INVESTIGATIONS INTO THE PRIMARY SEQUENCE OF IMMUNOGLOBULINS

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in the University of London

by

Nancy May Hogg

February, 1970

Wright-Fleming Institute of Microbiology
St. Mary's Hospital Medical School
London, W. 2.

ABSTRACT

In order to ascertain the extent of the heavy chain variable region, the sequences of the Fd fragments (residues 1-225) of two monoclonal proteins, named Daw and Cor, have been completed. They both belong to the IgG1 subclass and are both Gm(a+z+). They have a 77% identity in sequence from residue 1-99 but are completely different from residue 100-104. From 105-114 they have 50% identity and from 116-225 they are entirely identical.

A comparison with other heavy chains allowed a number of further conclusions to be drawn. Although Eu γ1 chain was identical with Daw and Cor from 115 onward, He γ1 chain had another residue in position 116. Therefore, the variable region must extend at least that far.

Comparison of the variable regions of Daw/Cor with Eu and a μ chain, Ou, allowed the division of these sequences into two subgroups - Daw/Cor/Ou representing one type and Eu, the other.

The grouping of two γ1 chains and a μ chain together indicated that the same sets of variable regions were shared across class lines, contrasting with the exclusiveness of the K and λ light chain variable regions for their own particular class of chain. This is evidence in

support of the "translocational hypothesis" which suggests that at least two genes are concerned in the synthesis of heavy and light chains.

Two regions of extensive variation were located in the heavy chain variable region. They are in analogous positions to the hypervariable areas of the light chains. It is postulated that these four areas could be drawn together by covalent linkages in each heavy and light chain in such a way as to form an area of high conformational individuality, which would be a good candidate for the antibody combining site.

ACKNOWLEDGEMENTS

I wish to thank Professor R. R. Porter for stimulating discussion, pertinent advice and patience.

Equally I thank Betty Press for expert guidance in the vagaries of protein sequencing and for her unflagging enthusiasm.

Finally I wish to indicate my debt to Mike Fried for his critical interest in this work.

I am grateful to the Medical Research Council of Canada for financial support during the course of these studies.

TABLE OF CONTENTS

Page No.

CHAPTER II Cont'd

Page No.

Page No.

LIST OF FIGURES

CHAPTER I

CHAPTER II

Page No.

LIST OF FIGURES Cont'd

CHAPTER III

Sequence determinations were carried out from the N-terminal end using the Dansyl-Edman technique $\longrightarrow$ and leucine aminopeptidase digestion $\longleftrightarrow$. They were carried out from the C-terminal end using carboxypeptidase A (CPA) digestion $\longleftarrow$ and hydrazinolysis $\longleftarrow\!\!+$.

Electrophoretic mobilities of peptides are given relative to lysine = +1.0, aspartic acid =-1.0 and glycine = 0. Mobilities in chromatography systems are measured relative to the solvent front (Rf value).

Page No.

LIST OF FIGURES Contd

CHAPTER IV

Page No.

LIST OF TABLES

LIST OF ABBREVIATIONS

The nomenclature for immunoglobulins published in the Bulletin of the World Health Organisation (1964) 30, 447 has been used throughout this thesis.

Asx	aspartic acid or asparagine
CHO	carbohydrate
Cmc	S-carboxyamidomethyl cysteine
CNBr	cyanogen bromide
DNS-Cl	5-dimethylamino-1-naphthalene sulphonyl chloride
$DNS-NH_2$	5-dimethylamino-1-naphthalene sulphonic acid
DNS-OH	5-dimethylamino-1-naphthalene sulphonamide
Glx	glutamic acid or glutamine
Hsr	homoserine
NH_3	0.88 sp. gr. ammonia
PCA	pyrrolid-2-one-5-carboxylic acid
PITC	phenylisothiocyanate
K	kappa type light chain
λ	lambda type light chain
α	alpha heavy chain (IgA)
γ	gamma heavy chain (IgG)
μ	mu heavy chain (IgM)
V region	region of variable sequence in heavy or light chain
C region	region of constant sequence in heavy or light chain

CHAPTER I.

GENERAL INTRODUCTION

I. GENERAL INTRODUCTION

Antibodies form part of a uniquely complex population of molecules, the immunoglobulins. The complexity is due to the fact that an individual is able to make a vast number of different antibodies against any given antigen. These antibodies may be directed against the same antigenic determinants, but they may vary both in the exactness with which they fit these determinants and in the class of immunoglobulin to which they belong.

In 1962, Porter proposed a general model for the structure of immunoglobulin IgG consisting of a pair of heavy chains held together by one or more disulphide bonds, each heavy chain being disulphide bonded to a single light chain (Figure I-1). Treatment with papain in the presence of cysteine splits the IgG molecule into two Fab fragments and one Fc fragment. Each Fab fragment contains an antibody combining site and is composed of a light chain and the amino (or N)-terminal half of a heavy chain. The Fc fragment is composed of the carboxyl (or C)-terminal halves of the two heavy chains held together by one or more disulphide bonds.

Neoplasms of immunoglobulin producing cells exist in a variety of

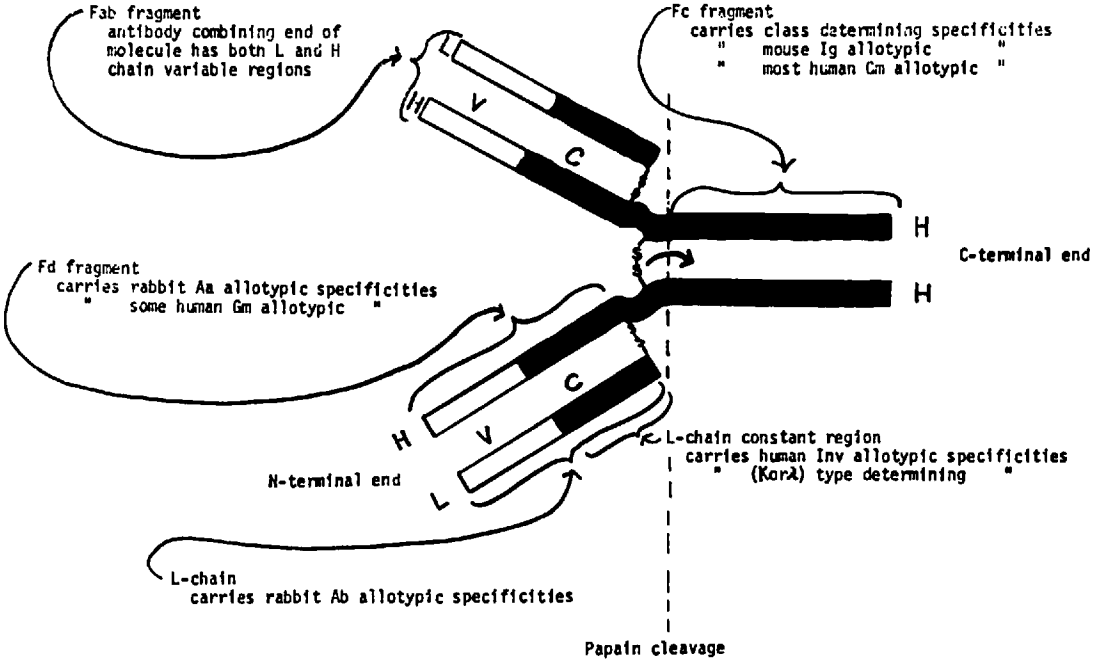


Figure I-1. Schematic representation of the structure of IgG.

animals and have been most widely studied in mouse and man (e.g. multiple myeloma). Each clone of these cancer cells produces a unique homogeneous immunoglobulin-like protein. Physical and chemical studies of these proteins have given most of the information on the structure of immunoglobulins.

The use of human myeloma proteins as representative of normal immunoglobulins has been considered justifiable on several grounds. They bear the same antigenic markers as normal immunoglobulins (Kunkel, 1963) and occur in proportions similar to their normal serum counterparts (review, Steinberg, 1969). N-terminal sequences of pooled light chains (Niall & Edman, 1967) and selected peptides from individuals and pools (Milstein et al, 1969) are a composite reflection of the sequences of single myeloma proteins. Combining activity has been reported against human immunoglobulin (Stone & Metzger, 1968; Kaplan & Metzger, 1969), streptolysin 0 (Zettervall et al, 1966; Waldenstrom and Winblad, 1964; Seligmann et al, 1968) and nitrophenyl derivatives (Eisen et al, 1967; Ashman & Metzger, 1967).

There is now considerable evidence which suggests that the specific affinity of an antibody for its antigen is due to the tertiary conformation

of the combining site on the antibody molecule. The experiments of Haber (1964), Witney and Tanford (1965), Freedman and Seka (1966) indicate that this in turn depends on the amino acid sequence of the constituent polypeptide chains. In these experiments extensively reduced Fab fragments or whole antibody was found to regain binding affinity upon reoxidation. Jaton et al (1968) have shown that under similar conditions and in the absence of light chains, polyalanylated heavy chain regains 59% of its activity toward DNP-lysine. Others have found that the activity of specific heavy chains is increased significantly upon noncovalent interaction with light chains, the association being greater if specific light chains are added (Grey & Mannik, 1965).

In affinity labelling experiments Singer's group (Singer & Thorpe, 1968; Good et al, 1968) have isolated tyrosine dipeptides from both heavy and light chains which they feel are homologous and situated in juxtaposition in a combining site. Fleet, Porter and Knowles (1969) have reported preliminary results using a phenylazide hapten, 4-azido-2-nitrophenyl (NAP), which inserts covalently into a carbon-hydrogen bond in the combining site after conversion to a reactive nitrene. They have obtained a heavy/light chain hapten labelling ratio of 3.5/1.

In apparent contrast to these results, Metzger and Potter (1968) have affinity labelled an IgA mouse myeloma (MOPC 315) in which all the label reacted with a tyrosine residue in the light chain.

Therefore, it would appear that both chains form part of the combining site. However, it is not yet clear whether the antibody combining site is composed of both heavy and light chains in each instance or whether it is present mainly in one or other of the chains.

Structural studies with Bence-Jones proteins assumed great significance following the discovery that they were identical to the light chains associated with myeloma proteins (Edelman & ~~Poulik~~ Gally, 1962). Peptide mapping of light chains from murine tumours (Potter et al, 1964; Bennett et al, 1965) and Bence-Jones proteins (Putnam & Easely, 1965) revealed that one part of the amino acid sequence was always identical while another part was specific for each protein. The explanation came in 1965 when Hilschmann and Craig compared partial sequences of two K Bence-Jones proteins, called Roy and Cu. They found the C-terminal half to be identical between the two proteins with the exception of a single residue while the N-terminal halves varied between the two. Titani et al (1965) were able to confirm this by comparison with a third protein, Ag, and its general significance was shown by the isolation of peptides

surrounding the cysteine residues of a number of Bence-Jones proteins (Milstein, 1966).

The striking heterogeneity of the N-terminal halves of light chains was candidate for contributing to the combining sites of antibody molecules. Peptide mapping studies indicated that part of the heavy chain was also variable (Frangione, Prelli & Franklin, 1967). Structural work on heavy chains was very much more difficult as they are twice as long as and not easily separable from light chains. Cyanogen bromide which cleaves polypeptide chains on the carboxyl side of methionine was used successfully by Givol and Porter (1965) and Press et al (1966) to cleave the heavy chain into a number of smaller fragments. For the first time the sequencing of heavy chain was approachable. It was the aim of this study to examine the variation occurring in heavy chains; that is, to discover how much sequence variation there was and how far it extended in the heavy chain. A comparison with light chains would then make possible some prediction as to what part of the sequence was involved in the antibody combining site. The sequence variation in the N-terminal halves of light chains has given rise to great speculation as to the genetic mechanisms which could produce such heterogeneity. It

was also our aim to ascertain whether the patterns of variability in heavy chains were similar to those in light chains.

II. STRUCTURAL STUDIES OF IMMUNOGLOBULINS

Immunoglobulins have been classified on the basis of antigenic specificities detected by serological means. The terms isotypic, allotypic and idiotypic were coined by Oudin (1960, 1966) to describe these specificities.

A. Structural Differences Defining Isotypic Specificities

Isotypic specificities are those which are found in all individuals of one species. They are often referred to as species specific determinants. The isotypic specificities of the constant regions of heavy chains define the immunoglobulin classes IgG, IgM, IgA, IgD, IgE and divisions of the main classes known as "subclasses" (Table I-1). Subclasses bear greater homology to one main class than to others and are thought to result from recent duplications of the gene specifying that particular class. The five classes of heavy chain are known as γ, μ, α, δ, and ϵ. (Only the first three classes will be discussed in the following sections as comparatively little is known about the

CLASS	IgG	IgA	IgM	IgD	IgE
Heavy chains:					
class	γ	α	μ	δ	ϵ
subclass	$\gamma 1$, $\gamma 2$, $\gamma 3$, $\gamma 4$	$\alpha 1$, $\alpha 2$			
mol. wt.	53,000	64,000	70,000		75,000
Light chains:					
mol. wt. 22,500	K, λ	Kλ	K, λ	K, λ	K, λ
mol. formula	$(K_2\gamma_2)$ or $(\lambda_2\gamma_2)$	$(K_2\alpha_2)_n$ or $(\lambda_2\alpha_2)_n$ n = 1, 2, 3,	$(K_2\mu_2)_5$ or $(\lambda_2\mu_2)_5$	$(K_2\delta_2)$ or $(\lambda_2\delta_2)$	$(K_2\epsilon_2)$ or $(\lambda_2\epsilon_2)$
$S_{20, w}$	6.5-7.0	7, 10, 13, 15	18-20	6.2-6.8	7.9
Mol. wt.	150,000	180,000-500,000	950,000	184,000	196,000
Carbohydrate%	2.9	7.5	11.8		10.7
Concentration in normal human serum, mg/100 ml	800-1,680	140-420	50-190	3-40	0.01-0.14

Table I-1. Classes of Human Immunoglobulins

δ and ϵ heavy chains.) Each combines with either of two types of light chain called K or λ. Isotypic specificities can be demonstrated only by immunization in another species.

1. IgG

Considerable progress has been made in the last several years towards the elucidation of the structure of IgG. Human IgG has been divided into four subclasses, IgG1-4, which differ antigenically in their heavy chain constant regions (Grey & Kunkel, 1964; Terry & Fahey, 1964). Some of their biological functions also differ. All subclasses but IgG-4 fix complement and all but IgG-2 fix to the skin (Terry, 1965; Ishizaka et al, 1967). They are found in varying amounts in serum, IgG-1 accounting for 70% of the total. Twenty-three allotypic antigenic markers (Muir & Steinberg, 1967; Natvig et al, 1969) called "Gm determinants" have been delineated on the heavy chains of human IgG. It was found that individual myeloma proteins were restricted according to Gm type and that these in turn were associated with a particular subclass (Table I-2). The IgG-4 subclass is presently without allotypic markers.

The comparative feature of the four subclasses studied in most detail is the nature of the disulphide bonding within and between the

Species	Polypeptide Chain	Light Chain	Location on chain: Fab	Location on chain: Fc
Man	Light chain	Inv(1)-Leu 191 Inv(3)-Val 191		
	K (Inv)			
	Heavy chain			
	γ (Gm)			
	γ 1		4(f), 17(z)	1(a) + (2, 5, 7, 9, 18, 20, 22)
	γ 2			23(n)
	γ 3			5(b) 13(b^3) 14(b^4) 21(g) + (6, 10, 11, 12, 15, 16)
	γ 4			
Rabbit	Light chain			
	K (b locus)	Ab 4, 5, 6, 9		
	Heavy chain			
	γ		Aa1, 2, 3	Aa8 & Aa10, Aa11, Aa12, Aa14, Aa15
	α		Aa1, 2, 3	Af1, 2, 3 (Conway et al, 1969)
	μ		Aa1, 2, 3	
Mouse	Heavy chain			
	γ			
	γ 1 (γ F)			Ig-4^b; Ig-$4^{acdefgh}$
	γ 2a (γ G)			Ig-1^a; Ig-1^b; Ig-1^c Ig-1^d; Ig-1^e; Ig-1^g; Ig-1
	γ 2b (γ H)			Ig-3^{ach}; Ig-3^b; Ig-3^g Ig-3^d; Ig-3^e Ig-3^f
	α			Ig-2^{ah}; Ig-2^b Ig-2^{cg}; Ig-2^{dc}; Ig-2^f

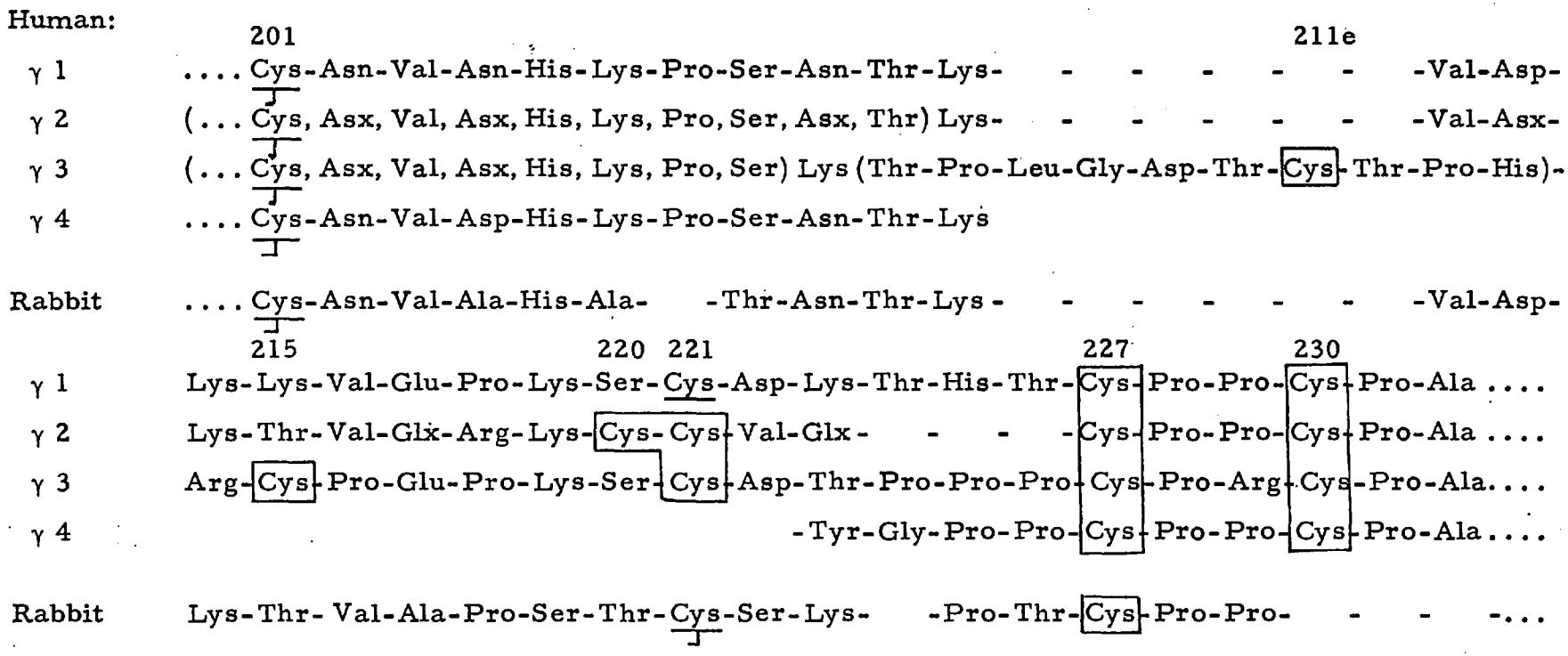
Table I-2. Major Allotypic Markers of Immunoglobulins.

heavy chains (Milstein et al, 1967; Steiner & Porter, 1967; Frangione et al, 1969). The intrachain bonds are striking in their periodic arrangement and appear to be a stable feature of the class. Three disulphide loops enclosing 58-60 residues appear symmetrically in each constant region (C region) of the γ chain. The amino acid sequences immediately adjacent to the cysteines forming the three disulphide bridges are almost completely constant for the four subclasses. The N-terminal quarter (V region) of the γ chain has a fourth disulphide bridge occurring in approximately the same position and containing 70-75 residues. This pattern has been taken as evidence of an evolutionary duplication of a smaller precursor (i.e. quarter length) to form the present heavy chain.

In contrast interchain covalent bonds and the surrounding amino acid sequences are very different for each subclass (Frangione et al, 1969; Frangione & Milstein, 1969) (Figure I-2). They are found within 21 residues surrounding the area of papain cleavage in the middle of the heavy chain - the "hinge region". The rabbit sequence can also be compared in this region (Cebra et al, 1968; Fruchter et al, 1969; O'Donnell et al, 1969). There are two γ 1 and γ4, four γ 2 and

five $\gamma 3$ interchain disulphide bonds (Figure I-2, 3). There is one interchain disulphide bond in the rabbit (Palmer & Nisonoff, 1964; O'Donnell et al, 1969). In the human $\gamma 1$ subclass the heavy-light chain bond is at residue 221 (Daw protein numbering), four residues away from the point of papain cleavage (225). The cysteines at this position in the $\gamma 2$ and $\gamma 3$ subclasses form interheavy chain bonds. The $\gamma 2$, $\gamma 3$, $\gamma 4$ and rabbit heavy-light chain bond is approximately 90 residues toward the N-terminus (position 133 - Daw numbering) (Figure I-3). This may not affect the conformation as radically as might be supposed because the disulphide bond 146-201 (134-220 in the rabbit) would bring these two positions fairly close together.

There is some difference in length in the hinge region. For maximum homology with $\gamma 1$ subclass, three positions must be deleted from $\gamma 2$ subclass (residues 223-226) and six residues inserted into $\gamma 3$ subclass (after residue 211). Frangione et al (1969) suggest that the number of covalent bonds in this region must have some effect on the flexibility of the Fab combining sites. Papain digestions cleaving in the hinge region, proceed at very different rates for each subclass, but do not seem to be correlated with the number of inter-chain disulphide



Key: Cys - heavy-light chain bond

Cys - inter-heavy chain bond

Cys - intra-heavy chain bond

Figure I-2. Amino Acid Sequences of Human and Rabbit γ Chains in the Region of the Inter-heavy chain Disulphide Bonds.

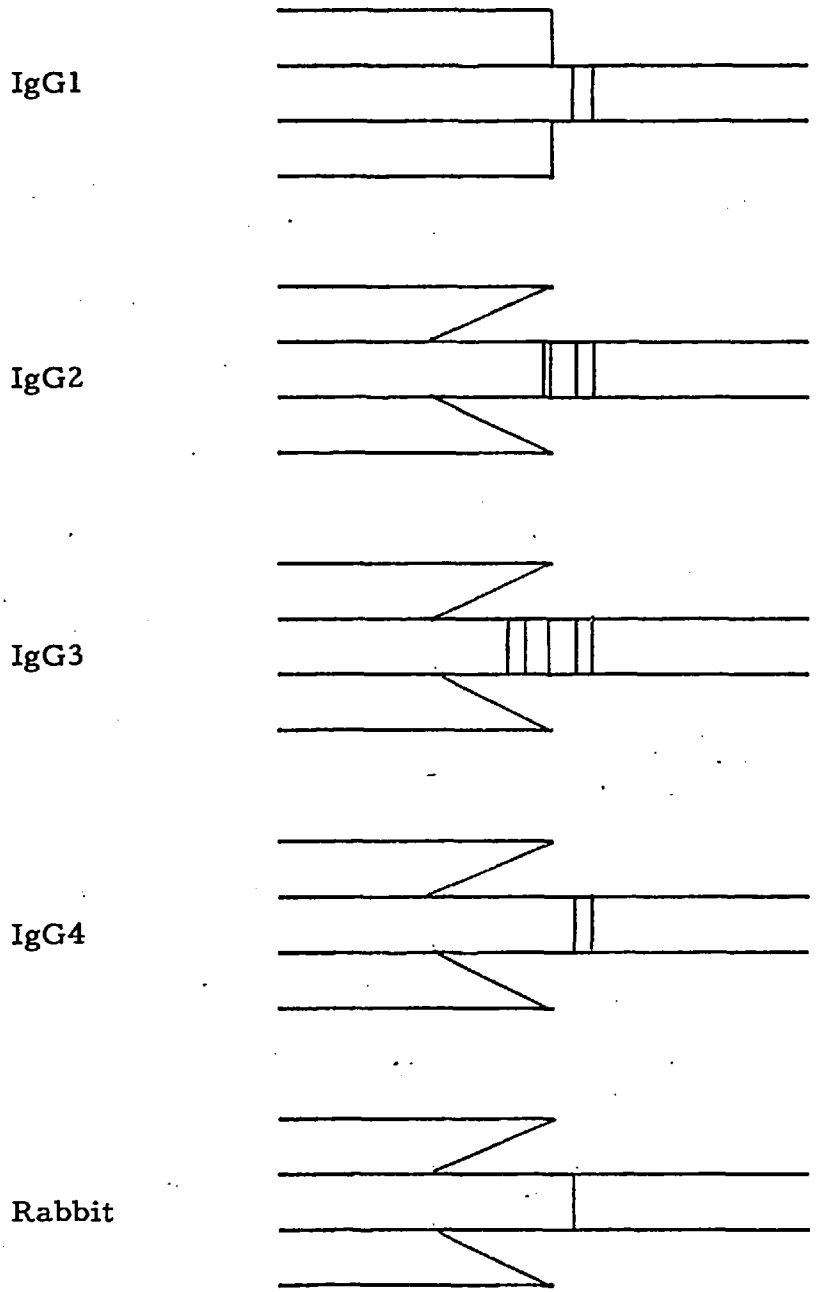


Figure I-3. Schematic representation of the arrangement of interchain disulphide bonds in human and rabbit IgG.

bonds (Jeffris et al, 1968). The γ 1 and γ 3 subclasses split easily in the absence of cysteine; whereas the digestion of the γ 4 subclass requires the presence of cysteine. The γ 2 subclass is fairly resistant to all forms of papain digestion.

Subdivision of the IgG class have been noted in many species distinguished mainly by electrophoretic and biological properties (see Milstein & Pink, 1969). For example the mouse has three IgG subclasses - IgG1, IgG2a and IgG2b. The first has a fast electrophoretic mobility and the latter two have slow mobility and bind complement. In addition, the IgG2a subclass was found to sensitise skin of heterologous species.

2. IgM

IgM (molecular weight of 900, 000) appears to be composed of five subunits (IgM_s) resembling IgG in structure and linked to each other by disulphide bonds (Miller & Metzger, 1965; Lamm & Small, 1966) (Table I-1). Another group has suggested that the basic subunit, IgM_s, consists of two heavy chains and three light chains (Suzuki & Deutsch, 1967). Each IgM heavy or μ chain appears to be involved in four interchain disulphide bonds. One of these bonds covalently links a μ chain and light chain (Miller & Metzger, 1965). The remaining

three bonds are intra- and inter-subunit. Using very mild reduction methods Morris and Inman (1968) were able to show that each IgM_s is bound to each of its neighbours by a single disulphide bond which has been localized in Fc_{μ} portion of IgM (Mihaesco & Seligmann, 1968; Onoue et al, 1968). There is a cysteine residue penultimate to the end of the μ chain (Doolittle et al, 1966; Abel & Grey, 1967; Wikler et al, 1969) but Beale & Feinstein (1969) suggest that it is involved in an intrasubunit bond and that the disulphide bond linking subunits is a carbohydrate containing peptide elsewhere in the Fc_{μ} fragment.

