. P., and Müller-Eberhard, H. J. (1975). *Proc. Natl. Acad. Sci. U.S.A.* **72,** 1989–1993.
Boman, H., Hermodson, M., Hammond, C. A., and Motulsky, A. G. (1976). *Clin. Genet.* **9,** 513–526.
Bornstein, D. L. (1982). *Ann. N.Y. Acad. Sci.* **389,** 323–337.
Bornstein, D. L., and Walsh, E. C. (1978). *J. Lab. Clin. Med.* **91,** 236–245.
Bouma, H., III, Kwan, S.-W., and Fuller, G. M. (1975). *Biochemistry* **14,** 4787–4792.
Brady, F. O. (1982). *Trends Biochem. Sci.* **7,** 143–145.
Braell, W. A., and Lodish, H. F. (1982). *J. Biol. Chem.* **257,** 4578–4582.
Braun, G. A., Marsh, J. B., and Drabkin, D. L. (1962a). *Metab., Clin. Exp.* **11,** 957–966.
Braun, G. A., Marsh, J. B., and Drabkin, D. L. (1962b). *Biochem. Biophys. Res. Commun.* **8,** 28–32.
Brightman, M. W. (1975). *In* "The Choroid Plexus in Health and Disease" (M. G. Netsky and S. Shuangshoti, eds.), pp. 86–112. Wright, Bristol, England.
Büchner, F. (1956). "Allgemeine Pathologie," 2nd ed., pp. 313–324. Urban & Schwarzenberg, Munich.
Buttyan, R., Olsson, C. A., Sheard, B., and Kallos, J. (1983). *J. Biol. Chem.* **258,** 14366–14370.
Cabezón, T., de Wilde, M., Herion, P., Loriau, R., and Bollen, A. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 6594–6598.
Cannon, J. G., and Kluger, M. J. (1983). *Science* **220,** 617–619.
Carrell, R. W., Owen, M. C., Brennan, S. O., and Vaughan, L. (1979). *Biochem. Biophys. Res. Commun.* **91,** 1032–1037.
Carrell, R. W., Boswell, D. R., Brennan, S. O., and Owen, M. C. (1980). *Biochem. Biophys. Res. Commun.* **93,** 399–402.
Celsus, A. C. (1713). *In* "De Medicina" (T. J. A. Almeloveen, ed.), Vol. 3. Joannem Wolters, Amsterdam.
Chader, G. J., and Westphal, U. (1968). *Biochemistry* **7,** 4272–4282.
Chanutin, A., Hortenstine, J. C., Cole, W. S., and Ludewig, S. (1938). *J. Biol. Chem.* **123,** 247–256.

Chasteen, N. D. (1977). *Coord. Chem. Rev.* **22,** 1–36

Chen, C.-L. C., and Feigelson, P. (1978). *J. Biol. Chem.* **253,** 7880–7885.

Cheresh, D. A., Haynes, D. H., and Distasio, J. A. (1984). *Immunology* **51,** 541–548.

Cheung, P., and Chan, L. (1983). *Nucleic Acids Res.* **11,** 3703–3715.

Chiu, K. M., Mortensen, R. F., Osmand, A. P., and Gewurz, H. (1977). *Immunology* **32,** 997–1005.

Chung, D. W., Rixon, M. W., MacGillivray, R. T. A., and Davie, E. W. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 1466–1470.

Chung, D. W., Que, B. G., Rixon, M. W., Mace, M., Jr., and Davie, E. W. (1983a). *Biochemistry* **22,** 3244–3250.

Chung, D. W., Chan, W.-Y., and Davie, E. W. (1983b). *Biochemistry* **22,** 3250–3256.

Church, W. R., Jernigan, R. L., Toole, J., Hewick, R. M., Knopf, J., Knutson, G. J., Nesheim, M. E., Mann, K. G., and Fass, D. N. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 6934–6937.

Ciejek, E. M., Tsai, M.-J., and O'Malley, B. W. (1983). *Nature (London)* **306,** 607–609.

Clamp, J. R. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 2, pp. 163–211. Academic Press, New York.

Clarke, H. G. M., Freeman, T., and Pryse-Phillips, W. (1971). *Clin. Sci.* **40,** 337–344.

Cleve, H., Alexander, K., Mitzkat, H. J., Nissen, P., and Salzmann, I. (1968). *Diabetologia* **4,** 48–55.

Cohen, A. S. (1983). *In* "Harrison's Principles of Internal Medicine" (R. G. Petersdorf, R. D. Adams, E. Braunwald, K. J. Isselbacher, J. B. Martin, and J. D. Wilson, eds.), 10th ed., pp. 368–371. McGraw-Hill, New York.

Colantuoni, V., Romano, V., Bensi, G., Santoro, C., Costanzo, F., Raugei, G., and Cortese, R. (1983). *Nucleic Acids Res.* **11,** 7769–7776.

Cole, T., Inglis, A., Nagashima, M., and Schreiber, G. (1985a). *Biochem. Biophys. Res. Commun.* **126,** 719–724.

Cole, T., Inglis, A. S., Roxburgh, C. M., Howlett, G. J., and Schreiber, G. (1985b). *FEBS Lett.* **182,** 57–61.

Cooper, E. H., and Stone, J. (1979). *Adv. Cancer Res.* **30,** 1–44.

Cooper, E. H., and Ward, A. M. (1979). *Invest. Cell Pathol.* **2,** 293–301.

Cooper, J. H. (1974). *Lab. Invest.* **31,** 232–238.

Costa, P. P., Figüeira, A. S., and Bravo, F. R. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 4499–4503.

Courtney, M., Buchwalder, A., Tessier, L.-H., Jaye, M., Benavente, A., Balland, A., Kohli, V., Lathe, R., Tolstoshev, P., and Lecocq, J.-P. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 669–673.

Courtney, M., Jallat, S., Tessier, L.-H., Benavente, A., Crystal, R. G., and Lecocq, J.-P. (1985). *Nature (London)* **313,** 149–151.

Courtoy, P. J., Lombart, C., Feldmann, G., Moguilevsky, N., and Rogier, E. (1981). *Lab. Invest.* **44,** 105–115.

Crabtree, G. R., and Kant, J. A. (1981). *J. Biol. Chem.* **256,** 9718–9723.

Crabtree, G. R., and Kant, J. A. (1982a). *Cell (Cambridge, Mass.)* **31,** 159–166.

Crabtree, G. R., and Kant, J. A. (1982b). *J. Biol. Chem.* **257,** 7277–7279.

Cromwell, S. (1963). *Protides Biol. Fluids* **11,** 484–486.

Cserr, H. F. (1971). *Physiol. Rev.* **51,** 273–311.

Dammacco, F., Miglietta, A., D'Addabbo, A., Fratello, A., Moschetta, R., and Bonomo, L. (1980). *Vox Sang.* **39,** 153–161.

Dancis, J., Braverman, N., and Lind, J. (1957). *J. Clin. Invest.* **36,** 398–404.

Daniels, J. C. (1975). *In* "Serum Protein Abnormalities: Diagnostic and Clinical Aspects" (S. E. Ritzmann and J. C. Daniels, eds.), pp. 213–241. Little, Brown, Boston, Massachusetts.

Daniels, J. C., Larson, D. L., Abston, S., and Ritzmann, S. E. (1974). *J. Trauma* **14,** 153–162.
Darcy, D. A. (1957). *Br. J. Cancer* **11,** 137–147.
Darcy, D. A. (1964). *Br. J. Exp. Pathol.* **45,** 281–293.
Darcy, D. A. (1966). *Br. J. Exp. Pathol.* **47,** 480–487.
Darlington, G. J., and Lachman, L. B. (1984). *J. Cell Biol.* **99,** 90a.
Davis, P. J., Spaulding, S. W., and Gregerman, R. I. (1970). *Endocrinology (Baltimore)* **87,** 978–986.
de Bruijn, M. H. L., and Fey, G. H. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 708–712.
Denko, C. W., and Gabriel, P. (1979). *J. Rheumatol.* **6,** 664–672.
Dente, L., Ciliberto, G., and Cortese, R. (1985). *Nucleic Acids Res.* **13,** 3941–3952.
Dickson, P. W. (1985). Ph.D. Thesis, University of Melbourne.
Dickson, P. W., Howlett, G. J., and Schreiber, G. (1982). *Eur. J. Biochem.* **129,** 289–293.
Dickson, P. W., Aldred, A. R., Marley, P. D., Tu, G.-F., Howlett, G. J., and Schreiber, G. (1985a). *Biochem. Biophys. Res. Commun.* **127,** 890–895.
Dickson, P. W., Howlett, G. J., and Schreiber, G. (1985b). *J. Biol. Chem.* **260,** 8214–8219.
Dickson, P. W., Aldred, A. R., Marley, P. D., Bannister, D., and Schreiber, G. (1986a). *J. Biol. Chem.*, **261,** 3475–3478.
Dickson, P. W., Bannister, D., and Schreiber, G. (1986b). *J. Trauma* (in press).
Diem, K., ed. (1960). "Geigy Scientific Tables," 6th ed. Geigy, A. G., Basel.
Dinarello, C. A. (1979). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **38,** 52–56.
Dinarello, C. A. (1981). *J. Exp. Med.* **153,** 1215–1224.
Dinarello, C. A., and Wolff, S. M. (1978). *N. Engl. J. Med.* **298,** 607–612.
Dinarello, C. A., Goldin, N. P., and Wolff, S. M. (1974). *J. Exp. Med.* **139,** 1369–1381.
Dinarello, C. A., Renfer, L., and Wolff, S. M. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 4624–4627.
Doolittle, R. F. (1973). *Adv. Protein Chem.* **27,** 1–109.
Doolittle, R. F. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 2, pp. 109–161. Academic Press, New York.
Doolittle, R. F. (1983). *Science* **222,** 417–419.
Doolittle, R. F. (1984). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 4, pp. 317–360. Orlando, Florida.
Drabkin, D. L., and Marsh, J. B. (1955). *J. Biol. Chem.* **212,** 623–631.
Duff, G. W., and Durum, S. K. (1983). *Nature (London)* **304,** 449–451.
Dugaiczyk, A., Law, S. W., and Dennison, O. E. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 71–75.
Durnam, D. M., Perrin, F., Gannon, F., and Palmiter, R. D. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77,** 6511–6515.
Durnam, D. M., Hoffman, J. S., Quaife, C. J., Benditt, E. P., Chen, H. Y., Brinster, R. L., and Palmiter, R. D. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 1053–1056.
Dwulet, F. E., and Benson, M. D. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 694–698.
Dwulet, F. E., and Putnam, F. W. (1981a). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 790–794.
Dwulet, F. E., and Putnam, F. W. (1981b). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 2805–2809.
Dyck, R. F., Lockwood, C. M., Kershaw, M., McHugh, N., Duance, V. C., Baltz, M. L., and Pepys, M. B. (1980). *J. Exp. Med.* **152,** 1162–1174.
Dziegielewska, K. M., Evans, C. A. N., Lai, P. C. W., Lorscheider, F. L., Malinowska, D. H., Møllgård, K., and Saunders, N. R. (1981). *Dev. Biol.* **83,** 193–200.
Eanes, E. D., and Glenner, G. G. (1968). *J. Histochem. Cytochem.* **16,** 673–677.
Eddington, C. L., Upchurch, H. F., and Kampschmidt, R. F. (1971). *Proc. Soc. Exp. Biol. Med.* **136,** 159–164.
Eddington, C. L., Upchurch, H. F., and Kampschmidt, R. F. (1972). *Proc. Soc. Exp. Biol. Med.* **139,** 565–569.

Edwards, A. J., Lee, M., and Harcourt, G. (1977). *Clin. Oncol.* **3,** 65–73.
Edwards, K., Fleischer, B., Dryburgh, H., Fleischer, S., and Schreiber, G. (1976a). *Biochem. Biophys. Res. Commun.* **72,** 310–318.
Edwards, K., Schreiber, G., Dryburgh, H., Millership, A., and Urban, J. (1976b). *Cancer Res.* **36,** 3113–3118.
Edwards, K., Schreiber, G., Dryburgh, H., Urban, J., and Inglis, A. S. (1976c). *Eur. J. Biochem.* **63,** 303–311.
Eisenberg, S., and Levy, R. I. (1975). *Adv. Lipid Res.* **13,** 1–89.
Eriksson, S. (1965). *Acta Med. Scand.* **177,** Suppl. 432.
Esnard, F. (1985). Ph.D. Thesis, Université François-Rabelais, Tours.
Esnard, F., and Gauthier, F. (1983). *J. Biol. Chem.* **258,** 12443–12447.
Esumi, H., Sato, S., Okui, M., Sugimura, T., and Nagase, S. (1979). *Biochem. Biophys. Res. Commun.* **87,** 1191–1199.
Esumi, H., Takahashi, Y., Sekiya, T., Sato, S., Nagase, S., and Sugimira, T. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 734–738.
Esumi, H., Takahashi, Y., Sato, S., Nagase, S., and Sugimura, T. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 95–99.
Farrar, J. J., and Hilfiker, M. L. (1982). *Fed. Proc, Fed. Am. Soc. Exp. Biol.* **41,** 263–268.
Feinberg, R. F., Sun, L.-H. K., Ordahl, C. P., and Frankel, F. R. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5042–5046.
Fink, D. J., Petz, L. D., and Black, M. B. (1967). *JAMA, J. Am. Med. Assoc.* **199,** 109–112.
Foucrier, J., Kraemer, M., Vassy, J., and Chalumeau, M. T. (1979). *Cell Differ.* **8,** 39–48.
Fowlkes, D. M., Mullis, N. T., Comeau, C. M., and Crabtree, G. R. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 2313–2316.
Frieden, E. (1979). *In* "Copper in the Environment" (J. O. Nriagu, ed.), Part II, pp. 241–284. Wiley, New York.
Friedman, M. J. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5421–5424.
Fries, E., Gustafsson, L., and Peterson, P. A. (1984). *EMBO J.* **3,** 147–152.
Fuller, G. (1983). *In* "Plasma Protein Secretion by the Liver" (H. Glaumann, T. Peters, Jr., and C. Redman, eds.), pp. 405–422. Academic Press, New York.
Fuller, G. M., Baglia, F. A., Rupp, R. G., Bohmfalk, J. F., Nickerson, J. M., Miller, R. L., and Milner, P. W. (1979). *J. Cell Biol.* **83,** 419a.
Fung, W.-P., and Schreiber, G. (1986). In preparation.
Gabr, M., El-Hawary, M. F. S., and El-Dali, M. (1971). *J. Trop. Med. Hyg.* **74,** 216–221.
Ganrot, P. O., and Bjerre, B. (1967). *Acta Obst. Gynecol. Scand.* **46,** 126–137.
Gauthier, F., and Mouray, H. (1976). *Biochem. J.* **159,** 661–665.
Gerbeck, C. M., Bezkorovainy, A., and Rafelson, M. E., Jr. (1967). *Biochemistry* **6,** 403–411.
Gery, I., and Waksman, B. H. (1972). *J. Exp. Med.* **136,** 143–155.
Gery, I., Gershon, R. K., and Waksman, B. H. (1971). *J. Immunol.* **107,** 1778–1780.
Gery, I., Gershon, R. K., and Waksman, B. H. (1972). *J. Exp. Med.* **136,** 128–142.
Giblett, E. (1974). *In* "Structure and Function of Plasma Proteins" (A. C. Allison, ed.), Vol. 1, pp. 55–72. Plenum, New York.
Giulian, D., and Lachman, L. B. (1985). *Science* **228,** 497–499.
Glaumann, H., Peters, T., Jr., and Redman, C., eds. (1983). "Plasma Protein Secretion by the Liver." Academic Press, New York.
Glenner, G. G. (1980). *N. Engl. J. Med.* **302,** 1283–1292.
Gofferje, H. (1978). *Med. Lab. (Behringwerke)* **5,** 38–44.
Goldberg, A. L., Baracos, V., Rodemann, P., Waxman, L., and Dinarello, C. (1984). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **43,** 1301–1306.
Goldstein, L. A., and Heath, E. C. (1984). *J. Biol. Chem.* **259,** 9212–9217.

Gonatas, N. K., Anderson, W., and Evangelista, I. (1967). *J. Neuropathol. Exp. Neurol.* **26,** 25–39.
Gordon, A. H. (1966). *Eur. J. Biochem.* **2,** 19–31.
Gordon, A. H. (1976). *Biochem. J.* **159,** 643–650.
Gordon, A. H., and Darcy, D. A. (1967). *Br. J. Exp. Pathol.* **48,** 81–89.
Gordon, A. H., and Limaos, E. A. (1979). *Br. J. Exp. Pathol.* **60,** 441–446.
Gordon, A. H., and Louis, L. N. (1969). *Biochem. J.* **113,** 481–488.
Gordon, H. W., Dixon, J., Rogers, J. C., Mittman, C., and Lieberman, J. (1972). *Hum. Pathol.* **3,** 361–370.
Gordon, J. I., Smith, D. P., Alpers, D. H., and Strauss, A. W. (1982). *J. Biol. Chem.* **257,** 8418–8423.
Gordon, S. G., and Cross, B. A. (1981). *J. Clin. Invest.* **67,** 1665–1671.
Gordon, T., Castelli, W. P., Hjortland, M. C., Kannel, W. B., and Dawber, T. R. (1977). *Am. J. Med.* **62,** 707–714.
Gorevic, P. D., and Franklin, E. C. (1981). *Annu. Rev. Med.* **32,** 261–271.
Grant, D. B., Hambley, J., Becker, D., and Pimstone, B. L. (1973). *Arch. Dis. Child.* **48,** 596–600.
Greenbaum, L. M. (1984). *Biochem. Pharmacol.* **33,** 2943–2944.
Grieninger, G., Hertzberg, K. M., and Pindyck, J. (1978). *Proc. Natl. Acad. Sci. U.S.A.* **75,** 5506–5510.
Grieninger, G., Plant, P. W., Liang, T. J., Kalb, R. G., Amrani, D., Mosesson, M. W., Hertzberg, K. M., and Pindyck, J. (1983). *Ann. N.Y. Acad. Sci.* **408,** 469–489.
Griffin, E. E., and Miller, L. L. (1973). *J. Biol. Chem.* **248,** 4716–4723.
Griffin, E. E., and Miller, L. L. (1974). *J. Biol. Chem.* **249,** 5062–5069.
Guenter, C. A., Welch, M. H., Russell, T. R., Hyde, R. M., and Hammarsten, J. F. (1968). *Arch. Intern. Med.* **122,** 254–257.
Guidotti, G., Clerici, E., Sambo, G., and Bazzano, E. (1959). *Experientia* **15,** 55–56.
Gutteridge, J. M. C. (1978). *Ann. Clin. Biochem.* **15,** 293–296.
Haddy, F. J., Emerson, T. E., Jr., Scott, J. B., and Daugherty, R. M., Jr., (1970). *Handb. Exp. Pharmakol.* [N.S.] **25,** 362–384.
Haider, M., and Tarver, H. (1969). *J. Nutr.* **99,** 433–445.
Haugen, T. H., Hanley, J. M., and Heath, E. C. (1981). *J. Biol. Chem.* **256,** 1055–1057.
Heide, K., and Schwick, H. G. (1973). *Angew. Chem., Int. Ed. Engl.* **12,** 721–733.
Heilmeyer, L., and Begemann, H. (1955). *In* "Lehrbuch der Inneren Medizin" (L. Heilmeyer, ed.), pp. 351–420. Springer-Verlag, Berlin and New York.
Heilmeyer, L., Keller, W., Vivell, O., Keiderling, W., Betke, K., Wöhler, F., and Schultze, H. E. (1961). *Dtsch. Med. Wochenschr.* **86,** 1745–1751.
Heim, W. G., and Ellenson, S. R. (1967). *Nature (London)* **213,** 1260–1261.
Heimburger, N. (1972). *In* "Pulmonary Emphysema and Proteolysis" (C. Mittman, ed.), pp. 307–310. Academic Press, New York.
Heimburger, N., and Haupt, H. (1965). *Clin. Chim. Acta* **12,** 116–118.
Heimburger, N., Heide, K., Haupt, H., and Schultze, H. E. (1964). *Clin. Chim. Acta* **10,** 293–307.
Henning, S. J. (1981). *Am. J. Physiol.* **241,** G199–G214.
Hertzberg, K. M., Pindyck, J., Mosesson, M. W., and Grieninger, G. (1981). *J. Biol. Chem.* **256,** 563–566.
Hitzig, W. H. (1977). "Plasmaproteine," 2nd ed. Springer-Verlag, Berlin and New York.
Hodgson, H. J. F., Potter, B. J., and Jewell, D. P. (1977). *Clin. Exp. Immunol.* **28,** 490–495.
Hoffenberg, R., and Ramsden, D. B. (1983). *Clin. Sci.* **65,** 337–342.
Hoffenberg, R., Black, E., and Brock, J. F. (1966). *J. Clin. Invest.* **45,** 143–152.
Hoffenberg, R., Gordon, A. H., and Black, E. G. (1971). *Biochem. J.* **122,** 129–134.