Electron micrographs corroborate this sort of structural picture. Chesebro et al (1968) describe intact IgM as being a "spider like structure" possessing five legs connected by a central ring. The central ring remains intact upon papain digestion releasing Fc_{μ} (Svehag et al, 1969). Reduction of a single Fc_{μ} intersubunit bridge seems to disrupt the arrangement completely suggesting a coiled or cyclic structure (Chesebro et al, 1968).

Reports concerning the valency of IgM have been contradictory, vacillating between five (Onoue et al, 1965; Metzger, 1967) and ten (Cooper, 1967; Merler et al, 1968) combining sites. Perhaps some

elucidation of the matter will come from work such as that of Onoue et al (1968) who reported an IgM against 1-azonapthalene-4-sulphonate which had ten combining sites, five with high and five with low affinity for the hapten. The IgM_s reacted in a similar manner but they could not establish whether the subunits contained one high and one low affinity site or were homogeneous in their affinities. They suggest that in many cases half the potential sites might not be detected because of this low affinity to antigen. Whether the inability to bind antigen was due to inadequate heavy-light bonding, features of the primary structure, or to steric factors inherent in conformation of the intact IgM molecule requires further study.

3. IgA

Serum IgA is for the main part a four chain molecule with a sedimentation coefficient of 6.9S (Table I-1). Approximately 10-15% is present as a 9-10S dimer (Heremans, 1959). Of the immunoglobulins in various secretions, IgA is present in highest concentration. Secretory IgA has a sedimentation coefficient of 11S and possesses antigenic determinants not present on serum IgA (Tomasi, 1965). They are found on a protein called secretory component or S component.

The structural model for secretory IgA consists of two molecules of serum IgA bound to one S component (Cebra & Small, 1967). There are conflicting reports about the nature of the binding. Cebra and Small (1967) found that the S component (consisting of two covalently bound chains with subunit molecular weight of 20, 500) could be dissociated away from rabbit colostral IgA by 5M guanidine hydrochloride. Hurlimann et al (1969) working with human salivary IgA found it necessary to reduce and alkylate before the S component (molecular weight 54, 000) could be separated away from IgA in any quantity.

Using fluorescein-conjugated antiserum, Tourville et al (1969) have shown that the S component is localized in epithelial cells lining glandular ducts while IgA is found in plasma cells underneath the epithelium. These workers postulate that after synthesis in the plasma cell, IgA passes along intercellular channels becoming bound to an S component. Rabbit 11S dimers are allotypically homogenous in vivo, although they can be hybridized by reduction and alkylation in vitro. This indicates that the dimer is probably made intracellularly (Lawton & Mage, 1969).

Three groups simultaneously demonstrated two subclasses of IgA

found in serum and in exocrine solutions, IgA1 and IgA2 (Vaerman & Heremans, 1966; Feinstein & Franklin, 1966; Kunkel & Prendergast, 1966). Abel and Grey (1968) found the IgA2 subclass to be peculiar in that it lacks heavy-light chain covalent bonding.

It was suggested that the $\alpha 2$ chains are shorter than $\alpha 1$ chains and lack the sequence that includes the cysteine involved in the heavy-light chain bond. After partially reducing molecules of both subclasses Grey et al (1968) found $\alpha 1$ heavy chain to contain seven moles and $\alpha 2$ only five moles of cysteine. As 75% of the light chains are present as dimers, they may lie in apposition between the heavy chains. The IgA2 subclass represents the main subclass of murine IgA and 6.6% of the total human IgA (Vaerman et al, 1968; Grey et al, 1968).

A C-terminal nonapeptide identical for $\alpha 1$ and $\alpha 2$ chains has been sequenced. The peptide was also isolated without the C-terminal tyrosine (Prahl & Grey, 1969). A comparison with the C-terminus of μ heavy chain (Wikler et al, 1969) shows five of nine residues to be identical.

$\alpha 1$ and $\alpha 2$ Met-Ala-Gln-Val-Asp-Gly-Thr-Cys-Tyr

μ Met;Ser-Asx-Thr-Ala(Gly, Thr)Cys-Tyr

The molecular weight of the α chain is 64, 000 and the μ chain values range from 60, 000-68, 000 (Suzuki et al, 1969) when carbohydrate is subtracted. γ chains have a molecular weight of 54, 000, a difference of 10-15, 000. Perhaps α and μ chains are extended by another "precursor segment" containing an intrachain disulphide bridge and an SH free to link subunits together. The penultimate cysteine would be in a good position to do this, but there is evidence in the case of the μ chain that it is involved in an inter-μ -chain bond within the IgM subunit (Beale & Feinstein, 1969) and Prahl and Gray (1969) feel that it forms half of an intrachain bond in the α chain.

4. Light chains

There are two types of light chain, called Kappa (K) and Lambda (λ). Light chains are conventionally grouped as "types" and heavy chains as "classes". Both types of light chain combine with all classes of heavy chain. They are of molecular weight 20-25, 000 and range from 208-216 amino acids in length. The dichotomy of the light chain into halves of constant (C region) and variable (V region) amino acid sequences has been mentioned in the General Introduction. The V region obeys some sequence restrictions but differs considerably from molecule to molecule

within a type. This heterogeneity will be discussed under "Idiotypy". The structure of light chains has remained fairly constant from the early vertebrates to man. They are similar in molecular weight, in electrophoretic heterogeneity, in amino acid sequence and in disulphide bond content. The light chains of man and mouse contain two intrachain disulphide loops, including about 60 amino acid residues each approximately in the center of the variable and constant regions, and one cysteine residue at the C-terminal end which forms a bond with the heavy chain (Milstein, 1964).

The two types can be readily distinguished by sequence both at the N-terminal and C-terminal ends (Hood et al, 1967). The N-terminus of K chains is aspartic or glutamic acid; the N-terminus of λ chains is blocked and most often assumed to be pyrrolidone carboxylic acid (PCA). The C-terminus of K chain is cysteine (exception being the pig - Cys-Glx-Ala (Franek, 1967)); the C-terminus of λ chains is cysteinyl-serine. This cysteine provides a link to the heavy chain. These similarities were recognised for an assortment of fifteen mammals and birds. Lamprey light chains have an aspartic N-terminus (Marchalonis & Edelman, 1968). Representative of the primitive fish classes, leopard

sharks and paddlefish both had K type N-terminal sequences (Suran & Papermaster, 1967; Pollara et al, 1968).

K and λ chains are present in varying amounts in each species (Hood et al, 1967). Man has a 3/2 ratio of K/λ chains. On the other hand the rabbit and mouse have approximately 90% K chains and the horse has only λ chains. There is no obvious reason for this distribution.

It is not known when the genome diverged to form both K and λ light chains. The 40% identity in sequence between them would place the common ancestral light chain 170 million years ago (Dayhoff et al, 1969). However, light chains with blocked N-terminals are suspected of being present in paddlefish immunoglobulins (Pollara et al, 1968). This would indicate that a λ type chain had evolved at least 400 million years ago (Romer, 1966).

B. Allotypy

Allotypy describes those antigenic specificities which are different in different groups of individuals within the same species. If two polypeptide chains are isotypic yet have distinct antigenic specificities, then their synthesis is thought to be controlled by allelic genes which are inherited in a simple Mendelian fashion. Allotypy can best be

demonstrated by intraspecies immunization, but can be shown by immunization of another species.

These alternative specificities have been most widely studied in man, mouse and rabbit and will be dealt with in that order (Table I-2). Any amino acid sequence correlations that are known will be noted.

1. Man

a. Heavy Chain Allotypy

The Gm factors are serological specificities associated with the γ chain C region and have been found on three IgG subclasses. Although each Gm factor is restricted to one subclass, it is found only on some chains of that subclass. In Caucasians γ 1 chains have either Gm(a) and (z) or Gm(a-) and (f) but not both sets; γ 2 chains are either Gm(n) or lack this factor; γ 3 are either Gm(b) including various Gm(b) factors or Gm(g) but not both (Kunkel et al, 1964, 1966; Terry et al, 1965; Natvig et al, 1967; Litwin & Kunkel, 1967). Recently antisera have been made to "non-g" and "non-z" markers (Natvig et al 1969). All Gm determinants are found on the Fc with the exception of Gm(z) and (f) which are located on the Fd. The γ heavy chain cistrons appear to be very closely linked so that various populations have γ gene

complexes. By studying population variation in the complexes and rare recombinations, Natvig et al (1969) were able to order the subclasses on the genome. Caucasians were found to have the following two combinations of γ heavy chain gene complexes:

	γ^4	γ^2	γ^3	γ^1
Gene Map		g^-a^-	ga^-	Z a
or				
		g^-a^-n	g^-a^-b	f a^-
Polypeptide Chain	Fd Fc	Fd Fc	Fd Fc	Fd Fc

A hybrid type of immunoglobulin has been found which appears to be a recombinant of γ 3 and γ1 subclasses. The serum contains normal quantities of both γ 2 and γ 4 subclasses. This is considered further evidence for the genome alignment of subclasses (Kunkel et al, 1969).

The sequence Asp-Glu-Leu-Thr-Lys has been correlated with Gm(a+) factor (Thorpe & Deutsch, 1966; Frangione et al, 1969) and the sequence Glu-Glu-Met-Thr-Lys has been found in Gm(a^-) IgG1 proteins in an analogous part of the chain and in IgG2 and IgG3 subclasses (Frangione et al, 1969; Wang & Fudenberg, 1969). The double amino acid difference in sequence is thought to be responsible for changed antigenicity. The γ 4 protein, Vin, contains the Glu-Glu-Met

sequence but is non-a unreactive (Frangione et al, 1969). This is perhaps because of an exchange of the Arg immediately preceding the tripeptide by a Gln in Vin protein. Another non-a unreactive but seemingly homologous peptide has been isolated from Old World Monkeys. It is Glu-Glu-Leu-Thr-Lys. Both Gm(a^+) and (a^-) peptides could be derived from it by two single step mutations (Wang. et al, 1969). The four sequences are:

Gm(a^+)	Arg-Asp-Glu-Leu-Thr-Lys
Gm(a^-)	Arg-Glu-Glu-Met-Thr-Lys
Vin(γ4)	Gln-Glu-Glu-Met-Thr-Lys
Old World Monkey	Glu-Glu-Leu-Thr-Lys

b. Light Chain Allotypy

The nine human K light chain constant regions which have been sequenced have been identical except for the alternative of Leu (Inv 1) or Val (Inv 3) at position 191 (Milstein, 1966; Baglioni et al, 1966). Inv 1 sera are normally Inv (1, 2) and very rarely Inv(1-2). There has been a shortage of antisera to test for the Inv (2) variant and consequently almost nothing is known about it but there may be three alleles at the Inv locus (Litwin & Kunkel, 1967). Lambda light chains

are either Oz (+) or (-). This correlates with either Lys or Arg at position 190. Ten individuals looked at by Ein (1968) had both λ type variants and there appeared to be a three fold excess of Oz(+) over Oz(-) in normal human serum. These findings favour more than one locus for the λ chain rather than a single locus with Oz factors indicating allelic alternatives. Moreover, of the seven λ chain constant regions completely sequenced other variants have been found. Ser-153 is replaced by Gly in Kern protein (Ponstingl et al, 1968) and Ala-144 is replaced by Val, Lys-172 is replaced by Asn in Mz protein (Milstein et al, 1967).

2. Mouse

Five immunoglobulin classes based on differences in heavy chain Fc have been defined in the mouse (Fahey et al, 1964). Four heavy chain gene loci, Ig-1, Ig-2, Ig-3, and Ig-4 correspond to classes IgG-2a, IgA, IgG-2b and IgG-1 respectively (Herzenberg et al, 1969). No locus has been identified for IgM and there are no known light chain polymorphisms.

Eight alleles ($Ig\text{-}1^{a\text{-}g}$) have been defined at the Ig-1 locus (Table I-2).

They consist of varying combinations of twelve distinct antigenic specificities which are all cross-reacting within the locus system. No specificity is unique to one allotype. It is postulated that multiple crossing over occurred in the evolution of these alleles. The specificies are known to be on the Fc fragment, but the chemical nature of the sites is unknown.

The Ig-2 locus controlling IgA synthesis has five alleles (Potter & Lieberman, 1967; Herzenberg, 1964) and they are specifically associated with Ig-1 locus alleles. They have been named in accordance with the Ig-1 alleles with which they segregate - Ig-2^{ah}, Ig-2^{b}, Ig-2^{cg}, Ig-2^{dc} and Ig-2^{f} (Table I-2).

A similar pattern is found at the Ig-3 locus (Potter & Lieberman, 1967; Warner & Herzenberg, 1966). Six alleles have been identified and have been found to associate with particular Ig-1 and Ig-2 alleles - Ig-3^{ach}; Ig-3^{b}; Ig-3^{g}; Ig-3^{d}; Ig-3^{e}; Ig-3^{f}. Only two alleles have been identified at the Ig-4 locus (Ig-4^{b} and Ig-$4^{acdefgh}$). They were distinguished by the difference in electrophoretic mobility of the Fc fragments of the IgG-4 molecules.

By analogy with the human γ chain system, lack of recombination

demonstrates that there is close linkage of the four mouse Ig loci (Potter & Lieberman, 1967; Herzenberg et al, 1969). Three specificities are held in common by IgG-2a and IgG-2b (Warner & Herzenberg, 1967). These similarities and peptide map comparisons indicate that the Ig-1 and Ig-3 must be the result of recent duplications and so be contiguous on the genome. Two rare recombinations found in wild mice permitted the ordering of the other two genes i. e. Ig 1-3-4-2 (Lieberman & Potter, 1969) or IgG-2a. IgG-2b. IgG-1. IgA

3. Rabbit

a. Heavy Chain Allotypy

There are three well-characterized loci involved in immunoglobulin synthesis. Aa1, Aa2, and Aa3 are alleles of the "a" locus (Oudin, 1960; Dray et al, 1962) and are located on heavy chain Fd fragments (Feinstein et al, 1963). Wilkinson (1969a) has found allotype related N-terminal sequence variation between Aa1 and Aa3 γchains. The "a" locus determinants are shared by the γ, α, and μ heavy chains (Todd & Inman, 1967; Pernis et al, 1968). These results could be most easily explained if the various classes of heavy chain were to share the N-terminal

sequences which give rise to the allelic specificities. Wilkinson (1969b) has found the same selection of N-terminal pronase peptides in IgG and IgA from rabbits homozygous for Aa1 and Aa3 allotypes.

A number of specificities has been located on the Fc of the γ chain. Aa8 and Aa10 are found on some, not all of γ chains bearing Aa1 specificities (Hamers, 1967). The Aa11 specificity segregates independently but is thought to be closely linked to the "a" locus (Zullo et al, 1968). It was therefore thought to be on the γ chain and was subsequently correlated with presence or absence of Met-226 (Prahl et al, 1969). Dubiski (1969) has reported a similar Fc marker, Aa14, found with equal frequency in "a" locus homo- or heterozygotes. The question of whether these Fc markers are controlled at a second locus has yet to be substantiated.

The "b" locus controls alleles Ab4, Ab5, Ab6 (Kelus & Gell, 1967) and Ab9 (Dubiski & Muller, 1967) and is not linked to the "a" locus (Todd, 1963; Feinstein, 1963). These specificities are found on K type chains. Each specificity involves two or three determinants (Mage, 1966) which may be partially accounted for by allotype related differences at the C-terminus of Ab4, Ab5 and Ab6 light chains (Appella et al, 1969; Frangione, 1969).

	Ab4	Asn-Arg-Gly-Asn-Cys (Asp)
Rabbit Light Chains	Ab5	Ser-Arg-Lys-Asn-Cys
	Ab6	Ser-Arg-Lys-Ser-Cys

Allotype Ac7 is found on b^- light chains and is controlled by a third locus, the "c" locus (Dray et al, 1963). These light chains constitute 5-20% of the rabbit light chain population and have a similar N- and C-terminal sequence to human λ chains (Appella et al, 1968).

C. Idiotypy

The term "idiotypy" was chosen to designate antigenic specificities of immunoglobulins which appear to be unique for each individual and for each set of antibodies produced by that individual. Antigenic specificities of this kind were observed in the rabbit by isoimmunization with antigen-antibody complexes (Oudin & Michel, 1963). At the same time Kunkel, Mannick and Williams (1963) found human myeloma proteins to have individual antigenic specificities. In a thorough study of rabbit antisera raised against idiotypes present on Salmonella typhi-anti S. typhi complexes, Oudin and Michel (1969) found that the number of idiotypes increased as a particular immunization progressed. Moreover, breeding experiments indicated that there was no pattern of inheritance involved so .....

that the range of idiotypic specificities is enormous. Significantly, antisera produced against the cross-reacting antigens S. typhi and S. typhimurium appeared to share some idiotypic determinants. This is in line with the finding that two BALB/C mouse myeloma proteins, S63 and S107, both showing combining specificity for the C polysaccharide of pneumococcus possessed idiotypic determinants in common (Cohn et al, 1969). Whatever the mechanism for generating variability in antibody structure, it is reasonable to think that single animals and closely inbred groups of animals would produce antibodies of a similar chemical nature toward the same or cross-reacting antigens.

Idiotypic specificities are found on the Fab fragment and IgG and IgM of a single rabbit antisera share idiotypic determinants (Kelus & Gell, 1968; Oudin & Michel, 1969). These two classes are known to share the same sets of heavy chain N-terminal variable sequences as well as sharing similar classes of light chains (Todd, 1963; Feinstein, 1963). Therefore, the structural differences responsible for idiotypic specificities must be located on the variable regions of the heavy and/or light polypeptide chains.

A considerable amount of sequence work has been done on the

N-terminal halves of light chains, especially on their pathological counterparts, the Bence-Jones proteins. No two have been found identical in sequence with the possible exception of two mouse proteins (RPC-20 & MOPC-104) which had identical fingerprints (Appella and Perham, 1967). Sequences have been published for 40 K and 26 λ human light chains (Figures I-4, 5). However, only 9 K and 7 λ light chains have been completely sequenced to residues 107(K) and 109 (λ), the concluding positions for light chain variable sequences. The others are N-terminal sequences of 22-27 residues and sequences surrounding the cysteins forming intra-chain loops.

It became apparent that the V regions could be placed into groups in which the heterogeneity between members was very much less than between groups. These groupings are called "subgroups" (Bull. Wld. Hlth. Org., 1969) of K and λ light chains (K subgroups - Smithies, 1967; Milstein, 1967; Niall & Edman, 1967: λ subgroups - Langer et al, 1968; Hood & Ein, 1968).

Three K subgroups have been defined, each with a basic sequence (Milstein, 1969a) (Figure I-4). Within a K subgroup the identity between any two members ranges from 70-85%; whereas, between a subgroup, identity decreases to 40-45%. The K-III subgroup is particularly homogeneous. Any two members have 84-87% identity.

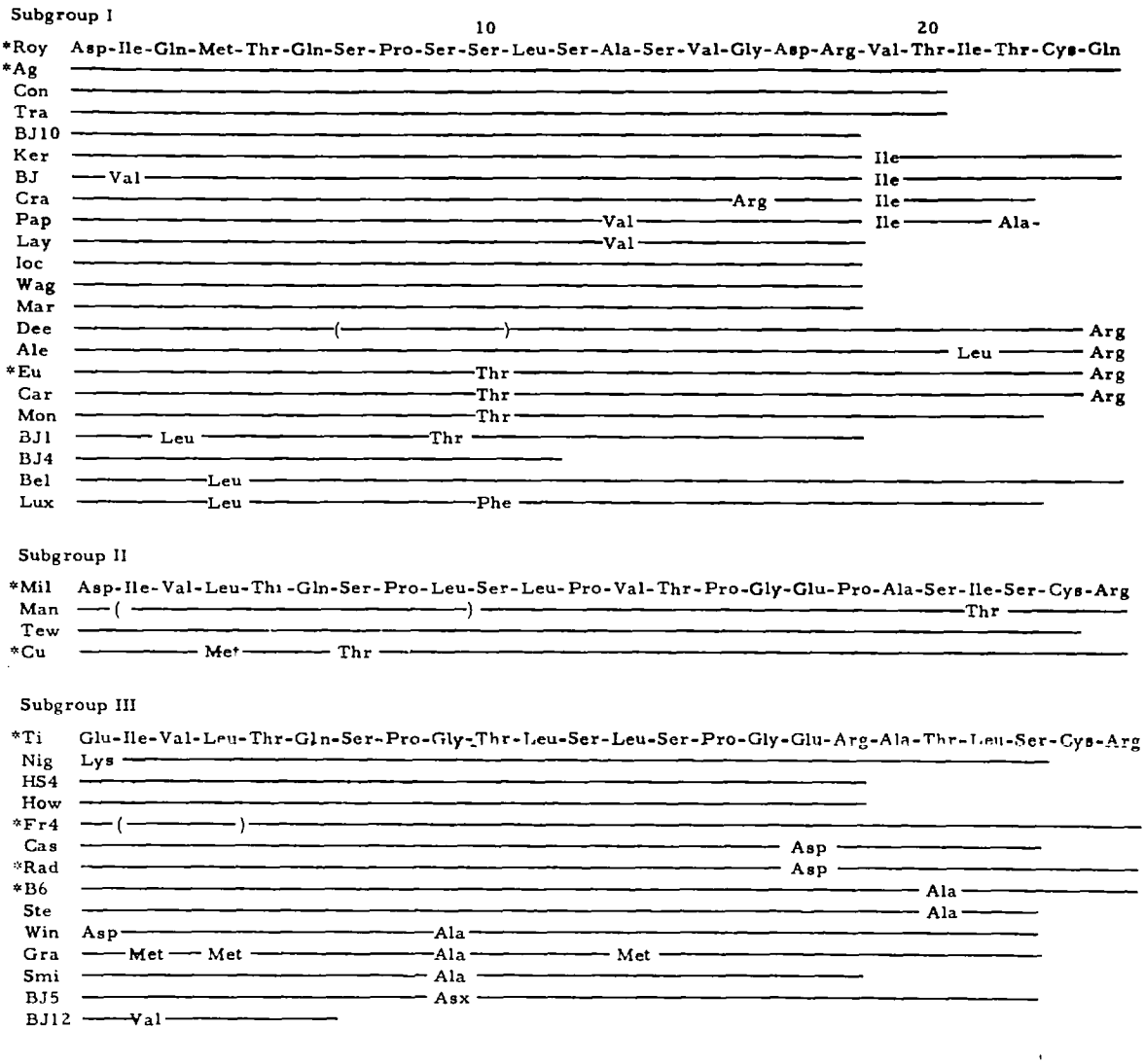
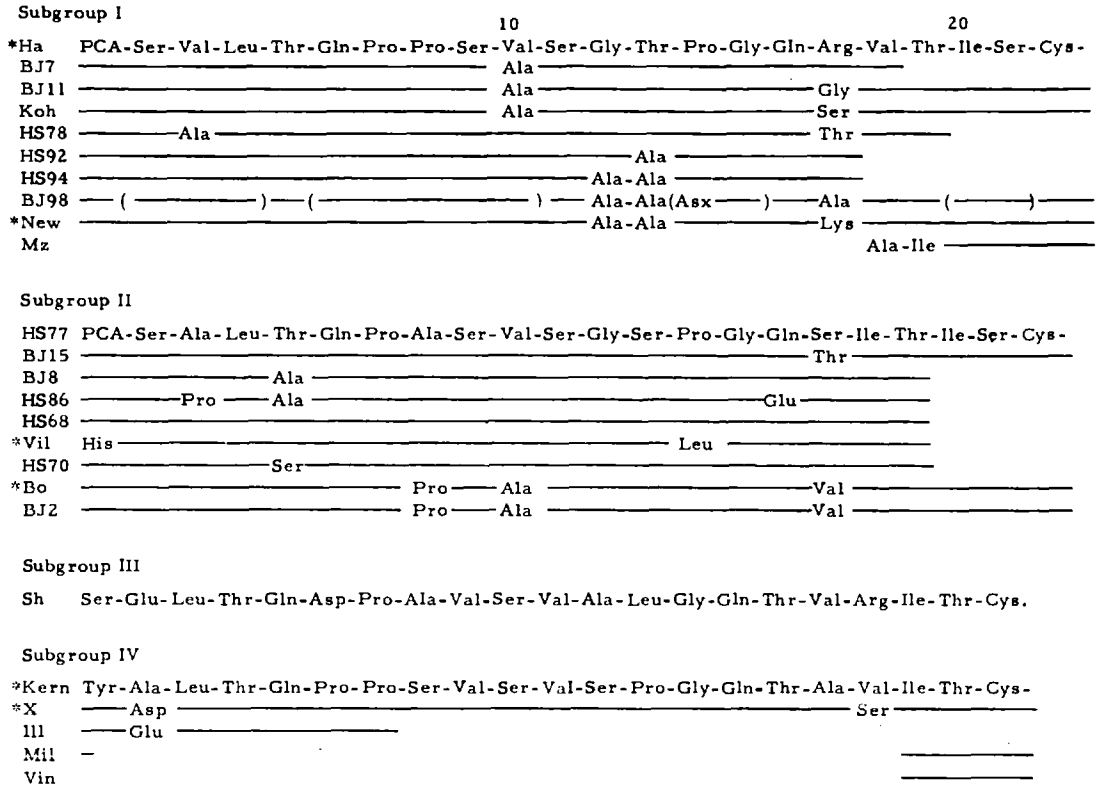


Figure I-4. N-Terminal Amino Acid Sequences of Human Kappa Light Chains.



* Light chains which have been completely sequenced.

Figure I-5. N-Terminal Amino Acid Sequences of Human Lambda Light Chains.

Kappa

Eu - Cunningham et al, 1968
Mil - Dreyer et al, 1967
Ste - Edman and Cooper, 1968
Roy, Cu - Hilschmann, 1967
BJ1, 3, 4, 5, 10, 12; HS4 - Hood et al, 1967
Lay, Ioc, Wag, Mar, How - Kaplan and Metzger, 1969
Day - Milstein, 1966a
Ker, BJ - Milstein, 1966b
Man - Milstein, 1967
Bel - Milstein, 1968
B6, Fr4, Rad - Milstein, 1969
Car, Dee, Ale - Milstein et al, 1969
Cra, Pap, Lux, Mon, Con, Tra, Nig, Win, Gra, Cas, Smi - Niall and Edman, 1967
Tew - Putnam, 1969
Ti - Suter et al, 1969
Ag - Titani et al, 1969

Lambda

BJ98 - Baglioni, 1967
HS68, 70, 77, 78, 86, 92, 94 - Hood and Ein, 1968
BJ2, 7, 8, 11, 15 - Hood et al, 1967
Koh - Kaplan and Metzger
New, III - Langer et al, 1968
X - Milstein et al, 1968
Mz, Mil - Milstein et al, 1967
Vin - Pink and Milstein, 1968
Kern - Ponstingl, 1967
Vil - Ponstingl and Hilschmann, 1969
Sh, Ha, Bo - Putnam et al, 1967
Daw - This thesis

The λ V region is at present divided into four subgroups (Figure I-5). With the exception of λ-III, two members of each subgroup have been fully sequenced. The percentage of identity between λ subgroup members is similar to that of K subgroups. It is 73-77%. Betweentsubgroups the homology drops to 55%. Sh protein, λ-III, bears a 60% homology to λ-IV and may be more closely related to it than to the other two subgroups.

Classification of the subgroups has been done solely on the basis of sequence. There is some indication that this could be done by specific antisera. Rabbit antisera detected two antigenic varieties of λ chains corresponding to subgroups I and IV (Tischendorf & Osserman, 1969).