Homburger, F. (1945). *J. Clin. Invest.* **24,** 43–45.
Hooper, D. C., Steer, C. J., Dinarello, C. A., and Peacock, A. C. (1981). *Biochim. Biophys. Acta* **653,** 118–129.
Home, C. H. W., Mallinson, A. C., Ferguson, J., and Goudie, R. B. (1971). *J. Clin. Pathol.* **24,** 464–466.
Housley, J. (1968). *J. Clin. Pathol.* **21,** 27–31.
Howlett, G., Nagashima, M., and Schreiber, G. (1981). *Biochem. Int.* **3,** 93–100.
Hsieh, H. S., and Frieden, E. (1975). *Biochem. Biophys. Res. Commun.* **67,** 1326–1331.
Huang, J. T. (1984). *Age* **7,** 63–65.
Hudig, D., and Sell, S. (1978). *Inflammation* **3,** 137–148.
Huerre, C., Uzan, G., Grzeschik, K. H., Weil, D., Levin, M., Hors-Cayla, M.-C., Boué, J., Kahn, A., and Junien, C. (1984). *Ann. Genet.* **27,** 5–10.
Hunt, L. T., and Dayhoff, M. O. (1980). *Biochem. Biophys. Res. Commun.* **95,** 864–871.
Hurley, J. V. (1983). "Acute Inflmmation," 2nd ed. Churchill-Livingstone, Edinburgh and London.
Hurley, J. V., Anderson, R. McD., and Sexton, P. T. (1981). *J. Pathol.* **134,** 57–70.
Hutchcraft, C. L., Gewurz, H., Hansen, B., Dyck, R. F., and Pepys, M. B. (1981). *J. Immunol.* **126,** 1217–1219.
Ingbar, S. H., and Woeber, K. A. (1981). *In* "Textbook of Endocrinology" (R. H. Williams, ed.), 6th ed., pp. 117–247. Saunders, Philadelphia, Pennsylvania.
Ingenbleek, Y. (1982). *In* "Marker Proteins in Inflammation" (R. C. Allen, J. Bienvenu, P. Laurent, and R. M. Suskind, eds.), pp. 405–414. de Gruyter, Berlin.
Ingenbleek, Y., De Visscher, M., and De Nayer, P. (1972). *Lancet* **2,** 106–108.
Ingenbleek, Y., Van Den Schrieck, H.-G., De Nayer, P., and De Visscher, M. (1975a). *Metab., Clin. Exp.* **24,** 633–641.
Ingenbleek, Y., Van Den Schrieck, H.-G., De Nayer, P., and De Visscher, M. (1975b). *Clin. Chim. Acta* **63,** 61–67.
Jamieson, G. A. (1965). *J. Biol. Chem.* **265,** 2019–2027.
Jamieson, J. C., Kaplan, H. A., Woloski, B. M. R. N. J., Hellman, M., and Ham, K. (1983). *Can. J. Biochem. Cell Biol.* **61,** 1041–1048.
Janoff, A. (1985). *Annu. Rev. Med.* **36,** 207–216.
Jeejeebhoy, K. N., Ho, J., Mehra, R., Jeejeebhoy, J., and Bruce-Robertson, A. (1977). *Biochem. J.* **168,** 347–352.
Jefferies, W. A., Brandon, M. R., Hunt, S. V., Williams, A. F., Gatter, K. C., and Mason, D. Y. (1984). *Nature (London)* **312,** 162–163.
Jeppsson, J.-O. (1978). *Laboratoriumsbl. Med. Diagn. E. v. Behring* **28,** 45–53.
Jeppsson, J.-O., Larsson, C., and Eriksson, S. (1975). *N. Engl. J. Med.* **293,** 576–579.
Johansson, B. G., Kindmark, C.-O., Trell, E. Y., and Wollheim, F. A. (1972). *Scand. J. Clin. Lab. Invest.* **29,** Suppl. 124, 117–126.
John, D. W., and Miller, L. L. (1969). *J. Biol. Chem.* **244,** 6134–6142.
Jörnvall, H., Carlström, A., Pettersson, T., Jacobsson, B., Persson, M., and Mutt, V. (1981). *Nature (London)* **291,** 261–263.
Kägi, J. H. R., and Nordberg, M. (1979). "Metallothionein." Birkhaeuser, Basel.
Kampschmidt, R. F. (1984). *J. Leuk. Biol.* **36,** 341–355.
Kampschmidt, R. F., and Upchurch, H. F. (1962). *Proc. Soc. Exp. Biol. Med.* **110,** 191–193.
Kampschmidt, R. F., and Upchurch, H. F. (1968). *Proc. Soc. Exp. Biol. Med.* **127,** 632–635.
Kampschmidt, R. F., and Upchurch, H. F. (1969). *Am. J. Physiol.* **216,** 1287–1291.
Kampschmidt, R. F., and Upchurch, H. F. (1970a). *Proc. Soc. Exp. Biol. Med.* **133,** 128–130.
Kampschmidt, R. F., and Upchurch, H. F. (1970b). *Proc. Soc. Exp. Biol. Med.* **134,** 1150–1152.

Kampschmidt, R. F., and Upchurch, H. F. (1974). *Proc. Soc. Exp. Biol. Med.* **146,** 904–907.
Kampschmidt, R. F., and Upchurch, H. F. (1980). *Proc. Soc. Exp. Biol. Med.* **164,** 537–539.
Kampschmidt, R. F., Upchurch, H. F., and Pulliam, L. A. (1973a). *Proc. Soc. Exp. Biol. Med.* **143,** 279–283.
Kampschmidt, R. F., Upchurch, H. F., Eddington, C. L., and Pulliam, L. A. (1973b). *Am. J. Physiol.* **224,** 530–533.
Kampschmidt, R. F., Upchurch, H. F., and Pulliam, L. A. (1982). *Ann. N.Y. Acad. Sci.* **389,** 338–353.
Kanda, Y., Goodman, D. S., Canfield, R. E., and Morgan, F. J. (1974). *J. Biol. Chem.* **249,** 6796–6805.
Kant, J. A., Lord, S. T., and Crabtree, G. R. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 3953–3957.
Kant, J. A., Fornace, A. J., Jr., Saxe, D., Simon, M. I., McBride, O. W., and Crabtree, G. R. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 2344–2348.
Karin, M., and Richards, R. I. (1982). *Nucleic Acids Res.* **10,** 3165–3173.
Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., and Beato, M. (1984). *Nature (London)* **308,** 513–519.
Kato, H., Enjyoji, K.-I., Miyata, T., Hayashi, I., Oh-ishi, S., and Iwanaga, S. (1985). *Biochem. Biophys. Res. Commun.* **127,** 289–295.
Katz, J., Bonorris, G., Okuyama, S., and Sellers, A. L. (1967). *Am. J. Physiol.* **212,** 1255–1260.
Katz, J., Sellers, A. L., and Bonorris, G. (1968). *In* "Stoffwechsel der isolierten perfundierten Leber" (W. Staib and R. Scholz, eds.), pp. 100–108. Springer-Verlag, Berlin and New York.
Kedinger, C., Gniazdowski, M., Mandel, J. L., Jr., Gissinger, F., and Chambon, P. (1970). *Biochem. Biophys. Res. Commun.* **38,** 165–171.
Kelman, L., Saunders, S. J., Frith, L., Wicht, S., and Corrigall, A. (1972a). *Am. J. Clin. Nutr.* **25,** 1174–1178.
Kelman, L., Saunders, S. J., Wicht, S., Frith, L., Corrigall, A., Kirsch, R. E., and Terblanche, J. (1972b). *Biochem. J.* **129,** 805–809.
Kerr, M. A. (1981). *Biochem. Educ.* **9,** 82–88.
Keyser, J. W. (1983). *In* "Plasma Protein Secretion by the Liver" (H. Glaumann, T. Peters, Jr., and C. Redman, eds.), pp. 464–465. Academic Press, New York.
Kindmark, C.-O., and Laurell, C.-B. (1972). *Scand. J. Clin. Lab. Invest.* **29,** Suppl. 124, 105–115.
Kingston, I. B., Kingston, B. L., and Putnam, F. W. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5377–5381.
Kirsch, R. E., Frith, L., Black, E., and Hoffenberg, R. (1968). *Nature (London)* **217,** 578–579.
Kirsch, R. E., Saunders, S. J., Frith, L., Wicht, S., Kelman, L., and Brock, J. F. (1969). *Am. J. Clin. Nutr.* **22,** 1559–1562.
Kisilevesky, R., Benson, M. D., Axelrad, M. A., and Boudreau, L. (1979). *Lab. Invest.* **41,** 206–210.
Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H., and Nakanishi, S. (1983). *Nature (London)* **305,** 545–548.
Kline, R. L., Scott, J. B., Haddy, F. J., and Grega, G. J. (1973). *Am. J. Physiol.* **225,** 1051–1056.
Knott, T. J., Robertson, M. E., Priestley, L. M., Urdea, M., Wallis, S., and Scott, J. (1984). *Nucleic Acids Res.* **12,** 3909–3915.
Koj, A. (1974). *In* "Structure and Function of Plasma Proteins" (A. C. Allison, ed.), Vol. 1, pp. 73–125. Plenum, New York.
Kozak, M. S., Hahn, H. H., Lennarz, W. J., and Wood, W. B., Jr. (1968). *J. Exp. Med.* **127,** 341–357.
Kraemer, M., Vassy, J., Fourcrier, J., Rigaut, J.-P., and Chalumeau, M.-T. (1981). *Biol. Cell* **40,** 103–108.
Kueppers, F., and Bearn, A. G. (1966). *Proc. Soc. Exp. Biol. Med.* **121,** 1207–1209.

Kurachi, K., Chandra, T., Friezner Degen, S. J., White, T. T., Marchioro, T. L., Woo, S. L. C., and Davie, E. W. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6826–6830.

Kurachi, K., Long, G. L., Leicht, M., Chandra, T., Mace, M., Jr., Woo, S. L. C., and Davie, E. W. (1982). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **41,** 517.

Kurtz, D. T. (1981). *Nature (London)* **291,** 629–631.

Kurtz, D. T., and Feigelson, P. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 4791–4795.

Kurtz, D. T., Sippel, A. E., and Feigelson, P. (1976). *Biochemistry* **15,** 1031–1036.

Kurtz, D. T., Chan, K.-M., and Feigelson, P. (1978). *J. Biol. Chem.* **253,** 7886–7890.

Kushner, I., Edgington, T. S., Trimble, C., Liem, H. H., and Müller-Eberhard, U. (1972). *J. Lab. Clin. Med.* **80,** 18–25.

Kushner, I., Gewurz, H., and Benson, M. D. (1981). *J. Lab. Clin. Med.* **97,** 739–749.

Lachman, L. B. (1982). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **41,** 369.

Lachman, L. B., Atkins, E., and Kampschmidt, R. F. (1983). *In* "Interleukins, Lymphokines and Cytokines" (J. J. Oppenheim and S. Cohen, eds.), p. 441. Academic Press, New York.

Larsen, G. L., and Henson, P. M. (1983). *Annu. Rev. Immunol.* **1,** 335–359.

Laurell, A.-B. (1979). *In* "Biological Functions of Proteinases" (H. Holzer and H. Tschesche, eds.), pp. 223–232. Springer-Verlag, Berlin and New York.

Laurell, C.-B. (1947). *Acta Physiol. Scand.* **14,** Suppl. 46, 1–129.

Laurell, C.-B., and Jeppsson, J.-O. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 229–264. Academic Press, New York.

Lawrence, H. S., and Landy, M., eds. (1969). "Mediators of Cellular Immunity." Academic Press, New York.

Lebreton, J. P., Joisel, F., Raoult, J. P., Lannuzel, B., Rogez, J. P., and Humbert, G. (1979). *J. Clin. Invest.* **64,** 1118–1129.

Ledford, B. E., and Davis, D. F. (1983). *J. Biol. Chem.* **258,** 3304–3308.

Leicht, M., Long, G. L., Chandra, T., Kurachi, K., Kidd, V. J., Mace, M., Jr., Davie, E. W., and Woo, S. L. C. (1982). *Nature (London)* **297,** 655–659.

Lentner, C., ed. (1981). "Geigy Scientific Tables," 8th ed. Ciba-Geigy, Basel.

Levin, M., Franklin, E. C., Frangione, B., and Pras, M. (1972). *J. Clin. Invest.* **51,** 2773–2776.

Levin, M., Pras, M., and Franklin, E. C. (1973). *J. Exp. Med.* **138,** 373–380.

Lewis, J. H., Iammarino, R. M., Spero, J. A., and Hasiba, U. (1978). *Blood* **51,** 129–137.

Liang, T. J., and Grieninger, G. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6972–6976.

Lieberman, J., Mittman, C., and Gordon, H. W. (1972). *Science* **175,** 63–65.

Lim, R. K. S. (1970). *Annu. Rev. Physiol.* **32,** 269–288.

Lingappa, V. R., Lingappa, J. R., and Blobel, G. (1979). *Nature (London)* **281,** 117–121.

Lodish, H. F., Kong, N., Snider, M., and Strous, G. J. A. M. (1983). *Nature (London)* **304,** 80–83.

Lomedico, P. T., Gubler, U., Hellmann, C. P., Dukovich, M., Giri, J. G., Pan, Y.-C. E., Collier, K., Semionow, R., Chua, A. O., and Mizel, S. B. (1984). *Nature (London)* **312,** 458–462.

Lorber, D., Tenenbaum, M., Thurston, S., Gander, G. W., and Goodale, F. (1971). *Proc. Soc. Exp. Biol. Med.* **137,** 896–901.

Lottspeich, F., Kellermann, J., Henschen, A., Rauth, G., and Müller-Esterl, W. (1984). *Eur. J. Biochem.* **142,** 227–232.

Luger, T. A., Sztein, M. B., Schmidt, J. A., Murphy, P., Grabner, G., and Oppenheim, J. J. (1983). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **42,** 2772–2776.

Lundwall, Å., Wetsel, R. A., Domdey, H., Tack, B. F., and Fey, G. H. (1984). *J. Biol. Chem.* **259,** 13851–13856.

McAdam, K. P. W. J., and Dinarello, C. A. (1980). *In* "Bacterial Endotoxins and Host Response" (M. K. Agarwal, ed.), p. 167. Elsevier/North-Holland Biomedical Press, Amsterdam.

McAdam, K. P. W. J., Elin, R. J., Sipe, J. D., and Wolff, S. M. (1978). *J. Clin. Invest.* **61,** 390–394.

McAdam, K. P. W. J., Li, J., Knowles, J., Foss, N. T., Dinarello, C. A., Rosenwasser, L. J., Selinger, M. J., Kaplan, M. M., and Goodman, R. (1982). *Ann. N.Y. Acad. Sci.* **389,** 126–136.

McFarlane, H., Ogbeide, M. I., Reddy, S., Adcock, K. J., Adeshina, H., Gurney, J. M., Cooke, A., Taylor, G. O., and Mordie, J. A. (1969). *Lancet* **1,** 392–394.

MacGillivray, R. T. A., Mendez, E., Sinha, S. K., Sutton, M. R., Lineback-Zins. J., and Brew, K. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 2504–2508.

Macintyre, S. S., Kushner, I., and Samols, D. (1985). *J. Biol. Chem.* **260,** 4169–4173.

Mackie, P. H., Crockson, R. A., and Stuart, J. (1979). *J. Clin. Pathol.* **32,** 1253–1256.

McKnight, G. S., Lee, D. C., and Palmiter, R. D. (1980a). *J. Biol. Chem.* **255,** 148–153.

McKnight, G. S., Lee, D. C., Hemmaplardh, D., Finch, C. A., and Palmiter, R. D. (1980b). *J. Biol. Chem.* **255,** 144–147.

McLean, J. W., Fukazawa, C., and Taylor, J. M. (1983). *J. Biol. Chem.* **258,** 8993–9000.

Majumdar, C., Tsukada, K., and Lieberman, I. (1967). *J. Biol. Chem.* **242,** 700–704.

March, C. J., Mosley, B., Larsen, A., Cerretti, D. P., Braedt, G., Price, V., Gillis, S., Henney, C. S., Kronheim, S. R., Grabstein, K., Conlon, P. J., Hopp, T. P., and Cosman, D. (1985). *Nature (London)* **315,** 641–647.

Marsh, J. B., and Drabkin, D. L. (1958). *J. Biol. Chem.* **230,** 1073–1081.

Marsh, J. B., and Drabkin, D. L. (1960).,*Metab., Clin. Exp.***9,** 946–955.

Marsh, J. B., Drabkin, D. L., Braun, G. A., and Parks, J. S. (1966). *J. Biol. Chem.* **241,** 4168–4174.

Masawe, A. E. J., and Rwabwogo-Atenyi, J. (1973). *Arch. Dis. Child.* **48,** 927–931.

Masuya, T., and Kozuru, M. (1963). *Kyushu J. Med. Sci.* **14,** 233–242.

Meek, R. L., Walsh, K. A., and Palmiter, R. D. (1982). *J. Biol. Chem.* **257,** 12245–12251.

Meiers, H. G., Beisenherz, D., Brüster, H., Strassburger, D., and Greuel, H. (1968). *Dtsch. Med. Wochenschr.* **93,** 1633–1636.

Melloni, E., Pontremoli, S., Salamino, F., Sparatore, B., Michetti, M., and Horecker, B. L. (1981). *Arch. Biochem. Biophys.* **208,** 175–183.

Menkin, V. (1936). *Proc. Soc. Exp. Biol. Med.* **34,** 570–572.

Menkin, V. (1937). *Arch. Pathol.* **24,** 65–82.

Menkin, V. (1940). *Am. J. Pathol.* **16,** 13–32.

Menkin, V. (1945). *Arch. Pathol.* **39,** 28–36.

Menkin, V. (1956). "Biochemical Mechanisms in Inflammation," 2nd ed. Thomas, Springfield, Illinois.

Mercer, J. F. B., and Hudson, P. (1982). *Bioscience Rep.* **2,** 761–768.

Merriman, C. R., Pulliam, L. A., and Kampschmidt, R. F. (1975). *Proc. Soc. Exp. Biol. Med.* **149,** 782–784.

Merriman, C. R., Pulliam, L. A., and Kampschmidt, R. F. (1977). *Proc. Soc. Exp. Biol. Med.* **154,** 224–227.

Merriman, C. R., Upchurch, H. F., and Kampschmidt, R. F. (1978). *Proc. Soc. Exp. Biol. Med.* **157,** 669–671.

Mester, L., and Szabados, L. (1968). *C.R. Hebd. Seances Acad. Sci., Ser. D* **266,** 34–36.

Miller, L. L. (1976). *Protides Biol. Fluids* **23,** 461–469.

Miller, L. L., and Bale, W. F. (1954). *J. Exp. Med.* **99,** 125–132.

Miller, L. L., and Griffin, E. E. (1975). *In* "Biochemical Action of Hormones" (G. Litwack, ed.), Vol. 3, pp. 159–186. Academic Press, New York.

Miller, L. L., Bly, C. G., Watson, M. L., and Bale, W. F. (1951). *J. Exp. Med.* **94,** 431–453.

Millership, A., Edwards, K., Chelladurai, M., Dryburgh, H., Inglis, A. S., Urban, J., and Schreiber, G. (1980). *Int. J. Pept. Protein Res.* **15,** 248–252.

Mita, S., Maeda, S., Shimada, K., and Araki, S. (1984). *Biochem. Biophys. Res. Commun.* **124,** 558–564.

Miura, K., Law, S. W. T., Nishi, S., and Tamaoki, T. (1979). *J. Biol. Chem.* **254,** 5515–5521.
Mizel, S. B., and Farrar, J. J. (1979). *Cell. Immunol.* **48,** 433–436.
Møllgård, K., Jacobsen, M., Jacobsen, G. K., Clausen, P. P., and Saunders, N. R. (1979). *Neurosci. Lett.* **14,** 85–90.
Moody, B. J. (1982). *Clin. Chim. Acta* **118,** 87–92.
Moore, F. D. (1965). *N. Engl. J. Med.* **273,** 567–577.
Morgan, E. H. (1981). *Mol. Aspects Med.* **4,** 1–123.
Morgan, E. H. (1983). *In* "Plasma Protein Secretion by the Liver" (H. Glaumann, T. Peters, Jr., and C. Redman, eds.), pp. 331–355. Academic Press, New York.
Morgan, E. H., and Peters, T., Jr. (1971a). *J. Biol. Chem.* **246,** 3500–3507.
Morgan, E. H., and Peters, T., Jr. (1971b). *J. Biol. Chem.* **246,** 3508–3511.
Morimoto, A., Watanabe, T., Ono, T., and Murakami, N. (1984). *In* "Thermal Physiology" (J. R. S. Hales, ed.), pp. 547–550. Raven Press, New York.
Morley, J. J., and Kushner, I. (1982). *Ann. N.Y. Acad. Sci.* **389,** 406–418.
Morrow, J. F., Stearman, R. S., Peltzman, C. G., and Potter, D. A. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 4718–4722.
Morse, J. O. (1978). *N. Engl. J. Med.* **299,** 1045–1048.
Mortensen, S. B., Sottrup-Jensen, L., Hansen, H. F., Petersen, T. E., and Magnusson,S. (1981). *FEBS Lett.* **135,** 295–300.
Morton, A., Hamilton, S. M., Ramsden, D. B., and Tavill, A. S. (1976). *In* "Plasma Protein Turnover" (R. Bianchi, G. Mariani, and A. S. McFarlane, eds.), pp. 165–177. University Park Press, Baltimore, Maryland.
Muglia, L., and Locker, J. (1984). *Nucleic Acids Res.* **12,** 6751–6762.
Müller-Eberhard, H. J. (1975a). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 393–432. Academic Press, New York.
Müller-Eberhard, H. J. (1975b). *Annu. Rev. Biochem.* **44,** 697–724.
Müller-Eberhard, H. J., and Nilsson, U. (1960). *J. Exp. Med.* **111,** 217–234.
Müller-Eberhard, H. J., Nilsson, U., and Aronsson, T. (1960). *J. Exp. Med.* **11,** 201–215.
Müller-Eberhard, U., and Liem, H. H. (1974). *In* "Structure and Function of Plasma Proteins" (A. C. Allison, ed.), Vol. 1, pp. 35–53. Plenum, New York.
Müller-Esterl, W., Fritz, H., Machleidt, W., Ritonja, A., Brzin, J., Kotnik, M., Turk, V., Kellermann, J., and Lottspeich, F. (1985). *FEBS Lett.* **182,** 310–314.
Murphy, P. A., Simon, P. L., and Willoughby, W. F. (1980). *J. Immunol.* **124,** 2498–2501.
Mutschler, L. E., and Gordon, A. H. (1966). *Biochim. Biophys. Acta* **130,** 486–492.
Myklebost, O., Williamson, B., Markham, A. F., Myklebost, S. R., Rogers, J., Woods, D. E., and Humphries, S. E. (1984). *J. Biol. Chem.* **259,** 4401–4404.
Nagase, S., Shimamune, K., and Shumiya, S. (1979). *Science* **205,** 590–591.
Nagashima, M., Urban, J., and Schreiber, G. (1980). *J. Biol. Chem.* **255,** 4951–4956.
Nagashima, M., Urban, J., and Schreiber, G. (1981). *J. Biol. Chem.* **256,** 2091–2093.
Nahon, J.-L., Gal, A., Erdos, T., and Sala-Trepat, J. M. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 5031–5035.
Nawa, H., Kitamura, H., Hirose, T., Asai, M., Inayama, S., and Nakanishi, S. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 90–94.
Nettleship, A. (1938). *Am. J. Clin. Pathol.* **8,** 398–404.
Nickerson, J. M., and Fuller, G. M. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 303–307.
Nieuwenhuizen, W., Emeis, J. J., and Hemmink, J. (1979). *Biochim. Biophys. Acta* **580,** 129–139.
Norman, M. E., Baehni, P. C., Tsai, C.-C., Stoller, N., McArthur, W. P., and Taichman, N. S. (1979). *J. Periodontal Res.* **14,** 361–369.
Northemann, W., Heisig, M., Kunz, D., and Heinrich, P. C. (1985). *J. Biol. Chem.* **260,** 6200–6205.

Norum, R. A., Bearn, A. G., Briscoe, W. A., and Briscoe, A. (1977). *Mt. Sinai J. Med.* **44,** 821–827.

Noyer, M., Dwulet, F. E., Hao, Y. L., and Putnam, F. W. (1980). *Anal. Biochem.* **102,** 450–458.

Nurmi, L., and Largen, M. T. (1982). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **41,** 318.

Ochi, Y., Fujiyama, Y., Hosoda, S., Myazaki, T., Yoshimura, M., Hachiya, T., and Kajita, Y. (1982a). *Clin. Chim. Acta* **122,** 145–160.

Ochi, Y., Fujiyama, Y., Hosoda, S., Hamazu, M., Kajita, Y., Myazaki, T., Hachiya, T., and Ishida, M. (1982b). *J. Immunol. Methods* **52,** 213–221.

Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H., and Sasaki, M. (1984). *Biochemistry* **23,** 5691–5697.

Okamoto, H., and Greenbaum, L. M. (1983a). *Biochem. Biophys. Res. Commun.* **112,** 701–708.

Okamoto, H., and Greenbaum, L. M. (1983b). *Biochem. Pharmacol.* **32,** 2637–2638.

Okamoto, H., and Greenbaum, L. M. (1983c). *Life Sci.* **32,** 2007–2013.

Okubo, H., Miyanaga, O., Nagano, M., Ishibashi, H., Kudo, J., Ikuta, T., and Shibata, K. (1981). *Biochim. Biophys. Acta* **668,** 257–267.

Oliveira, E. B., Gotschlich, E. C., and Liu, T.-Y. (1979). *J. Biol. Chem.* **254,** 489–502.

Oppenheim, J. J., Stadler, B. M., Siraganian, R. P., Mage, M., and Mathieson, B. (1982). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **41,** 257–262.

Ortel, T. L., Takahashi, N., and Putnam, F. W. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 4761–4765.

Osaki, S., Johnson, D. A., and Frieden, E. (1966). *J. Biol. Chem.* **241,** 2746–2751.

Osborne, J. C., Jr., and Brewer, H. B., Jr. (1977). *Adv. Protein Chem.* **31,** 253–337.

Osmand, A. P., Friedenson, B., Gewurz, H., Painter, R. H., Hofmann, T., and Shelton, E. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 739–743.

Owen, J. A. (1967). *Adv. Clin. Chem.* **9,** 1–41.

Owen, J. S., and McIntyre, N. (1982). *Trends Biochem. Sci.* **7,** 95–98.

Owen, M. C., Brennan, S. O., Lewis, J. H., and Carrell, R. W. (1983). *N. Engl. J. Med.* **309,** 694–698.

Pagano, M., Engler, R., Esnard, F., and Gauthier, F. (1984a). *In* "Marker Proteins in Inflammation" (P. Arnaud, J. Bienvenu, and P. Laurent, eds.), Vol. 2, pp. 203–205. de Gruyter, Berlin.

Pagano, M., Esnard, F., Engler, R., and Gauthier, F. (1984b). *Biochem. J.* **220,** 147–155.

Palmer, W. G. (1976). *J. Reticuloendothel. Soc.* **19,** 301–310.

Pekarek, R. S., and Beisel, W. R. (1971). *Proc. Soc. Exp. Biol. Med.* **138,** 728–732.

Pekarek, R. S., Wannemacher, R. W., Jr., and Beisel, W. R. (1972a). *Proc. Soc. Exp. Biol. Med.* **140,** 685–688.

Pekarek, R. S., Powanda, M. C., and Wannemacher, R. W., Jr. (1972b). *Proc. Soc. Exp. Biol. Med.* **141,** 1029–1031.

Pepys, M. B., Dash, A. C., Markham, R. E., Thomas, H. C., Williams, B. D., and Aviva, P. (1978). *Clin. Exp. Immunol.* **32,** 119–124.

Pepys, M. B., Baltz, M., Gomer, K., Davies, A. J. S., and Doenhoff, M. (1979a). *Nature (London)* **278,** 259–261.

Pepys, M. B., Dyck, R. F., de Beer, F. C., Skinner, M., and Cohen, A. S. (1979b). *Clin. Exp. Immunol.* **38,** 284–293.

Peters, T., Jr. (1970). *Adv. Clin. Chem.* **13,** 37–111.

Peters, T., Jr. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 133–181. Academic Press, New York.

Peters, T., Jr., and Peters, J. C. (1972). *J. Biol. Chem.* **247,** 3858–3863.

Peterson, P. A. (1971). *J. Biol. Chem.* **246,** 34–43.

Phong, T. L., Muller, M. T., and Mortensen, R. F. (1982). *J. Immunol.* **129,** 665–672.

Poncin, J. E., Martial, J. A., and Gielen, J. E. (1984). *Eur. J. Biochem.* **140,** 493–498.

Porter, R. R., and Reid, K. B. M. (1979). *Adv. Protein. Chem.* **33,** 1–71.
Poulik, M. D., and Weiss, M. L. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 2, pp. 51–108. Academic Press, New York.
Pras, M., Prelli, F., Franklin, E. C., and Frangione, B. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 539–542.
Price, D. L., Whitehouse, P. J., and Struble, R. G. (1985). *Annu. Rev. Med.* **36,** 349–356.
Putnam, F. W. (1975a). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 88–89. Academic Press, New York.
Putnam, F. W. (1975b). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 92–93. Academic Press, New York.
Putnam, F. W. (1975c). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 265–316. Academic Press, New York.
Ramsden, D. B., Princé, H. P., Burr, W. A., Bradwell, A. R., Black, E. G., Evans, A. E., and Hoffenberg, R. (1978). *Clin. Endocrinol.* **8,** 109–122.
Ramsey, E. M. (1975). "The Placenta of Laboratory Animals and Man." Holt, New York.
Reid, K. B. M., and Porter, R. R. (1981). *Annu. Rev. Biochem.* **50,** 433–464.
Reinke, R., and Feigelson, P. (1985). *J. Biol. Chem.* **260,** 4397–4403.
Ricca, G. A., and Taylor, J. M. (1981). *J. Biol. Chem.* **256,** 11199–11202.
Ricca, G. A., Hamilton, R. W., McLean, J. W., Conn, A., Kalinyak, J. E., and Taylor, J. M. (1981). *J. Biol. Chem.* **256,** 10362–10368.
Ritchie, D. G., and Fuller, G. M. (1981). *Inflammation* **5,** 275–287.
Ritchie, D. G., and Fuller, G. M. (1983). *Ann. N.Y. Acad. Sci.* **408,** 490–502.
Rixon, M. W., Chan, W.-Y., Davie, E. W., and Chung, D. W. (1983). *Biochemistry* **22,** 3237–3244.
Robbins, S. L., and Cotran, R. S. (1979). "Pathologic Basis of Disease," 2nd ed. Saunders, Philadelphia, Pennsylvania.
Robinson, S. I., Small, D., Idzerda, R., McKnight, G. S., and Vogelstein, B. (1983). *Nucleic Acids Res.* **11,** 5113–5130.
Roll, D. E., and Glew, R. H. (1981). *J. Biol. Chem.* **256,** 8190–8196.
Rosen, A. M., and Geller, D. M. (1977). *Biochem. Biophys. Res. Commun.* **78,** 1060–1066.
Rosenberg, S., Barr, P. J., Najarian, R. C., and Hallewell, R. A. (1984). *Nature (London)* **312,** 77–80.
Rosenthal, C. J., and Franklin, E. C. (1975). *J. Clin. Invest.* **55,** 746–753.
Rosenthal, C. J., Franklin, E. C., Frangione, B., and Greenspan, J. (1976). *J. Immunol.* **116,** 1415–1418.
Rosenwasser, L. J., Dinarello, C. A., and Rosenthal, A. S. (1979). *J. Exp. Med.* **150,** 709–714.
Rothschild, M. A., Oratz, M., Wimer, E., and Schreiber, S. S. (1961). *J. Clin. Invest.* **40,** 545–554.
Rothschild, M. A., Oratz, M., Mongelli, J., and Schreiber, S. S. (1968). *J. Clin. Invest.* **47,** 2591–2599.
Rothschild, M. A., Oratz, M., Mongelli, J., and Schreiber, S. S. (1969a). *Am. J. Physiol.* **216,** 1127–1130.
Rothschild, M. A., Oratz, M., and Schreiber, S. S. (1969b). *Am. J. Dig. Dis.* **14,** 711–744.
Rothschild, M. A., Oratz, M., Mongelli, J., and Schreiber, S. S. (1971). *J. Clin. Invest.* **50,** 1812–1818.
Rothschild, M. A., Oratz, M., and Schreiber, S. S. (1972a). *N. Engl. J. Med.* **286,** 748–757.
Rothschild, M. A., Oratz, M., and Schreiber, S. S. (1972b). *N. Engl. J. Med.* **286,** 816–821.
Rothschild, M. A., Oratz, M., and Schreiber, S. S. (1973). *Gastroenterology* **64,** 324–337.
Rothschild, M. A., Oratz, M., and Schreiber, S. S. (1975). *Annu. Rev. Med.* **26,** 91–104.
Rowett, H. G. Q. (1960). "The Rat as a Small Mammal," 2nd ed. Murray, London.

Rupp, R. G., and Fuller, G. M. (1979a). *Exp. Cell Res.* **118,** 23–30.
Rupp, R. G., and Fuller, G. M. (1979b). *Biochem. Biophys. Res. Commun.* **88,** 327–334.
Russell, J. H., and Geller, D. M. (1975). *J. Biol. Chem.* **250,** 3409–3413.
Rydén, L. (1972). *Eur. J. Biochem.* **26,** 380–386.
Saraiva, M. J. M., Birken, S., Costa, P. P., and Goodman, D. S. (1984). *Ann. N.Y. Acad. Sci.* **435,** 86–100.
Sargent, T. D., Yang, M., and Bonner, J. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 243–246.
Schade, R., Götz, F., Porstmann, B., Friedrich, A., and Nugel, E. (1982). *Agents Actions Suppl.* **10,** 213–231.
Schmid, K. (1975). *In* "The Plasma Proteins" (F. W. Putnam, ed.), 2nd ed., Vol. 1, pp. 183–228. Academic Press, New York.
Schmid, K., Kaufmann, H., Isemura, S., Bauer, F., Emura, J., Motoyama, T., Ishiguro, M., and Nanno, S. (1973). *Biochemistry* **12,** 2711–2724.
Schoene, W. C. (1979). *In* "Pathologic Basis of Disease" (S. L. Robbins and R. S. Cotran, eds.), 2nd ed., pp. 1530–1598. Saunders, Philadelphia, Pennsylvania.
Schreiber, G. (1979). *In* "Biological Functions of Proteinases" (H. Holzer and H. Tschesche, eds.), pp. 150–164. Springer-Verlag, Berlin and New York.
Schreiber, G., and Howlett, G. (1983). *In* "Plasma Protein Secretion by the Liver" (H. Glaumann, T. Peters, Jr., and C. Redman, eds.), pp. 423–449. Academic Press, New York.
Schreiber, G., and Schreiber, M. (1972). *J. Biol. Chem.* **247,** 6340–6346.
Schreiber, G., and Schreiber, M. (1973). *Subcell. Biochem.* **2,** 307–353.
Schreiber, G., and Urban, J. (1978). *Rev. Physiol. Biochem. Pharmacol.* **82,** 27–95.
Schreiber, G., Boutwell, R. K., Potter, V. R., and Morris, H. P. (1966). *Cancer Res.* **26,** Part 1, 2357–2361.
Schreiber, G., Urban, J., Zähringer, J., Reutter, W., and Frosch, U. (1971). *J. Biol. Chem.* **246,** 4531–4538.
Schreiber, G., Dryburgh, H., Millership, A., Matsuda, Y., Inglis, A., Phillips, J., Edwards, K., and Maggs, J. (1979). *J. Biol. Chem.* **254,** 12013–12019.
Schreiber, G., Dryburgh, H., Weigand, K., Schreiber, M., Witt, I., Seydewitz, H., and Howlett, G. (1981). *Arch. Biochem. Biophys.* **212,** 319–328.
Schreiber, G., Howlett, G., Nagashima, M., Millership, A., Martin, H., Urban, J., and Kotler, L. (1982). *J. Biol. Chem.* **257,** 10271–10277.
Schreiber, G., Aldred, A. R., Thomas, T., Birch, H. E., Dickson, P. W., Tu, G.-F., Heinrich, P. C., Northemann, W., Howlett, G. J., de Jong, F. A., and Mitchell, A. (1986). *Inflammation* **10,** 59–66.
Schultze, H. E., and Heremans, J. F. (1966). "Molecular Biology of Human Proteins," Vol. 1. Elsevier, Amsterdam.
Schultze, H. E., Heide, K., and Haupt, H. (1962). *Clin. Chim. Acta* **7,** 854–868.
Schumacher, G., and Schlumberger, H. D. (1963). *Dtsch. Med. Wochenschr.* **88,** 645–651.
Schwandt, P. (1982). *Klin. Wochenschr.* **60,** 637–649.
Schwandt, P., Fateh-Moghadam, A., Richter, W., and Sandel, P. (1979). *Lancet* **2,** 794.
Sharp, H. L., Bridges, R. A., Krivit, W., and Freier, E. F. (1969). *J. Lab. Clin. Med.* **73,** 934–939.
Sharpe, C. R., Sidoli, A., Shelley, C. S., Lucero, M. A., Shoulders, C. C., and Baralle, F. E. (1984). *Nucleic Acids Res.* **12,** 3917–3932.
Shetty, P. S., Watrasiewicz, K. E., Jung, R. T., and James, W. P. T. (1979). *Lancet* **2,** 230–232.
Shirahama, T., Skinner, M., Westermark, P., Rubinow, A., Cohen, A. S., Brun, A., and Kemper, T. L. (1982). *Am. J. Pathol.* **107,** 41–50.
Sipe, J. D. (1978). *Br. J. Exp. Pathol.* **59,** 305–310.
Sipe, J. D., Ignaczak, T. F., Pollock, P. S., and Glenner, G. G. (1976). *J. Immunol.* **116,** 1151–1156.
Sipe, J. D., McAdam, K. P. W. J., and Uchino, F. (1978). *Lab. Invest.* **38,** 110–114.

Sipe, J. D., Vogel, S. N., Ryan, J. L., McAdam, K. P. W. J., and Rosenstreich, D. L. (1979). *J. Exp. Med.* **150,** 597–606.

Sipe, J. D., Vogel, S. N., Sztein, M. B., Skinner, M., and Cohen, A. S. (1982). *Ann. N.Y. Acad. Sci.* **389,** 137–150.

Sipe, J. D., Colten, H. R., Goldberger, G., Edge, M. D., Tack, B. F., Cohen, A. S., and Whitehead, A. S. (1985). *Biochemistry* **24,** 2931–2936.

Sletten, K., Westermark, P., and Natvig, J. (1976). *J. Exp. Med.* **143,** 993–998.

Smith, F. R., Goodman, D. S., Zaklama, M. S., Gabr, M. K., El Maraghy, S., and Patwardhan, V. N. (1973). *Am. J. Clin. Nutr.* **26,** 973–981.

Smith, L. C., Pownall, H. J., and Gotto, A. M., Jr. (1978). *Annu. Rev. Biochem.* **47,** 751–777.

Smithies, O. (1965). *Science* **150,** 1595–1598.

Snyder, S., and Coodley, E. L. (1976). *Arch. Intern. Med.* **136,** 778–781.

Sottrup-Jensen, L., Stepanik, T. M., Kristensen, T., Lønblad, P. B., Jones, C. M., Wierzbicki, D. M., Magnusson, S., Domdey, H., Wetsel, R. A., Lundwall, Å., Tack, B. F., and Fey, G. H. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 9–13.

Staruch, M. J., and Wood, D. D. (1982). *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **41,** 317.

Stein, P. D., Leu, J. D., Welch, M.H., and Guenter, C. A. (1971). *Circulation* **43,** 227–239.

Stevens, P. M., Hnilica, V. S., Johnson, P. C., and Bell, R. L. (1971). *Ann. Intern. Med.* **74,** 672–680.

Strauss, A. W., Bennett, C. D., Donohue, A. M., Rodkey, J. A., and Alberts, A. W. (1977a). *J. Biol. Chem.* **252,** 6846–6855.

Strauss, A. W., Donohue, A. M., Bennett, C. D., Rodkey, J. A., and Alberts, A. W. (1977b). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 1358–1362.

Stuart, G. W., Searle, P. F., Chen, H. Y., Brinster, R. L., and Palmiter, R. D. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 7318–7322.

Sueyoshi, T., Enjyoji, K., Shimada, T., Kato, H., Iwanaga, S., Bando, Y., Kominami, E., and Katunuma, N. (1985). *FEBS Lett.* **182,** 193–195.

Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H., and Itakura, K. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6613–6617.