The heterogeneity is scattered throughout the length of the variable region. Each individual protein has a number of variants of a seemingly [illegible] sequence. There are 49 conservative residues common to the K chain (42 for λ chains) and more common to each subgroup (Milstein & Pink, 1969). These latter positions do not necessarily coincide from one subgroup to another. For example, at position 17 in λ-I there are five variants in eight proteins; whereas, nine λ-II proteins have Ser-17 and two λ-IV proteins have Thr-17 (Figure I-5). Gaps and insertions must

be made in the basic subgroup sequences to maximize homologies between them. The K-II sequence is three residues longer than K-I and K-III after position 30 (Figure I-6). Single insertions and deletions also occur within subgroups. New protein (λ -I) has a deletion at position 29 (Figure I-6); Cu protein (K-II) has an addition of a Glu to the N-terminus (Figure I-4). This sort of evidence indicates that the subgroups are probably not due to polymorphism at a single V region locus but that there are structural genes for each subgroup. Milstein et al (1969) were able to find peptides characteristic of K subgroups I and III in the light chains of seventeen normal donors, thus demonstrating the non-allelic nature of the basic sequences.

In Figures I-4, 5, 6 and 7 three sorts of variation are noticeable. Firstly, there is the single mutation occurring in an otherwise conservative residue, e.g. K-I (BJ Val-2; Ale Leu-21) λ -II (Vil His-1) (Figures I-4, 5). In addition there are hypervariable positions known as "hot spots". These occur after the two cysteines (23 and 88) which form the N-terminal intrachain loop (Figures I-6, 7). Residues 28-31 are variable in K and λ subgroups. Moreover, the deletions and insertions typical of subgroups are localized between positions 30 and 31 adding

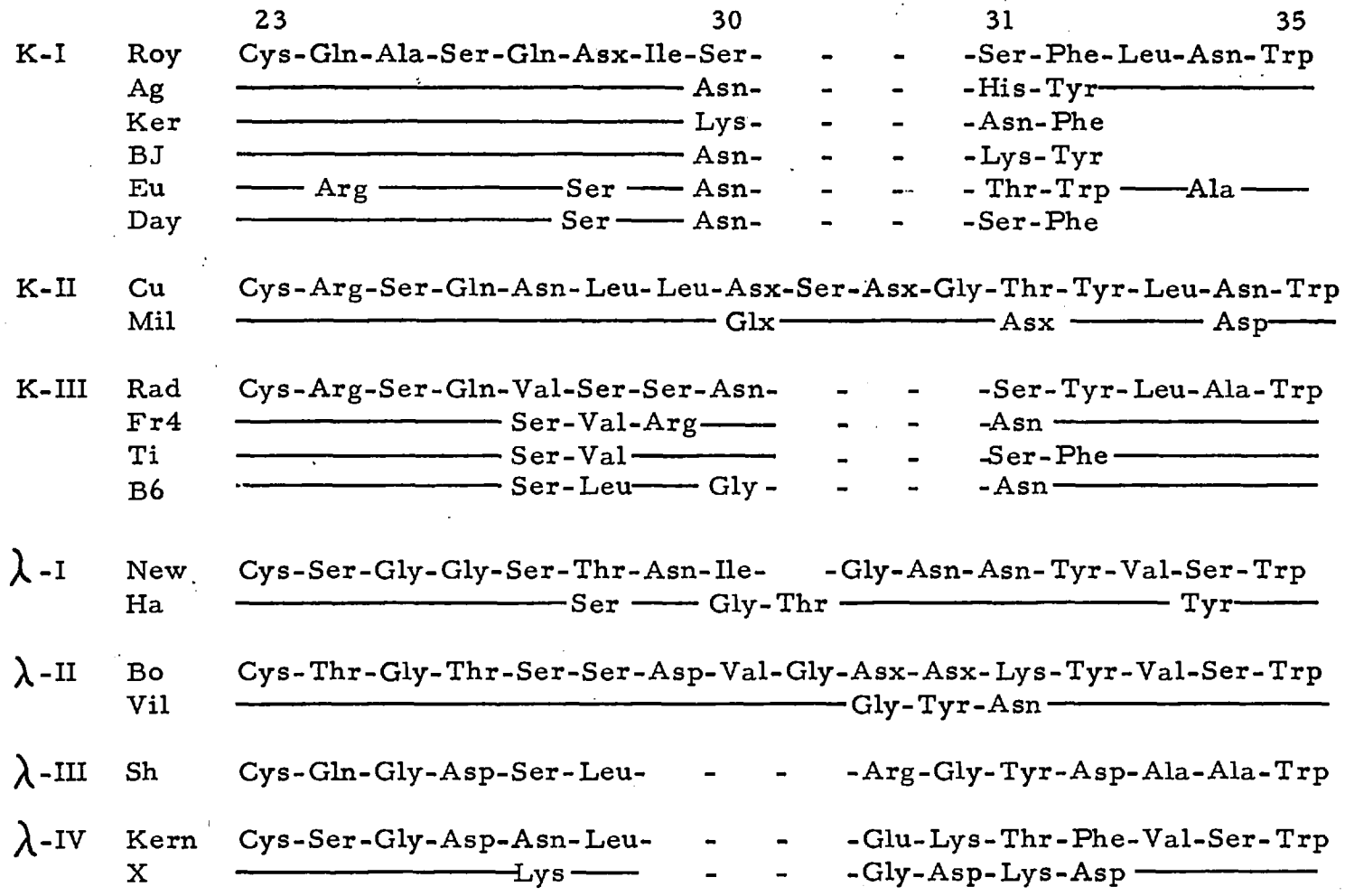


Figure I-6. Sequences following Cys-23 in Kappa and Lambda light chains.

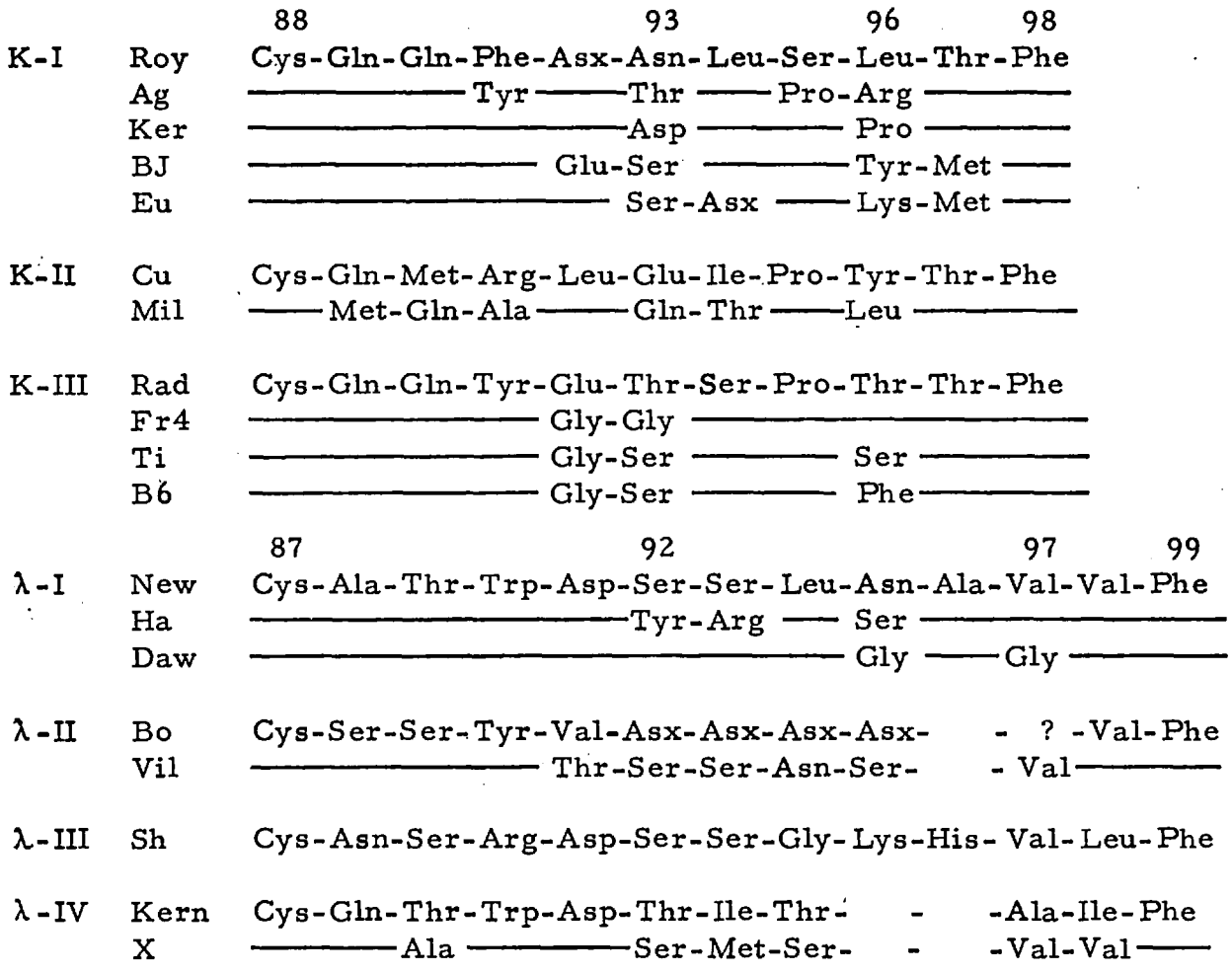


Figure I-7. Sequences Following Cys-88 in Kappa and Lambda light chains.

heterogeneity in length. The Cys at K-88 is followed by two hypervariable positions at 93 and 96 (Figure I-6). For example, in K-I, 13 proteins have 10 different residues at position 93 (Milstein & Pink, 1969). The variability following the Cys at λ -87 extends from residues 92-97 including differences in length (Figure I-7). Thus there is apparently random scattered heterogeneity and two locations where there is sequence variation in each light chain regardless of subgroup.

The third sort of variation occurring within a subgroup are the concerted changes found in a small number of subgroup members. For example in λ -I subgroup, three proteins have Ala and five have Gly at position 12; four have Ala and three Thr at position 13 (Figure I-5); in K-I subgroup - four Ile, twelve Val at position 19, four Arg and six Gln at position 24 (Figure I-4). There does not appear to be any evidence for linkage of the alternate forms. In Figure I-4, Thr-10, Arg-24 and Ser-28 overlap to some extent but would not be said to demonstrate linkage. Milstein et al (1969) found a K-I peptide containing Arg-24 in seventeen normal individuals thus ruling out allelism as the cause of these alternate amino acids. Most replacements can be accounted for by a change of a single nucleotide in the DNA codon. The evolutionary implications of this sequence variation will be examined in Chapter IV.

CHAPTER II

MATERIALS AND METHODS

A. Preparation of Fragment Daw 2a'

1. Materials:

The monoclonal immunoglobulin Daw was present at 7-8 g/100 ml in the idiopathic hypergammaglobulinemic serum, Daw 28. It was prepared as described by Press, Piggot and Porter (1966).

Papain was purchased from the Worthington Biochemical Corporation, Freehold, N.J.

Cyanogen bromide was purchased from Eastman Organic Chemicals, Rochester 3, N.Y.

Dithiothreitol (DTT or Cleland's Reagent) was purchased from Calbiochem, Los Angeles, and stored at 4^{o}.

Iodoacetamide was recrystallised from ethanol and washed with petroleum ether ($60-80^{o}$).

Iodo(1-^{14}C) acetamide was purchased from the Radiochemical Centre, Amersham, Bucks. and stored at -20^{o} as a 0.2 mC/ml solution in water.

2. Methods:

(a) Papain Digestion

A papain digest of Daw immunoglobulin was carried out using the

conditions of Press, Piggot and Porter (1966). The product was chromatographed at 4° on Sephadex G-100 in 0.15 M NaCl-0.025 M NaH_2PO_4 buffer, pH 7.0 in order to separate 10-15% of differentially digested material. The main peak of the elution profile which consisted of Fab and Fc was concentrated under pressure in Visking tubing (23/32 inch) while it was dialysed against water. It was subsequently freeze-dried in a round bottomed flask.

(b) Cyanogen Bromide Cleavage

The cyanogen bromide cleavage reaction was done as described by Press, Piggot and Porter (1966). For example, a 500 mg sample of Daw papain Fab/Fc mixture was dissolved in 12.5 ml of 70% (V/V) formic acid to which was added 1.25 gm CNBr. The mixture was incubated at 25° for 24 hrs. It was then diluted with 9 vol. of water and freeze dried to remove the CNBr, which was collected in an acetone-solid CO_2 cold trap and destroyed with caustic soda. The lyophilised powder was dissolved in a small volume (5-10 ml) of 6M urea/0.2 M sodium formate buffer, pH 3.3 and immediately fractionated on a Sephadex G-100 column (5 x 75 cm) of the same buffer.

The elution profile is seen in Figure II-1. The "Fab" fragment which remains intact because of intrachain disulphide linkages is

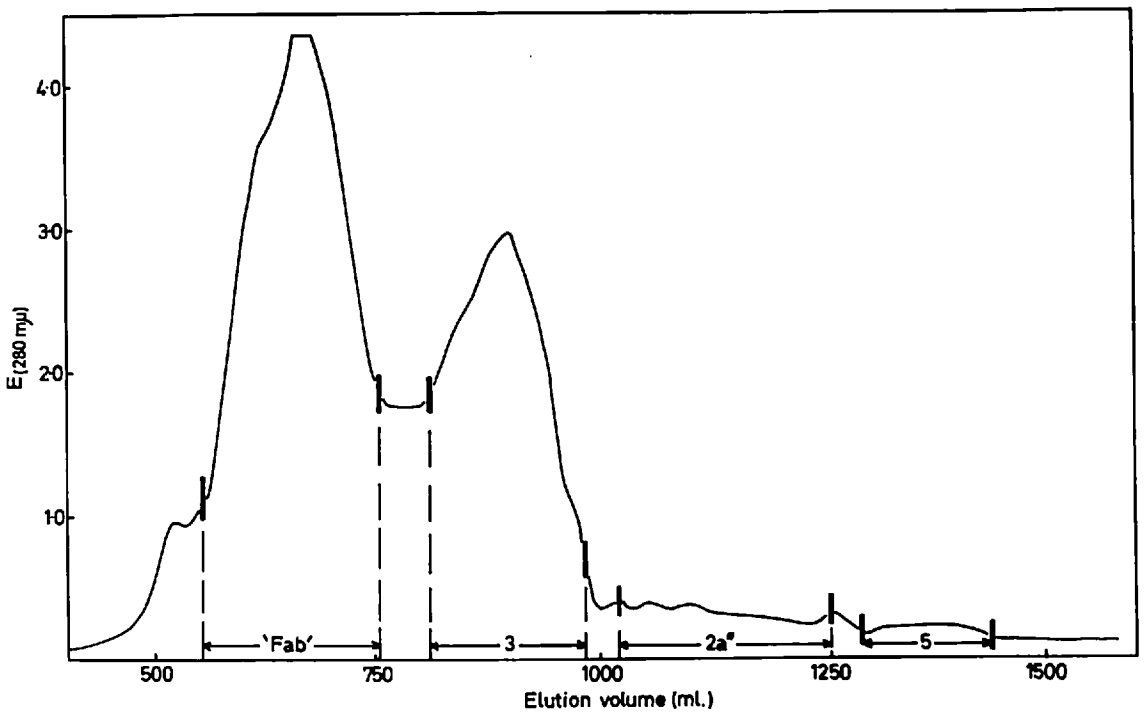


Figure II-1. Fractionation of cyanogen bromide treated Dav IgG on a column (5 :: 75 cm) of Sephadex G-100 in 6M urea 0.2 M Sodium formate, pH 3.3.

separated in this step from the Fc which has been cleaved into three pieces called, 2a'', 3, and 5. Fragments "Fab" and 3 were located by noting the absorption of column effluents at 280 mμ. They were then pooled, desalted on coarse Sephadex G-25 in 1N acetic acid and lyophilysed. Because fragments 2a'' and 5 contained no aromatic residues, they could not be located in the above manner. However, systematic pooling of fractions after fragment 3 located both fragments 2a'' and 5. They were desalted on coarse Sephadex G-25 in 0.05 N NH_3. The yield of the "Fab" fragment was usually 80-85%.

(c) Reduction and Alkylation of "Fab" fragment

The freeze-dried "Fab" fragment was dissolved in 6M guanidine 0.4 M tris-HCl buffer, containing 2 mM EDTA (40 mg fragment/ml buffer) and reduced with 0.05 M DTT (a 10 fold excess relative to "Fab") for 4 hrs at 37°, followed by alkylation with 1.2 fold molar excess of iodoacetamide at 0° for 1 hour. An aqueous solution of iodo(1-^{14}C) acetamide was added before incubation in amounts sufficient to label each half cystine with 10^5 cpm/ μmole (17 μl from a stock solution of 0.5 mC/2.5 ml). Glacial acetic acid was added to acidify the reaction mixture which was then put directly onto a Sephadex G-100 (5 x 75 cm) column in 6M urea/0.2 M sodium formate buffer, pH 3.3.

The fractionation pattern of reduced and alkylated "Fab" is seen in Figure II-2. The light chain fraction is followed immediately by fragment 2a' and a third peak contains fragments 4 and 2b which can be separated by gel filtration on Sephadex G-50 in 0.05 N NH_3. The three fractions were desalted on coarse Sephadex G-25 in 1N acetic acid. The yield of fragment 2a' was generally 80%. Therefore, the overall yield of Daw 2a' from Daw IgG was approximately 43%.

B. Enzymatic Digestions of Peptides

1. Materials:

Trypsin, chymotrypsin and carboxypeptidase A (CPA) were purchased from the Worthington Biochemical Corporation, Freehold, N.J. The CPA had been pretreated with di-isopropylfluorophosphate.

Chymotrypsin and CPA were used without further treatment. Trypsin was reacted with L-1-(p-toluenesulphonyl) amino-2-phenylethyl chloromethyl ketone (TPCK) in order to inactivate any chymotryptic contamination (Kostka & Carpenter, 1964).

Trypsin, chymotrypsin and CPA were dissolved in 0.1 M NH_4HCO_3 (2 mg/ml) and stored at -20°.

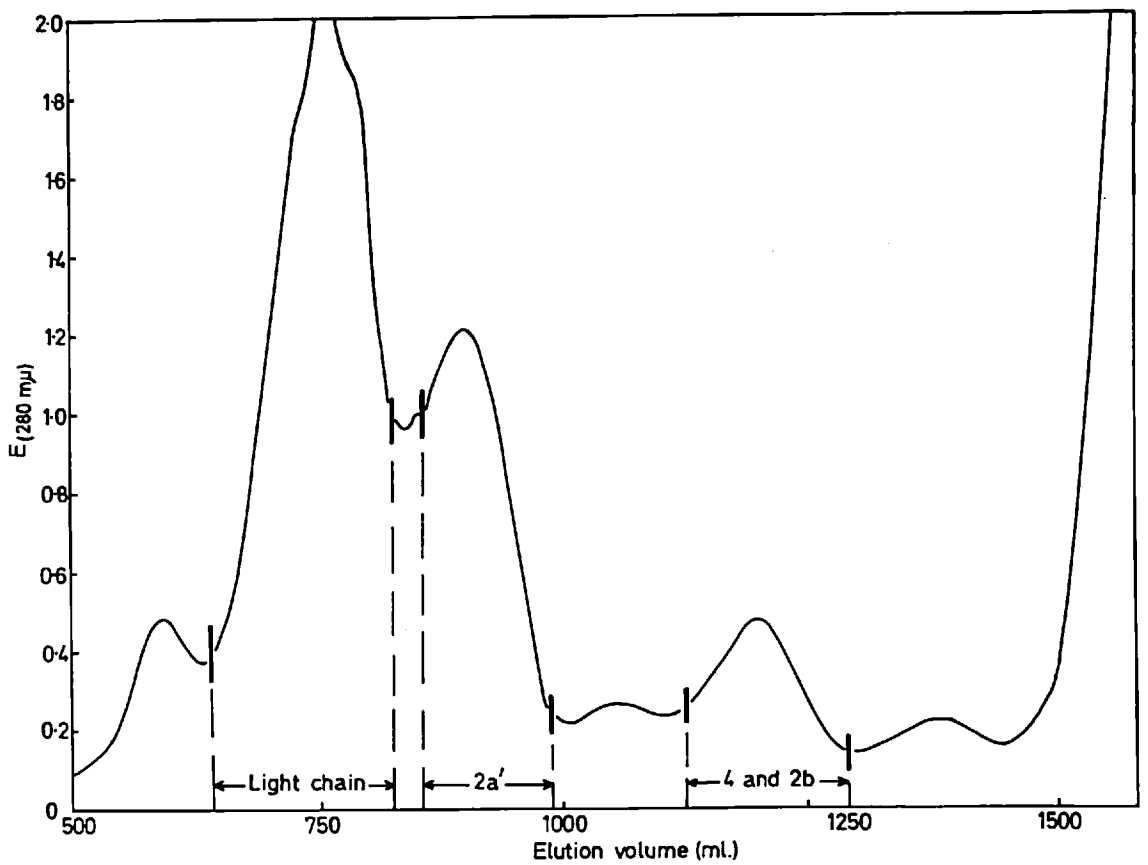


Figure II-2. Fractionation of reduced and alkylated cyanogen bromide fragment Daw "Fab" on a column (5 x 75 cm) of Sephadex G-100 in 6M urea 0.2M Sodium formate, pH 3.3.

Leucine aminopeptidase (LAP) was prepared from hog kidneys by Patrick Piggot (Press et al, 1966) and was dialysed before use against 0.01 M tris-HCl buffer, pH 8 containing 0.027 M $MgCl_2$ in order to remove traces of free amino acids.

2. Methods:

The conditions for tryptic and chymotryptic digestion were varied in order to obtain maximum hydrolysis for different peptides. The tryptic digestion of 2a' was carried out at 1/70: enzyme/substrate ratio. 20 mg peptide/ml of 0.05 M NH_4HCO_3, pH 8.0 were digested for 4 hours at 37°. Tryptic digestions of small peptides were done at peptide concentration of 0.125 μ mole/ml in 0.05 M NH_4HCO_3, pH 8.0 using 30 μ g of trypsin and incubating for 3-4 hours at 37°.

The conditions for chymotryptic digests varied considerably depending on the relative number of susceptible bonds and the solubility of the peptide. The chymotryptic digestion of the very insoluble T-1 peptide was carried out in 0.05 N NH_3 at pH 8 in a Radiometer pH stat. Conditions for digestion of individual peptides will be given in Chapter III.

Carboxypeptidase digestions were uniformly carried out. 0.01 μ mole of peptide was dissolved in 0.02 M NH_4HCO_3, pH 8.0 to which was added 20 μ l of CPA (1 mg/ml in 0.1 M NH_4HCO_3). The digest

was incubated at 37° for lengths of time varying from 15 minutes to 18 hours depending on the ease with which the C-terminal amino acids could be hydrolysed by CPA. After the digest was completed the sample was dried and without further preparation, put directly onto the amino acid analyser.

In the only leucine aminopeptidase digestion done, 0.1 μ mole of peptide in 1 ml of 0.05 M NH_4HCO_3, pH 8 was added to 0.1 ml of 0.02 M $MnCl_2$ and 0.1 mg of LAP. The reaction proceeded at 37° for 3 hours.

C. Column Chromatography of Enzymatic Digests

1. Materials:

Whatman Glass Fibre Paper - 2.5 cm GF/A circles

2, 5-diphenyloxazole (PPO)

1, 4-bis-2-(4-methyl-5-phenyloxazolyl)-benzene (dimethyl POPOP)

Scintillation Fluid:

5.0 gm PPO

0.3 gm dimethyl POPOP

1 litre toluene (analar grade)

2. Methods:

Enzymatic digests were generally fractionated on columns of

Sephadex G-50 in 0.05 N NH_3 or Sephadex G-25 in 0.02 N NH_3. The ammonia buffers were used as they are extremely volatile and do not absorb at the wave lengths used to assay column fractions - 280 and 215 mμ. The latter wave length was used for detecting peptide material that did not contain aromatic amino acids. In the case of "hot" digests, 10-20μl samples were pipetted onto glass fibre discs which were counted in a Unilux scintillation counter (Nuclear Chicago Corporation).

When the spectrophotometer readings at 280 and 215 mμ were plotted and the 'counts' superimposed, an elution profile was obtained. This profile was divided into a number of "fractions" and the contents of the appropriate test tubes were pooled.

D. Purification of Peptides

1. Materials:

i. Buffers for electrophoresis: (made up to 5 litres)

pH 1.9 100 ml 90% formic acid, 400 ml glacial acetic acid

pH 3.5 16.6 ml pyridine, 166.0 ml glacial acetic acid

pH 6.5 800.0 ml pyridine, 20.0 ml glacial acetic acid

pH 9.1 1% ammonium carbonate

Coolant - white spirit 115 (Esso Petroleum Co. Ltd.)

ii. Buffers for Chromatography:

PAW pyridine/amyl alcohol/water: 35/35/30 by volume

BAW butan-1-ol/acetic acid/water: 24/6/10 by volume

iii. Cadmium-ninhydrin solution (Heilmann et al, 1957):

a. 5 gm cadmium acetate

250 ml of glacial acetic acid

500 ml of water

b. add 15 ml 'a' to 85 ml of 1% ninhydrin in acetone

c. dip paper and dry in oven

d. ninhydrin positive spots will be shades of yellow, orange and pink on a white background.

2. Methods:

a. Separation of Peptides:

Aliquots of Sephadex column fractions were subjected to high-voltage electrophoresis at either pH 1.9, 3.5, 6.5 or 9.1 in order to see which pH or combination thereof would give the maximum separation of the fraction's component peptides. The high-voltage electrophoresis was carried out essentially by the method of Katz, Dreyer and Anfinsen (1959). Electrophoresis at pH 9.1 was done on a Locarte cold plate. Peptides not resolved by electrophoresis were separated by descending chromatography in either PAW or BAW.

Analytical runs and preparative separations of small amounts of peptide were done on Whatman No. 1 chromatography paper, loading 0.01-0.03 μ moles per cm. For large scale preparative runs, Whatman No. 3MM chromatography paper was used. Sample, approximately $2\frac{1}{2}$ times more concentrated, was applied to the paper but never for more than a 15 cm strip.

A guide-strip of standard amino acid mixture was run on either side of the paper in parallel to the peptide sample. Methyl green was often included in the marker mixture as it has a fast cathodic mobility and could be used as a visible indication of the progress of the run. After the run, a guide-strip was cut from either side of the paper which included the amino acid standard and a small strip of the separated peptide mixture. This was then stained and the positions of the various peptides located. Cutting lines were drawn across the unstained portion to indicate the positions of the peptides. These strips were cut out and eluted with either 0.05 N NH_3 for acidic and neutral peptides or 0.1 N acetic acid for basic peptides.

Analytical electrophoresis was routinely carried out for 45 minutes at 4,000 volts. Chromatography papers were left until the solvent reached the bottom of the sheet, which usually meant leaving them overnight. Times for the preparative runs were adjusted according to the known locations of peptides.

The charge of peptides containing aspartic or glutamic residues was determined at pH 6.5 according to the method of Offord (1966). The mobility of the peptide was measured against lysine for basic residues and aspartic acid for acidic peptides, using the position of glycine as the true origin.

b. Detection of Peptides on Paper:

After electrophoresis or chromatography an analytical paper was immediately dried in the chromatography oven (90°) and stained with cadmium-ninhydrin solution. Colour was allowed to develop for about three minutes in the oven before removal for a more gradual development at room temperature. Preparative runs were allowed to dry entirely at room temperature in order to prevent excessive adherence of sample to the paper.