Sundelin, J., Laurent, B. C., Anundi, H., Trägårdh, L., Larhammar, D., Björk, L., Eriksson, U., Åkerström, B., Jones, A., Newcomer, M., Peterson, P. A., and Rask, L. (1985a). *J. Biol. Chem.* **260,** 6472–6480.

Sundelin, J., Melhus, H., Das, S., Eriksson, U., Lind, P., Trägårdh, L., Peterson, P. A., and Rask, L. (1985b). *J. Biol. Chem.* **260,** 6481–6487.

Sutherland, R. L., and Brandon, M. R. (1976). *Endocrinology (Baltimore)* **98,** 91–98.

Sutton, H. E. (1970). *Prog. Med. Genet.* **7,** 163–216.

Sztein, M. B., Vogel, S. N., Sipe, J. D., Murphy, P. A., Mizel, S. B., Oppenheim, J. J., and Rosenstreich, D. L. (1981). *Cell. Immunol.* **63,** 164–176.

Takahashi, N., Ortel, T. L., and Putnam, F. W. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 390–394.

Takahashi, N., Takahashi, Y., and Putnam, F. W. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 73–77.

Takio, K., Towatari, T., Katunuma, N., Teller, D. C., and Titani, K. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 3666–3670.

Tanaka, T., Ohkubo, H., and Nakanishi, S. (1984). *J. Biol. Chem.* **259,** 8063–8065.

Termine, J. D., Eanes, E. D., Ein, D., and Glenner, G. G. (1972). *Biopolymers* **11,** 1103–1113.

Thomas, T., and Schreiber, G. (1985). *Inflammation* **9,** 1–7.

Thomas, T., and Schreiber, G. (1986). In preparation.

Thompson, W. L., Abeles, F. B., Beall, F. A., Dinterman, R. E., and Wannemacher, R. W., Jr. (1976). *Biochem. J.* **156,** 25–32.

Tietz, N. W. (1976). *In* "Fundamentals of Clinical Chemistry" (N. W. Tietz, ed.), pp. 873–944. Saunders, Philadelphia, Pennsylvania.

Tillett, W. S., and Francis, T., Jr. (1930). *J. Exp. Med.* **52,** 561–571.

Toh, H., Hayashida, H., Kikuno, R., Yasunaga, T., and Miyata, T. (1985). *Nature (london)* **314,** 199.

Tokunaga, M., Loranger, J. M., and Wu, H. C. (1984). *J. Cell. Biochem.* **24,** 113–120.

Tomlinson, B. E., Blessed, G., and Roth, M. (1970). *J. Neurol. Sci.* **11,** 205–242.

Tracht, M. E., Tallal, L., and Tracht, D. G. (1967). *Life Sci.* **6,** 2621–2628.

Trägårdh, L., Anundi, H., Rask, L., Sege, K., and Peterson, P. A. (1980). *J. Biol. Chem.* **255,** 9243–9248.

Travis, J., and Salvesen, G. S. (1983). *Annu. Rev. Biochem.* **52,** 655–709.

Travis, J., Owen, M., George, P., Carrell, R., Rosenberg, S., Hallewell, R. A., and Barr, P. J. (1985). *J. Biol. Chem.* **260,** 4384–4389.

Tu, G.-F., de Jong, F., Apostolopoulos, J., Nagashima, M., Fidge, N., Schreiber, G., and Howlett, G. (1986). *Inflammation* (in press).

Unterman, R. D., Lynch, K. R., Nakhasi, H. L., Dolan, K. P., Hamilton, J. W., Cohn, D. V., and Feigelson, P. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 3478–3482.

Urban, J., Inglis, A. S., Edwards, K., and Schreiber, G. (1974). *Biochem. Biophys. Res. Commun.* **61,** 494–501.

Urban, J., Chelladurai, M., Millership, A., and Schreiber, G. (1976). *Eur. J. Biochem.* **67,** 477–485.

Urban, J., Chan, D., and Schreiber, G. (1979). *J. Biol. Chem.* **254,** 10565–10568.

Urban, J., Nagashima, M., and Schreiber, G. (1982a). *Biochem. Int.* **4,** 75–82.

Urban, J., Chelladurai, M., and Schreiber, G. (1982b). *Biochem. Int.* **4,** 177–185.

Uzan, G., Frain, M., Park, I., Besmond, C., Maessen, G., Trépat, J. S., Zakin, M. M., and Kahn, A. (1984). *Biochem. Biophys. Res. Commun.* **119,** 273–281.

Van Damme, J., De Ley, M., Opdenakker, G., Billiau, A., De Somer, P., and Van Beeumen, J. (1985). *Nature (London)* **314,** 266–268.

Van Gool, J., and Ladiges, N. C. J. (1969). *J. Pathol.* **97,** 115–126.

Vannice, J. L., Taylor, J. M., and Ringold, G. M. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 4241–4245.

Virchow, R. (1860). "Cellular Pathology" (translated from the 2nd edition of the original by F. Chance). Churchill, London.

Waldmann, T. A., Wochner, R. D., and Strober, W. (1969). *Am. J. Med.* **46,** 275–285.

Wannemacher, R. W., Jr., Pekarek, R. S., Thompson, W. L., Curnow, R. T., Beall, F. A., Zenser, T. V., DeRubertis, F. R., and Beisel, W. R. (1975). *Endocrinology (Baltimore)* **96,** 651–661.

Wasserman, K., Joseph, J. D., and Mayerson, H. S. (1956). *Am. J. Physiol.* **184,** 175–182.

Waterlow, J. C. (1969). *In* "Mammalian Protein Metabolism" (H. N. Munro and J. B. Allison, eds.), Vol. 3, pp. 325–390. Academic Press, New York.

Weeke, B., and Jarnum, S. (1971). *Gut* **12,** 297–302.

Weidner, N., Ittyerah, T. R., Wochner, R. D., and Sherman, L. A. (1979). *Thromb. Res.* **15,** 651–661.

Weimer, H. E., and Benjamin, D. C. (1965). *Proc. Soc. Exp. Biol. Med.* **122,** 1112–1114.

Weimer, H. E., Humelbaugh, C., and Roberts, D. M. (1967). *Am. J. Physiol.* **213,** 418–424.

Weisner, B., and Kauerz, U. (1983). *J. Neurol. Sci.* **61,** 27–35.

Weisner, B., and Röthig, H.-J. (1983). *Eur. Neurol.* **22,** 96–105.

Werner, M. (1969). *Clin. Chim. Acta* **25,** 299–305.

Werner, M., and Cohnen, G. (1967). *Helv. Med. Acta, Suppl.* **47,** 143.

Wetsel, R. A., Lundwall, Å., Davidson, F., Gibson, T., Tack, B. F., and Fey, G. H. (1984). *J. Biol. Chem.* **259,** 13857–13862.

Whitehead, A. S., Bruns, G. A. P., Markham, A. F., Colten, H. R., and Woods, D. E. (1983). *Science* **221,** 69–71.

Wieland, T. (1972). *Naturwissenschaften* **59,** 225–231.

Woeber, K. A., and Ingbar, S. H. (1968). *J. Clin. Invest.* **47,** 1710–1721.

Wolowski, B. M. R. N. J., and Fuller, G. M. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 1443–1447.

Wright, J. R., Calkins, E., Breen, W. J., Stolte, G., and Schultz, R. T. (1969). *Medicine (Baltimore)* **48,** 39–60.

Wright, W. W., Musto, N. A., Mather, J. P., and Bardin, C. W. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 7565–7569.

Yang, F., Brune, J. L., Baldwin, W. D., Barnett, D. R., and Bowman, B. H. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5875–5879.

Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., van Bragt, P. H., Baldwin, W. D., and Bowman, B. H. (1984). *Proc. Natl. Acad. Sci. U.S..A* **81,** 2752–2756.

Yeoh, G. C. T., and Morgan, E. H. (1974). *Biochem. J.* **144,** 215–224.

Yssing, M., Jensen, H., and Jarnum, S. (1969). *Acta Paediatr. Scand.* **58,** 109–115.

Yu, S., and Redman, C. (1977). *Biochem. Biophys. Res. Commun.* **76,** 469–476.

Yu, S., Redman, C. M., Goldstein, J., and Blombäck, B. (1980). *Biochem. Biophys. Res. Commun.* **96,** 1032–1038.

Yu, S., Sher, B., Kudryk, B., and Redman, C. M. (1984). *J. Biol. Chem.* **259,** 10574–10581.

Zalka, E. v. (1928). *Virchow's Arch. Pathol. Anat. Physiol.* **267,** 379–412.

Zannis, V. I., McPherson, J., Goldberger, G., Karathanasis, S. K., and Breslow, J. L. (1984). *J. Biol. Chem.* **259,** 5495–5499.

Zeineh, R. A., and Kukral, J. C. (1970). *J. Trauma* **10,** 493–498.

Zilva, J. F., and Pannall, P. R. (1979). "Clinical Chemistry in Diagnosis and Treatment," 3rd ed. Lloyd-Luke, London.

Appendix: Sequences of Plasma Proteins

Winona C. Barker

National Biomedical Research Foundation
Georgetown University Medical Center
Washington, D.C. 20007

I. Introductory Remarks

In the Appendix to "The Plasma Proteins," Volume IV, the sequences of 56 plasma proteins were spelled out, most of them complete sequences of mature human proteins. Some of the sequences represented complete precursors; only a few were fragmentary or derived from another mammal. These data had accumulated in the scientific literature over a span of about twenty years and were determined mainly by direct peptide sequencing. In this Appendix to the current volume, 28 plasma protein sequences are presented, fully half of which are complete precursors translated from nucleotide sequences. Only a third of these recently determined sequences are principally derived by protein sequencing methods. Six of the sequences are from other mammals; four are large fragments or one chain of a larger molecule. Four that were fragmentary in Volume IV are now completed and shown here. Others (see Table I) that have been revised slightly are not spelled out again. Most of the sequences were abstracted from the Protein Sequence Database of August 1985, maintained at the Protein Identification Resource of the National Biomedical Research Foundation at Georgetown University in Washington, D.C., under the sponsorship of the Division of Research Resources of the National Institutes of Health.

The Plasma Proteins, Vol. V

TABLE I

Revisions to Sequences since Volume IV

Protein	No. of added residues	No. of differences	References[a]
α_1-Antitrypsin precursor	24	3	Bollen *et al.* (1983)
β_2-Microglobulin precursor	20	1 deleted	Rosa *et al.* (1983); Suggs *et al.* (1981)
Proapolipoprotein A-I precursor	24	3	Shoulders *et al.* (1983); Cheung and Chan (1983); Law and Brewer (1984)
Apolipoprotein E precursor	18	0	Paik *et al.* (1985); McLean *et al.* (1984)
Apolipoprotein C-II	1	4	Hospattankar *et al.* (1984)
Fibrinogen α chain precursor	19	5	Rixon *et al.* (1984)
Fibrinogen β chain precursor	30	5	Chung *et al.* (1983a)
Fibrinogen γ chain precursor	26	0	Chung *et al.* (1983b)
Retinol-binding protein precursor	17	1	Colantuoni *et al.* (1983)
Connective-tissue activating peptide III (was β-thromboglobulin)	4	0	Castor *et al.* (1983)
Cystatin C (was post-γ-globulin)	0	0	Turk *et al.* (1983)

[a]Key to references: Bollen, A., Herzog, A., Cravador, A., Herion, P., Chuchana, P., Vander Straten, A., Loriau, R., Jacobs, P., and Van Elsen, A. (1983). *DNA* **2,** 255–264. Castor, C. W., Miller, J. W., and Walz, D. A. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 765–769. Cheung, P., and Chan, L. (1983). *Nucleic Acids Res.* **11,** 3703–3715. Chung, D. W., Que, B. G., Rixon, M. W., Mace, M., Jr., and Davie, E. W. (1983a). *Biochemistry* **22,** 3244–3250. Chung, D. W., Chan, W.-Y., and Davie, E. W. (1983b). *Biochemistry* **22,** 3250–3256. Colantuoni, V., Romano, V., Bensi, G., Santoro, C., Costanzo, F., Raugei, G., and Cortese, R. (1983). *Nucleic Acids Res.* **11,** 7769–7776. Hospattankar, A. V., Fairwell, T., Ronan, R., and Brewer, H. B., Jr. (1984). *J. Biol. Chem.* **259,** 318–322. Law, S. W., and Brewer, H. B., Jr. (1984). *Proc. Natl. Acad. Sci. U.S.A.* **81,** 66–70. McLean, J. W., Elshourbagy, N. A., Chang, D. J., Mahley, R. W., and Taylor, J. M. (1984). *J. Biol. Chem.* **259,** 6498–6504. Paik, Y.-K., Chang, D. J., Reardon, C. A., Davies, G. E., Mahley, R. W., and Taylor, J. M. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 3445–3449. Rixon, M. W., Chan, W.-Y., Davie, E. W., and Chung, D. W. (1983). *Biochemistry* **22,** 3237–3244. Rosa, F., Berissi, H., Weissenbach, J., Maroteaux, L., Fellous, M., and Revel, M. (1983). *EMBO J.* **2,** 239–243. Shoulders, C. C., Kornblihtt, A. R., Munro, B. S., and Baralle, F. E. (1983). *Nucleic Acids Res.* **11,** 2827–2837. Suggs, S. V., Wallace, R. B., Hirose, T., Kawashima, E. H., and Itakura, K. (1981). *Proc. Natl. Acad. Sci. U.S.A.* **78,** 6613–6617. Turk, V., Brzin, J., Longer, M., Ritonja, A., Eropkin, M., Borchart, U., and Machleidt, W. (1983). *Hoppe-Seyler's Z. Physiol. Chem.* **364,** 1487–1496.

The sequences are presented in the one-letter amino acid code shown in Table II, and the molecular weights of the unmodified protein chains are calculated automatically from the amino acid molecular weights shown there. A slash at the beginning or end of the sequence indicates that there are unsequenced segments at the amino or carboxyl end, respectively. Otherwise, with one exception, the sequences shown here have been completely determined. Principal references to

the literature reporting the sequence determinations are given in the text of each sequence entry. Minor discrepancies between the sequences reported from different laboratories are not noted; more complete documentation is available from the Protein Identification Resource. Covalent modifications to the sequences, such as disulfide bonds, bound carbohydrate, and cleavages during activation, are noted, as are active-site residues. Most of this information appears in feature tables, which also include internal duplications, functional domains, and homologies with other proteins.

In the Protein Sequence Database, sequences that can be shown to be related to one another are grouped into superfamilies. Many recently determined sequences represent large molecules with several functional domains. These complex proteins are encoded by multiexonic genes that appear to have formed from fusions

TABLE II

One-Letter Amino Acid Abbreviations and Molecular Weights

Abbreviation	Amino acid	Molecular weight[a]
A	Alanine	89.09
C	Cysteine	121.15
D	Aspartic acid	133.10
E	Glutamic acid	147.13
F	Phenylalanine	165.19
G	Glycine	75.07
H	Histidine	155.16
I	Isoleucine	131.17
K	Lysine	146.19
L	Leucine	131.17
M	Methionine	149.21
N	Asparagine	132.12
P	Proline	115.13
Q	Glutamine	146.15
R	Arginine	174.20
S	Serine	105.09
T	Threonine	119.12
V	Valine	117.15
W	Tryptophan	204.23
Y	Tyrosine	181.19
B	Aspartic acid or asparagine	132.61
Z	Glutamic acid or glutamine	146.64
X	Undetermined amino acid	128.16
	Water molecule	18.015

[a]Based on C = 12.011, H = 1.0079, O = 15.9994, N = 14.0067, and S = 32.06.

of separate genes or gene segments. Thus a protein may contain sequence segments that are respectively related to several unrelated proteins. For such proteins, the sequence segment upon which the principal superfamily classification is based is indicated; homologies with proteins in other superfamilies are noted in the feature tables. A list of proteins whose sequences are included in Section II of this Appendix is given in Table III. Unless noted, the complete human sequence is shown. Sequences in the same superfamily are grouped together.

Elsewhere in this volume appear sequences of several proteins that are not listed in this Appendix or that of Volume IV: human vitamin D-binding protein Gc2 (Chapter 1, Fig. 24); human secretory component, which corresponds to the

TABLE III

Protein Sequences

List of sequences	Source	Identifier[a]	Page
Coagulation factor VIII precursor	Human	EZHU	370
Complement subcomponent C1r (EC 3.4.21.41), b chain	Human	C1HURB	372
Haptoglobin-2 precursor	Human	HPHU2	373
Factor X (EC 3.4.21.6) precursor	Human (fragment)	EXHU	374
Protein Z	Bovine	KXBOZ	375
Complement factor D	Human	DBHU	376
Complement factor B (in EC 3.4.21.47)	Human	BBHU	377
Factor XII (EC 3.4.21.38)	Human	KFHU12	378
Inter-α-trypsin inhibitor (BPI type)	Human	TIHUBI	379
α_1-Antichymotrypsin precursor	Human	ITHUC	380
Angiotensinogen precursor	Human	ANHU	381
Complement C3 precursor	Human	C3HU	382
Complement C4A	Human	C4HU	384
Complement C5	Human (fragment)	C5HU	386
Complement C9 precursor	Human (fragment)	C9HU	387
Fibronectin	Human	FNHU	388
Kininogen, HMW I precursor	Bovine	KGBOH1	391
Kininogen, HMW II precursor	Bovine	KGBOH2	392
Kininogen, LMW I precursor	Human	KGHUL1	393
Kininogen, LMW II precursor	Bovine	KGBOL2	394
Apolipoprotein A-IV precursor	Rat	LPRTA4	395
Serum amyloid P-component	Human	YLHUP	396
α_{2u}-Globulin precursor	Rat	UART	397
Leucine-rich α_2-glycoprotein	Human	GPHUA2	398
α_2-HS-glycoprotein B chain	Human	WOHUB	399
α-Fetoprotein precursor	Human	FPHU	400
Hemopexin	Human	OQHU	401
Erythropoietin precursor	Human	ZUHU	402

[a]The identifier is an arbitrary 4- to 6-character identification code for purposes of computer retrieval.

extracellular portion of the poly-Ig receptor (Chapter 2, Fig. 32); human apolipoprotein IV (Chapter 3, Fig. 4); the carboxyl end of human apolipoprotein B (Chapter 3, Fig. 6); a partial sequence of human pregnancy zone protein (Chapter 4, Fig. 4); and rat major acute-phase protein, also called T-kininogen (Chapter 5, Fig. 4).

II. Amino Acid Sequences

EZHU
Coagulation factor VIII precursor - Human

Wood, W.I., Capon, D.J., Simonsen, C.C., Eaton, D.L., Gitschier, J., Keyt, B., Seeburg, P.H., Smith, D.H., Hollingshead, P., Wion, K.L., Delwart, E., Tuddenham, E.G.D., Vehar, G.A., and Lawn, R.M., Nature 312, 330-337, 1984 (Sequence translated from the mRNA and DNA sequences)

Factor VIIIa, along with calcium and phospholipid, acts as a cofactor for factor IXa when it converts factor X to the activated form, factor Xa.

Superfamily: 1-739,1713-2046/ceruloplasmin

Residues	Feature
1-19	Domain: signal sequence
20-740,1668-2351	Protein: factor VIIIa
20-356	Domain: A1
357-740	Domain: A2
741-1667	Domain: B
1668-2046	Domain: A3
2047-2200	Domain: C1
2201-2351	Domain: C2
20-356,399-739, 1713-2046	Duplication: homology with ceruloplasmin domains

Mol. wt. unmod. chain = 267,007 Number of residues = 2351

```
                5         10        15        20        25        30
   1 M Q I E L S T C F F L C L L R F C F S A T R R Y Y L G A V E
  31 L S W D Y M Q S D L G E L P V D A R F P P R V P K S F P F N
  61 T S V V Y K K T L F V E F T D H L F N I A K P R P P W M G L
  91 L G P T I Q A E V Y D T V V I T L K N M A S H P V S L H A V
 121 G V S Y W K A S E G A E Y D D Q T S Q R E K E D D K V F P G
 151 G S H T Y V W Q V L K E N G P M A S D P L C L T Y S Y L S H
 181 V D L V K D L N S G L I G A L L V C R E G S L A K E K T Q T
 211 L H K F I L L F A V F D E G K S W H S E T K N S L M Q D R D
 241 A A S A R A W P K M H T V N G Y V N R S L P G L I G C H R K
 271 S V Y W H V I G M G T T P E V H S I F L E G H T F L V R N H
 301 R Q A S L E I S P I T F L T A Q T L L M D L G Q F L L F C H
 331 I S S H Q H D G M E A Y V K V D S C P E E P Q L R M K N N E
 361 E A E D Y D D D L T D S E M D V V R F D D D N S P S F I Q I
 391 R S V A K K H P K T W V H Y I A A E E E D W D Y A P L V L A
 421 P D D R S Y K S Q Y L N N G P Q R I G R K Y K K V R F M A Y
 451 T D E T F K T R E A I Q H E S G I L G P L L Y G E V G D T L
 481 L I I F K N Q A S R P Y N I Y P H G I T D V R P L Y S R R L
 511 P K G V K H L K D F P I L P G E I F K Y K W T V T V E D G P
 541 T K S D P R C L T R Y Y S S F V N M E R D L A S G L I G P L
 571 L I C Y K E S V D Q R G N Q I M S D K R N V I L F S V F D E
 601 N R S W Y L T E N I Q R F L P N P A G V Q L E D P E F Q A S
 631 N I M H S I N G Y V F D S L Q L S V C L H E V A Y W Y I L S
```

```
 661 I G A Q T D F L S V F F S G Y T F K H K M V Y E D T L T L F
 691 P F S G E T V F M S M E N P G L W I L G C H N S D F R N R G
 721 M T A L L K V S S C D K N T G D Y Y E D S Y E D I S A Y L L
 751 S K N N A I E P R S F S Q N S R H P S T R Q K Q F N A T T I
 781 P E N D I E K T D P W F A H R T P M P K I Q N V S S S D L L
 811 M L L R Q S P T P H G L S L S D L Q E A K Y E T F S D D P S
 841 P G A I D S N N S L S E M T H F R P Q L H H S G D M V F T P
 871 E S G L Q L R L N E K L G T T A A T E L K K L D F K V S S T
 901 S N N L I S T I P S D N L A A G T D N T S S L G P P S M P V
 931 H Y D S Q L D T T L F G K K S S P L T E S G G P L S L S E E
 961 N N D S K L L E S G L M N S Q E S S W G K N V S S T E S G R
 991 L F K G K R A H G P A L L T K D N A L F K V S I S L L K T N
1021 K T S N N S A T N R K T H I D G P S L L I E N S P S V W Q N
1051 I L E S D T E F K K V T P L I H D R M L M D K N A T A L R L
1081 N H M S N K T T S S K N M E M V Q Q K K E G P I P P D A Q N
1111 P D M S F F K M L F L P E S A R W I Q R T H G K N S L N S G
1141 Q G P S P K Q L V S L G P E K S V E G Q N F L S E K N K V V
1171 V G K G E F T K D V G L K E M V F P S S R N L F L T N L D N
1201 L H E N N T H N Q E K K I Q E E I E K K E T L I Q E N V V L
1231 P Q I H T V T G T K N F M K N L F L L S T R Q N V E G S Y D
1261 G A Y A P V L Q D F R S L N D S T N R T K K H T A H F S K K
1291 G E E E N L E G L G N Q T K Q I V E K Y A C T T R I S P N T
1321 S Q Q N F V T Q R S K R A L K Q F R L P L E E T E L E K R I
1351 I V D D T S T Q W S K N M K H L T P S T L T Q I D Y N E K E
1381 K G A I T Q S P L S D C L T R S H S I P Q A N R S P L P I A
1411 K V S S F P S I R P I Y L T R V L F Q D N S S H L P A A S Y
1441 R K K D S G V Q E S S H F L Q G A K K N N L S L A I L T L E
1471 M T G D Q R E V G S L G T S A T N S V T Y K K V E N T V L P
1501 K P D L P K T S G K V E L L P K V H I Y Q K D L F P T E T S
1531 N G S P G H L D L V E G S L L Q G T E G A I K W N E A N R P
1561 G K V P F L R V A T E S S A K T P S K L L D P L A W D N H Y
1591 G T Q I P K E E W K S Q E K S P E K T A F K K K D T I L S L
1621 N A C E S N H A I A A I N E G Q N K P E I E V T W A K Q G R
1651 T E R L C S Q N P P V L K R H Q R E I T R T T L Q S D Q E E
1681 I D Y D D T I S V E M K K E D F D I Y D E D E N Q S P R S F
1711 Q K K T R H Y F I A A V E R L W D Y G M S S S P H V L R N R
1741 A Q S G S V P Q F K K V V F Q E F T D G S F T Q P L Y R G E
1771 L N E H L G L L G P Y I R A E V E D N I M V T F R N Q A S R
1801 P Y S F Y S S L I S Y E E D Q R Q G A E P R K N F V K P N E
1831 T K T Y F W K V Q H H M A P T K D E F D C K A W A Y F S D V
1861 D L E K D V H S G L I G P L L V C H T N T L N P A H G R Q V
1891 T V Q E F A L F F T I F D E T K S W Y F T E N M E R N C R A
1921 P C N I Q M E D P T F K E N Y R F H A I N G Y I M D T L P G
1951 L V M A Q D Q R I R W Y L L S M G S N E N I H S I H F S G H
1981 V F T V R K K E E Y K M A L Y N L Y P G V F E T V E M L P S
2011 K A G I W R V E C L I G E H L H A G M S T L F L V Y S N K C
2041 Q T P L G M A S G H I R D F Q I T A S G Q Y G Q W A P K L A
2071 R L H Y S G S I N A W S T K E P F S W I K V D L L A P M I I
2101 H G I K T Q G A R Q K F S S L Y I S Q F I I M Y S L D G K K
2131 W Q T Y R G N S T G T L M V F F G N V D S S G I K H N I F N
2161 P P I I A R Y I R L H P T H Y S I R S T L R M E L M G C D L
2191 N S C S M P L G M E S K A I S D A Q I T A S S Y F T N M F A
2221 T W S P S K A R L H L Q G R S N A W R P Q V N N P K E W L Q
2251 V D F Q K T M K V T G V T T Q G V K S L L T S M Y V K E F L
2281 I S S S Q D G H Q W T L F F Q N G K V K V F Q G N Q D S F T
2311 P V V N S L D P P L L T R Y L R I H P Q S W V H Q I A L R M
2341 E V L G C E A Q D L Y
```

C1HURB
Complement subcomponent C1r (EC 3.4.21.41), b chain - Human

Arlaud, G.J., and Gagnon, J., Biochemistry 22, 1758-1764, 1983

C1r is a dimer of identical chains, each of which is activated by cleavage into two chains, a and b, connected by disulfide bonds.

This protein is a serine protease that combines with C1q and C1s to form C1, the first component of the classical pathway of the complement system. C1r activates C1s so that it can, in turn, activate C2 and C4.

Superfamily: trypsin

Residues	Feature
51,118	Binding site: carbohydrate (Asn)
39	Active site: His
94	Active site: Asp
191	Active site: Ser
157-176,187-217	Disulfide bonds: (by homology)
114	Disulfide bonds: to a chain (putative)

Mol. wt. unmod. chain = 27,095 Number of residues = 242

```
          5        10        15        20        25        30
  1 I I G G Q K A K M G N F P W Q V F T N I H G R G G G A L L G
 31 D R W I L T A A H T L Y P K E H E A Q S N A S L D V F L G H
 61 T N V E E L M K L G N H P I R R V S V H P D Y R Q D E S Y N
 91 F E G D I A L L E L E N S V T L G P N L L P I C L P D N D T
121 F Y D L G L M G Y V S G F G V M E E K I A H D L R F V R L P
151 V A N P Q A C E N W L R G K N R M D V F S Q N M F C A G H P
181 S L K Q D A C Q G D S G G V F A V R D P N T D R W V A T G I
211 V S W G I G C S R G Y G F Y T K V L N Y V D W I K K E M E E
241 E D
```

HPHU2
Haptoglobin-2 precursor - Human

Maeda, N., J. Biol. Chem. 260, 6698-6709, 1985 (Sequence translated from the DNA sequence)

Yang, F., Brune, J.L., Baldwin, W.D., Barnett, D.R., and Bowman, B.H., Proc. Nat. Acad. Sci. USA 80, 5875-5879, 1983 (Sequence translated from the mRNA sequence)

Raugei, G., Bensi, G., Colantuoni, V., Romano, V., Santoro, C., Costanzo, F., and Cortese, R., Nucl. Acids Res. 11, 5811-5819, 1983 (Sequence of residues 3-406 translated from the mRNA sequence)

Kurosky, A., Barnett, D.R., Lee, T.-H., Touchstone, B., Hay, R.E., Arnott, M.S., Bowman, B.H., and Fitch, W.M., Proc. Nat. Acad. Sci. USA 77, 3388-3392, 1980

If position 71 is Lys, then position 130 is Glu, and vice versa, in the f-s chain shown.

Superfamily: 162-406/trypsin

Residues	Feature
1-18	Domain: signal sequence
19-160,162-406	Protein: haptoglobin-2, alpha-2 and beta chains
29-88,89-146	Duplication:
52-86,111-145	Disulfide bonds: intra-alpha chain
33,92	Disulfide bonds: inter-alpha chain

Mol. wt. unmod. chain = 45,205 Number of residues = 406

```
          5        10        15        20        25        30
  1 M S A L G A V I A L L L W G Q L F A V D S G N D V T D I A D
 31 D G C P K P P E I A H G Y V E H S V R Y Q C K N Y Y K L R T
 61 E G D G V Y T L N D K K Q W I N K A V G D K L P E C E A D D
 91 G C P K P P E I A H G Y V E H S V R Y Q C K N Y Y K L R T E
121 G D G V Y T L N N E K Q W I N K A V G D K L P E C E A V C G
151 K P K N P A N P V Q R I L G G H L D A K G S F P W Q A K M V
181 S H H N L T T G A T L I N E Q W L L T T A K N L F L N H S E
211 N A T A K D I A P T L T L Y V G K K Q L V E I E K V V L H P
241 N Y S Q V D I G L I K L K Q K V S V N E R V M P I C L P S K
271 D Y A E V G R V G Y V S G W G R N A N F K F T D H L K Y V M
301 L P V A D Q D Q C I R H Y E G S T V P E K K T P K S P V G V
331 Q P I L N E H T F C A G M S K Y Q E D T C Y G D A G S A F A
361 V H D L E E D T W Y A T G I L S F D K S C A V A E Y G V Y V
391 K V T S I Q D W V Q K T I A E N
```

EXHU
Factor X (EC 3.4.21.6) precursor - Human (fragment)

Alternate names: Stuart factor

Fung, M.R., Hay, C.W., and MacGillivray, R.T.A., Proc. Nat. Acad. Sci. USA 82, 3591-3595, 1985 (Sequence translated from the mRNA sequence)

Leytus, S.P., Chung, D.W., Kisiel, W., Kurachi, K., and Davie, E.W., Proc. Nat. Acad. Sci. USA 81, 3699-3702, 1984 (Sequence of residues 103-476 translated from the mRNA sequence)

Factor Xa is a vitamin K-dependent glycoprotein that converts prothrombin to thrombin in the presence of factor Va, Ca++, and phospholipid during blood clotting.

Superfamily: 223-476/trypsin

Residues	Feature
1-28	Domain: carboxyl end of signal sequence
29-167,171-476	Protein: factor X, light and heavy chains
29-73	Domain: calcium-binding
29-167,223-476	Protein: factor Xa
34,35,42,44,47, 48,53,54,57, 60,67	Modified residue: gamma-carboxyglutamic acid
91	Modified residue: beta-hydroxyaspartic acid
78-109	Domain: type A homology with EGF
117-152	Domain: type B homology with EGF
209,219	Binding site: carbohydrate (Asn) (possible)
264	Active site: His
310	Active site: Asp
407	Active site: Ser
223-476	Domain: serine protease

Mol. wt. unmod. chain = 53,473 Number of residues = 476

	5				10				15				20				25				30
1/S	L	A	G	L	L	L	L	G	E	S	L	F	I	R	R	E	Q	A	N	N	I	L	A	R	V	T	R	A	N
31	S	F	L	E	E	M	K	K	G	H	L	E	R	E	C	M	E	E	T	C	S	Y	E	E	A	R	E	V	F	E
61	D	S	D	K	T	N	E	F	W	N	K	Y	K	D	G	D	Q	C	E	T	S	P	C	Q	N	Q	G	K	C	K
91	D	G	L	G	E	Y	T	C	T	C	L	E	G	F	E	G	K	N	C	E	L	F	T	R	K	L	C	S	L	D
121	N	G	D	C	D	Q	F	C	H	E	E	Q	N	S	V	V	C	S	C	A	R	G	Y	T	L	A	D	N	G	K
151	A	C	I	P	T	G	P	Y	P	C	G	K	Q	T	L	E	R	R	K	R	S	V	A	Q	A	T	S	S	S	G
181	E	A	P	D	S	I	T	W	K	P	Y	D	A	A	D	L	D	P	T	E	N	P	F	D	L	L	D	F	N	Q
211	T	Q	P	E	R	G	D	N	N	L	T	R	I	V	G	G	Q	E	C	K	D	G	E	C	P	W	Q	A	L	L
241	I	N	E	E	N	E	G	F	C	G	G	T	I	L	S	E	F	Y	I	L	T	A	A	H	C	L	Y	Q	A	K
271	R	F	K	V	R	V	G	D	R	N	T	E	Q	E	E	G	G	E	A	V	H	E	V	E	V	V	I	K	H	N
301	R	F	T	K	E	T	Y	D	F	D	I	A	V	L	R	L	K	T	P	I	T	F	R	M	N	V	A	P	A	C
331	L	P	E	R	D	W	A	E	S	T	L	M	T	Q	K	T	G	I	V	S	G	F	G	R	T	H	E	K	G	R
361	Q	S	T	R	L	K	M	L	E	V	P	Y	V	D	R	N	S	C	K	L	S	S	S	F	I	I	T	Q	N	M
391	F	C	A	G	Y	D	T	K	Q	E	D	A	C	Q	G	D	S	G	G	P	H	V	T	R	F	K	D	T	Y	F
421	V	T	G	I	V	S	W	G	E	S	C	A	R	K	G	K	Y	G	I	Y	T	K	V	T	A	F	L	K	W	I
451	D	R	S	M	K	T	R	G	L	P	K	A	K	S	H	A	P	E	V	I	T	S	S	P	L	K				

KXBOZ
Protein Z - Bovine

Hojrup, P., Jensen, M.S., and Petersen, T.E., FEBS Lett. 184, 333-338, 1985

Protein Z is a single-chain plasma glycoprotein of unknown function. Although homologous with the vitamin K-dependent clotting factors, it has lost two of the essential catalytic residues and has no enzymatic activity.

Superfamily: 143-396/trypsin

Residues	Feature
1-46	Domain: calcium-binding
7,8,11,15,17, 20,21,26,27, 30,33,36,40	Modified site: gamma-carboxyglutamic acid
51-82	Domain: type A homology with EGF
89-125	Domain: type B homology with EGF
64	Modified site: beta-hydroxyaspartic acid
59,191,289	Binding site: carbohydrate (Asn)
388	Binding site: carbohydrate (Thr)
143-396	Domain: homology with trypsin

Mol. wt. unmod. chain = 43,112 Number of residues = 396

```
          5        10        15        20        25        30
  1 A G S Y L L E E L F E G H L E K E C W E E I C V Y E E A R E
 31 V F E D D E T T D E F W R T Y M G G S P C A S Q P C L N N G
 61 S C Q D S I R G Y A C T C A P G Y E G P N C A F A E S E C H
 91 P L R L D G C Q H F C Y P G P E S Y T C S C A R G H K L G Q
121 D R R S C L P H D R C A C G T L G P E C C Q R P Q G S Q Q N
151 L L P F P W Q V K L T N S E G K D F C G G V L I Q D N F V L
181 T T A T C S L L Y A N I S V K T R S H F R L H V R G V H V H
211 T R F E A D T G H N D V A L L D L A R P V R C P D A G R P V
241 C T A D A D F A D S V L L P Q P G V L G G W T L R G R E M V
271 P L R L R V T H V E P A E C G R A L N A T V T T R T S C E R
301 G A A A G A A R W V A G G A V V R E H R G A W F L T G L L G
331 A A P P E G P G P L L L I K V P R Y A L W L R Q V T Q Q P S
361 R A S P R G D R G Q G R D G E P V P G D R G G R W A P T A L
391 P P G P L V
```

DBHU
Complement factor D - Human

Alternate names: C3 convertase activator

Niemann, M.A., Bhown, A.S., Bennett, J.C., and Volanakis, J.E., Biochemistry 23, 2482-2486, 1984

Factor D cleaves factor B when the latter is complexed with factor C3b, activating the C3bBb complex, which then becomes the C3 convertase of the alternate pathway.

Superfamily: trypsin

Residues	Feature
26-42,124-184, 155-165,174-199	Disulfide bonds: (by homology)
41	Active site: His
88	Active site: Asp
178	Active site: Ser

Mol. wt. unmod. chain = 23,817 Number of residues = 222

	5				10				15				20				25				30
1	I	L	G	G	R	E	A	E	A	H	A	R	P	Y	M	A	S	V	Q	L	N
31	G	A	H	L	C	G	G	V	L	V	A	E	Q	W	V	L	S	A	A	H	C
61	L	E	D	A	A	D	G	K	V	Q	V	L	L	G	A	T	H	L	P	Q	P
91	E	P	X	X	X	I	T	I	E	V	L	R	A	V	P	H	P	D	S	Q	P
121	D	T	I	D	H	D	L	L	L	L	Q	L	S	E	K	A	T	L	G	P	A
151	V	R	P	L	P	W	Q	R	V	D	R	D	V	A	P	G	T	L	C	D	V
181	A	G	W	G	I	V	N	H	A	G	R	R	P	D	S	L	Q	H	V	L	L
211	P	V	L	D	R	A	T	C	R	L	Y	D	V	L	R	L	M	C	A	E	S
241	N	R	R	D	S	C	K	G	D	S	G	G	P	L	V	C	G	G	V	L	E
271	G	V	V	V	T	S	G	S	R	V	C	G	N	R	K	K	P	G	I	Y	T
301	R	V	A	T	Y	A	A	W	I	D	H	V	L								

BBHU
Complement factor B (in EC 3.4.21.47) - Human

Alternate names: C3 proactivator

Includes: Bb fragment of C3/C5 convertase

Mole, J.E., Anderson, J.K., Davison, E.A., and Woods, D.E., J. Biol. Chem. 259, 3407-3412, 1984

Christie, D.L., and Gagnon, J., Biochem. J. 209, 61-70, 1983 (Sequence of residues 235-739)

Campbell, R.D., and Porter, R.R., Proc. Nat. Acad. Sci. USA 80, 4464-4468, 1983 (Sequence of residues 321-739 translated from the DNA sequence)

Factor B is cleaved by factor D into two fragments, Ba and Bb. Bb, a serine protease, then combines with complement factor 3b to generate the C3 or C5 convertase. This is part of the alternate pathway of the complement system.

Superfamily: 235-739/trypsin

Residues	Feature
1-234	Peptide: Ba fragment
9-74,75-134, 137-194	Duplication: homology with domains of beta-2-glycoprotein I
97,117,260,353	Binding site: carbohydrate (Asn)
235-739	Peptide: Bb fragment, serine protease
501	Active site: His
551	Active site: Asp
674	Active site: Ser

Mol. wt. unmod. chain = 83,000 Number of residues = 739

```
          5         10        15        20        25        30
  1 T P W S L A R P Q G S C S L E G V E I K G G S F R L L Q E G
 31 Q A L E Y V C P S G F Y P Y P V Q T R T C R S T G S W S T L
 61 K T Q D Q K T V R K A E C R A I H C P R P H D F E N G E Y W
 91 P R S P Y Y N V S D E I S F H C Y D G Y T L R G S A N R T C
121 Q V N G R W S G Q T A I C D N G A G Y C S N P G I P I G T R
151 K V G S Q Y R L E D S V T Y H C S R G L T L R G S Q R R T C
181 Q E G G S W S G T E P S C Q D S F M Y D T P Q E V A E A F L
211 S S L T E T I E G V D A E D G H G P G E Q Q K R K I V L D P
241 S G S M N I Y L V L D G S D S I G A S N F T G A K K C L V N
271 L I E K V A S Y G V K P R Y G L V T Y A T Y P K I W V K V S
301 E A D S S N A D W V T K Q L N E I N Y E D H K L K S G T N T
331 K K A L Q A V Y S M M S W P D D V P P E G W N R T R H V I I
361 L M T D G L H N M G G D P I T V I D E I R D L L Y I G K D R
391 K N P R E D Y L D V Y V F G V G P L V N Q V N I N A L A S K
421 K D N E Q H V F K V K D M E N L E D V F Y Q M I D E S Q S L
451 S L C G M V W E H R K G T D Y H K Q P W Q A K I S V I R P S
481 K G H E S C M G A V V S E Y F V L T A A H C F T V D D K E H
511 S I K V S V G G E K R D L E I E V V L F H P N Y N I N G K K
541 E A G I P E F Y D Y D V A L I K L K N K L K Y G Q T I R P I
571 C L P C T E G T T R A L R L P P T T T C Q Q Q K E E L L P A
601 Q D I K A L F V S E E E K K L T R K E V Y I K N G D K K G S
631 C E R D A Q Y A P G Y D K V K D I S E V V T P R F L C T G G
661 V S P Y A D P N T C R G D S G G P L I V H K R S R F I Q V G
691 V I S W G V V D V C K N Q K R Q K Q V P A H A R D F H I N L
721 F Q V L P W L K E K L Q D E D L G F L
```

KFHU12
Factor XII (EC 3.4.21.38) - Human

Alternate names: Hageman factor

McMullen, B.A., and Fujikawa, K., J. Biol. Chem. 260, 5328-5341, 1985 (Sequence of residues 1-360)

Fujikawa, K., and McMullen, B.A., J. Biol. Chem. 258, 10924-10933, 1983 (Sequences of residues 335-343 and 354-596)

Factor XII is a serum glycoprotein that participates in the initiation of blood coagulation, fibrinolysis, and the generation of bradykinin and angiotensin.