It was found that according to the colour of the cadmium-ninhydrin staining, predictions could be made of the peptide's N-terminal amino acid. Glycine was yellow turning to orange; threonine - yellow only; serine - orange only; aspartic acid - rust; valine and isoleucine - faint pink becoming darker with time.

The sodium hypochlorite-starch iodide stain was employed mainly

for locating peptides with blocked N-terminals and was done according to the method of Pan and Deutscher (1956).

The Pauly stain for histidine and the α-nitroso-β-naphthol stain for tyrosine were used as outlined by Easely (1965). The tyrosine stain worked extremely well on top of the cadmium-ninhydrin stain.

The Sakaguchi stain for arginine and the Ehrlich stain for tryptophane were those of Smith (1950).

c. Amino Acid Analysis:

After elution from the preparative paper, an aliquot of peptide was taken for amino acid analysis in order that its composition, quantity and purity might be ascertained.

The aliquot was placed in an acid washed thick-walled rimless Pyrex tube (16 x 150 mm) and dried. 0.5 ml of constant boiling HCl (3 x glass-distilled) was added. If it was suspected that tyrosine was present in the peptide, 20 μl of 0.1 M phenol was added to the tube (Sanger & Thompson, 1963). The tube was evacuated to a pressure of 0.02 mm Hg. This pressure was maintained while the acid mixture which had been frozen in an ethanol-CO_2 bath was allowed to freeze and thaw twice in order that any dissolved gases could escape. The tube was sealed and placed in an oven at 110° for 22 hours after which time

the tube was broken open and the contents dried down on a Rotary Evapomix.

In samples suspected of possessing homoserine, the homoserine lactone which forms under acidic conditions was reconverted to homoserine by treatment of the dried hydrolysate with pyridine acetate buffer, pH 6.5 at 100° for 1 hour (Ambler, 1965).

Samples were dissolved in 0.5 ml of 0.2 N sodium citrate-HCl buffer, pH 2.2 for amino acid analysis on either of two Beckman-Spinco analysers operated essentially according to the method of Spackman, Moore and Stein (1958). The "fast"analyser separated acidic and neutral amino acids in 3 hours 20 minutes and the basic amino acids in 70 minutes. The times for "slow" analyser runs were respectively $6\frac{1}{2}$ hours and $3\frac{1}{2}$ hours. The "slow" analyser was used to separate Glu from Hsr. In all other runs 10% methanol was added to the starting buffer to aid in Thr/Ser separation (Thompson & Miles, 1964). Corrections of 9% for Ser and 4% for Thr were made to compensate for partial destruction during the hydrolysis. As Trp was completely destroyed, its presence was calculated by considering the 280 mμ absorption minus the amount of tyrosine present. The molar extinction

coefficient at 280 mμ of tyrosine was taken to be 1.3 and of tryptophane, 5.5 (Beavan & Holiday, 1952). Compositions of peptides were calculated on the assumption that the lowest average of the amounts of different amino acids present was integral; amino acids present in trace amounts were disregarded.

E. Sequence Analysis of Peptides

1. Materials:

1-dimethylaminonaphthalene sulphonyl chloride (DNS-Cl) was dissolved in analar acetone (2.5 mg/ml) and kept at -20°.

Phenylisothiocyanate (PITC) was redistilled 'in vacuo' using a water pump (2 mm Hg)-b.p. 102°. It was made up as a 5% solution in pyridine and stored at -20°.

Pyridine and n-butyl acetate were redistilled before use (b.p. 114-116°, 126-127° respectively).

Trifluoracetic acid (TFA) was stored at room temperature over sulphuric acid.

Silica Gel G was purchased from E. Merck A.-G., Darmstadt, Germany.

Polyamide sheets were purchased from Micro-Bio Laboratories Ltd., London, W.11.

Solvents for Separation of DNS-amino acids:

i. Solvent A (Morse & Horecker, 1966)

Benzene:pyridine:acetic acid - 80:20:2 (V/V)

ii. Solvent A (Seiler & Wiechmann, 1964)

Methyl acetate:isopropanol:conc. ammonia - 45:35:20 (V/V)

2. Methods:

a. N-terminal analysis of polypeptide chains and very large peptides (more than 60 residues in length) was done using the 1-fluoro-2, 4-dinitrobenzene technique of Sanger, as described by Porter (1957).

b. For smaller peptides the Dansylation procedure, in conjunction with the Edman degradation technique was used for determining the peptide sequence. Both techniques were done essentially as described by Gray (1967).

Dansyl Technique:

0.005-0.01 μmole peptide solution was placed in a small Pyrex tube (10 x 75 mm) and dried in a desiccator over phosphorus pentoxide, (P_2O_5).

The peptide was then suspended in 10 μl of 0.2 M $NaHCO_3$ and

10 µl of DNS-Cl solution. The tube was covered with parafilm and incubated at 37° for 1 hour. The contents were again dried 'in vacuo'. 50 µl of constant boiling HCl was added to the tube which was sealed and placed in a 110° oven for 16-18 hours. The tube was evacuated when Cmc was suspected and incubated for only 8 hours if Pro was suspected.

After the incubation period the tube was broken open, dried 'in vacuo' over NaOH pellets and extracted with 50 µl of wet ethyl acetate. The supernatant was removed to another tube and was dried in a nitrogen jet as was the residue. The contents of both tubes were dissolved in 10 µ l of 2 N NH_3 from which a 2 µl sample was taken for identification. It is necessary to look at both phases as DNS- ϵ-Lys, -Arg, -α His, -0-Tyr remain in the residue phase. Such amino acids as DNS-Asp, -Glu, -Ser, -Thr distribute themselves between both extract phase and residue.

Identification of DNS-amino acids:

Chromatography on 250µ layers of silica gel G was used to identify DNS-amino acids. A standard mixture of amino acids made by the method of Boulton and Bush (1964) was run on each "thin layer plate". Chromatography solvents A, B, C, D, of Morse and Horecker (1966) and Solvent A

of Seiler and Wiechmann (1964) were used to develop the plates. The most useful combination was found to be solvent A of Morse and Horecker (1966), which gave a good spread of all DNS-amino acids, but did not differentiate between Asp/Cmc; Ser/Thr/Glu or Ile/Leu, followed by Solvent A of Seiler and Wiechmann (1964) which separated Asp and Cmc; Ser, Thr and Glu (Figure II-3). The latter solvent also separated α-Lys, ϵ-Lys, Arg, $^{\alpha}$-His and O-Tyr. Fortunately the problem of distinguishing between Ile and Leu did not arise very frequently. It could usually be circumvented by an Edman degradation step followed by amino acid analysis which determined which of the two amino acids had been removed.

After chromatography for 30 minutes in the first Solvent A, the plate was viewed under a portable Ultraviolet lamp (Camag TL900) at 350 m μ. It was necessary to view the plate while still wet as fluorescence faded quite rapidly upon drying. The plate was then dried in the chromatography oven for 15 minutes and placed in the second solvent A for one hour. As this solvent was more alkaline the sensitivity of detecting the fluorescence was increased (Hartley & Massey, 1956).

A second dansylation procedure was used for samples of approx-

imately 0.001 μ mole or 1 nmole. It was essentially identical to the method just described except that all reagent volumes were scaled to one quarter. The samples were placed in soda glass tubes (5 x 50 mm) and incubated for ½ hour or until the sample was colourless. The ethyl acetate extraction was omitted and the dried hydrolysate dissolved in 2 μ l of 2 N NH_3. The Woods and Wang (1967) method of detection on polyamide sheets was used. Samples were placed on both sides of the sheet. On one face was applied 1 μl of amino acid "marker" mixture consisting of 10 μ mole/ml of DNS-Pro, -Ile, -Phe, -Gly, -Glu, -Ser, and -Arg. This was a gift from Dr. Ian O'Donnell. Then 1 μ l of sample was placed on both faces. The sheet was developed for 30 minutes in the first solvent (1.5% formic acid), it was removed and dried about ten minutes under a blow dryer. It was then run in a second solvent for one hour at 90° to the first direction of chromatography. The second solvent was benzene/acetic acid (9:1). To distinguish between DNS-Thr/-Ser and -Asp/-Glu a third solvent was employed which was ethyl acetate/methanol/acetic acid (20:1:1 by volume) (Crowshaw et al, 1967). The sheet was run in the same direction as for Solvent 2 (Figure II-3).

The advantages of the second technique over the first are several. Firstly the second technique is fivefold more sensitive, 1 nmole being

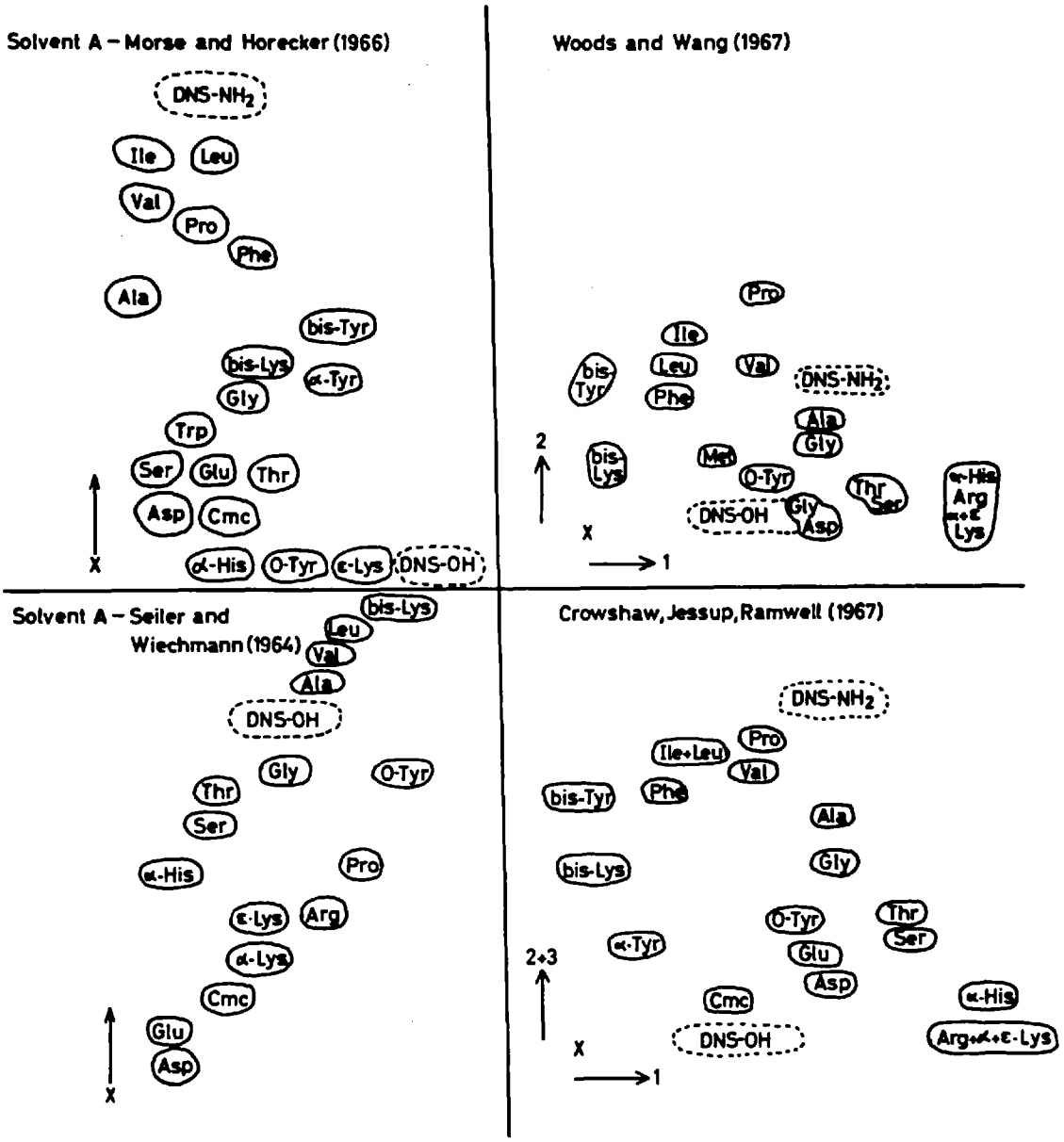


Figure II-3. Schematic representation of the separation of DNS-amino acids on layers of silica gel (Morse & Horecker, 1966; Seiler & Wiechmann, 1964) and polyamide (Woods & Wang, 1967; Crowshaw et al, 1967).

easily seen. Moreover the DNS-spots remain visible for several days after the chromotography has taken place; whereas the silica gel DNS-spots begin to fade almost immediately. The polyamide plates will last for at least fifteen runs if washed well with IN NH_3/ Acetone (50:50 by volume). Care must be taken against piercing the coating. As one obtains a wide spread or fingerprint of DNS-amino acids, one can better judge between the pertinent amino acids and the contaminants.

However in a silica gel separation, one has the advantage of being able to run more than one sample, usually six samples and two standards simultaneously. Moreover, it is perhaps slightly easier to separate the basic amino acids on silica gel using Seiler and Wiechmann's Solvent A than using the polyamide method which never separates α and ϵ-lysine or arginine from each other.

Edman Degradation Technique:

The Dansyl technique was used in conjunction with the stepwide Edman degradation technique (Gray, 1967). The peptide was taken to dryness in a stoppered Exelo tube. It was then dissolved in 0.1 ml water, 0.1 ml pyridine, and 0.2 ml of 5% PITC in pyridine. The tube

was flushed with nitrogen and incubated for 60 minutes at 45°. Excess reagents were removed in a vacuum desiccator over conc. H_2SO_4. The N-terminal amino acid was cleaved off as a phenylthiohydantoin by incubation at 45° for 30 minutes with the cyclising agent, 0.2 ml of TFA; the TFA was subsequently removed in vacuo over NaOH pellets. The sample was dissolved in 0.15 ml water and extracted three times with 1.5 ml redistilled n-butyl acetate in order to remove the phenylthiohydantoin. The aqueous residue was dried, 100 µl water added and an aliquot taken for identification of the N-terminal amino acid by the Dansyl method. The remainder of the peptide solution (made up to 100 µ l with water) was ready for the next Edman degradation step. This was only done after making certain that the Dansyl identification had been successful.

CHAPTER III

DETERMINATION OF THE SEQUENCES OF DAW 2a' AND COR 5b

I. THE AMINO ACID SEQUENCE OF DAW 2a'

A. Tryptic Peptides of Daw 2a'

Daw 2a' was prepared as described in the previous section. Its amino acid composition is given in Table III-1. Asx was found to be the N-terminal residue of Daw 2a' by the fluoro-dinitro benzene method. The sequence was determined by the following procedures.

1. Tryptic Digestion

Daw 2a' was digested with trypsin (360 μg/μmole of fragment in 0.6 ml of 0.05 M NH_4HCO_3, pH 8.1) for 4 hours at 37°. The trypsin was added in three equal aliquots at hourly intervals. A precipitate formed which accounted for one half of the total optical absorption at 280 mμ. The precipitate was collected by centrifugation and washed twice with 0.05 N NH_3. The supernatant and washings were fractionated on a column of Sephadex G-50 in 0.05 N NH_3. The elution profile is given in Figure III-1. Four fractions, named T1, T2, T3, and T4 were obtained and from these fractions pure peptides were isolated. The amino acid compositions of Daw 2a' and the component tryptic peptides are shown in Table III-1.

Table III-1: TRYPTIC PEPTIDES OF DAW 2a'

Composition (moles of amino acid residue/mole of peptide

Amino Acid	T1	T2.5	T2.p	T3.1	T3.2	T3.3	T3.4	T3.6	T.4	Sum of Peptides	Fragment 2a'
Lys	2	0	1.1	1.0	1.1	1.1	1.2	2.1	0	9	8.0
His	2	0	0	0	0	0	0	0	1.0	3	2.4
Arg	0	1.1	0	0	0	0	0	0	0	1	0.8
Asp	5	1.8	1.3	1.0	1.0	0	0	0	0	10	10.0
Thr	7	2.4	1.9	0	0	1.9	0	0	1.0	15	14.0
Ser	11	0	5.1	1.0	0	2.2	3.2	0	0	22	21.0
Glu	3	0	2.0	0	0	0	0	1.2	0	6	6.0
Pro	5	1.3	0	0	0	0	2.4	0.8	0	10	10.0
Gly	4	2.2	2.9	0	0	3.2	1.2	0	0	13	13.0
Ala	2	2.2	1.3	0	0	1.8	1.1	0	0	8	8.0
Val	8	1.2	2.0	0	0.9	1.0	1.0	0.9	0	15	14.0
Ile	1	0	0.9	0	0	0	0	0	0	2	1.9
Leu	5	0	1.0	0	0	1.8	1.0	0	0	9	9.0
Tyr	3	1.8	1.6	0	0	0	0	0	0	7	6.0
Phe	2	0	0.9	0	0	0	0.9	0	0	4	3.7
Cmc	1	1.	1.	1.	0	1	0	0	0	5	3.4
Trp	1	0	1	0	0	0	0	0	0	2	2.0
Total Residues	62	15	24	4	3	14	12	5	2	141	133.2

Cmc composition of fragment 2a' is calculated from the amino acid analysis.

Cmc residues in the tryptic peptides are calculated from the radioactivity of the peptides.

Tryptophane content in fragment 2a' is calculated from adsorption at 280 mμ.

Tryptophane content of the peptides is based on staining of the peptides on paper with the Ehrlich reagent.

Composition of 2a' is based in 8 alanine residues/mole fragment 2a'.

Composition of T1 is based on the sum of the residues of the constituent chymotryptic peptides - Table III-5.

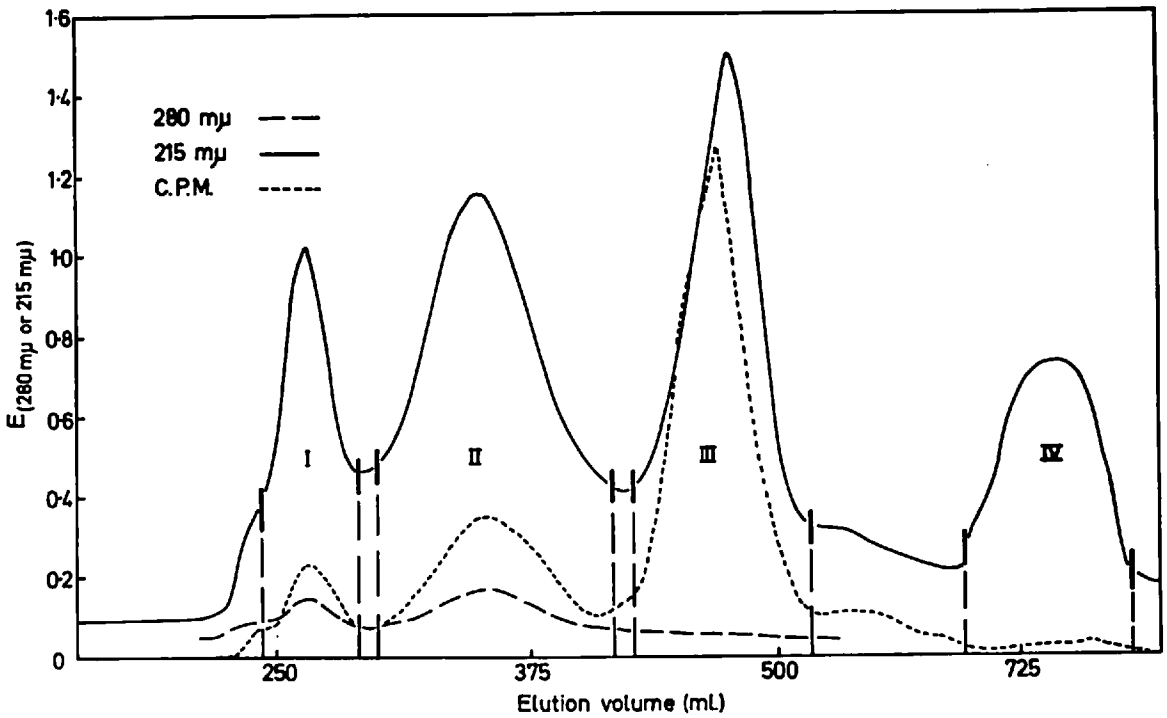


Figure III-1. Fractionation of the tryptic digest of Dav 2a' on a column (2.0 x 145 cm) of Sephadex G-50 in 0.05 N NH_3.

2. Amino Acid Sequences of Tryptic Peptides

a. Fraction T1

By amino acid composition, this fraction appeared to contain soluble portions of or the same peptide as the precipitate, named T1, which formed during tryptic digestion. Nothing more was done with this fraction as it was less than 10% of T1.

b. Fraction T2

Two peptides T2.5 and T2.p were isolated from this fraction.

T2.p had a very low solubility and was impractical to purify by paper electrophoresis as it could not be obtained in more than a 10% yield. It was, however, possible to precipitate it by concentration of the gel filtration fraction to 0.5 μ mole peptide/ml followed by addition of glacial acetic acid to a concentration of 1N and leaving at 2° overnight. The precipitate was centrifuged and washed with 0.1 N acetic acid at 0°. By this method the peptide was obtained in 60% yield. Its composition is given in Table III-1.

The N-terminal amino acid was found to be serine by the Dansyl-Edman technique. Since it was a peptide of 24 residues, it was digested with chymotrypsin (800 μ g/μ mole peptide in 0.4 ml of 0.05 M NH_4HCO_3,

pH 8.1) at 37° for 4 hours. Insoluble material was removed by centrifugation. Three major peptides, T2.p.C1, C3, C4 and one minor peptide, C5 were separated by paper electrophoresis at pH 3.5. The mobilities and amino acid compositions of these peptides are given in Table III-2 in which the sum of the three major peptides is comparable with that of T2.p. Figure III-2 shows the method used to sequence the three peptides. T2.p.C3 was difficult to sequence as the second residue, Cmc, became very easily blocked after the first Edman step (often 75% blockage). This arises as a result of condensation of the α-amino group with the carboxamide methyl side chain (Piggot, 1966). A 20 minute CPA digestion removed the C-terminal four amino acids, the serine in amounts slightly less than molar. A chymotryptic digest of T2.p.C1 was carried out from which peptides T2.p.C1a and C1b were isolated by electrophoresis at pH 6.5 and sequenced. By composition, T2.p.C5 appeared to be the C-terminal half of T2.p.C4 and was useful for completing the sequence of that peptide.

A peptide C6.4 isolated from the chymotryptic digest of Daw 2a' established the overlap between T2.p.C3 and T2.p.C1 placing T2.p.C3 in a position N-terminal to T2.p.C1. As T2.p.C3 has serine N-terminal,

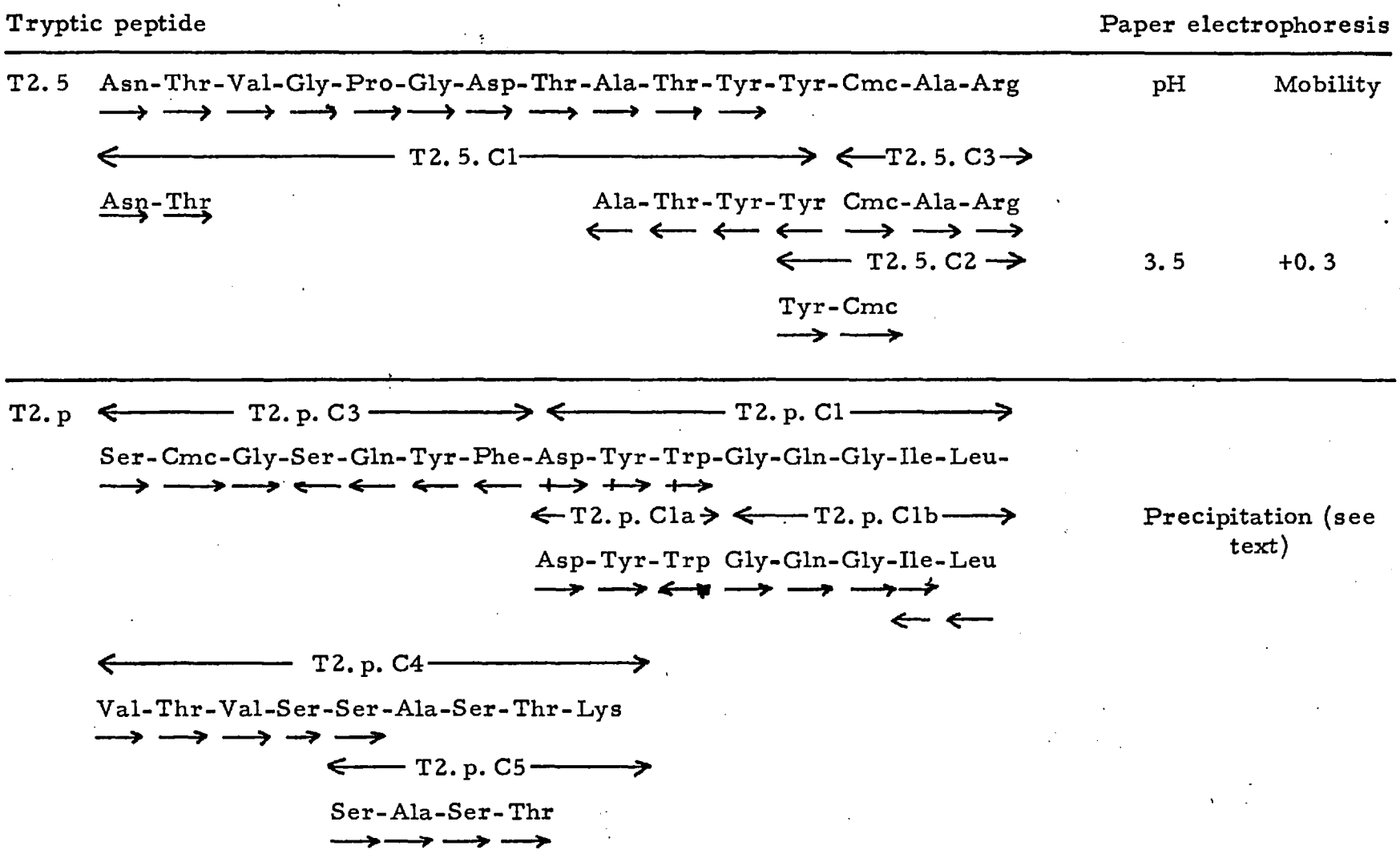


Figure III-2. Amino Acid sequences of and Isolation Procedures for Peptides from Tryptic Fraction Two.

Table III-2: METHOD OF ISOLATION AND COMPOSITION OF CHYMOTRYPTIC PEPTIDES OF DAW 2a' TRYPTIC PEPTIDE T2.p

Amino Acid	T2.p	T2.p.C1	T2.p.C3	T2.p.C4	T2.p.C5	T2.p.C1a	T2.p.C1b
Lys	1.1	0	0	1.2	1.1	0	0
Asp	1.3	1.4	0	0	0	1.0	0
Thr	1.9	0	0	1.7	0.7	0	0
Ser	5.1	0	2.4	3.0	2.2	0	0
Glu	2.0	1.0	0.9	0	0	0	1.1
Gly	2.9	2.2	1.2	0	0	0	2.2
Ala	1.3	0	0	1.2	1.0	0	0
Val	2.0	0	0	1.9	0	0	0
Ile	0.9	0.7	0	0	0	0	0.7
Leu	1.0	0.8	0	0	0	0	1.0
Tyr	1.6	0.9	0.7	0	0	1.0	0
Phe	0.9	0	0.8	0	0	0	0
CMC	1.	0	1	0	0	0	0
Trp	1	1	0	0	0	1	0
Total	24	8	7	9	5	3	5
Electrophoretic mobility at pH 3.5		-0.5	0	+0.14	+0.35	-0.3	0

Cmc content of the peptides is based on radioactivity and Trp content of the peptides was assumed to be 1 for those peptides which stained with Ehrlich's reagent on paper.

Mobility is given relative to lysine = +1.0, aspartic acid = -1.0 and glycine = 0.

Table III-3. Method of Isolation and Composition of Chymotryptic Peptides of Daw 2a' Tryptic Peptide T2.5.

Amino acid	T2.5	T2.5.C1	T2.5.C2	T2.5.C3
Arg	1.1	0	1.0	1.0
Asp	1.8	1.9	0	0
Thr	2.4	2.6	0	0
Pro	1.3	1.2	0	0
Gly	2.2	2.3	0	0
Ala	2.2	1.1	1.1	1.0
Val	1.2	1.1	0	0
Tyr	1.8	1.8	0.9	0
Cmc	1.	0	1.	1.
Total	15	12	4	3

Electrophoretic mobility at pH 3.5:

+0.3 -0.33 +0.55 +0.65

Cmc content of the peptides is based on radioactivity. Mobility is given relative to lysine = +1.0, aspartic acid = -1.0 and glycine = 0.

it must be the first peptide of T2. p. T2. p. C4 contained a C-terminal lysine and was, therefore, the C-terminal peptide of T2. p. Thus, the complete sequence was established as shown in Figure III-2.

As T2. p. C3 is neutral at pH 6. 5, the Glx residue is Gln. T2. p. C1, containing both a Glx and an Asx residue bears one net negative charge at pH 6. 5. The component chymotryptic peptides, T2. p. C1a and T2. p. C1b, had one negative charge and no charge respectively; therefore, the Asx must be Asp and the Glx must be Gln.

T2. 5 The supernatant of Fraction T2 contained peptide T2. 5 which was isolated in a pure state after paper electrophoresis at pH 3. 5.

The partial sequence of this peptide was determined by 11 steps of the Dansyl-Edman technique (Figure III-2). The sequence was completed by the isolation of peptides T2. 5. C1, C2 and C3 from a chymotryptic digest of T2. 5. The peptides were isolated by electrophoresis at pH 3. 5. The mobilities and compositions of the peptides are given in Table III-3 and their sequence determination in Figure III-2.

The electrophoretic mobility of T2. 5 at pH 6. 5 was neutral. After removal of the N-terminal Asx, there was very little change in mobility. Therefore, there must be an Asn in the N-terminal position and an Asp in the seventh position.

C. Fraction T3

After an electrophoretic separation at pH 6.5, five pure peptides were obtained from Fraction T3. They were named T3.1, T3.2, T3.3, T3.4 and T3.6. Their compositions are given in Table III-1 and they were sequenced in a straightforward manner (Figure III-3).

The position of the Cmc in T3.1 was determined by the 80% decrease in radioactivity which occurred after the second step of the Edman degradation. As T3.1 and T3.2 were neutral at pH 6.5, both peptides contained aspartic acid residues. T3.6 had a mobility of +0.4 at pH 6.5; therefore, the glutamic residue must be present as glutamic acid.

The partial sequence of T3.3 was determined by 10 steps of the Dansyl-Edman technique. The sequence was completed by the isolation of peptides T3.3.C1 and T3.3.C2 from a chymotryptic digest of T3.3. These peptides were purified by electrophoresis at pH 3.5. Their mobilities and compositions are given in Table III-4 and the sequence determination in Figure III-3. The Cmc in T3.3.C2 was positioned by the 80% decrease in radioactivity which occurred after the second step of the Edman degradation.

Tryptic Peptide		Paper electrophoresis pH	Mobility
T3.1	Ser-Cmc-Asp-Lys → → →	6.5	-0.04
T3.2	Val-Asp-Lys → →	6.5	0
T3.3	Ser-Thr-Ser-Gly-Gly-Thr-Ala-Ala-Leu-Gly-Cmc-Leu-Val-Lys → → → → → → → → → →	6.5	+0.2
	← T3.3.C1 → ← T3.3.C2 →		
	Ser-Thr → →		
	Ala-Ala-Leu Gly-Cmc-Leu-Val ← ← ← → → → →		
T3.4	Gly-Pro-Ser-Val-Phe-Pro-Leu-Ala-Pro-Ser-Ser-Lys → → → → → → → → → → →	6.5	+0.3
T3.6	Lys-Val-Glu-Pro-Lys → → → → →	6.5	+0.4
T4	Thr-His →	3.5	+0.5

Figure III-3. Amino Acid Sequences of and Isolation Procedures for Peptides from Tryptic Fractions Three and Four.

Table III-4. Method of Isolation and Composition of Chymotryptic Peptides of Daw 2a' Tryptic Peptide T3. 3.

Amino acid	T3. 3	T3. 3. C1	T3. 3. C2
Lys	1. 1	0	1. 0
Thr	1. 9	1. 8	0
Ser	2. 2	2. 0	0
Gly	3. 2	2. 0	1. 2
Ala	1. 8	1. 6	0
Val	1. 0	0	0. 9
Leu	1. 8	0. 8	0. 9
Cmc	1.	0	1.
Total	14	9	5
Electrophoretic mobility at pH 3. 5			
	+0. 15	0	+0. 4

Cmc content of the peptides is based on radioactivity. Mobility is given relative to lysine = +1. 0, aspartic acid = -1. 0 and glycine = 0.

d. Fraction T4

The only relevant peptide that this fraction contained was

Thr. His

Its sequence is included in Figure III-3.

e. Peptide T1

The insoluble product of the tryptic digestion of Daw 2a' proved to be a peptide of 62 amino acids in length, named T1. The N-terminal amino acid was identified as Asx by the fluorodinitrobenzene method and the sequence was determined by the isolation of peptides from a chymotryptic digest (700 µ g. enzyme/µ mole peptide for 3 hours at 37° in a Radiometer pH-stat). The supernatant from the digest was fractionated on a column of Sephadex G-50 in 0.05 N NH_3. The compositions of and the isolation procedures for the nine component peptides of T1 are found in Tables III-5 and 6 respectively. The sequences of the nine peptides are seen in Figure III-4. They were determined by a combination of both Dansyl-Edman technique and CPA digestions.

As tripeptide T1. C2. 4a appeared to comprise the C-terminal part of T1. C2. 6a, it has been placed in Figure III-4 in such a position.

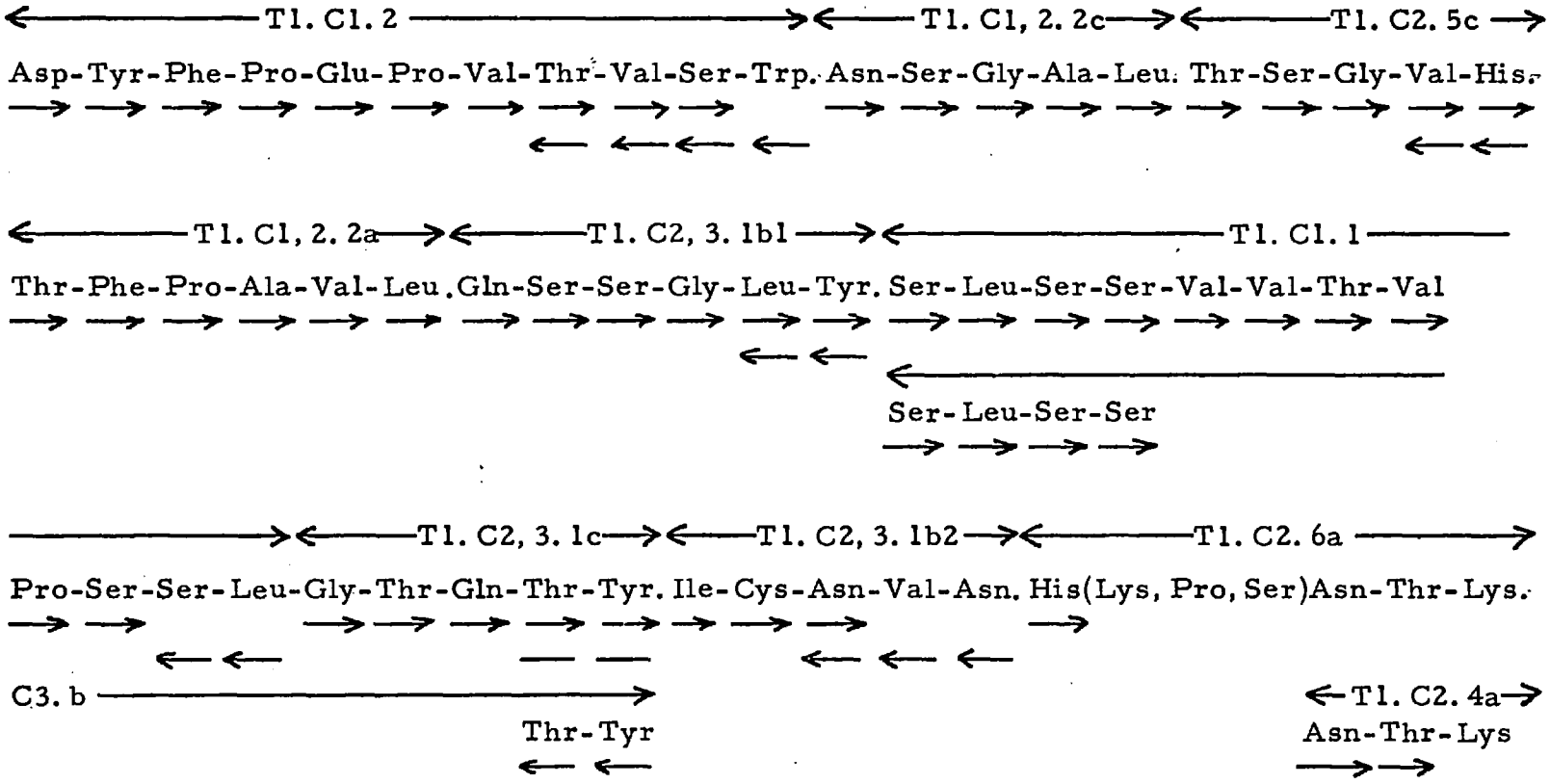


Figure III-4. Amino Acid Sequence of Tryptic Peptide, T1.

Table III-5. Chymotryptic Peptides of Daw tryptic Peptide 2a' T1

Composition (moles of amino acid residue/mole of peptide)

Amino Acid	T1. C1. 1	T1. C1. 2	T1. C1, 2. 2a	T1. C1, 2. 2c	T1. C2. 5c	T1. C2. 6a	T1. C2, 3. 1b1	T1. C2, 3-1b2	T1. C2, 3. 1c	Sum of Peptides	Peptide T1
Lys	0	0	0	0	0	2. 0	0	0	0	2	1. 2
His	0	0	0	0	0. 9	0. 8	0	0	0	2	2. 1
Arg	0	0	0	0	0	0	0	0	0	0	0
Asp	0	1. 1	0	1. 0	0	1. 1	0	2. 1	0	5	4. 4
Thr	0. 8	0. 9	0. 9	0	0. 9	0. 9	0	0	1. 8	7	7. 8
Ser	5. 0	1. 0	0	1. 0	1. 1	1. 2	1. 9	0	0	11	11. 6
Glu	0	1. 1	0	0	0	0	1. 0	0	1. 0	3	4. 0
Pro	1. 1	1. 9	0. 9	0	0	1. 0	0	0	0	5	5. 2
Gly	0	0	0	1. 0	1. 1	0	1. 2	0	1. 3	4	6. 3
Ala	0	0	1. 1	1. 0	0	0	0	0	0	2	2. 0
Val	2. 9	1. 9	1. 2	0	1. 0	0	0	1. 1	0	8	9. 0
Ile	0	0	0	0	0	0	0	0. 8	0	1	0. 8
Leu	2. 2	0	1. 1	1. 0	0	0	1. 0	0	0	5	6. 0
Tyr	0	1. 0	0	0	0	0	0. 9	0	0. 9	3	4. 2
Phe	0	1. 1	0. 8	0	0	0	0	0	0	2	2. 4
Cmc	0	0	0	0	0	0	0	1	0	1	0. 7
Trp	0	1	0	0	0	0	0	0	0	1	1
Total Residues	12	11	6	5	5	7	6	5	5	62	68. 7

Composition of peptide T1 is based on 2 Alanine residues/mole peptide.
Cmc composition of peptide T1 is calculated from the amino acid analysis.
Cmc residues in other peptides were calculated from radioactivity of the peptides.

Table III-6. Isolation Procedures for Component Peptides of Daw 2a' Tryptic Peptide T1.

Tryptic peptides	*Elution Volume	Paper Electrophoresis	
		pH or Chromatography System $^{\phi}$	Mobility
T1. C1. 1	1.5	3.5	0
T1. C1. 2	1.5	3.5	-0.6
		9.1	-0.57
T1. C1, 2. 2a	1.6	3.5	0
		9.1	-0.21
T1. C1, 2. 2c	1.6	3.5	0
		9.1	-0.42
T1. C2. 5c	1.7	3.5	+0.54
		1.9	-0.35
T1. C2. 6a	1.7	3.5	+0.87
		6.5	+0.63
T1. C2, 3. 1b1	1.8	3.5	0
		9.5	-0.43
		BAW	0.45
T1. C2, 3. 1b2	1.8	3.5	0
		9.5	-0.43
		BAW	0.4
T1. C2, 3. 1c	1.8	3.5	0
		9.1	-0.52

* The elution volumes refer to fractionation on Sephadex G - 50 in 0.05 N NH_3

ϕ Chomatography was done in BAW (butan-1-ol/acetic acid/water: 24/6/10).

Electrophoretic mobilities are given relative to lysine = +1.0, aspartic acid = 1.0, and glycine = 0. Chromatographic mobilities are taken relative to the solvent front (Rf value).

The N-terminal part of the sequence has been ordered by comparison with a peptic peptide involving one of the intrachain disulphide bridges of the Fd of IgG1 - Tyr-Ile-Cmc-Asn-Val-Asn-His-Lys (Pro. Ser. Asn. Thr) (Frangione et al, 1969).

As peptides T1. C1, 2. 2c; T1. C2, 3. 1b1; T1,C2, 3. 1b2 and T1. C2, 3. 1c were neutral at pH 6. 5, their respective glutamic and aspartic residues are in the amide form. Peptide T1. C1. 2 had a mobility of -0. 6 at pH 6. 5 equal to 2 negative charges; therefore, the peptide contains both aspartic and glutamic acid. Peptide T1. C2. 4a had a mobility of +0. 47 at pH 6. 5; the peptide must contain asparagine.

The ordering of the peptides was partially accomplished by means of three overlapping sequences obtained from the chymotryptic digest of Daw 2a'. The peptide C3. d establishes the overlap between tryptic peptide T3. 3 and T1. C1. 2 and thus places the latter as the N-terminal peptide of T1. Peptide C3. b overlaps peptides T1. C1. 1 and T1. C2, 3. 1c. The last overlapping peptide, C4. h, places T1. C2. 6a before tryptic peptide T3. 2 making the former peptide the C-terminus of T1.

The remainder of the peptides were put in order by a comparison of T1 with the homologous sequence from rabbit heavy chain (Fruchter et al, 1969) (Figure III-5). The six full stops in Figures III-4, 5 indicate

	150										160										170
Daw 2a' T1	Asp-	Tyr-	Phe-	Pro-	Glu-	Pro-	Val-	Thr-	Val-	Ser-	Trp.	Asn-	Ser-	Gly-	Ala-	Leu.	Thr-	Ser-	Gly-	Val-	His.
Rabbit heavy chain*	Gly-	Tyr-	Leu-	Pro-	Glu-	Pro-	Val-	Thr-	Val-	Thr-	Trp-	Asn-	Ser-	Gly-	Thr-	Leu-	Thr-	Asp-	Gly-	Val-	Arg-

									180												190
Daw 2a' T1	Thr-	Phe-	Pro-	Ala-	Val-	Leu.	Gln-	Ser-	Ser-	Gly-	Leu-	Tyr.	Ser-	Leu-	Ser-	Ser-	Val-	Val-	Thr-	Val-	Pro-
Rabbit heavy chain*	Thr-	Phe-	Pro-	Ser-	Val-	Arg-	Gln-	Ser-	Ser-	Gly-	Leu-	Tyr-	Ser-	Val-	Pro-	Ser-	Thr-	Val-	Ser-	Val-	-

									200												210
Daw 2a' T1	Ser-	Ser-	Leu-	Gly-	Thr-	Gln-	Thr-	Tyr.	Ile-	Cys-	Asn-	Val-	Asn.	His	(Lys,	Pro,	Ser)	Asn-	Thr-	Lys	
Rabbit heavy chain*	-	-	-Ser-	Glx-	Pro	(Pro,	Ser)	Thr-	Cys-	Asn-	Val-	Ala-	His-	Ala	(	,	Thr,	Asx)	Thr-	Lys	

*Fuchter et al, 1969

Figure III-5. Comparative Sequences of Daw 2a' T1 and Rabbit Heavy Chain from Residue 150-211.

the positions where overlapping sequences were not available in Daw 2a' T1 and the alignment was deduced by comparison with the rabbit heavy chain.

In Table III-1 the sum of the amino acid compositions of the nine tryptic peptides is compared with the composition of fragment 2a' and it can be seen that these tryptic peptides do satisfactorily account for the composition of the fragment.

B. Chymotryptic Peptides of Daw 2a'

1. Chymotryptic Digestion

To establish the order of the tryptic peptides of Daw 2a', sequences overlapping the tryptic peptides were required. For this purpose a chymotryptic digest of Daw 2a' (500 μg. enzyme/μ mole fragment) was carried out in a pH-stat for $4\frac{1}{2}$ hours at 37°. Eighty per cent of the digest was recovered and passed through Sephadex G-50 (0.05 N NH_3). Six fractions were collected of which C3, C4, C5, and C6 contained relevant peptides. The elution profile is shown in Figure III-6.

2. Amino Acid Sequences of Chymotryptic Peptides

In Table III-7 are found the amino acid compositions of the chymotryptic peptides which establish the tryptic peptide overlaps. The

Figure III-6. Fractionation of the chymotryptic digest of Dav 2a' on a column (1.8 × 145 cm) of Sephadex G-50 in 0.05 N NH_3.

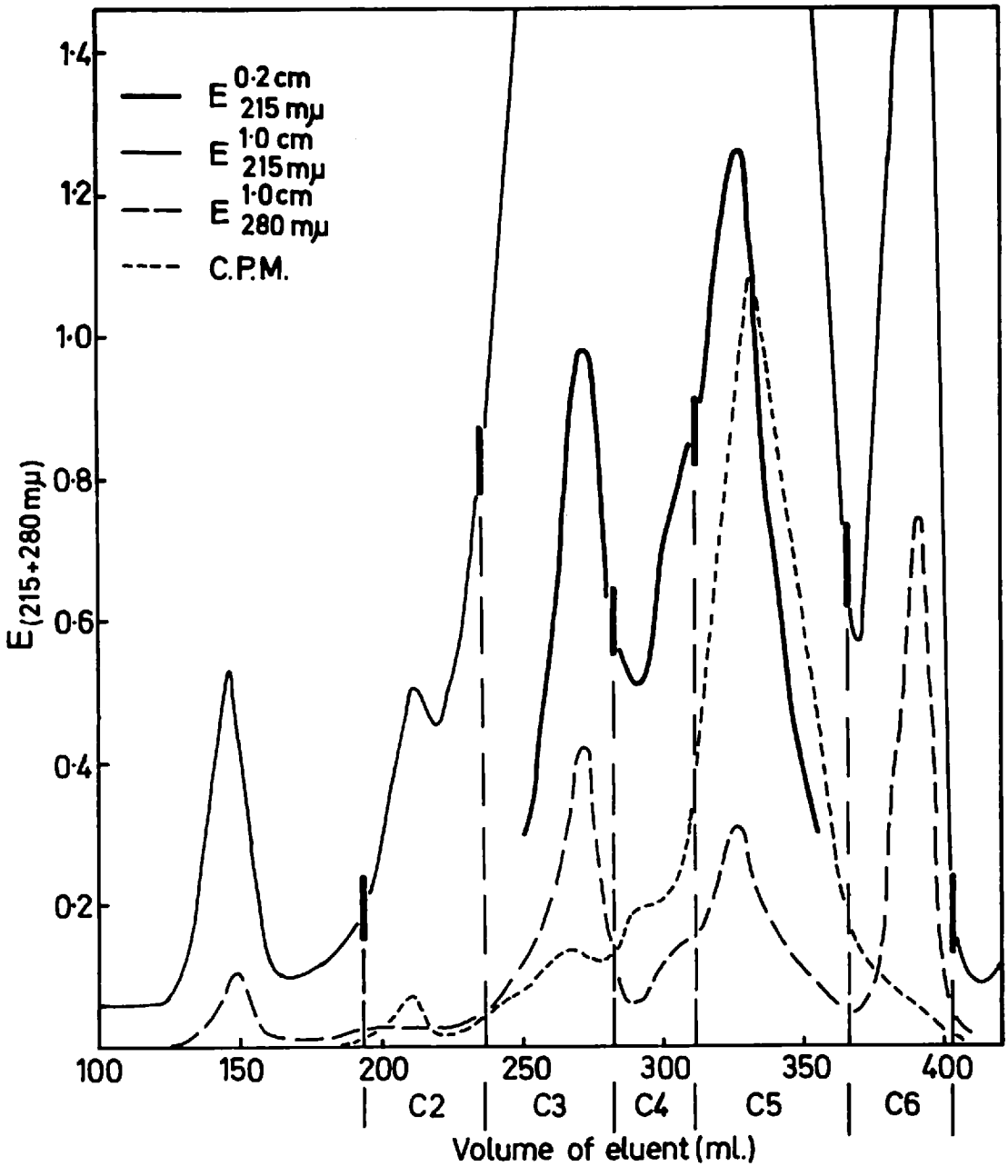


Table III-7. Chymotryptic Peptides of Daw 2a' which Established Tryptic Peptide Overlaps
(Composition: moles of amino acid/mole of peptide)

Amino Acid	C3.a	C3.b	C3.d	C4.f1	C4.f4	C4.h	C5.d7	C5.d9	C5.e8	C5.g1	C5.h	C6.d
Lys	0	0	1.0	1.3	1.1	3.5	0	0	0	0	1.1	0
His	0	0	0	0	0	0	0	0	0	0	0.9	0
Arg	0	0	0	0	0	0	0	0	0	0.8	0	0
Asp	2.0	0	1.2	0	0	1.2	0	0	0	0	1.0	1.2
Thr	2.6	2.1	0.9	2.0	1.0	0	0	0.9	0	0	0.9	0
Ser	0	5.0	1.5	4.1	3.1	0	0	0	0	2.3	1.1	0
Glu	0	1.2	1.0	0	0	1.1	0	0	1.1	1.0	0	0
Pro	1.0	1.2	2.0	1.0	2.3	1.1	0	0	0	0	0	0
Gly	2.1	1.5	0	1.8	1.4	0	1.0	0	1.9	1.2	0	0
Ala	1.0	0	0	2.5	1.1	0	0	0	0	1.0	0	0
Val	1.3	3.1	2.7	0	1.0	2.0	0	2.1	0	0	0	0
Ile	0	0	0	0	0	0	0	0	0.9	0	0	0
Leu	0	2.1	0	1.3	1.2	0	1.0	0	1.1	0	0	0
Tyr	1.0	0.8	0.8	0	0	0	0	0	0	0.7	0	0.9
Phe	0	0	0.9	0	0.8	0	0	0	0	0	0	0.9
Cmc	0	0	0	0	0	0	(1)	0	0	(2)	(1)	0
Trp	0	0	(1)	0	0	0	0	0	0	0	0	(1)
Total	11	17	13	14	13	9	3	3	5	9	6	4

The Trp content of the peptides was assumed to be 1 for those peptides which stained with Ehrlich's reagent on paper. The Cmc composition was based on radioactivity.

chymotryptic peptides from T1, which were reisolated and characterised have not been included in Table III-7, but they have been included in the listing of the methods of sequencing and isolation of the chymotryptic peptides of 2a', Figure III-7, because the methods of isolation were changed from those used to obtain peptides from the chymotryptic digest of T1. They have also been included in Figure III-8 as have been a number of peptides resulting from minor chymotryptic cleavages e.g. C6.c2 and C5.a.

In Figure III-8, C3.a has been placed at the N-terminus of 2a' because it consists of the first eleven residues of tryptic peptide T2.5 and has an N-terminal asparagine. Daw 2a' was determined to have an aspartic N-terminal residue and the only tryptic peptides having an N-terminal aspartic are T2.5 and T1. As the latter peptide follows T3.3 (overlap peptide C3.d), the N-terminal tryptic peptide must be T2.5

The next tyrosine, C6.c1, was cleaved free by chymotrypsin and the C-terminal of T2.5 - Cys.Ala.Arg initiated C5.g1 which overlapped T2.5 and T2.p. This was confirmed by a tryptic digest of C5.g1 from which the peptides C5.g1.Ta and C5.g1.Tc were isolated (see Figure III-7).

Chymotryptic Peptides		pH or Chrom. system	Mobility
C3. a	Asp-Thr-Val-Gly-Pro-Gly-Asp Thr-Ala-Thr-Tyr	3. 5	-0. 3
C6. c1	Tyr	3. 5 BAW	-0. 2 0. 4
C5. g1	Cmc-Ala. Arg-Ser-Cmc-Gly. Ser-Gln-Tyr C5. g1. Tc — C5. g1. Ta Ser Ser-Gln-Tyr	6. 5 1. 9	+0. 28 -1. 0
C6. c2	Phe	3. 5 BAW	-0. 2 0. 62
C5. a	Asp-Tyr-Trp-Gly-Gln-Gly-Ile-Leu	6. 5	-0. 24
C6. d	Phe-Asp-Tyr-Trp	3. 5 BAW	-0. 1 0. 8
C5. e8	Gly-Gln-Gly-Ile-Leu	6. 5 BAW	0 0. 7
C5. d9	Val-Thr-Val	6. 5 BAW	0 0. 75
C4. fd	Ser-Ser-Ala-Ser-Thr-Lys-Gly-Pro-Ser- Val-Phe-Pro-Leu	3. 5 PAW	+0. 45 0. 75
C4. f2a	Ala-Pro-Ser-Ser-Lys-Ser-Thr-Ser-Gly- Gly-Thr-Ala-Ala-Leu	3. 5	+0. 3
C5. d7	Gly-Cys-Leu	6. 5 BAW	0 0. 6
C3. d	Val-Lys-Asp-Tyr-Phe-Pro-Glu-Pro-Val- Thr-Val-Ser-Trp	3. 5	-0. 09
C5. c	Asn-Ser-Gly-Ala-Leu	6. 5	-0. 05
C5. i1	Thr-Ser-Gly-Val-His	6. 5 1. 9	+0. 39 -0. 38

C4. d5	Thr-Phe-Pro-Ala-Val-Leu → → → → → →	3. 5 PAW	0
C5. d6	Gln-Ser-Ser-Gly-Leu-Tyr → → → → ⇄	6. 5 BAW	0 0. 47
C5. d8	Ser-Leu → →	6. 5 BAW	0 0. 7
C3. c	Ser-Ser-Val-Val-Thr-Val-Pro-Ser-Ser-Leu → → → → → → → → ← ←	3. 5 BAW	0 0. 53
C5. e4	Gly-Thr-Gln-Thr-Tyr → → ⇄ ⇄	6. 5 BAW	0 0. 32
C3b	Ser-Leu-Ser-Ser-Val-Val-Thr-Val-Pro-Ser- → → → Ser-Leu-Gly-Thr-Gln-Thr-Tyr ← ←	3. 5 BAW	-0. 1 0. 53
C5. d5	Ile-Cys-Asn-Val-Asn → ← ← ←	6. 5 BAW	0 0. 53
C5. i	His(Lys, Pro, Ser)Asn-Val → ← ←	6. 5	+0. 4
C4. h	Lys-Val-Asp-Lys-Lys- Val-Glu-Pro-Lys → → → → → → →	3. 5	+0. 82
C5. h	Ser-Cys-Asp-Lys-Thr-His → → → ←	6. 5 1. 9	+0. 33 0
C3. e	Lys- Val-Asp-Lys-Lys-Val-Glu-Pro-Lys → → → Ser-Cys-Asp-Lys-Thr-His ←	3. 5	+0. 82

Chromatography was done either in BAW (Butan-1-ol/ acetic acid/water: 24/6/10) or PAW (Pyridine/amyl alcohol/water: 35/35/30)

Figure III-7. Amino Acid Sequences of and Isolation Procedures for Chymotryptic Peptides of Daw 2a'.