Factor XII, prekallikrein, and HMW kininogen form a complex bound to an anionic surface. Prekallikrein is cleaved by factor XII to form kallikrein, which then cleaves factor XII first to alpha-factor XIIa and then to beta-factor XIIa. Alpha-factor XIIa activates factor XI to factor XIa.

Superfamily: 354-396/trypsin

Residues	Feature
1-353,354-596	Protein: alpha-factor XIIa
28-69	Domain: type II homology with fibronectin
79-111	Domain: type A homology with EGF
116-151	Domain: type I homology with fibronectin
159-190	Domain: type A homology with EGF
198-276	Domain: kringle
230	Binding site: carbohydrate (Asn)
280,286,309, 310,318	Binding site: carbohydrate (Thr) (possible)
289	Binding site: carbohydrate (Ser) (possible)
354-596	Domain: serine protease
335-343,354-596	Protein: beta-factor XIIa

Mol. wt. unmod. chain = 65,759 Number of residues = 596

```
          5        10        15        20        25        30
  1 I P P W E A P K E H K Y K A E E H T V V L T V T G E P C H F
 31 P F Q Y H R Q L Y H K C T H K G R P G P Q P W C A T T P N F
 61 D Q D Q R W G Y C L E P K K V K D H C S K H S P C Q K G G T
 91 C V N M P S G P H C L C P Q H L T G N H C Q K E K C F E P Q
121 L L R F F H K N E I W Y R T E Q A A V A R C Q C K G P D A H
151 C Q R L A S Q A C R T N P C L H G G R C L E V E G H R L C H
181 C P V G Y T G P F C D V D T K A S C Y D G R G L S Y R G L A
211 R T T L S G A P C Q P W A S E A T Y R N V T A E Q A R N W G
241 L G G H A F C R N P D N D I R P W C F V L N R D R L S W E Y
271 C D L A Q C Q T P T Q A A P P T P V S P R L H V P L M P A Q
301 P A P P K P Q P T T R T P P Q S Q T P G A L P A K R E Q P P
331 S L T R N G P L S C G Q R L R K S L S S M T R V V G G L V A
361 L R G A H P Y I A A L Y W G H S F C A G S L I A P C W V L T
391 A A H C L Q D R P A P E D L T V V L G Q E R R N H S C E P C
421 Q T L A V R S Y R L H E A F S P V S Y Q H D L A L L R L Q E
451 D A D G S C A L L S P Y V Q P V C L P S G A A R P S E T T L
481 C Q V A G W G H Q F E G A E E Y A S F L Q E A Q V P F L S L
511 E R C S A P D V H G S S I L P G M L C A G F L E G G T D A C
541 Q G D S G G P L V C E D Q A A E R R L T L Q G I I S W G S G
571 C G D R N K P G V Y T D V A Y Y L A W I R E H T V S
```

TIHUBI
Inter-alpha-trypsin inhibitor (BPI type) - Human

Wachter, E., and Hochstrasser, K., Hoppe-Seyler's Z. Physiol. Chem. 362, 1351-1355, 1981

Superfamily: basic protease inhibitor

Residues	Feature
1-21	Domain: glycopeptide
22-77	Domain: inactive
78-143	Domain: active
26-76,35-59, 51-72,82-132, 91-115,107-128	Disulfide bonds:
22-77,78-133	Duplication: homology with basic protease inhibitor
92	Reactive site: Arg (trypsin)
10	Binding site: carbohydrate (Ser)
45	Binding site: carbohydrate (Asn)

Mol. wt. unmod. chain = 15,541 Number of residues = 143

```
          5         10        15        20        25        30
  1 A V L P Q E E E G S G G G Q L V T E V T K K E D S C Q L G Y
 31 S A G P C M G M T S R Y F Y N G T S M A C E T F Q Y G G C M
 61 G N G N N F V T E K E C L Q T C R T V A A C N L P V I R G P
 91 C R A F I Q L W A F D A V K G K C V L F P Y G G C Q G N G N
121 K F Y S E K E C R E Y C G V P G D E D E E L L
```

ITHUC
Alpha-1-antichymotrypsin precursor - Human

Chandra, T., Stackhouse, R., Kidd, V.J., Robson, K.J.H., and Woo, S.L.C., Biochemistry 22, 5055-5061, 1983 (Sequence translated from the mRNA sequence)
It is uncertain whether Met-1 or Met-4 is the initiator.

The concentration of this plasma protease inhibitor, which is synthesized in the liver, increases in the acute phase of inflammation or infection. It can inhibit neutrophil cathepsin G and mast cell chymase, both of which can convert angiotensin I to the active angiotensin II.

Superfamily: antithrombin-III

Residues	Feature
1-25	Domain: signal sequence
383	Reactive site: Leu (chymotrypsin)

Mol. wt. unmod. chain = 48,636 Number of residues = 433

```
          5         10        15        20        25        30
  1 M E R M L P L L A L G L L A A G F C P A V L C H P N S P L D
 31 E E N L T Q E N Q D R G T H V D L G L A S A N V D F A F S L
 61 Y K Q L V L K A L D K N V I F S P L S I S T A L A F L S L G
 91 A H N T T L T E I L K A S S S P H G D L L R Q K F T Q S F Q
121 H L R A P S I S S S D E L Q L S M G N A M F V K E Q L S L L
151 D R F T E D A K R L Y G S E A F A T D F Q D S A A A K K L I
181 N D Y V K N G T R G K I T D L I K D P D S Q T M M V L V N Y
211 I F F K A K W E M P F D P Q D T H Q S R F Y L S K K K W V M
241 V P M M S L H H L T I P Y F R D E E L S C T V V E L K Y T G
271 N A S A L F I L P D Q D K M E E V E A M L L P E T L K R W R
301 D S L E F R E I G E L Y L P K F S I S R D Y N L N D I L L Q
331 L G I E E A F T S K A D L S G I T G A R N L A V S Q V V H K
361 V V S D V F E E G T E A S A A T A V K I T L L S A L V E T R
391 T I V R F N R P F L M I I V P T D T Q N I F F M S K V T N P
421 S K P R A C I K Q W G S Q
```

ANHU
Angiotensinogen precursor - Human

Includes: angiotensin I; angiotensin II

Kageyama, R., Ohkubo, H., and Nakanishi, S., Biochemistry 23, 3603-3609, 1984 (Sequence translated from the mRNA sequence)
It is uncertain whether Met-1 or Met-10 is the initiator.

Superfamily: antithrombin-III

Residues	Feature
1-33	Domain: signal sequence
34-43	Peptide: angiotensin I
34-41	Peptide: angiotensin II
47,170,304,328	Binding site: carbohydrate (Asn) (possible)

Mol. wt. unmod. chain = 53,154 Number of residues = 485

```
            5         10        15        20        25        30
  1 M R K R A P Q S E M A P A G V S L R A T I L C L L A W A G L
 31 A A G D R V Y I H P F H L V I H N E S T C E Q L A K A N A G
 61 K P K D P T F I P A P I Q A K T S P V D E K A L Q D Q L V L
 91 V A A K L D T E D K L R A A M V G M L A N F L G F R I Y G M
121 H S E L W G V V H G A T V L S P T A V F G T L A S L Y L G A
151 L D H T A D R L Q A I L G V P W K D K N C T S R L D A H K V
181 L S A L Q A V Q G L L V A Q G R A D S Q A Q L L L S T V V G
211 V F T A P G L H L K Q P F V Q G L A L Y T P V V L P R S L D
241 F T E L D V A A E K I D R F M Q A V T G W K T G C S L M G A
271 S V D S T L A F N T Y V H F Q G K M K G F S L L A E P Q E F
301 W V D N S T S V S V P M L S G M G T F Q H W S D I Q D N F S
331 V T Q V P F T E S A C L L L I Q P H Y A S D L D K V E G L T
361 F Q Q N S L N W M K K L S P R T I H L T M P Q L V L Q G S Y
391 D L Q D L L A Q A E L P A I L H T E L N L Q K L S N D R I R
421 V G E V L N S I F F E L E A D E R E P T E S T Q Q L N K P E
451 V L E V T L N R P F L F A V Y D Q S A T A L H F L G R V A N
481 P L S T A
```

C3HU
Complement C3 precursor - Human

Includes: complement C3a anaphylatoxin; C3dk

de Bruijn, M.H.L., and Fey, G.H., Proc. Nat. Acad. Sci. USA 82, 708-712, 1985 (Sequence translated from the mRNA sequence)

Complement C3 contains two chains, formed by removal of residues 668-671 and linked by a disulfide bond. Its activation by a C3 convertase, which is the central reaction in both classical and alternative complement pathways, releases the anaphylatoxin C3a from the amino end of the alpha chain and generates C3b.

C3a is a vasoactive peptide and a mediator of inflammation. C3b, with its highly reactive thiol group, binds to the surface of foreign particles where it is a cofactor in the formation of C5 convertase and facilitates phagocytosis. The activity of C3b is regulated by proteolytic cleavage involving factors H and I. Its degradation products can also be biologically active.

Superfamily: alpha-2-macroglobulin

Residues	Feature
1-22	Domain: signal sequence
23-667,672-1663	Protein: complement C3, beta and alpha chains
85,939	Binding site: carbohydrate (Asn)
672-748	Peptide: C3a anaphylatoxin
23-667,749-1663	Protein: C3b
1617	Binding site: carbohydrate (Asn) (putative)
1010,1013	Thiolester bonds:
946-1303	Peptide: C3dk fragment
955-1303	Peptide: C3dg fragment
955-1001	Peptide: C3g fragment
1002-1303	Peptide: C3d fragment

Mol. wt. unmod. chain = 187,163 Number of residues = 1663

	5					10					15					20					25					30
1	M	G	P	T	S	G	P	S	L	L	L	L	L	T	H	L	P	L	A	L	G	S	P	M	Y	S	I	I	T
31	P	N	I	L	R	L	E	S	E	E	T	M	V	L	E	A	H	D	A	Q	G	D	V	P	V	T	V	T	V	H
61	D	F	P	G	K	K	L	V	L	S	S	E	K	T	V	L	T	P	A	T	N	H	M	G	N	V	T	F	T	I
91	P	A	N	R	E	F	K	S	E	K	G	R	N	K	F	V	T	V	Q	A	T	F	G	T	Q	V	V	E	K	V
121	V	L	V	S	L	Q	S	G	Y	L	F	I	Q	T	D	K	T	I	Y	T	P	G	S	T	V	L	Y	R	I	F
151	T	V	N	H	K	L	L	P	V	G	R	T	V	M	V	N	I	E	N	P	E	G	I	P	V	K	Q	D	S	L
181	S	S	Q	N	Q	L	G	V	L	P	L	S	W	D	I	P	E	L	V	N	M	G	Q	W	K	I	R	A	Y	Y
211	E	N	S	P	Q	Q	V	F	S	T	E	F	E	V	K	E	Y	V	L	P	S	F	E	V	I	V	E	P	T	E
241	K	F	Y	Y	I	Y	N	E	K	G	L	E	V	T	I	T	A	R	F	L	Y	G	K	K	V	E	G	T	A	F
271	V	I	F	G	I	Q	D	G	E	Q	R	I	S	L	P	E	S	L	K	R	I	P	I	E	D	G	S	G	E	V
301	V	L	S	R	K	V	L	L	D	G	V	Q	N	L	R	A	E	D	L	V	G	K	S	L	Y	V	S	A	T	V
331	I	L	H	S	G	S	D	M	V	Q	A	E	R	S	G	I	P	I	V	T	S	P	Y	Q	I	H	F	T	K	T
361	P	K	Y	F	K	P	G	M	P	F	D	L	M	V	F	V	T	N	P	D	G	S	P	A	Y	R	V	P	V	A
391	V	Q	G	E	D	T	V	Q	S	L	T	Q	G	D	G	V	A	K	L	S	I	N	T	H	P	S	Q	K	P	L
421	S	I	T	V	R	T	K	K	Q	E	L	S	E	A	E	Q	A	T	R	T	M	Q	A	L	P	Y	S	T	V	G

```
 451 N S N N Y L H L S V L R T E L R P G E T L N V N F L L R M D
 481 R A H E A K I R Y Y T Y L I M N K G R L L K A G R Q V R E P
 511 G Q D L V V L P L S I T T D F I P S F R L V A Y Y T L I G A
 541 S G Q R E V V A D S V W V D V K D S C V G S L V V K S G Q S
 571 E D R Q P V P G Q Q M T L K I E G D H G A R V V L V A V D K
 601 G V F V L N K K N K L T Q S K I W D V V E K A D I G C T P G
 631 S G K D Y A G V F S D A G L T F T S S S G Q Q T A Q R A E L
 661 Q C P Q P A A R R R R S V Q L T E K R M D K V G K Y P K E L
 691 R K C C E D G M R E N P M R F S C Q R R T R F I S L G E A C
 721 K K V F L D C C N Y I T E L R R Q H A R A S H L G L A R S N
 751 L D E D I I A E E N I V S R S E F P E S W L W N V E D L K E
 781 P P K N G I S T K L M N I F L K D S I T T W E I L A V S M S
 811 D K K G I C V A D P F E V T V M Q D F F I D L R L P Y S V V
 841 R N E Q V E I R A V L Y N Y R Q N Q E L K V R V E L L H N P
 871 A F C S L A T T K R R H Q Q T V T I P P K S S L S V P Y V I
 901 V P L K T G L Q E V E V K A A V Y H H F I S D G V R K S L K
 931 V V P E G I R M N K T V A V R T L D P E R L G R E G V Q K E
 961 D I P P A D L S D Q V P D T E S E T R I L L Q G T P V A Q M
 991 T E D A V D A E R L K H L I V T P S G C G E Q N M I G M T P
1021 T V I A V H Y L D E T E Q W E K F G L E K R Q G A L E L I K
1051 K G Y T Q Q L A F R Q P S S A F A A F V K R A P S T W L T A
1081 Y V V K V F S L A V N L I A I D S Q V L C G A V K W L I L E
1111 K Q K P D G V F Q E D A P V I H Q E M I G G L R N N N E K D
1141 M A L T A F V L I S L Q E A K D I C E E Q V N S L P G S I T
1171 K A G D F L E A N Y M N L Q R S Y T V A I A G Y A L A Q M G
1201 R L K G P L L N K F L T T A K D K N R W E D P G K Q L Y N V
1231 E A T S Y A L L A L L Q L K D F D F V P P V V R W L N E Q R
1261 Y Y G G G Y G S T Q A T F M V F Q A L A Q Y Q K D A P D H Q
1291 E L N L D V S L Q L P S R S S K I T H R I H W E S A S L L R
1321 S E E T K E N E G F T V T A E G K G Q G T L S V V T M Y H A
1351 K A K D Q L T C N K F D L K V T I K P A P E T E K R P Q D A
1381 K N T M I L E I C T R Y R G D Q D A T M S I L D I S M M T G
1411 F A P D T D D L K Q L A N G V D R Y I S K Y E L D K A F S D
1441 R N T L I I Y L D K V S H S E D D C L A F K V H Q Y F N V E
1471 L I Q P G A V K V Y A Y Y N L E E S C T R F Y H P E K E D G
1501 K L N K L C R D E L C R C A E E N C F I Q K S D D K V T L E
1531 E R L D K A C E P G V D Y V Y K T R L V K V Q L S N D F D E
1561 Y I M A I E Q T I K S G S D E V Q V G Q Q R T F I S P I K C
1591 R E A L K L E E K K H Y L M W G L S S D F W G E K P N L S Y
1621 I I G K D T W V E H W P E E D E C Q D E E N Q K Q C Q D L G
1651 A F T E S M V V F G C P N
```

C4HU
Complement C4A - Human

Includes: complement C4a anaphylatoxin

Belt, K.T., Carroll, M.C., and Porter, R.R., Cell 36, 907-914, 1984 (Sequence translated from the mRNA sequence)

Gln-991 is deamidated after translation.

Complement C4 is cleaved after translation into three chains: beta, alpha, and gamma. Residues 1424 or 1427 may be the carboxyl end of the alpha chain.

There are at least two genes coding for C4, C4A and C4B. Each gene has many alleles.

Superfamily: alpha-2-macroglobulin

Residues	Feature
1-656,661-1424, 1432-1722	Protein: complement C4, beta, alpha, and gamma chains
991-994	Thiolester bonds:
661-737	Peptide: C4a anaphylatoxin

Mol. wt. unmod. chain = 190,083 Number of residues = 1722

```
          5         10        15        20        25        30
  1 K P R L L L F S P S V V H L G V P L S V G V Q L Q D V P R G
 31 Q V V K G S V F L R N P S R N N V P C S P K V D F T L S S E
 61 R D F A L L S L Q V P L K D A K S C G L H Q L L R G P E V Q
 91 L V A H S P W L K D S L S R T T N I Q G I N L L F S S R R G
121 H L F L Q T D Q P I Y N P G Q R V R Y R V F A L D Q K M R P
151 S T D T I T V M V E N S H G L R V R K K E V Y M P S S I F Q
181 D D F V I P D I S E P G T W K I S A R F S D G L E S N S S T
211 Q F E V K K Y V L P N F E V K I T P G K P Y I L T V P G H L
241 D E M Q L D I Q A R Y I Y G K P V Q G V A Y V R F G L L D E
271 D G K K T F F R G L E S Q T K L V N G Q S H I S L S K A E F
301 Q D A L E K L N M G I T D L Q G L R L Y V A A A I I E S P G
331 G E M E E A E L T S W Y F V S S P F S L D L S K T K R H L V
361 P G A P F L L Q A L V R E M S G S P A S G I P V K V S A T V
391 S S P G S V P E A Q D I Q Q N T D G S G Q V S I P I I I P Q
421 T I S E L Q L S V S A G S P H P A I A R L T V A A P P S G G
451 P G F L S I E R P D S R P P R V G D T L N L N L R A V G S G
481 A T F S H Y Y Y M I L S R G Q I V F M N R E P K R T L T S V
511 S V F V D H H L A P S F Y F V A F Y Y H G D H P V A N S L R
541 V D V Q A G A C E G K L E L S V D G A K Q Y R N G E S V K L
571 H L E T D S L A L V A L G A L D T A L Y A A G S K S H K P L
601 N M G K V F E A M N S Y D L G C G P G G G D S A L Q V F Q A
631 A G L A F S D G D Q W T L S R K R L S C P K E K T T R K K R
661 N V N F Q K A I N E K L G Q Y A S P T A K R C C Q D G V T R
691 L P M M R S C E Q R A A R V Q Q P D C R E P F L S C C Q F A
721 E S L R K K S R D K G Q A G L Q R A L E I L Q E E D L I D E
751 D D I P V R S F F P E N W L W R V E T V D R F Q I L T L W L
781 P D S L T T W E I H G L S L S K T K G L C V A T P V Q L R V
811 F R E F H L H L R L P M S V R R F E Q L E L R P V L Y N Y L
841 D K N L T V S V H V S P V E G L C L A G G G G L A Q Q V L V
871 P A G S A R P V A F S V V P T A A A A V S L K V V A R G S F
901 E F P V G D A V S K V L Q I E K E G A I H R E E L V Y E L N
```

```
 931 P L D H R G R T L E I P G N S D P N M I P D G D F N S Y V R
 961 V T A S D P L D T L G S E G A L S P G G V A S L L R L P R G
 991 C G E Q T M I Y L A P T L A A S R Y L D K T E Q W S T L P P
1021 E T K D H A V D L I Q K G Y M R I Q Q F R K A D G S Y A A W
1051 L S R D S S T W L T A F V L K V L S L A Q E Q V G G S P E K
1081 L Q E T S N W L L S Q Q Q A D G S F Q D P C P V L D R S M Q
1111 G G L V G N D E T V A L T A F V T I A L H H G L A V F Q D E
1141 G A E P L K Q R V E A S I S K A N S F L G E K A S A G L L G
1171 A H A A A I T A Y A L S L T K A P V D L L G V A H N N L M A
1201 M A Q E T G D N L Y W G S V T G S Q S N A V S P T P A P R N
1231 P S D P M P Q A P A L W I E T T A Y A L L H L L L H E G K A
1261 E M A D Q A S A W L T R Q G S F Q G G F R S T Q D T V I A L
1291 D A L S A Y W I A S H T T E E R G L N V T L S S T G R N G F
1321 K S H A L Q L N N R Q I R G L E E E L Q F S L G S K I N V K
1351 V G G N S K G T L K V L R T Y N V L D M K N T T C Q D L Q I
1381 E V T V K G H V E Y T M E A N E D Y E Y D E L P A K D D P D
1411 A P L Q P V T P L Q L F E G R R N R R R R E A P K V V E E Q
1441 E S R V H Y T V C I W R N G K V G L S G M A I A D V T L L S
1471 G F H A L R A D L E K L T S L S D R Y V S H F E T E G P H V
1501 L L Y F D S V P T S R E C V G F E A V Q E V P V G L V Q P A
1531 S A T L Y D Y Y N P E R R C S V F Y G A P S K S R L L A T L
1561 C S A E V C Q C A E G K C P R Q R R A L E R G L Q D E D G Y
1591 R M K F A C Y Y P R V E Y G F Q V K V L R E D S R A A F R L
1621 F E T K I T Q V L H F T K D V K A A A N Q M R N F L V R A S
1651 C R L R L E P G K E Y L I M G L D G A T Y D L E G H P Q Y L
1681 L D S N S W I E E M P S E R L C R S T R Q R A A C A Q L N D
1711 F L Q E Y G T Q G C Q V
```

C5HU
Complement C5 - Human (fragment)

Includes: complement C5a anaphylatoxin

Lundwall, A.B., Wetsel, R.A., Kristensen, T., Whitehead, A.S., Woods, D.E., Ogden, R.C., Colten, H.R., and Tack, B.F., J. Biol. Chem. 260, 2108-2112, 1985 (Sequence translated from the mRNA sequence)

Activation of C5 by a C5 convertase initiates the spontaneous assembly of the late complement components, C5-9, into the membrane attack, or lytic, complex. The convertase also releases the anaphylatoxin C5a from the amino end of the alpha chain and generates C5b. C5a has potent spasmogenic and chemotactic activity. C5b has a transient binding site for C6. The C5b,6 complex is the foundation upon which the lytic complex is assembled.

Superfamily: alpha-2-macroglobin

Residues	Feature
1-262,267-491	Protein: complement C5 (fragment), beta chain (carboxyl end) and alpha chain (amino end)
330	Binding site: carbohydrate (Asn)
267-340	Peptide: C5a anaphylatoxin
1-262,341-491	Protein: C5b (fragment)

Mol. wt. unmod. chain = 54,280 Number of residues = 491

```
          5        10        15        20        25        30
  1/R V D D G V A S F V L N L P S G V T V L E F N V K T D A P D
 31 L P E E N Q A R E G Y R A I A Y S S L S Q S Y L Y I D W T D
 61 N H K A L L V G E H L N I I V T P K S P Y I D K I T H Y N Y
 91 L I L S K G K I I H F G T R E K F S D A S Y Q S I N I P V T
121 Q N M V P S S R L L V Y Y I V T G E Q T A E L V S D S V W L
151 N I E E K C G N Q L Q V H L S P D A D A Y S P G Q T V S L N
181 M A T G M D S W V A L A A V D S A V Y G V Q R G A K K P L E
211 R V F Q F L E K S D L G C G A G G G L N N A N V F H L A G L
241 T F L T N A N A D D S Q E N D E P C K E I L R P R R T L Q K
271 K I E E I A A K Y K H S V V K K C C Y D G A C V N N D E T C
301 E Q R A A R I S L G P R C I K A F T E C C V V A S Q L R A N
331 I S H K D M Q L G R L H M K T L L P V S K P E I R S Y F P E
361 S W L W E V H L V P R R K Q L Q F A L P D S L T T W E I Q G
391 I G I S N T G I C V A D T V K A K V F K D V F L E M N I P Y
421 S V V R G E Q I Q L K G T V Y N Y R T S G M Q S L A L S P R
451 L E C N G K I S G H C K L R L P G S S D S P A S A S Q V A G
481 I T G T H H H A Q P T/
```

C9HU
Complement C9 precursor - Human (fragment)

Stanley, K.K., Kocher, h.-P., Luzio, J.P., Jackson, P., and Tschopp, J., EMBO J. 4, 375-382, 1985 (Sequence translated from the mRNA sequence)

DiScipio, R.G., Gehring, M.R., Podack, E.R., Kan, C.C., Hugli, T.E., and Fey, G.H., Proc. Nat. Acad. Sci. USA 81, 7298-7302, 1984 (Sequence of residues 3-559 translated from the mRNA sequence)

Thrombin cleaves factor C9 between His-266 and Gly-267 to produce C9a and C9b.

C9 is the final component of the complement system to be added in the assembly of the membrane attack complex. It is able to enter lipid bilayers, forming transmembrane channels.

Superfamily: complement C9

Residues	Feature
1-266	Protein: carboxyl end of C9a
267-559	Protein: C9b
278,415	Binding site: carbohydrate (Asn) (probable)
513-546	Domain: type A homology with EGF

Number of residues = 559

```
                5         10        15        20        25        30
  1/S M S A C R S F A V A I C I L E I S I L T A Q Y T T S Y D P
 31 E L T E S S G S A S H I D C R M S P W S E W S Q C D P C L R
 61 Q M F R S R S I E V F G Q F N G K R C T D A V G D R R Q C V
 91 P T E P C E D A E D D C G N D F Q C S T G R C I K M R L R C
121 N G D N D C G D F S D E D D C E S E P R P P C R D R V V E E
151 S E L A R T A G Y G I N I L G M D P L S T P F D N E F Y N G
181 L C N R D R D G N T L T Y Y R R P W N V A S L I Y E T K G E
211 K N F R T E H Y E E Q I E A F K S I I Q E K T S N F N A A I
241 S L K F T P T E T N K A E Q C C E E T A S S I S L H G K G S
271 F R F S Y S K N E T Y Q L F L S Y S S K K E K M F L H V K G
301 E I H L G R F V M R N R D V L T T T F V D D I K A L P T T Y
331 E K G E Y F A F L E T Y G T H Y S S S G S L G G L Y E L I Y
361 V L D K A S M K R K G V E L K D I K R C L G Y H L D V S L A
391 F S E I S V G A E F N K D D C V K R G E G R A V N I T S E N
421 L I D D V V S L I R G G T R K Y A F E L K E K L L R G T V I
451 D V T D F V N W A S S I N D A P V L I S Q K L S P I Y N L V
481 P V K M K N A H L K K Q N L E R A I E D Y I N E F S V R K C
511 H T C Q N G G T V I L M D G K C L C A C P F K F E G I A C E
541 I S K Q K I S E G L P A L E F P N E K
```

FNHU
Fibronectin - Human

Kornblihtt, A.R., Umezawa, K., Vibe-Pedersen, K., and Baralle, F.E., EMBO J. 4, 1755-1759, 1985 (Sequences of residues 1-2050 and 2082-2355 translated from the mRNA sequence)
This sequence differs from that shown in lacking residue 1305. The nucleotide sequence is not given in this paper.

Kornblihtt, A.R., Vibe-Pedersen, K., and Baralle, F.E., Nucl. Acids Res. 12, 5853-5868, 1984 (Sequences of residues 942-2050 and 2082-2355 translated from the mRNA sequence)

Bernard, M.P., Kolbe, M., Weil, D., and Chu, M.-L., Biochemistry 24, 2698-2704, 1985 (Sequence of residues 1563-2355 translated from the mRNA sequence)

Each of the "extra domain" and the connecting strand 3 are present in some forms of fibronectin and absent in others. These differences are probably due to alternate splicing of the mRNA.

Superfamily: fibronectin

Residues	Feature
1	Modified residue: pyrrolidone carboxylic acid
21-241	Domain: fibrin- and heparin-binding region
277-577	Domain: collagen-binding region
779-1054	Domain: DNA-binding region
1410-1517	Domain: cell-attachment domain
1600-1689	Domain: extra domain
1690-1960	Domain: second heparin-binding region
1961-2071	Domain: connecting strand 3
2175-2306	Domain: second fibrin-binding
21-65,66-109, 110-154,155-199, 200-241,277-313, 439-486,487-529, 530-577,2175-2219, 2220-2263,2264-2306	Duplication: type I homology regions ("fingers")
314-373,374-438	Duplication: type II homology regions
578-669,688-778, 779-874,875-964, 965-1054,1055-1141, 1142-1234,1235-1325, 1326-1415,1416-1509, 1510-1599,1600-1689, 1690-1781,1782-1870, 1871-1960,2072-2151	Duplication: type III homology regions

Mol. wt. unmod. chain = 259,621 Number of residues = 2355

```
          5        10        15        20        25        30
 1 Q A Q Q M V Q P Q S P V A V S Q S K P G C Y D N G K H Y Q I
31 N Q Q W E R T Y L G N V L V C T C Y G G S R G F N C E S K P
61 E A E E T C F D K Y T G N T Y R V G D T Y E R P K D S M I W
```

```
  91 D C T C I G A G R G R I S C T I A N R C H E G G Q S Y K I G
 121 D T W R R P H E T G G Y M L E C V C L G N G K G E W T C K P
 151 I A E K C F D H A A G T S Y V V G E T W E K P Y Q G W M M V
 181 D C T C L G E G S G R I T C T S R N R C N D Q D T R T S Y R
 211 I G D T W S K K D N R G N L L Q C I C T G N G R G E W K C E
 241 R H T S V Q T T S S G S G P F T D V R A A V Y Q P Q P H P Q
 271 P P P Y G H C V T D S G V V Y S V G M Q W L K T Q G N K Q M
 301 L C T C L G N G V S C Q E T A V T Q T Y G G N S N G E P C V
 331 L P F T Y N G R T F Y S C T T E G R Q D G H L W C S T T S N
 361 Y E Q D Q K Y S F C T D H T V L V Q T Q G G N S N G A L C H
 391 F P F L Y N N H N Y T D C T S E G R R D N M K W C G T T Q N
 421 Y D A D Q K F G F C P M A A H E E I C T T N E G V M Y R I G
 451 D Q W D K Q H D M G H M M R C T C V G N G R G E W T C Y A Y
 481 S Q L R D Q C I V D D I T Y N V N D T F H K R H E E G H M L
 511 N C T C F G Q G R G R W K C D P V D Q C Q D S E T G T F Y Q
 541 I G D S W E K Y V H G V R Y Q C Y C Y G R G I G E W H C Q P
 571 L Q T Y P S S S S G P V E V F I T E T P S Q P N S H P I Q W N
 601 A P Q P S H I S K Y I L R W R P K N S V G R W K E A T I P G
 631 H L N S Y T I K G L K P G V V Y E G Q L I S I Q Q Y G H Q E
 661 V T R F D F T T S S T P V T S N T V T G E T T P F S P L
 691 V A T S E S V T E I T A S S F V V S W V S A S D T V S G F R
 721 V E Y E L S E E G D E P Q Y L D L P S T A T S V N I P D L L
 751 P G R K Y I V N V Y Q I S E D G E Q S L I L S T S Q T T A P
 781 D A P P D P T V D Q V D D T S I V V R W S R P Q A P I T G Y
 811 R I V Y S P S V E G S S T E L N L P E T A N S V T L S D L Q
 841 P G V Q Y N I T I Y A V E E N Q E S T P V V I Q Q E T T G T
 871 P R S D T V P S P R D L Q F V E V T D V K V T I M W T P P E
 901 S A V T G Y R V D V I P V N L P G E H G Q R L P I S R N T F
 931 A E V T G L S P G V T Y Y F K V F A V S H G R E S K P L T A
 961 Q Q T T K L D A P T N L Q F V N E T D S T V L V R W T P P R
 991 A Q I T G Y R L T V G L T R R G Q P R Q Y N V G P S V S K Y
1021 P L R N L Q P A S E Y T V S L V A I K G N Q E S P K A T G V
1051 F T T L Q P G S S I P P Y N T E V T E T T I V I T W T P A P
1081 R I G F K L G V R P S Q G G E A P R E V T S D S G S I V V S
1111 G L T P G V E Y V Y T I Q V L R D G Q E R D A P I V N K V V
1141 T P L S P P T N L H L E A N P D T G V L T V S W E R S T T P
1171 D I T G Y R I T T T P T N G Q Q G N S L E E V V H A D Q S S
1201 C T F D N L S P G L E Y N V S V Y T V K D D K E S V P I S D
1231 T I I P A V P P P T D L R F T N I G P D T M R V T W A P P P
1261 S I D L T N F L V R Y S P V K N E E D V A E L S I S P S D N
1291 A V V L T N L L P G T E Y V V S V S S V Y E Q H E S T P L R
1321 G R Q K T G L D S P T G I D F S D I T A N S F T V H W I A P
1351 R A T I T G Y R I R H H P E H F S G R P R E D R V P H S R N
1381 S I T L T N L T P G T E Y V V S I V A L N G R E E S P L L I
1411 G Q Q S T V S D V P R D L E V V A A T P T S L L I S W D A P
1441 A V T V R Y Y R I T Y G E T G G N S P V Q E F T V P G S K S
1471 T A T I S G L K P G V D Y T I T V Y A V T G R G D S P A S S
1501 K P I S I N Y R T E I D K P S Q M Q V T D V Q D N S I S V K
1531 W L P S S S P V T G Y R V T T T P K N G P G P T K T K T A G
1561 P D Q T E M T I E G L Q P T V E Y V V S V Y A Q N P S G E S
1591 Q P L V Q T A V T N I D R P K G L A F T D V D V D S I K I A
1621 W E S P Q G Q V S R Y R V T Y S S P E D G I H E L F P A P D
1651 G E E D T A E L Q G L R P G S E Y T V S V V A L H D D M E S
1681 Q P L I G T Q S T A I P A P T D L K F T Q V T P T S L S A Q
1711 W T P P N V Q L T G Y R V V T P K E K T G P M K E I N L A
1741 P D S S S V V V S G L M V A T K Y E V S V Y A L K D T L T S
1771 R P A Q G V V T T L E N V S P P R R A R V T D A T E T T I T
1801 I S W R T K T E T I T G F Q V D A V P A N G Q T P I Q R T I
1831 K P D V R S Y T I T G L Q P G T D Y K I Y L Y T L N D N A R
```

```
1861 S S P V V I D A S T A I D A P S N L R F L A T T P N S L L V
1891 S W Q P P R A R I T G Y I I K Y E K P G S P P R E V V P R P
1921 R P G V T E A T I T G L E P G T E Y T I Y V I A L K N N Q K
1951 S E P L I G R K K T D E L P Q L V T L P H P N L H G P E I L
1981 D V P S T V Q K T P F V T H P G Y D T G N G I Q L P G T S G
2011 Q Q P S V G Q Q M I F E E H G F R R T T P P T T A T P I R H
2041 R P R P Y P P N V G E E I Q I G H I P R E D V D Y H L Y P H
2071 G P G L N P N A S T G Q E A L S Q T T I S W A P F Q D T S E
2101 Y I I S C H P V G T D E E P L Q F R V P G T S T S A T L T G
2131 L T R G A T Y N I I V E A L K D Q Q R H K V R E E V V T V G
2161 N S V N E G L N Q P T D D S C F D P Y T V S H Y A V G D E W
2191 E R M S E S G F K L L C Q C L G F G S G H F R C D S S R W C
2221 H D N G V N Y K I G E K W D R Q G E N G Q M M S C T C L G N
2251 G K G E F K C D P H E A T C Y D D G K T Y H V G E Q W Q K E
2281 Y L G A I C S C T C F G G Q R G W R C D N C R R P G G E P S
2311 P E G T T G Q S Y N Q Y S Q R Y H Q R T N T N V N C P I E C
2341 F M P L D V Q A D R E D S R E
```

KGBOH1
Kininogen, HMW I precursor - Bovine

Includes: bradykinin (kallidin I)

Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H., and Nakanishi, S., Nature 305, 545-549, 1983 (Sequence translated from the mRNA sequence)

Bradykinin is released from kininogen by kallikrein.

Superfamily: cystatin

Residues	Feature
1-22	Domain: signal sequence (probable)
19-135,136-257, 258-379	Duplication: homology with cystatin
380-388	Peptide: bradykinin

Mol. wt. unmod. chain = 68,890 Number of residues = 621

```
            5         10        15        20        25        30
  1 M K L I T I L F L C S R L L P S L T Q E S S Q E I D C N D Q
 31 D V F K A V D A A L T K Y N S E N K S G N Q F V L Y R I T E
 61 V A R M D N P D T F Y S L K Y Q I K E G D C P F Q S N K T W
 91 Q D C D Y K D S A Q A A T G E C T A T V A K R G N M K F S V
121 A I Q T C L I T P A E G P V V T A Q Y E C L G C V H P I S T
151 K S P D L E P V L R Y A I Q Y F N N N T S H S H L F D L K E
181 V K R A Q R Q V V S G W N Y E V N Y S I A Q T N C S K E E F
211 S F L T P D C K S L S S G D T G E C T D K A H V D V K L R I
241 S S F S Q K C D L Y P V K D F V Q P P T R L C A G C P K P I
271 P V D S P D L E E P L S H S I A K L N A E H D G A F Y F K I
301 D T V K K A T V Q V V A G L K Y S I V F I A R E T T C S K G
331 S N E E L T K S C E I N I H G Q I L H C D A N V Y V V P W E
361 E K V Y P T V N C Q P L G Q T S L M K R P P G F S P F R S V
391 Q V M K T E G S T T V S L P H S A M S P V Q D E E R D S G K
421 E Q G P T H G H G W D H G K Q I K L H G L G L G H K H K H D
451 Q G H G H H G S H G L G H G H Q K Q H G L G H G H K H G H G
481 H G K H K N K G K N N G K H Y D W R T P Y L A S S Y E D S T
511 T S S A Q T Q E K T E E T T L S S L A Q P G V A I T F P D F
541 Q D S D L I A T V M P N T L P P H T E S D D D W I P D I Q T
571 E P N S L A F K L I S D F P E T T S P K C P S R P W K P V N
601 G V N P T V E M K E S H D F D L V D A L L
```

KGBOH2
Kininogen, HMW II precursor - Bovine

Includes: bradykinin (kallidin I)

Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H., and Nakanishi, S., Nature 305, 545-549, 1983 (Sequence translated from the mRNA sequence)

Bradykinin is released from kininogen by kallikrein.

Superfamily: cystatin

Residues	Feature
1-22	Domain: signal sequence (probable)
19-135,136-256, 257-377	Duplication: homology with cystatin
378-386	Peptide: bradykinin
400	Binding site: carbohydrate (Ser)

Mol. wt. unmod. chain = 68,710 Number of residues = 619

```
          5         10        15        20        25        30
  1 M K L I T I L F L C S R L L P S L T Q E S S Q E I D C N D Q
 31 D V F K A V D A A L T K Y N S E N K S G N Q F V L Y R I T E
 61 V A R M D N P D T F Y S L K Y Q I K E G D C P F Q S N K T W
 91 Q D C D Y K D S A Q A A T G Q C T A T V A K R G N M K F S V
121 A I Q T C L I T P A E G P V V T A Q Y E C L G C V H P I S T
151 K S P D L E P V L R Y A I Q Y F N N N T S H S H L F D L K E
181 V K R A Q K Q V V S G W N Y E V N Y S I A Q T N C S K E E F
211 S F L T P D C K S L S S G D T G E C T D K A H V D V K L R I
241 S S F S Q K C D L Y P G E D F L P P M V C V G C P K P I P V
271 D S P D L E E A L N H S I A K L N A E H D G T F Y F K I D T
301 V K K A T V Q V V G G L K Y S I V F I A R E T T C S K G S N
331 E E L T K S C E I N I H G Q I L H C D A N V Y V V P W E E K
361 V Y P T V N C Q P L G Q T S L M K R P P G F S P F R S V Q V
391 M K T E G S T T V S L P H S A M S P V Q D E E R D S G K E Q
421 G P T H G H G W D H G K Q I K L H G L G L G H K H K H D Q G
451 H G H H R S H G L G H G H Q K Q H G L G H G H K H G H G H G
481 K H K N K G K N N G K H Y D W R T P Y L A S S Y E D S T T S
511 S A Q T Q E K T E E T T L S S L A Q P G V A I T F P D F Q D
541 S D L I A T V M P N T L P P H T E S D D D W I P D I Q T E P
571 N S L A F K L I S D F P E T T S P K C P S R P W K P V N G V
601 N P T V E M K E S H D F D L V D A L L
```

KGHUL1
Kininogen, LMW I precursor - Human

Alternate names: alpha-2-thiol proteinase inhibitor

Includes: bradykinin

Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H., and Sasaki, M., Biochemistry 23, 5691-5697, 1984 (Sequence translated from the mRNA sequence)

In the presence of kallikrein this protein is converted into two chains, heavy and light, held together by a disulfide bond and bradykinin is released.