Peptides C6.d, C5.e8 and C5.d9 (also C6.c2 and C5.a which were isolated in lower yields) have been shown to belong to peptide T2.p (see Figure III-7 and 8). C4.f4 overlaps the C-terminus of T2.p and the N-terminus of T3.4; peptide C4.f1 overlaps the C-terminus of T3-4 and the N-terminus of T3.3. The C-terminal part of T3.3 is found in two chymotryptic peptides, C5.d7 and C3.d. The latter peptide overlaps T3.3 and T1.

Only one overlapping sequence was obtained to aid in the alignment of the chymotryptic peptides within T1. Peptide C3.b, mentioned previously, allowed peptides C5.d8, C3.c and C5.e4 to be put in order.

Peptide C3.e (and C4.h) contained N-terminal lysine and the position and partial sequence (Figure III-7) of these two peptides enabled the four small tryptic peptides, T3.1, T3.2, T3.6 and T4 to be aligned C-terminally to T1. Peptide C4.h placed T3.2 in front of T3.6; peptide C5.h placed T3.1 in front of T4. Peptide C3.e aligned the four peptides in the order T3.2, T3.6, T3.1 and T4. These peptides have been reported by Steiner and Porter (1967).

Therefore, in Figure III-8 the sequence of Daw 2a' extending from residue 85-225 is shown. The peptides isolated from the tryptic and chymotryptic digests are indicated.

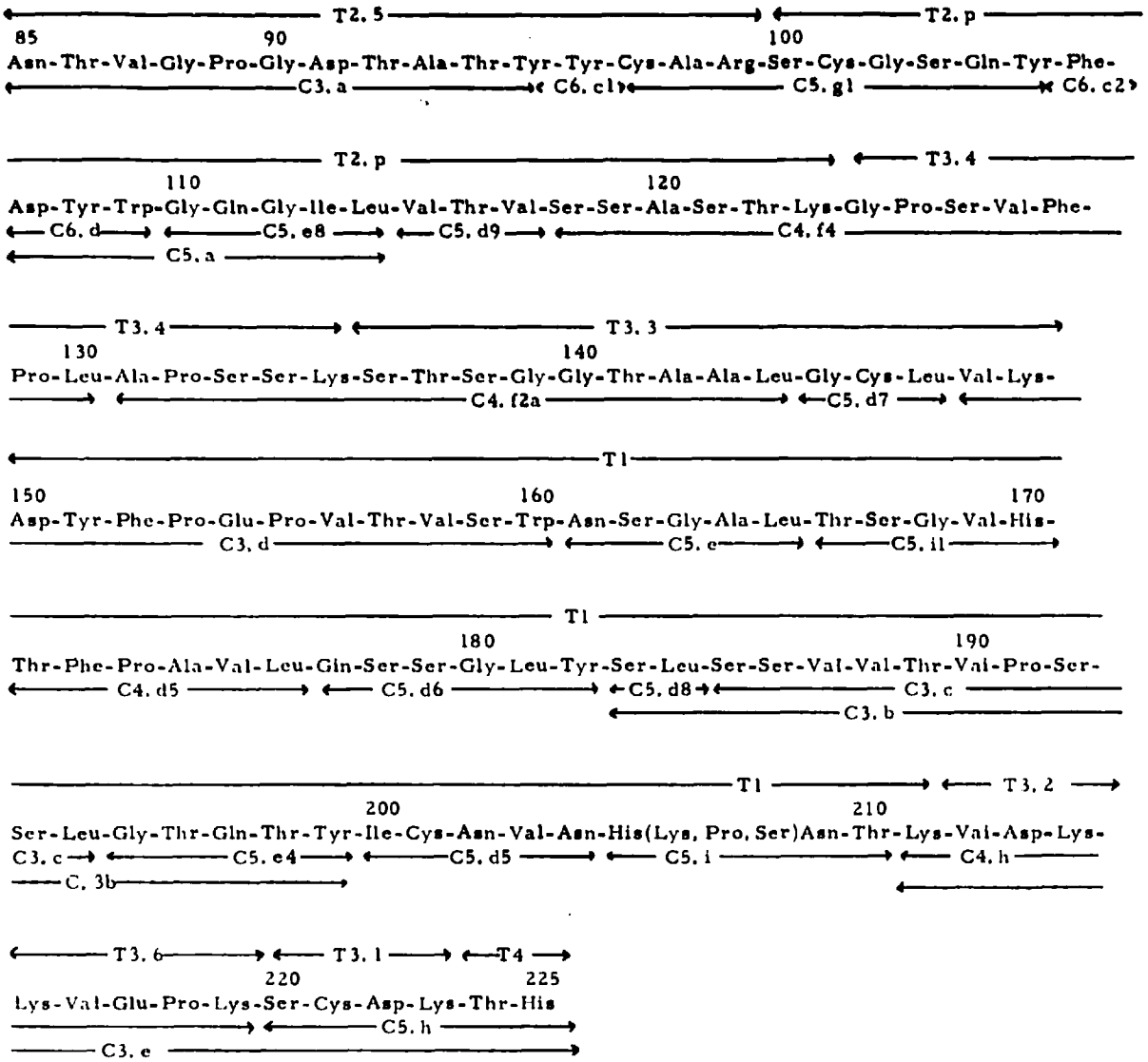


Figure III-8. Amino acid sequence of Daw 2a' (85-225) showing alignment of tryptic peptides and chymotryptic peptides.

II. THE AMINO ACID SEQUENCE OF COR 5b

The sequence of the 34 amino acids of the N-terminal fragment of Cor protein, Cor 5b was done in order that a comparison might be made of the complete Fd fragments of Cor and Daw proteins.

A. Isolation of Cor Fragment 5b

Fragment 5b was prepared from Cor heavy chain by digestion with CNBr followed by fractionation on a column of Sephadex G-100 in 6M urea 0.2M sodium formate as described for cleavage products of Daw 2a'. Fraction 5 (elution volume 2.5) was totally reduced as described by Cebra, Givol and Porter (1967), alkylated with iodo-(1-^{14}C) acetamide and rerun on the same column. Fragment 5b was separated from both 5a and 6a in this step (Press & Hogg, 1969b). The amino acid composition of Cor5b is given in Table III-8.

B. Chymotryptic Peptides of Cor 5b

1. Chymotryptic Digestion

1.5 μmoles of Cor 5b (5 mg/ml in 0.1 M NH_4HCO_3, pH 8.1) were digested with chymotrypsin (80 μg. enzyme/μmole fragment) for 3 hours at 37°. The digest was centrifuged to remove insoluble material and the

Table III-8. Chymotryptic Peptides of Cor 5b

(Composition: moles of amino acid/ mole of peptide)

Amino acid	C2. a	C2. b	C3. a	C3. c	C4. b	Sum of peptides	Fragment Cor 5b
Arg	0	1.0	0	0	0	1.0	1.0
Lys	0	0.9	0	0	0	0.9	1.0
Thr	1.0	1.6	0.9	2.9	0	6.5	6.6
Ser	3.0	1.3	0	0	1.0	5.3	5.1
Glu	0	2.1	1.1	0	0	3.2	3.6
Pro	0	1.8	0	0	0	1.8	2.0
Gly	1.1	1.3	0	0	1.1	3.5	3.2
Ala	0	1.2	0	0	0	1.2	1.4
Val	0	0.9	1.0	0	0	1.9	2.0
Leu	1.0	1.9	1.0	1.2	0	5.0	4.7
Phe	0	0	0	1.2	0.9	2.1	1.8
Hsr	0.9	0	0	0	0	0.9	0.7
Cmc	0	0	0	0.7	0	0.7	0.9
Total	7	14	4	6	3	34	34

Mobility -

pH 1.9 +0.35 +0.53 +0.10 +0.40 +0.50

The composition of Cor 5b was based on 2 residues of valine/mole Cor 5b. Cmc residues were based on the radioactivity of the peptides. The electrophoretic mobility at pH 1.9 was determined relative to the mobility of lysine.

supernatant was fractionated on a column of Sephadex G-25 in 0.02 N NH_3. The elution profile consisting of four fractions is shown in Figure III-9. The first fraction appeared to contain undigested material and was disregarded. Fractions 2, 3 and 4 each contained peptides which were purified by paper electrophoresis at pH 1.9.

The amino acid compositions of Cor 5b and the five component chymotryptic peptides are given in Table III-8.

2. Amino Acid Sequences of the Five Chymotryptic Peptides

Peptide C3.a stained with the hypochlorite-starch-iodide stain but not with ninhydrin. The amino acid composition was the same as that of Cor protein N-terminal tripeptide PCA.Val.Thr (Press & Piggot, 1967) but included an additional Leu. CPA digestion removed 1 mole of Leu and 0.9 mole Thr. Therefore, this peptide has the sequence

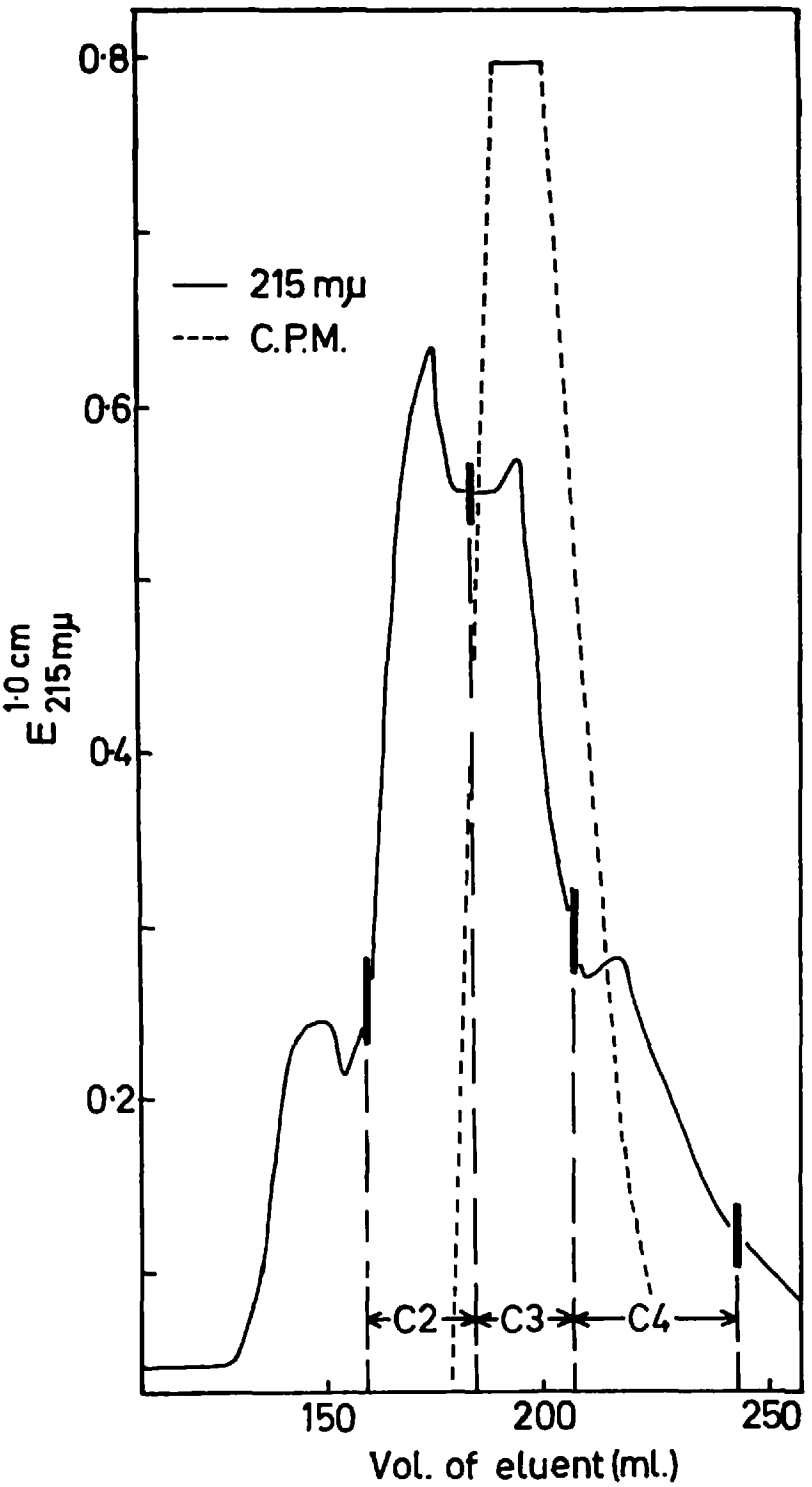
PCA-Val-Thr-Leu

and represents the N-terminal peptide of Cor 5b and of Cor heavy chain.

Peptide C2.b was the largest peptide to be isolated. Its sequence was determined by eleven steps of the Dansyl-Edman technique and by a $1\frac{1}{2}$ hour CPA digestion which removed 1 mole Leu, 0.9 mole Thr and 0.2 mole Gln. The sequence is

Arg-Glu-Ser-Gly-Pro-Ala-Leu-Val-Lys-Pro-Thr-Gln-Thr-Leu

Figure III-9. Fractionation of the chymotryptic digest of Cor 5b on a column (2.0 x 145 cm) of Sephadex G-25 in 0.02 N NH_3.



The second residue was deduced to be in the acid form because the peptide had one positive charge at pH 6.5 and the other glutamic residue near the end of the peptide was removed by CPA as glutamine.

Peptide C3.3 was the only peptide to contain Cmc. The sequence was determined by five Dansyl-Edman steps and a $1\frac{1}{2}$ hour CPA digestion which removed 1.0 mole Phe and 0.9 mole Thr. The sequence is

$$\overrightarrow{\text{Thr}}\text{-}\overrightarrow{\text{Leu}}\text{-}\overrightarrow{\text{Thr}}\text{-}\overrightarrow{\text{Cmc}}\text{-}\underset{\leftarrow}{\overrightarrow{\text{Thr}}}\text{-}\underset{\leftarrow}{\text{Phe}}$$

Peptide C4.b was the smallest peptide to be isolated and its sequence completed by three Dansyl-Edman steps is

$$\overrightarrow{\text{Ser}}\text{-}\overrightarrow{\text{Gly}}\text{-}\overrightarrow{\text{Phe}}$$

Peptide C2.a contained homoserine and was, therefore, the C-terminal peptide of Cor 5b. The sequence was determined by six Dansyl-Edman steps to be

$$\overrightarrow{\text{Ser}}\text{-}\overrightarrow{\text{Leu}}\text{-}\overrightarrow{\text{Ser}}\text{-}\overrightarrow{\text{Gly}}\text{-}\overrightarrow{\text{Glu}}\text{-}\overrightarrow{\text{Thr}}\text{-}\overrightarrow{\text{Met}}$$

3. Ordering of the Peptides of Cor 5b

Peptides C3.a and C2.a were known to be the N- and C-terminal peptides respectively because of the presence of a blocked N-terminal amino acid in C3.a and of a homoserine C-terminal residue in C2.a. Fragments C2.b, C3.c and C4.b were ordered by comparison with the

sequence of Daw protein, positions 5-27. In this region the sequence of Cor 5b is identical with Daw protein except for a substitution of an Arg at Daw 13 by a Lys in C2.b at the same position. The C-terminal peptide, C2.a differs from the Daw sequence at three consecutive positions, corresponding to Daw 31-33. Figure III-10 gives the completed sequence of Cor 5b.

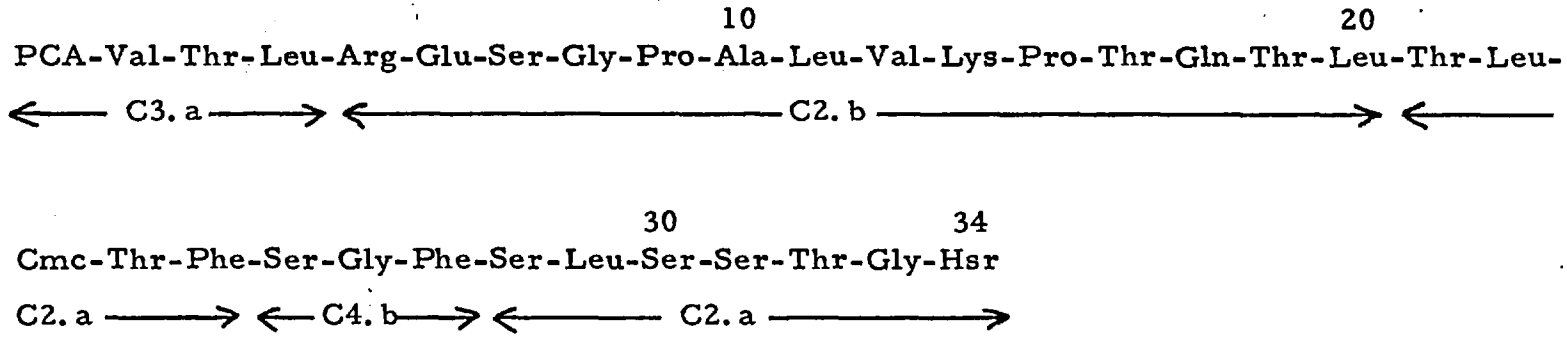


Figure III-10. Amino acid sequence of Cor 5b showing chymotryptic peptides derived from it.

III. A PROBLEM ENCOUNTERED IN SEQUENCING DAW 2a': A Partial Sequence of Daw Lambda Light Chain.

The sequencing of Daw fragment 2a' (85-225) progressed without exceptional difficulty until there remained a gap between what are now known to be positions 100 and 124 (Figure III-8). The tryptic digest of Daw 2a' was done with 2u moles of fragment and in excess to the peptides known by yield to belong to the sequence, a considerable number of small peptides were isolated. As sequencing progressed it became obvious that the peptides were unrelated to each other and obtained in yields which did not compare to the major Daw 2a' peptides. Moreover, since Cor protein had 20 residues in the "gap", it was unlikely that the Daw sequence would be 2-3 times longer in that region, although we did have reason to suspect (by analogy with the light chain) that this was the area where heterogeneity in sequence and in length might occur.

In short, the extra peptides came from a fragment of Daw light chain which had been contaminating the preparations of Daw 2a' to the extent of 20%. Since no peptides were found which could be placed before residue 66, it was considered that the contamination came from a fragment of light chain, 148 amino acids in length

(position 66-214). The residue immediately before position 66 is a lysine, present in both of the proteins sequenced in this area - New (Langer et al, 1968) and Ha (Putnam et al, 1967). It is possible that papain cleavage occurred at this point in 20% of the molecules. As Daw 2a' is 140 amino acids long the two fragments would not have been separated in the gel filtration step following the reduction and alkylation which did separate intact light chain from Daw 2a'.

When this difficulty was recognised the work proceeded quickly to conclusion. The missing tryptic peptide, T2.p, which was 24 residues in length had been recovered in yields lower than the light chain peptides because of adherence to the preparative paper electrophoresis sheets. It was subsequently isolated in 60% yield by a precipitation technique (see text).

In Figure III-11 the partial sequence of Daw lambda light chain is compared to New and Ha of lambda subgroup I. Forty-four variable positions can be compared. New and Daw have only 5 differences (or 89% identity), 2 of which occur in the hypervariable region at positions 95 and 97. Ha protein has fewer identities with both proteins (Ha/Daw-68%; Ha/New-73%). Daw light chain has 50-59%

identity with λ subgroups II, III and IV in the region from 66-109. Peptides from the C-region which could be compared by sequence or by composition 110-131; 159-168; 192-214 were identical in all three proteins.

Daw light chain fits quite well into λ subgroup I. There does not appear to be any unusual homology in the sequence or identities in the pattern of variation between Daw light and heavy chains. No very obvious homologies could however be expected, since although Daw and Cor heavy chains are very similar, their light chains must be very dissimilar, as they are of two different types, K and λ.

```
          66                  70
Daw  Thr-Gly-Thr-Ser-Ala-Thr-Leu. Gly-Ile-Thr-Gly-Leu-Thr-Thr-

New  Ser ———————————————————————————————————————— Arg ———
Ha   Ser ——————————— Ser ——— Ala ——————————— Arg-Ser-

          80                                          90
Daw  Gly-Asx(Glx, Ala, Asx, Tyr)Tyr. Cys(Gly, Tyr)Trp. Asx-Ser-Ser-

New  ————————————————————————————————————————————————————
Ha   Glu———————— His——— His ———————————————— Tyr-Arg-

          95       97            100
Daw  Leu-Gly-Ala-Gly-Val-Phe. Gly-Gly-Gly-Thr-Lys. Leu-Thr-Val-

New  —— Asn —— Val ———————————————————— Val ————————
Ha   ——Ser —— Val ———————————————————— Gln ————————

Daw  Leu-Gly-Gln-Pro-Lys. Ala-Ala (Pro, Ser, Val, Thr)(Leu, Phe, Pro,

New  ————————————————————————————————————————————————————
Ha   ____ Arg_____________________________________________

                                        130
Daw  Pro, Ser, Ser, Glu, Glu, Leu)(Gln, Ala, Asn, Lys)

New  ———————————————————————————
Ha

          159                                   168
Daw  Ala-Gly-Val-Glu-Thr-Thr-Thr-Pro-Ser-Lys

New  ——————————————————————————————————
Ha

          192
Daw  Ser-Tyr-Ser-Cys-Glx-Val-Thr-His-Glx-Gly-Ser-Thr-Val-Glu-

New  ————————————————————————————————————————————————————
Ha

                                              214
Daw  Lys. Thr-Val-Ala-Pro(Thr, Glu, Cys)Ser.

New  ———————————————————————————————————
Ha
```

Figure III-11: A partial sequence of Daw lambda light chain, residues (66-131; 159-168; 192-214) and comparison with lambda light chains, New and Ha.

CHAPTER IV

DISCUSSION

I. Introduction

Both heavy and light chains are known to be heterogeneous at the N-terminal end by the results of peptide mapping and in the case of the latter, by extensive sequence work. As antibody activity has been shown to reside in the Fab fragment, it has been assumed that the highly variable sequences are directly related to antibody specificity. The aim of this study has been to determine the extent of the variable region in the heavy chain and thus to gain some insight into immunoheterogeneity and its structural basis.

The study was carried out using two monoclonal immunoglobulin proteins called Daw and Cor. Daw protein was obtained from a Caucasian with idiopathic hypergammaglobulinemia; Cor protein from a Negro with multiple myeloma. Both proteins belong to the IgG1 subclass and both are Gm(a+z+). It was considered reasonable to chose two proteins that were as alike as possible so as to eliminate differences due to subclass or allotype. It was hoped that the sequence variation remaining would be candidate for constituting the antibody combining site.

Cyanogen bromide fragments of Cor heavy chain were isolated by a method similar to that used for Daw protein (Press & Piggot, 1967).

As Cor contains five methionine residues compared to the four of Daw, one more Cor fragment was isolated. A comparison of fragments produced by cyanogen bromide cleavage of Daw and Cor heavy chains is shown in Figure IV-1. Daw 1-84, that is, fragments 2b and 4, have been reported by Press and Piggot (1967) and in this study the sequence has been extended to residue 225, the C-terminal residue of the Fd fragment. This length of sequence, residues 85-225 was contained in cyanogen bromide fragment 2a'. Steiner and Porter (1967) have reported the 28 residue sequence of Daw 2a''; therefore, the sequence of Daw has been completed from residue 1-253. Moreover, the C-terminal 18 residues, fragment 5, have been sequenced by Press et al (1966). Miss E. M. Press has sequenced Cor fragments 5a, 6a, and 2a' in conjunction with the work presented in this thesis (Press & Hogg, 1969a, b). The author has sequenced Cor 5b in order that the two immunoglobulins, Daw and Cor might be complete from residue 1-225. A comparison of their sequences is given in Figure IV-2. The two sequences have been aligned to give maximum homology. As it was decided to base the numbering on Daw protein, an adjustment of Cor protein was necessary. It involved a deletion in Cor of three residues, 88-90 and an insertion of four residues after position 104.

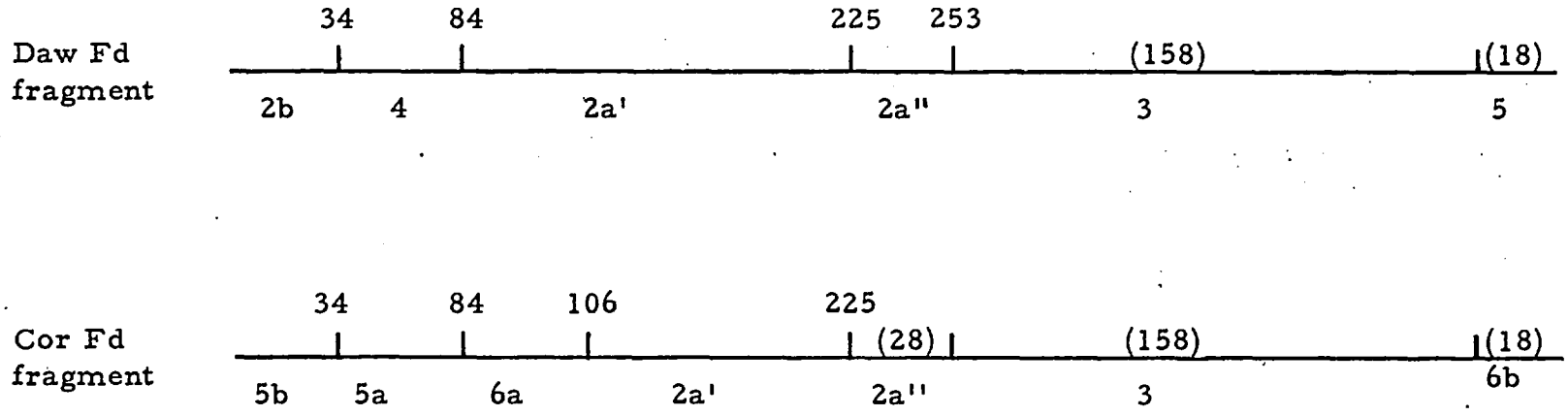


Figure IV-1. Comparison of the Fragments Produced by Cyanogen Bromide Cleavage of the Heavy Chains of Daw and Cor Proteins.

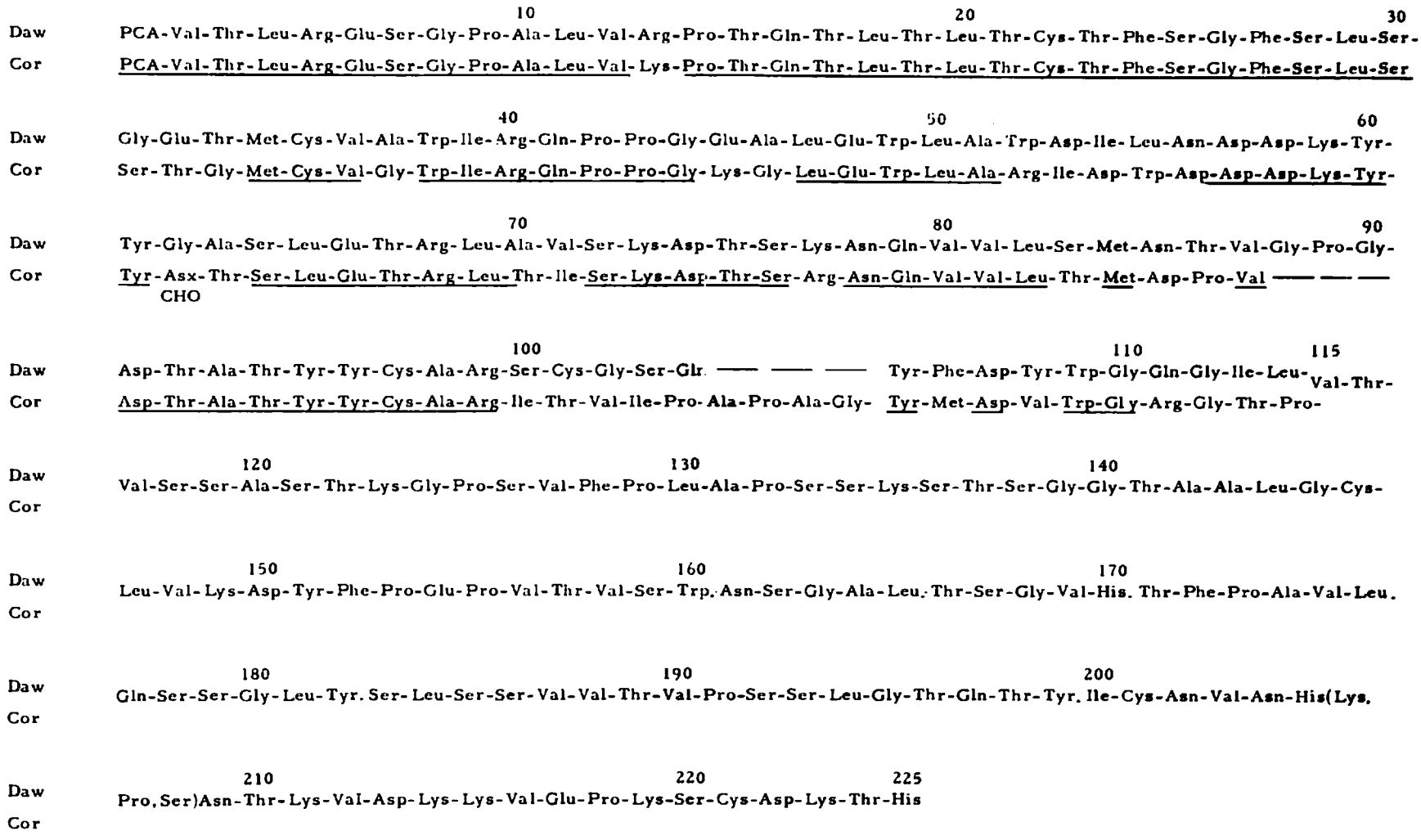


Figure IV-2. Comparison of the Amino Acid Sequences of the Fd fragments of Daw and Cor proteins. The long dashes indicate deletions and the full stops indicate that the overlapping peptides at these points have not been isolated. Identical residues are underlined.

An even better homology could be obtained if two tryptophanes were deleted from the comparative sequence - Daw 52, Cor 55a.

II. Comparison of Daw and Cor Proteins

A comparison of Daw and Cor from residue 1-99 will show that there are 76 residues in similar positions, i.e. a 77% identity in sequence. The sequence from 1-30 is identical except for a difference of a lysine in Cor for an arginine in Daw at position 13. Examples of shorter stretches of identical sequence are 38-44 and 91-99.

Between 100 and 104 there are no identical residues between the two proteins. Cor has an extra four residues in this region which have been placed after Daw 104.

From residue 105 to 114, five of the ten positions are shared by identical residues. Therefore, the homology (50%) is somewhat lower than that of residues 1-99 (77%).

From residue 115-225 the sequence is identical between the two proteins.

Two disulphide bonds were a common feature of several myeloma Fd fragments studied by Frangione and Milstein (1967). Cor protein has two disulphide bonds linking Cys 22-97 in the V-region and Cys

146-201 in the C-region (Press & Hogg, 1969b). It also contains an unpaired Cys-35. Daw protein is exceptional in having three disulphide bonds in the Fd (Fig. IV-3). If, by analogy with Cor, Cys 22-97 and Cys 146-201 are linked, then Cys 35 must be bonded by Cys 101. This seems possible as the CNBr fragments 2b, 4 were joined to 2a' before reduction.

III. Comparison of Daw and Cor with other Heavy Chains

Edelman et al (1969) have recently published a complete sequence of an IgG1 heavy chain, Eu, and twenty-one residues of a second one, He, which extends from residues 105-126 (Daw numbering). The N-terminal 24 and 9 additional residues of a third γ1 chain, Ste, have been sequenced by Fisher et al (1969). All three proteins are Gm (a-f+). A μ chain, Ou, has been sequenced from 1-106 (Daw numbering) by Wikler et al (1969). These sequences are compared to Daw and Cor (Figure IV-4).

The sequence of Eu becomes identical to that of Daw and Cor at position 115, but He has an alanine in this position. Therefore, the variation in γ1 chains extends at least to position 116. The μ chain sequence does not extend far enough to make this comparison.

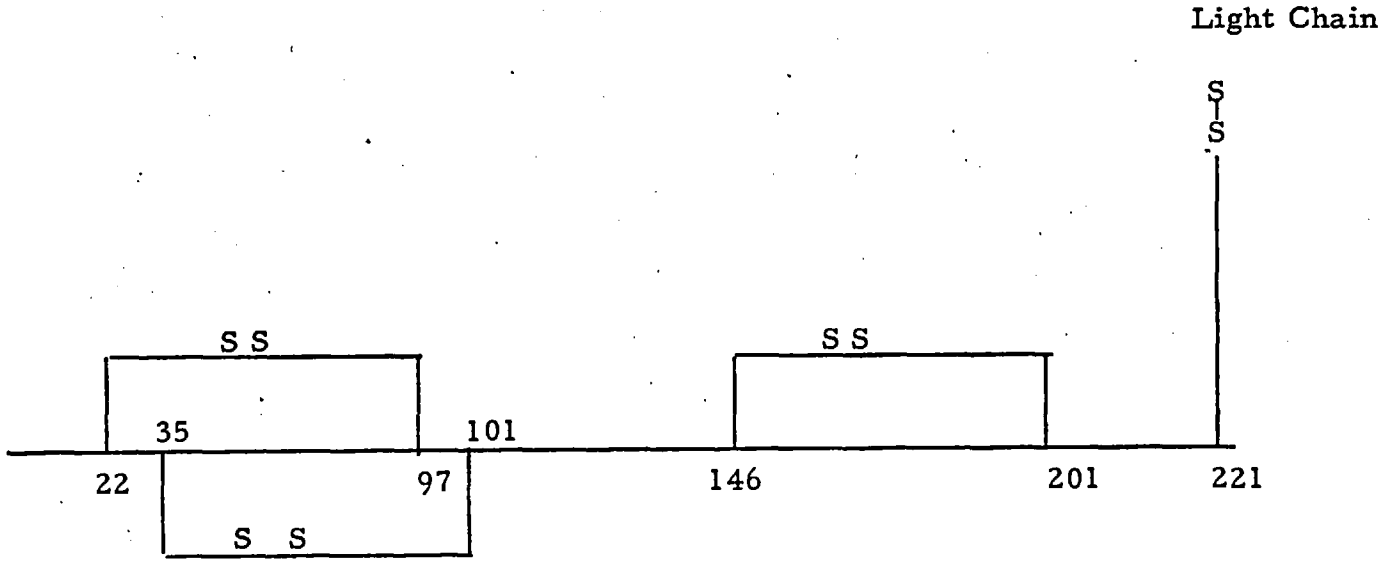


Figure IV-3. The Intra- and Interchain Disulphide Bonds of Daw Fd fragment.

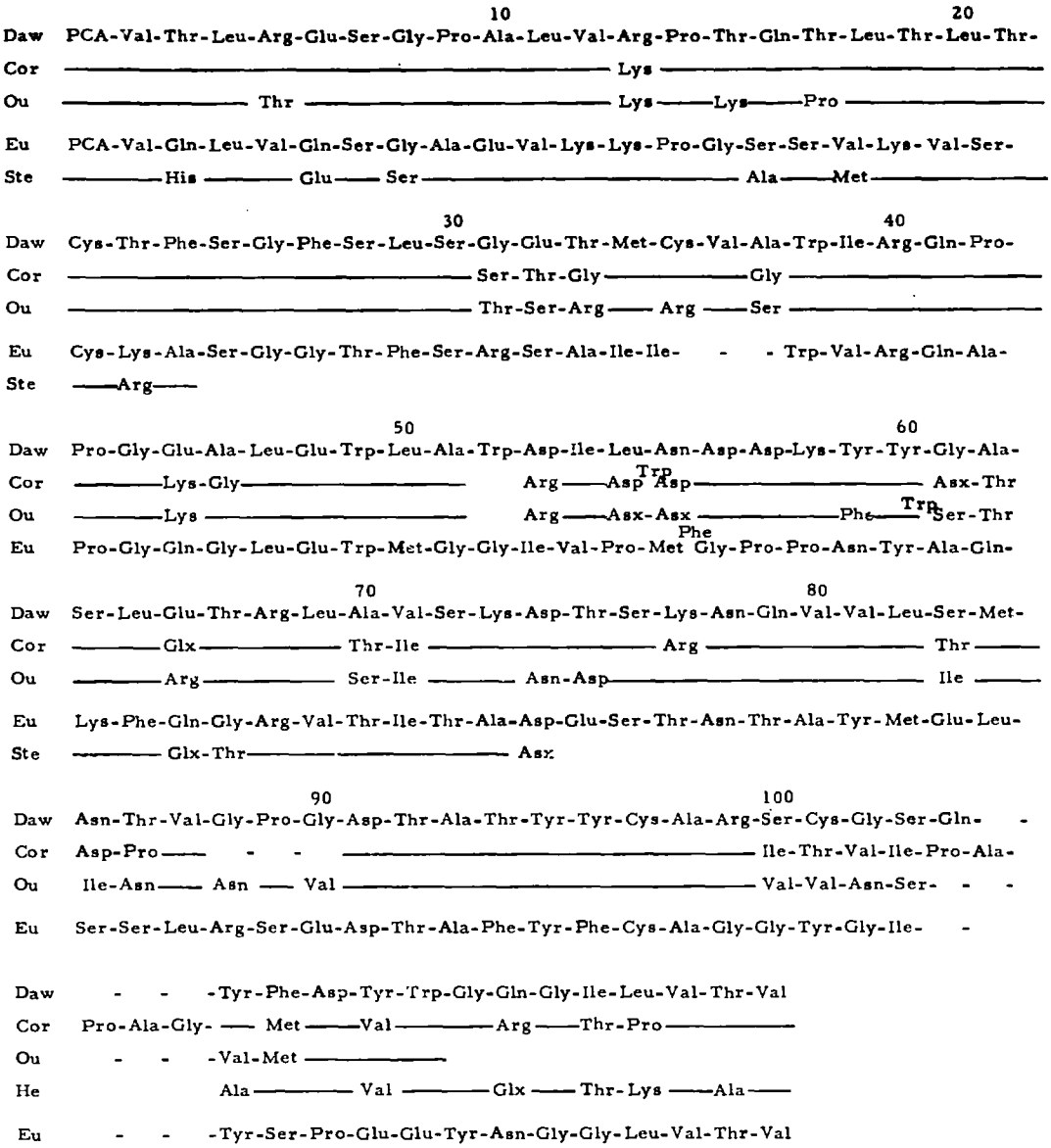


Figure IV. 4. A comparison between variable region Sequences of γ1 heavy chains Daw, Cor and Eu, the μ chain Ou and partial sequences of γ1 chains Ste and He.

The homology between the sequence of the μ chain, Ou, and that of Daw and Cor is quite remarkable (Fig. IV-4). Between 1-99 there is a 70% identity with Daw and a 76% identity with Cor, approximately the same percentage identity as between the two γ1 chains, Daw and Cor.

When the sequence of Eu was compared between 1-99 with Daw, Cor, and Ou, there was only 29, 28, and 27% identity respectively. These similarities invariably occurred within the blocks of sequence identical between Daw, Cor, and Ou (Fig. IV-4).

Although similarities and differences are quite distinct from residue 1-99, after position 99 this separation is not so obvious. We can fully compare only Daw, Cor and Eu in this region as the sequence of Ou stops at 106 and the twenty-one residue peptide, He, begins at 105 and continues into the constant area. From position 100-104 the sequence between the four proteins varies almost completely one from the other. This region includes an addition of four residues into Cor after 104 and a deletion of one residue position 104 in Ou and Eu.

From 104 to 116 there are a number of similarities between Daw, Cor and He; 5 of 12 residues are identical between all three (42%) and two more identical between any two (58%). Eu protein has only two of

the five shared identities (16.5%) and has 3, 2, 0 further residues shared with Daw, Cor and He respectively.

It is proposed to place the heavy chain variable regions of Daw, Cor and Ou in one subgroup called Daw-type and to place Eu in a second subgroup called Eu-type.

Ste protein has a 75% identity with Eu from 1-24, but only 34% identity with Daw in this region (Fig. IV-4). Therefore Ste fits quite well into the Eu-type subgroup.

There are indications that other subgroups might exist (Fig. IV-5). N-terminal peptides have been sequenced for a variety of monoclonal heavy chains representing γ, μ and α chain classes. It is imprudent to postulate subdivisions on such limited sequences. However, most light chains can be placed in their appropriate subgroups on the basis of the first four residues (Figs. I-4, 5). Therefore, the heavy chain sequences have been placed into four subgroups.

Daw and Cor were chosen for comparison because they both belonged to subclass IgG1 and were both Gm (a+z+). It was hoped in this way to minimize amino acid changes related to particular polymorphic forms. Since Gm (z+) and (f+) are the only factors to be

I DAW-type			
Daw	γ1	PCA-Val-Thr-Leu-Arg...	Press and Piggott, 1967
Cor	γ1	PCA-Val-Thr-Leu-Thr...	Press and Hogg, 1969b
Ou	μ	PCA-Val-Thr-Leu-Thr...	Wikler et al, 1969
Io	μ	PCA-Val-Thr-Leu	Montgomery et al, 1969
II EU-type			
Eu	γ1	PCA-Val-Gln-Leu...	Edelman et al, 1969
Fi	γ	PCA-Val-Gln-Leu	Montgomery et al, 1969
Vu	α	PCA-Val-Gln-Leu	Montgomery et al, 1969
Ste	γ1	PCA-Val-His-Leu...	Fisher et al, 1969
III VIN-type			
Vin	γ4	Gln-Val-Gln-Leu...	Pink and Milstein, 1968
Wag	μ	Gln-Val-Gln-Leu	Kaplan and Metzger, 1969
Koh	μ	Gln-Val-Gln-Leu	Kaplan and Metzger, 1969
IV Bennett-type			
Dos	μ	PCA-Ser-Val-Ala-Asx (Gly, Thr, Leu) Glx	Bennett, 1968
Bus	μ	PCA-Ser-Val-Leu-Asx-Gly	
Bal	μ	PCA-Ser-Val-Ala-Glx (Gly, Leu)	
Dau	μ	PCA-Ser-Val-Leu-Asx (Ala, Thr)	

Figure IV-5. Human Heavy Chain Variable Region Subgroups.

expressed on the Fd part of the heavy chain, it was felt that they might be associated with the N-terminal sequences in the same manner as the rabbit allotypes (Wilkinson, 1969a). Therefore, the question was asked whether the subgrouping and the Gm (z) and (f) factors could be correlated. Daw and Cor are Gm (a+z+) and Eu, Ste and He are Gm (a-f+). However, the μ chain, Ou, does not express Gm specificities and He has tentatively been placed in the Daw-type subgroup. This would mean that the Gm factors may not be associated with subgroup and, therefore, not with the V region.

In the C region of the Fd of Daw/Cor and Eu there is only one amino acid difference which is at position 215 (Daw/Cor-Lys; Eu-Arg). This could be responsible for the Gm (z) and (f) factors ... lysine denoting Gm (z) and arginine denoting Gm (f). Further sequences will decide this point.

The existence of subgroups shared by μ chains and by $\gamma 1$ chains of different Gm specificities gives support to the hypothesis that two genes are involved in the synthesis of one heavy chain; one, the C region class specific gene and the other coming from shared sets of V region genes. The idea was proposed to account for the fact that

the very heterogeneous light chain V regions associated with C regions which were constant in sequence. That the same allelic markers in the V regions of rabbit heavy chains are associated with all three non-allelic class markers in the constant region was suggestive that the same process was happening in heavy chains. Moreover, the hybrid IgG1-G3 immunoglobulin found in the serum of a patient missing both IgG1 and IgG3 subclasses provides strong evidence for single C region genes for each subclass (Kunkel et al, 1969). It is highly unlikely that multiple genes could be involved in such a recombinational event. The demonstration of a close identity of sequence in the V regions of two gamma and one mu chain makes it most likely that two pieces of genetic information are put together at some point in the differentiation process to form one heavy chain. The separation of the two regions entails some step linking them at the DNA, RNA or protein level. A translational device can be eliminated as both heavy and light chains are made as single units on the ribosomes (Knopf et al, 1967; Fleischman, 1967). Arguments have been presented for the fusion occurring at the DNA level. Dreyer, Gray and Hodd (1967) proposed a copy-splice mechanism by analogy with the E. coli replication

system (Cairns, 1963). They envisaged that a copy of a particular V region is built up or spliced onto a C region gene in a conservative fashion. A subtely different analogy is made in the "translocational hypothesis" proposed first by Dreyer and Bennett (1965) and developed by Lennox and Cohn (1967), which suggests that some mechanism akin to the integration of the bacteriophage lambda (Campbell, 1962) or P22 episome (Smith & Levine, 1967) into the E. coli genome could be operating to cause a V-gene carrying a pairing region to recombine with a C-gene containing the same specific pairing region. The recombination would cause a fused VC-gene which would become integrated into the chromosome.

The pairing region may not be translated as there appears to be no length of sequence close to the switchover point which differs sufficiently between κ and λ chains to serve as a distinct recognition point or pairing region. Within the subclasses of IgG the first known difference occurs at position 134 (18 residues after the proposed beginning of the C region at position 116) (Frangione et al, 1969). It will be of interest to see whether the sequence preceding this point will be common to all heavy chain classes. The pairing region of lambda bacteriophage, b2, is 10^4 bases long, although the actual DNA homology between the

bacteriophage and bacterial sites of attachment have not yet been rigorously shown (Smith & Levine, 1967). Such a length of sequence might not be necessary for the accurate pairing of immunoglobulin chains. In fact, perhaps one should be looking for a non-specific recognition region in sequences of identity between all three types of chain (see section IV). That each set of V regions fuses solely to its particular type of C region could be due to chromosomal proximity. It is known that the four human IgG subclasses are contiguous on the genome (Natvig et al, 1967) and that the three IgG subclasses and IgA subclass are linked in the mouse (Herzenberg et al, 1968). Conversely K and λ light chain types appear to be unlinked to heavy chain genes and to each other (Mage et al, 1969) and this may be the reason for the exclusiveness of their distinctive V region pools.

IV. Comparison of Daw and Cor Proteins with Light Chains

There are homologies to be found between the heavy-light chain pattern of variation. The variable region of heavy chains is approximately the same length as that of light chains. The variation in K chains extends for 107 residues; the λ light chain region is often somewhat longer -

108-112 residues. There is a sequence - Pro-Ser-Val-(Phe) - found near the beginning of the constant region in both types of light chains and the γ1 heavy chains (Fig. IV-6). Making many adjustments in order to produce maximum homology, a mere 25% identity is obtained between the 107 residues of the light chain C regions and the 110 residues comprising the constant area of the Fd. The intrachain disulphide bonds are in homologous positions in the two chains (light chains: 23-88 and 136-195 and heavy chains: 22-97 and 146-201). Identities in sequence are grouped in particular in the region of the two sets of C region cysteines.

Within light chain subgroups there is 70-85% identity of sequence and between subgroups there is 40-55% identity. As heavy chain similarities and differences fall into the same range of percentages it would seem justifiable to group the sequences as has been described. However, heavy chain subgrouping is not completely homologous with the light chain subgrouping because heavy chain V regions appear to be shared across class lines rather than being restricted to one particular class of C region as is found in light chains in which the K and λ C regions each have an exclusive set of V regions. A possible explanation

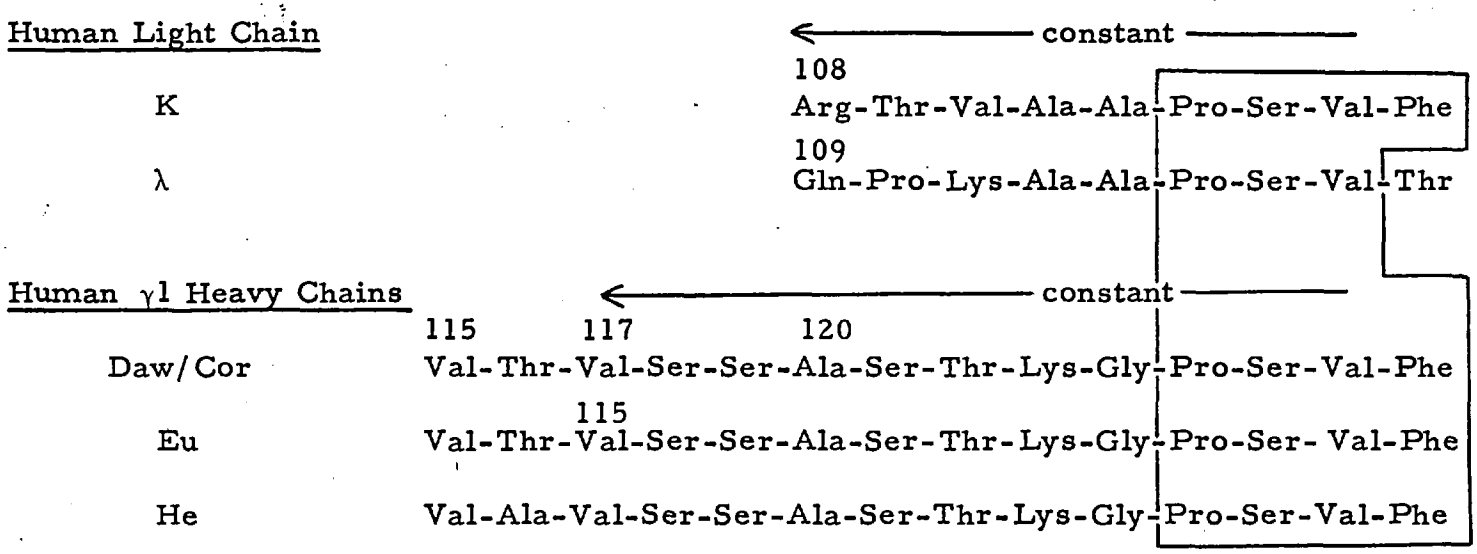


Figure IV-6. The N-terminal sequences of the constant regions of light and heavy chains.

for this phenomenon was given in the preceding section.

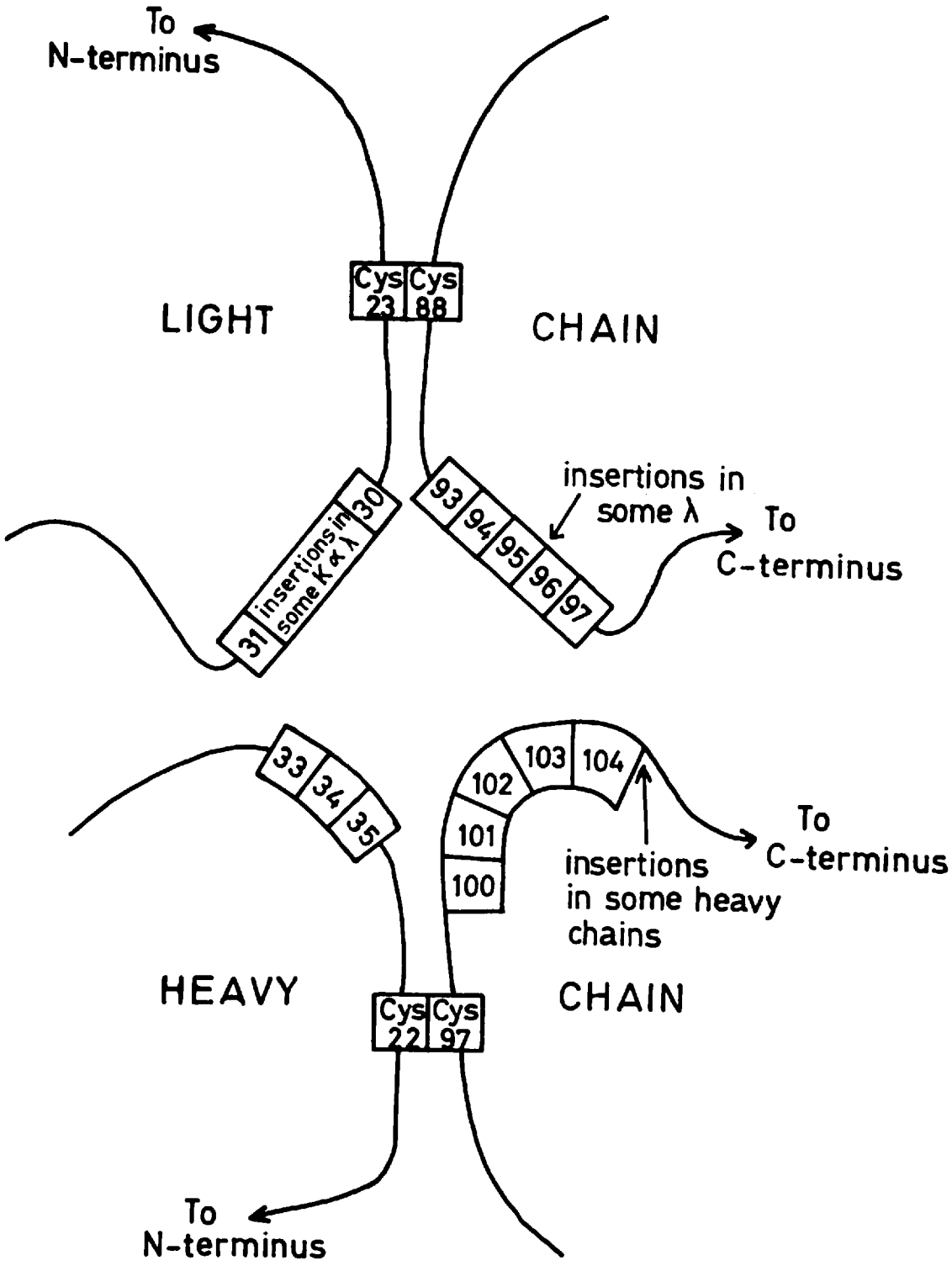
One of the major aims of this study was to compare the sequences of heavy and light chains in order to discover the amino acids which may be involved in the amino acid combining site. The patterns of variation are much the same in both heavy and light chains. Sequence variations most often involve one variant of 3-4 residues at any one position (e.g. Figure I-4 Position 20: Rad/Fr4/Ti-Thr; B6-Ala) except in the areas after the two cysteines which form the V region intrachain bond where extensive variation occurs (Figures I-6, 7). This has been described under "Idiotypy" in the Introduction. A parallel for this sort of heterogeneity is to be found in the heavy chains, Daw, Cor and Ou. The blocks of identical sequence are interspersed with single and double positions in which variation has occurred in one of the three proteins (e.g. Position 45: Cor/Ou-Lys; Daw-Glu. Position 46: Daw/Ou-Ala; Cor-Gly) (Figure IV-4). The positions where all three residues vary are much more restricted - 31 to 33 and 100 to 104 are the most hypervariable areas. Although containing invariant Met-84 and Val-87, the area 83-90 is also variable and includes a deletion of three residues from Cor (88-90). The variant areas would be brought in close proximity to each other by the formation of the V region intrachain

loop - Cys 22-97 (Cys 22-88 in light chains). As it is thought that the N-terminal sequences of light and heavy chains lie parallel to each other (Feinstein & Rowe, 1965; Valentine & Green, 1967), the hypervariable areas in both chains may be in juxtaposition, thus fulfilling the requirements for a specific combining site incorporating both chains (Figure IV-7). Other variable residues if not directly involved could act to modulate the conformation of the amino acids in the site. Affinity labelling studies (Fleet et al, 1969) and X-ray crystallographic studies of Fab fragments (Terry et al, 1968; Rossie & Nisonoff, 1968; Avey et al, 1968) will help to pinpoint the combining site.