Superfamily: cystatin

Residues	Feature
1-18	Domain: signal sequence
19	Modified site: pyrrolidone carboxylic acid
381-389	Peptide: bradykinin
48,169,205,294	Binding site: carbohydrate (Asn) (putative)
19-136,137-258, 259-380	Duplication: homology with cystatin

Mol. wt. unmod. chain = 47,883 Number of residues = 427

```
          5        10        15        20        25        30
  1 M K L I T I L F L C S R L L L S L T Q E S Q S E E I D C N D
 31 K D L F K A V D A A L K K Y N S Q N Q S N N Q F V L Y R I T
 61 E A T K T V G S D T F Y S F K Y E I K E G D C P V Q S G K T
 91 W Q D C E Y K D A A K A A T G E C T A T V G K R S S T K F S
121 V A T Q T C Q I T P A E G P V V T A Q Y D C L G C V H P I S
151 T Q S P D L E P I L R H G I Q Y F N N N T Q H S S L F M L N
181 E V K R A Q R Q V V A G L N F R I T Y S I V Q T N C S K E N
211 F L F L T P D C K S L W N G D T G E C T D N A Y I D I Q L R
241 I A S F S Q N C D I Y P G K D F V Q P P T K I C V G C P R D
271 I P T N S P E L E E T L T H T I T K L N A E N N A T F Y F K
301 I D N V K K A R V Q V V A G K K Y F I D F V A R E T T C S K
331 E S N E E L T E S C E T K K L G Q S L D C N A E V Y V V P W
361 E K K I Y P T V N C Q P L G M I S L M K R P P G F S P F R S
391 S R I G E I K E E T T S H L R S C E Y K G R P P K A G A E P
421 A S E R E V S
```

KGBOL2
Kininogen, LMW II precursor - Bovine

Nawa, H., Kitamura, N., Hirose, T., Asai, M., Inayama, S., and Nakanishi, S., Proc. Nat. Acad. Sci. USA 80, 90-94, 1983 (Sequence translated from the mRNA sequence)
Nine disulfide bonds are present.
Potential carbohydrate-binding sites are Asn-47, Asn-87, Asn-168, Asn-169, Asn-197, Asn-204, and Asn-280.

Superfamily: cystatin

Residues	Feature
1-22	Domain: signal sequence (probable)
19-135,136-256, 257-377	Duplication: homology with cystatin
378-386	Peptide: bradykinin

Mol. wt. unmod. chain = 48,148 Number of residues = 434

	5							10							15							20							25							30
1	M	K	L	I	T	I	L	F	L	C	S	R	L	L	P	S	L	T	Q	E	S	S	Q	E	I	D	C	N	D	Q
31	D	V	F	K	A	V	D	A	A	L	T	K	Y	N	S	E	N	K	S	G	N	Q	F	V	L	Y	R	I	T	E
61	V	A	R	M	D	N	P	D	T	F	Y	S	L	K	Y	Q	I	K	E	G	D	C	P	F	Q	S	N	K	T	W
91	Q	D	C	D	Y	K	D	S	A	Q	A	A	T	G	Q	C	T	A	T	V	A	K	R	G	N	M	K	F	S	V
121	A	I	Q	T	C	L	I	T	P	A	E	G	P	V	V	T	A	Q	Y	E	C	L	G	C	V	H	P	I	S	T
151	K	S	P	D	L	E	P	V	L	R	Y	A	I	Q	Y	F	N	N	N	T	S	H	S	H	L	F	D	L	K	E
181	V	K	R	A	Q	K	Q	V	V	S	G	W	N	Y	E	V	N	Y	S	I	A	Q	T	N	C	S	K	E	E	F
211	S	F	L	T	P	D	C	K	S	L	S	S	G	D	T	G	E	C	T	D	K	A	H	V	D	V	K	L	R	I
241	S	S	F	S	Q	K	C	D	L	Y	P	G	E	D	F	L	P	P	M	V	C	V	G	C	P	K	P	I	P	V
271	D	S	P	D	L	E	E	A	L	N	H	S	I	A	K	L	N	A	E	H	D	G	T	F	Y	F	K	I	D	T
301	V	K	K	A	T	V	Q	V	V	G	G	L	K	Y	S	I	V	F	I	A	R	E	T	T	C	S	K	G	S	N
331	E	E	L	T	K	S	C	E	I	N	I	H	G	Q	I	L	H	C	D	A	N	V	Y	V	V	P	W	E	E	K
361	V	Y	P	T	V	N	C	Q	P	L	G	Q	T	S	L	M	K	R	P	P	G	F	S	P	F	R	S	V	Q	V
391	M	K	T	E	G	S	T	T	T	H	V	K	S	C	E	Y	K	G	R	P	Q	E	A	G	A	E	P	A	P	Q
421	G	E	V	S	L	P	A	E	S	P	Q	L	A	R																				

LPRTA4
Apolipoprotein A-IV precursor - Rat

Boguski, M.S., Elshourbagy, N., Taylor, J.M., and Gordon, J.I., Proc. Nat. Acad. Sci. USA 81, 5021-5025, 1984 (Sequence translated from the mRNA sequence)

This apoprotein is a major component of HDL and chylomicrons, but unlike other apoproteins, approximately 50% of the plasma apoA-IV is not associated with the classical lipoproteins.

Nine of the thirteen 22-amino acid tandem repeats (each 22-mer is actually a tandem array of two, A and B, related 11-mers) occurring in this sequence are predicted to be highly alpha-helical, and many of these helices are amphipathic. They may therefore serve as lipid-binding domains with lecithin:cholesterol acyltransferase (LCAT) activating abilities.

Superfamily: apolipoprotein

Residues	Feature
1-20	Domain: signal sequence
33-54,60-81, 82-103,115-136, 137-158,159-180, 181-202,203-224, 225-246,247-268, 269-286,287-308, 309-330	Duplication: tandem repeats of 22-amino acid unit
104-114	Duplication: a tandem repeat of the B group of 11-mers

Mol. wt. unmod. chain = 44,465 Number of residues = 391

	5					10					15					20					25					30
1	M	F	L	K	A	V	V	L	T	V	A	L	V	A	I	T	G	T	Q	A	E	V	T	S	D	Q	V	A	N	V
31	M	W	D	Y	F	T	Q	L	S	N	N	A	K	E	A	V	E	Q	L	Q	K	T	D	V	T	Q	Q	L	N	T
61	L	F	Q	D	K	L	G	N	I	N	T	Y	A	D	D	L	Q	N	K	L	V	P	F	A	V	Q	L	S	G	H
91	L	T	K	E	T	E	R	V	R	E	E	I	Q	K	E	L	E	D	L	R	A	N	M	M	P	H	A	N	K	V
121	S	Q	M	F	G	D	N	V	Q	K	L	Q	E	H	L	R	P	Y	A	T	D	L	Q	A	Q	I	N	A	Q	T
151	Q	D	M	K	R	Q	L	T	P	Y	I	Q	R	M	Q	T	T	I	Q	D	N	V	E	N	L	Q	S	S	M	V
181	P	F	A	N	E	L	K	E	K	F	N	Q	N	M	E	G	L	K	G	Q	L	T	P	R	A	N	E	L	K	A
211	T	I	D	Q	N	L	E	D	L	R	S	R	L	A	P	L	A	E	G	V	Q	E	K	L	N	H	Q	M	E	G
241	L	A	F	Q	M	K	K	N	A	E	E	L	H	T	K	V	S	T	N	I	D	Q	L	Q	K	N	L	A	P	L
271	V	E	D	V	Q	S	K	L	K	G	N	T	E	G	L	Q	K	S	L	E	D	L	N	K	Q	L	D	Q	Q	V
301	E	V	F	R	R	A	V	E	P	L	G	D	K	F	N	M	A	L	V	Q	Q	M	E	K	F	R	Q	Q	L	G
331	S	D	S	G	D	V	E	S	H	L	S	F	L	E	K	N	L	R	E	K	V	S	S	F	M	S	T	L	Q	K
361	K	G	S	P	D	Q	P	L	A	L	P	L	P	E	Q	V	Q	E	Q	V	Q	E	Q	V	Q	P	K	P	L	E
391	S																													

YLHUP
Serum amyloid P-component - Human

Alternate names: 9.5S alpha-1-glycoprotein

Frangione, B., submitted to the Protein Sequence Database, June 1985

Anderson, J.K., and Mole, J.E., Ann. N.Y. Acad. Sci. 389, 216-234, 1982
This sequence differs considerably from that shown.

Superfamily: C-reactive protein

Residues	Feature
36-95	Disulfide bonds:

Mol. wt. unmod. chain = 23,268 Number of residues = 204

	5				10				15				20				25				30
1	H	T	D	L	S	G	K	V	F	V	F	P	R	E	S	V	T	D	H	V	N
31	Q	N	F	T	L	C	F	R	A	Y	S	D	L	S	R	A	Y	S	L	F	S
61	L	L	V	Y	K	E	R	V	G	E	Y	S	L	Y	I	G	R	H	K	V	T
91	P	V	H	I	C	V	S	W	E	S	S	S	G	I	A	E	F	W	I	N	G
121	Q	G	Y	F	V	E	A	Q	P	K	I	V	L	G	Q	E	Q	D	S	Y	G
151	V	G	E	I	G	D	L	Y	M	W	D	S	V	L	P	P	E	N	I	L	S
181	N	I	L	D	W	Q	A	L	N	Y	E	I	R	G	Y	V	I	I	K	P	L

1	H	T	D	L	S	G	K	V	F	V	F	P	R	E	S	V	T	D	H	V	N	L	I	T	P	L	E	K	P	L
31	Q	N	F	T	L	C	F	R	A	Y	S	D	L	S	R	A	Y	S	L	F	S	Y	N	T	Q	G	R	D	N	E
61	L	L	V	Y	K	E	R	V	G	E	Y	S	L	Y	I	G	R	H	K	V	T	P	K	V	I	E	K	F	P	A
91	P	V	H	I	C	V	S	W	E	S	S	S	G	I	A	E	F	W	I	N	G	T	P	L	V	K	K	G	L	R
121	Q	G	Y	F	V	E	A	Q	P	K	I	V	L	G	Q	E	Q	D	S	Y	G	G	K	F	D	R	S	Q	S	F
151	V	G	E	I	G	D	L	Y	M	W	D	S	V	L	P	P	E	N	I	L	S	A	Y	Q	G	T	P	L	P	A
181	N	I	L	D	W	Q	A	L	N	Y	E	I	R	G	Y	V	I	I	K	P	L	V	W	V							

UART
Alpha-2u-globulin precursor - Rat

Dolan, K.P., Unterman, R., McLaughlin, M., Nakhasi, H.L., Lynch, K.R., and Feigelson, P., J. Biol. Chem. 257, 13527-13534, 1982 (Sequence of residues 34-181 translated from the mRNA sequence)

Drickamer, K., Kwoh, T.J., and Kurtz, D.T., J. Biol. Chem. 256, 3634-3636, 1981 (Sequence of residues 1-65 translated from the mRNA sequence)

Superfamily: alpha-2u-globulin

Residues	Feature
1-19	Domain: signal sequence (probable)
54	Binding site: carbohydrate (Asn) (probable)

Mol. wt. unmod. chain = 20,737 Number of residues = 181

```
            5         10        15        20        25        30
  1 M K L L L L L C L G L T L V C G H A E E A S S T R G N L D
 31 V A K L N G D W F S I V V A S N K R E K I E E N G S M R V F
 61 M Q H I D V L E N S L G F K F R I K E N G E C R E L Y L V A
 91 Y K T P E D G E Y F V E Y D G G N T F T I L K T D Y D R Y V
121 M F H L I N F K N G E T F Q L M V L Y G R T K D L S S D I K
151 E K F A K L C E A H G I T R D N I I D L T K T D R C L Q A R
181 G
```

GPHUA2
Leucine-rich alpha-2-glycoprotein - Human

Takahashi, N., Takahashi, Y., and Putnam, F.W., Proc. Nat. Acad. Sci. USA 82, 1906-1910, 1985

This sequence contains eight 24-residue segments that strongly resemble the consensus sequence PPGLLQGLPQLRXLDLSGNXLESL and five segments that are much less similar to this pattern.

The function of this plasma protein is not known.

Superfamily: leucine-rich alpha-2-glycoprotein

Residues	Feature
2	Binding site: carbohydrate (Thr)
44,151,234,290	Binding site: carbohydrate (Asn)
271	Binding site: carbohydrate (Asn) (possible)
8-21,268-294	Disulfide bonds:

Mol. wt. unmod. chain = 34,346 Number of residues = 312

```
          5        10        15        20        25        30
  1 V T L S P K D C Q V F R S D H G S S I S C Q P P A E I P G Y
 31 L P A D T V H L A V E F F N L T H L P A N L L Q G A S K L Q
 61 E L H L S S N G L E S L S P E F L R P V P Q L R V L D L T R
 91 N A L T G L P P G L F Q A S A T L D T L V L K E N Q L E V L
121 E V S W L H G L K A L G H L D L S G N R L R K L P P G L L A
151 N F T L L R T L D L G E N Q L E T L P P D L L R G P L Q L E
181 R L H L E G N K L Q V L G K D L L L P Q P D L R Y L F L N G
211 N K L A R V A A G A F Q G L R Q L D M L D L S N N S L A S V
241 P E G L W A S L G Q P N W D M R D G F D I S G N P W I C D Q
271 N L S D L Y R W L Q A Q K D K M F S Q N D T R C A G P E A V
301 K G Q T L L A V A K S Q
```

WOHUB
Alpha-2-HS-glycoprotein B chain - Human

Gejyo, F., Chang, J.-L., Burgi, W., Schmid, K., Offner, G.D., Troxler, R.F., Van Halbeek, H., Dorland, L., Gerwig, G.J., and Vliegenthart, F.G., J. Biol. Chem. 258, 4966-4971, 1983

Superfamily: alpha-2-HS-glycoprotein

Residues	Feature
6	Binding site: carbohydrate (Ser)
18	Disulfide bonds: to A chain

Mol. wt. unmod. chain = 2,740 Number of residues = 27

```
          5         10        15        20        25
1 T V V Q P S V G A A A G P V V P P C P G R I R H F K V
```

FPHU
Alpha-fetoprotein precursor - Human

Morinaga, T., Sakai, M., Wegmann, T.G., and Tamaoki, T., Proc. Nat. Acad. Sci. USA 80, 4604-4608, 1983 (Sequence translated from the mRNA sequence)

Superfamily: serum albumin

Residues	Feature
1-18	Domain: signal sequence (probable)
25-217,218-409, 410-609	Duplication:

Mol. wt. unmod. chain = 68,677 Number of residues = 609

```
                5         10        15        20        25        30
  1 M K W V E S I F L I F L L N F T E S R T L H R N E Y G I A S
 31 I L D S Y Q C T A E I S L A D L A T I F F A Q F V Q E A T Y
 61 K E V S K M V K D A L T A I E K P T G D E Q S S G C L E N Q
 91 L P A F L E E L C H E K E I L E K Y G H S D C C S Q S E E G
121 R H N C F L A H K K P T P A S I P L F Q V P E P V T S C E A
151 Y E E D R E T F M N K F I Y E I A R R H P F L Y A P T I L L
181 W A A R Y D K I I P S C C K A E N A V E C F Q T K A A T V T
211 K E L R E S S L L N Q H A C A V M K N F G T R T F Q A I T V
241 T K L S Q K F T K V N F T E I Q K L V L D V A H V H E H C C
271 R G D V L D C L Q D G E K I M S Y I C S Q Q D T L S N K I T
301 E C C K L T T L E R G Q C I I H A E N D E K P E G L S P N L
331 N R F L G D R D F N Q F S S G E K N I F L A S F V H E Y S R
361 R H P Q L A V S V I L R V A K G Y Q E L L E K C F Q T E N P
391 L E C Q D K G E E E L Q K Y I Q E S Q A L A K R S C G L F Q
421 K L G E Y Y L Q N A F L V A Y T K K A P Q L T S S E L M A I
451 T R K M A A T A A T C C Q L S E D K L L A C G E G A A D I I
481 I G H L C I R H E M T P V N P G V G Q C C T S S Y A N R R P
511 C F S S L V V D E T Y V P P A F S D D K F I F H K D L C Q A
541 Q G V A L Q T M K Q E F L I N L V K Q K P Q I T E E Q L E A
571 V I A D F S G L L E K C C Q G Q E Q E V C F A E E G Q K L I
601 S K T R A A L G V
```

OQHU
Hemopexin - Human

Alternate names: beta-1B-glycoprotein

Takahashi, N., Takahashi, Y., and Putnam, F.W., Proc. Nat. Acad. Sci. USA 82, 73-77, 1985

Frantikova, V., Borvak, J., Kluh, I., and Moravek, L., FEBS Lett. 178, 213-216, 1984 (Sequence of residues 1-232)

Hemopexin is a serum glycoprotein that binds heme and transports it to the liver for breakdown, after which the free hemopexin returns to the circulation.

Superfamily: hemopexin

Residues	Feature
9-210,217-439	Duplication:
27-208,126-131, 165-177,234-437, 343-385,395-412	Disulfide bonds:
1	Binding site: carbohydrate (Thr)
41,164,217,223,430	Binding site: carbohydrate (Asn)

Mol. wt. unmod. chain = 49,295 Number of residues = 439

```
          5        10        15        20        25        30
  1 T P L P P T S A H G N V A E G E T K P D P D V T E R C S D G
 31 W S F D A T T L D D N G T M L F F K G E F V W K S H K W D R
 61 E L I S E R W K N F P S P V D A A F R Q G H N S V F L I K G
 91 D K V W V Y P P E K K E K G Y P K L L Q D E F P G I P S P L
121 D A A V E C H R G E C Q A E G V L F F Q G D R E W F W D L A
151 T G T M K E R S W P A V G N C S S A L R W L G R Y Y C F Q G
181 N Q F L R F D P V R G E V P P R Y P R D V R D Y F M P C P G
211 R G H G H R N G T G H G N S T H H G P E Y M R C S P H L V L
241 S A L T S D N H G A T Y A F S G T H Y W R L D T S R D G W H
271 S W P I A H Q W P Q G P S A V D A A F S W E E K L Y L V Q G
301 T Q V Y V F L T K G G Y T L V S G Y P K R L E K E V G T P H
331 G I I L D S V D A A F I C P G S S R L H I M A G R R L W W L
361 D L K S G A Q A T W T E L P W P H E K V D G A L C M E K S L
391 G P N S C S A N G P G L Y L I H G P N L Y C Y S D V E K L N
421 A A K A L P Q P Q N V T S L L G C T H
```

ZUHU
Erythropoietin precursor - Human

Jacobs, K., Shoemaker, C., Rudersdorf, R., Neill, S.D., Kaufman, R.J., Mufson, A., Seehra, J., Jones, S.S., Hewick, R., Fritsch, E.F., Kawakita, M., Shimizu, T., and Miyake, T., Nature 313, 806-810, 1985 (Sequence translated from the mRNA and DNA sequences)

The carboxyl-terminal four amino acids (T-G-D-R) may be removed in processing erythropoietin.

At least one disulfide bond is present.

Erythropoietin is produced by kidney or liver of adult mammals and by liver of fetal or neonatal mammals.

Superfamily: erythropoietin

Residues	Feature
1-27	Domain: signal sequence
51,110	Binding site: carbohydrate (Asn)
65	Binding site: carbohydrate (Asn) (possible)

Mol. wt. unmod. chain = 21,307 Number of residues = 193

```
          5        10        15        20        25        30
  1 M G V H E C P A W L W L L L S L L S L P L G L P V L G A P P
 31 R L I C D S R V L E R Y L L E A K E A E N I T T G C A E H C
 61 S L N E N I T V P D T K V N F Y A W K R M E V G Q Q A V E V
 91 W Q G L A L L S E A V L R G Q A L L V N S S Q P W E P L Q L
121 H V D K A V S G L R S L T T L L R A L G A Q K E A I S P P D
151 A A S A A P L R T I T A D T F R K L F R V Y S N F L R G K L
181 K L Y T G E A C R T G D R
```

Index

F

G

Y