V. Evolutionary Patterns of Specificity Regions

The question of how V region heterogeneity comes about has yet to be answered and a most lively debate is being waged in the literature. Essentially two opposing theories have been proposed. The first suggests that the enormous number of specificity regions are carried in the germ line; the heterogeneity of the V regions reflects the heterogeneity of the genome. Talmage (1965), Dreyer and Bennett (1965), Hood et al (1967) and Putnam et al (1967) have expressed the

Figure IV-7. Schematic representation of an antibody combining site including parts of the variable regions of both heavy and light chains.



view that V regions show a pattern of evolutionary homology analogous to that seen when other polypeptide chains are compared e.g. haemoglobin; that is, a model based on gene duplication and subsequent mutation.

The conventional model for the evolution of proteins is summarized by Jukes (1969). This model states that base replacements in DNA occur continually. They are distributed on a random basis (Muller, 1947). The majority of replacements give rise to single amino acid substitutions in proteins which are screened by natural selection, the deleterious ones disappear, the rare beneficial ones are rapidly incorporated into the genome of the species and a small proportion of the neutral ones are also incorporated into the genome (Kimura, 1968; King & Jukes, 1969; Laird et al, 1969). Immunoglobulins may differ from other proteins because a large number of mutational changes may be adaptive or beneficial rather than neutral or harmful. It is immunologically advantageous to have a large available assortment of different antibodies to cope with various antigens. Jukes (1969) has examined the replacements in immunoglobulins, haemoglobins, and cytochrome c's form randomness by means of the Poisson Distribution excluding invariant residues thought to be essential for the functioning of the protein. If

eight residues are excluded from the calculations the distribution of changes in the V regions of light chains is random. He also found that K/λ light chain comparisons between mouse/human and human/human are the same for both V regions and C regions (55-60% identity). A similar comparison of α and β haemoglobin chains falls in the same range (53, 54%). Therefore, by genetic analysis the immunoglobins have been evolving in the normal way, at least in a way analogous to that of the haemoglobins.

Most perplexing is the great multiplicity of homologous V regions. Considering the number of light chains which have been examined by sequencing and by peptide maps, and the failure to find two which are identical, there must be more than 10^4 V-genes in the genome (Quattrochi et al, 1969). Britten and Kohne (1968) have discovered by DNA-hybridization experiments that more than $\frac{1}{3}$ of the DNA of higher organisms is made up of sequences repeated from 1000-10^6 times, resulting from "saltatory replications". For these duplications to be integrated into chromosomes and dissembled throughout the species is probably a rare event and may have to do with the evolution of the species.

If many DNA sequences are similar to each other and adjacent on chromosomes, high rates of unequal crossing over might be expected to occur. There is evidence for this but it is not considered common (Bridges, 1935; Thomas, 1966). Precise control mechanisms would be needed to protect higher organisms from the lethal effects that families of repeated sequences might induce. It is known that there are 100's and 1000's of similar ribosomal genes. It is suspected that repeated sequences occur in structural genes and could easily form part of the immune system.

All known light chain sequences can be arranged into an evolutionary tree (Hood & Talmage, 1969). Each branching involves a duplication and subsequent mutational event. The older or more established the mutation is the more proteins in which it will appear (Figure I-5: λ I-Ala, Thr-13); the more recent the mutation, the more unique (Figure I-5: II Vil-His-1). Furthermore, in order to obtain maximum homology between members of subgroups it is necessary in every case to make deletions and insertions. These facts can be easily understood if there are a multiplicity of V genes in the germ line.

There are, however, a number of observations which cannot be

satisfactorily explained by the germ line theory and another proposal has been put forward. It is suggested that there are a limited number of V genes, perhaps one per subgroup, carried in the genome and that some sort of mechanism acts on them during cellular differentiation to produce the myriad of sequences that we find. The hypothesis calls for mechanisms such as intragenic crossing over prior to transcription (Smithies, 1963, 1967), recombination of a small number of tandemly duplicated genes (Edelman & Gally, 1967) and introduction of errors into DNA during repair (Brenner & Milstein, 1967) after which a cell would be fixed in its ability to produce one particular sequence. The latter proposal would predict some pattern to the heterogeneity which is not found. Milstein and Pink (1969) have combined the first two suggestions to explain many observations on V region heterogeneity. There is almost certainly a number of genes coding for each set of V regions. They suggest that for each cistron there is an allelic gene pair between which crossing over occurs. The repeated variance would then be explained if the alleles differed from each other at certain positions. However, one should find evidence of linked groups of alternatives. This does not appear to be the case,

at least not in the N-terminal 24 residues of light chains.

Although there is evidence for similar subgroups in light chains of human, mouse and guinea pig, the rabbit has different predominating sequences. One would then postulate that there must have been simpler ancestral V regions at some point along the way from which the V regions of higher species evolved and proliferated. However, heterogeneity seems to have been the earmark of the immune response from the beginning. Although there has been reason to think that light chain banding may in part be artifactual (Melchers et al, 1966) the light chain starch gel spread in lampreys is comparable to that found in mammals (Marchalonis & Edelman, 1968). If one wishes to support the germ line theory one has to postulate that as each species evolved, a few selected V region genes were selected in face of the vast number performing an antibody function.

Furthermore, there are possible species specific residues in human and mouse K chains which can be explained most easily by a simple inheritance. All human K chains that have so far been sufficiently sequenced (4K-I, 2K-II, 4K-III) have exclusively Leu-33, Tyr-36 and Thr-72. The two mouse K Bence-Jones, M41 and M70, have Ser /Met-32, Leu /Phe-36 and Ser-72. They become identical at

position 95 which no two human K chains do. In fact, human K position 96 is intensely variable. The simplest explanation is that the basic K subgroups have developed by duplication and mutation of a single sequence in each species.

The problem of the rabbit allotypes is similar. The "a" locus markers are found on the Fd of heavy chains. They are inherited in a simple Mendelian fashion. Approximately eight residues scattered throughout the N-terminal 35 residues may be involved in "a" locus allotypy (Wilkinson, 1969a). It is difficult to understand why these residues should be maintained in an unchanged state in multiple copies of the rabbit V-gene unless they fulfilled some critical structural role which is not immediately apparent. If one assumes that polymorphism took place after the formation of the alleles, one must propose a concomitant duplication of all three alleles; if one assumes that polymorphism took place before allelism, then similar allelic mutations must have occurred in each polymorph.

Moreover, the presence of idiotypic determinants lends support to an individual diversifying mechanism. Idiotypy is unique for each animal and for each sort of antibody produced by that animal. It is based somehow on the V region sequences. This is supported by the finding of similar idiotypic determinants on rabbit IgG and IgM (Oudin & Michel, 1969). These workers failed to find any pattern of inheritance of idiospecificities

from parent to offspring. This presents a very strong argument for individual generation of specificities.

In conclusion, the presence of species specific residues suggests that subgroups of V regions and their members and subclasses of C regions are very recent and perhaps coincident duplications. There appears to be nothing unusual about the sort and percentage of mutations occurring. Very little is known about the method or extent of biochemical mutations. Perhaps in a system where heterogeneity is an advantage, certain controls and selections pressures are minimal. It is calculated that fibrinopeptides are mutating 30 times faster than Cytochromes C (McLaughlin & Dayhoff, 1969). Immunoglobulins may be mutating several times faster than fibrinopeptides. The necessity in the germ line theory for having this multiple generation of V regions so many times during evolution, i.e. for each species that has been studied - makes the system much more explicable if a limited number of prototypes are inherited and acted upon somatically in a way which mimics the sort of mutations we have come to consider as being associated with evolutionary time. After examining the idiotypic profiles of two successive bleedings on one rabbit, Oudin and Michel (1969) have made the observation that the idiopattern carried by one sort of molecule in the first bleeding was carried by two sorts of molecules in the second bleeding. Could this be evidence for somatic mutation and differentiation in progress?

REFERENCES

Abel, C. A. & Grey, H. M. (1967). Science 156: 1609.

Abel, C. A. & Grey, H. M. (1968). Biochemistry 7: 2682

Ambler, R. P. (1965). Biochem. J. 96: 32p.

Appella, E., Mage, R., Dubiski, S. & Reisfeld, R. A. (1968). Proc. Nat. Acad. Sci. 60: 975.

Appella, E. & Perham, R. N. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 37.

Appella, E., Rejnek, J. & Reisfeld, R. A. (1969). J. Mol. Biol. 41: 473.

Ashman, R. F. & Metzger, H. (1967). J. Biol. Chem. 242: 3405.

Atfield, G. N. & Morris, C. J. O. R. (1961). Biochem. J. 81: 606.

Avey, H. P., Poljak, R. J., Rossi, G. & Nisonoff, A. (1968). Nature 220: 1248.

Baglioni, C. (1967). Biochem. Biophys. Res. Commun. 26: 82.

Baglioni, C., Alescio, Z. L., Cioli, D. & Carbonara, A. (1966). Science 152: 1519.

Beale, D. & Feinstein, A. (1969). Biochem. J. 112: 187.

Beavan, G. H. & Holiday, E. R. (1952). Adv. Protein Chem. 7: 319.

Bennett, J. C. (1968). Biochemistry 7: 3340.

Bennett, J. C., Hood, L., Dreyer, W. J. & Potter, M. (1965). J. Mol. Biol. 12: 81.

Boulton, A. A. & Bush, I. E. (1964). Biochem. J. 92: 11p.

Brenner, S. & Milstein, C. (1966). Nature 211: 242.

Bridges, C. B. (1935). J. Heredity 26: 60.

Britten, R. J. & Kohne, D. E. (1968). Science 161: 529.

Bull. Wld. Hlth. Org. (1965). 33: 721.

Bull. Wld. Hlth. Org. (1969). In press.

Cairns, J. (1963). Cold Spring Harbor Symp. Quant. Biol. 28: 43.

Campbell, A. (1962). Adv. in Genetics 11: 101.

Cebra, J. J., Givol, D. & Porter, R. R. (1968). Biochem. J. 107: 69.

Cebra, J. J. & Small, P. A. (1967). Biochemistry 6: 503.

Chesebro, B., Bloth, B. & Svehag, S. E. (1968). J. Exp. Med. 127: 399.

Cohn, M., Notani, G. & Rice, S. A. (1969). Immunochemistry 6: 111.

Conway, T. P., Dray, S. & Lichter, E. A. (1969). J. Immunol. 102: 544.

Cooper, A. G. (1967). Science 157: 933.

Crowshaw, K., Jessup, S. J. & Ramwell, P. W. (1967). Biochem. J. 103: 79.

Cunningham, B. A., Gottlieb, P. D., Konisberg, W. H. & Edelman, G. M. (1968). Biochemistry 7: 1983.

Dayhoff, M. O., Sochard, M. R. & McLaughlin, P. J. (1969). Atlas of Protein Sequence and Structure 4: 25.

Doolittle, R. F., Singer, S. J. & Metzger, H. (1966). Science 154: 1561.

Dray, S., Dubiski, S., Kelus, A., Lennox, E. S. & Oudin, J. (1962). Nature, 195: 785.

Dreyer, W. J. & Bennett, J. C. (1965). Proc. Nat. Acad. Sci. 54: 864.

Dreyer, W. J., Gray, W. R. & Hood, L. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 353.

Dubiski, S. (1969). J. Immunol. 103: 120.

Dubiski, S. & Muller, P. J. (1967). Nature 214: 696.

Easely, C. W. (1965). Biochim. Biophys. Acta 107: 386.

Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U. & Waxdal, M. J. (1969). Proc. Nat. Acad. Sci. 63: 78.

Edelman, G. M. & Gally, J. A. (1967). Proc. Nat. Acad. Sci. 57: 353.

Edelman, G. M. & ~~Poulik~~ Gally J. A., ~~M. D.~~ (1962). J. Exp. Med. 116: 207.

Edman, P. & Cooper, A. G. (1968). FEBS Letters 2: 33.

Ein, D. (1968). Proc. Nat. Acad. Sci. 60: 982.

Eisen, H. N., Little, J. R., Osterland, C. K. & Simms, E. S. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 75.

Fahey, J. L., Wunderlich, J. & Mishell, R. (1964). J. Exp. Med. 120: 223.

Feinstein, A. (1963). Nature 199: 1197.

Feinstein, A. & Franklin, E. C. (1966). Nature 212: 1496.

Feinstein, A. & Rowe, A. J. (1965). Nature 205: 147.

Fisher, C. E., Palm, W. H. & Press, E. M. (1969). FEBS Letters 5: 20.

Fleet, G. W. J., Porter, R. R. & Knowles, J. R. (1969). Nature 224: 511.

Fleischman, J. B. (1967). Biochemistry 6: 1319.

Franek, F. (1967). Nobel Symposium 3: Gamma Globulins. p. 221 Ed. Killander, J.: Interscience Publishers.

Frangione, B. (1969). FEBS Letters 3: 341.

Frangione, B. & Milstein, C. (1967). Nature 216: 939.

Frangione, B. & Milstein, C. (1969). Nature 224: 597.

Frangione, B. & Milstein, C. & Pink, J. R. L. (1969). Nature 221: 145.

Frangione, B., Prelli, F. & Franklin, E. C. (1967). Immunochemistry 4: 95.

Freedman, M. H. & Sela, M. (1966). J. Biol. Chem. 241: 5225.

Fruchter, R. G., Jackson, S. A., Mole, L. E. & Porter, R. R. (1969). Biochem. J. In Press.

Givol, D. & Porter, R. R. (1965). Biochem. J. 97: 32p.

Good, A. H. Ovary, Z. & Singer, S. J. (1968). Biochemistry 7: 1304.

Gottlieb, P. D., Cunningham, B. A., Waxdal, M. J., Konigsberg, W. H. & Edelman, G. M. (1968). Proc. Nat. Acad. Sci. 61: 168.

Gray, W. R. (1967). In Methods in Enzymology. Vol. 11. Ed. by Colowick, S. P. & Kaplan, N. O. New York: Academic Press Inc.

Grey, H. M., Abel, C. A., Yount, W. J. & Kunkel, H. G. (1968). J. Exp. Med. 128: 1223.

Grey, H. M. & Kunkel, H. G. (1964). J. Exp. Med. 120: 253.

Grey, H. M. & Mannik, M. (1965). J. Exp. Med. 122: 619.

Haber, E. (1964). Proc. Nat. Acad. Sci. 52: 1099.

Hamers, R. & Hamers-Casterman, C. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 129.

Hartley, B. S. & Massey, V. (1956). Biochim. Biophys. Acta 21: 58.

Heilmann, J., Barollier, J. & Watzke, E. (1957). Hoppe Seyler's Z. Physiol. Chem. 309: 219.

Heremans, J. F. (1959). Clin. Chim. Acta 4: 96.

Herzenberg, L. A. (1964). Cold Spring Harbor Symp. Quant. Biol. 29: 455.

Herzenberg, L. A., McDevitt, H. O. & Herzenberg, L. A. (1968). Ann. Rev. Genetics 2: 209.

Hilschmann, N. (195). Hoppe-Seyler's Z. Physiol. Chem. 348: 1077.

Hilschmann, N. & Craig, L. C. (1965). Proc. Nat. Acad. Sci. 53: 1403.

Hood, L. & Ein, D. (1968). Nature 220: 764.

Hood, L., Gray, W. R., Sanders, B. G. & Dreyer, W. J. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 133.

Hood, L. & Talmage, D. W. (1969). Symposium on Developmental Aspects of Antibody Formation and Structure, Prague-Czechoslovak Academy of Sciences in Prague. In press.

Hurlimann, J., Waldesbuhl, M. & Zuber, C. (1969). Biochim. Biophys. Acta 181: 393.

Ishizaka, T., Ishizaka, K., Salmon, S. & Fudenberg, H. H. (1967). J. Immunol. 99: 82.

Jaton, J.-C., Klinman, N. R., Givol, D. & Sela, M. (1968). Biochemistry 7: 4185.

Jeffris, R., Weston, P. D., Stanworth, D. R. & Clamp, J. R. (1968). Nature 219: 666.

Jukes, T. H. (1969). Biochem. Genetics 3: 109.

Kaplan, A. P. & Metzger, H. (1969). Biochemistry 8: 3944.

Katz, A. M., Dreyer, W. J. & Anfinsen, C. B. (1959). J. Biol. Chem. 234: 2897.

Kelus, A. S. & Gell, P. G. H. (1967). Progress In Allergy 11: 141.

Kelus, A. S. & Gell, P. G. H. (1968). J. Exp. Med. 127: 215.

Kimura, M. (1968). Nature 217: 624.

King, J. L. & Jukes, T. H. (1969). Science. In press.

Knopf, P. M., Parkhouse, R. M. E. & Lennox, E. S. (1967). Proc. Nat. Acad. Sci. 58: 2288.

Kostka, V. & Carpenter, F. H. (1964). J. Biol. Chem. 239: 1799.

Kunkel, H. G. (1963). Harvey Lectures, Ser. 59: 219.

Kunkel, H. G., Allen, J. C. & Grey, H. M. (1964). Cold Spring Harbor Symp. Quant. Biology 29: 443.

Kunkel, H. G., Mannick, M. & Williams, R. C. (1963). Science 140: 1218.

Kunkel, H. G., Natvig, J. B. & Joslin, F. G. (1969). Proc. Nat. Acad. Sci. 62: 144.

Kunkel, H. G. & Prendergast, R. A. (1966). Proc. Soc. Exp. Biol. Med. 122: 910.

Kunkel, H. G., Yount, W. J. & Litwin, S. D. (1966). Science 154: 1041.

Laird, C. D., McConaughy, B. L. & McCarthy, B. J. (1969). Nature 224: 149.

Lamm, M. E. & Small, P. A. (1966). Biochemistry 5: 267.

Langer, B., Steinmetz-Kayne, M. & Hilschmann, N. (1968). Hoppe-Seyler's Z. Physiol. Chem. 349: 945.

Lawton, A. R. & Mage, R. G. (1969). J. Immunol. 102: 693.

Lennox, E. S. & Cohn, M. (1967). Ann. Rev. Biochem. 36: 365.

Lieberman, R. & Potter, M. (1969). J. Exp. Med. 130: 519.

Litwin, S. D. & Kunkel, H. G. (1967). J. Exp. Med. 125: 847.

Mage, R. G., Reisfeld, R. A. & Dray, S. (1966). Immunochemistry 3: 299.

Mage, R. G., Young, G. O., Rejnek, J., Reisfeld, R. A. & Appella, E. (1969). 17th Colloq. on Protides of the Biological Fluids, Bruges. In press.

Marchalonis, J. J. & Edelman, G. M. (1968). J. Exp. Med. 127: 891.

McLaughlin, P. J. & Dayhoff, M. O. (1969). Atlas of Protein Sequence and Structure 4: 39.

Melchers, F., Lennox, E. S. & Facon, M. (1966). Biochem. Biophys. Res. Comm. 243: 386.

Merler, E., Karlin, L. & Matsumoto, S. (1968). J. Biol. Chem. 243: 386.

Metzger, H. (1967). Proc. Nat. Acad. Sci. 570: 1490.

Metzger, H. & Potter, M. (1968). Science 162: 1398.

Mihaesco, C. & Seligmann, M. (1968). J. Exp. Med. 127: 431.

Miller, F. & Metzger, H. (1965). J. Biol. Chem. 240: 4740.

Milstein, C. (1964). J. Mol. Biol. 9: 836.

Milstein, C. (1966a). Nature 209: 370.

Milstein, C. (1966b). Biochem. J. 101: 352.

Milstein, C. (1967). Nature 216: 330.

Milstein, C. (1969). FEBS Letters 2: 301.

Milstein, C. (1968). 5th FEBS Meeting, Prague, Symp. III: Academic Press, N. Y. In press.

Milstein, C., Clegg, J. B. & Jarvis, J. M. (1968). Biochem. J. 110: 63.

Milstein, C., Frangione, B. & Pink, J. R. L. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 63.

Milstein, C., Milstein, C. P. & Feinstein, A. (1969). Nature 221: 151.

Milstein, C. & Pink, J R. L. (1969). Progress in Biophys. and Molec. Biol. In press.

Montgomery, P. C., Bello, A. C. & Rockey, J. H. (1969). Biochim. Biophys. Acta. In press.

Morse, D. & Horecker, B. L. (1966). Analyt. Biochem. 14: 429.

Morris, J. E. & Inman, F. P. (1968). Biochemistry 7: 2851.

Muir, W. G. & Steinberg, A. G. (1967). In Seminars in Hematology. 4: 156. Ed. Miescher, P. A.: Grune and Shatton, N. Y.

Muller, H. J. (1947). Proc. Roy. Soc. (B) 134: 1.

Natvig, J. B., Kunkel, H. G. & Joslin, F. B. (1969). J. Immunol. 102: 611.

Natvig, J. B., Kunkel, H. G. & Litwin, S. P. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 173.

Niall, H. D. & Edman, P. (1967). Nature 216: 262.

O'Donnell, I. J., Frangione, B. & Porter, R. R. (1969). Biochem. J. In press.

Offord, R. E. (1966). Nature 211: 591.

Onoue, K., Grossberg, A. L., Yagi, Y. & Pressman, D. (1968). Science 162: 574.

Onoue, K. Kishimoto, T. & Yamamura, Y. (1968). J. Immunol. 100: 238.

Onoue, K., Yagi, Y., Grossberg, A. L. & Pressman, D. (1965). Immunochemistry 2: 401.

Oudin, J. (1960). J. Exp. Med. 112: 107, 125.

Oudin, J. (1966). Proc. R. Soc. (B) 166: 207.

Oudin, J. & Michel, M. (1963). C. R. Acad. Sci., Paris 257: 805.

Oudin, J. & Michel, M. (1969). J. Exp. Med. 130: 619.

Palmer, J. L. & Nisonoff, A. (1964). Biochemistry 3: 863.

Pan, S. C. & Deutscher, J. D. (1956). Analyt. Chem. 28: 836.

Permis, B., Torrigiani, G., Amante, L., Kelus, A. S. & Cebra, J. J. (1968). Immunology 14: 445.

Piggot, P. J. (1966). Ph. D. Thesis, University of London, p. 46.

Pink, J. R. L. & Milstein, C. (1968). 5th EEBS Meeting, Prague, Symp. III: Academic Press, N. Y. In press.

Pollara, B., Suran, A., Finstad, J. & Good, R. A. (1968). Proc. Nat. Acad. Sci. 59: 1307.

Ponstingl, H., Hess, M. & Hilschmann, N. (1968). Hoppe-Seyler's Z. Physiol. Chem. 349: 867.

Ponstingl, H., Hess, M., Langer, B., Steinmetz-Kayne, M. & Hilschmann, N. (1967). Hoppe-Seyler's Z. Physiol. Chem. 348: 1213.

Ponstingl, H. & Hilschmann, N. (1969). Hoppe-Seyler's Z. Physiol. Chem. 350: 1148.

Porter, R. R. (1957). In Methods in Enzymology, Vol. 4: 221. Ed. Colowick, S. P. & Kaplan, N. O. N.Y.: Academic Press Inc.

Porter, R. R. (1962). In Basic Problems of Neoplastic Disease, p. 177. Ed. Gellhorn, A. & Hirschberg, E.: Columbia Univ. Press.

Potter, M., Dreyer, W. J., Kuff, E. L. & McIntire, K. R. (1964). J. Mol. Biol. 8: 814.

Potter, M. & Lieberman, R. (1967). Advan. Immunol. 7: 92.

Prahl, J. W. & Grey, H. M. (1969). Pacific Slopes Biochem. Conf. In press.

Prahl, J. W., Mandy, W. J., David, G. S., Steward, M. W. & Todd, C. W. (1969). 17th Colloq. on Protides of the Biol. Fluids, Bruges In press.

Press, E. M., Givol, D., Piggot, P. J., Porter, R. R. & Wilkinson, J. M. (1966). Proc. R. Soc. (B) 166: 150.

Press, E. M & Hogg, N. M. (1969a). Nature 223: 807.

Press, E. M. & Hogg, N. M. (1969b). Biochem. J. In press.

Press, E. M. & Piggot, P. J. (1969). Cold Spring Harbor Symp. Quant. Biol. 32: 45.

Press, E. M., Piggot, P. J. & Porter, R. R. (1966). Biochem. J. 99: 356.

Putnam, F. W. (1969). Science, 163: 633.

Putnam, F. W. & Easely, C. W. (1965). J. Biol. Chem. 240: 1626.

Putnam, F. W., Shinoda, T., Titani, K. & Wikler, M. (1967). Science 157: 1050.

Putnam, F. W., Titani, K., Wikler, M. & Shinoda, T. (1967). Cold Spring Harbor Symp. Quant. Biol. 32: 9.

Quattrochi, R., Cioli, D. & Baglioni, C. (1969). J. Exp. Med. 130: 401.

Romer, A. S. (1966). In Vertebrate Paleontology: 3rd Ed. Univ. Chicago Press, Chicago.

Rossi, G. & Nisonoff, A. (1968). Biochem. Biophys. Res. Comm. 31: 914.

Sanger, F. & Thompson, E. O. P. (1963). Biochim. Biophys. Acta 71: 468.

Seiler, N. & Wiechmann, J. (1964). Experientia 20: 559.

Seligmann, M., Danon, F., Basch, A. & Bernard, J. (1968). Nature 220: 711.

Singer, S. J. & Thorpe, N. O. (1968). Proc. Nat. Acad. Sci. 60: 1371.

Smith, H. O. & Levine, M. (1967). Virology 31: 207.

Smith, I. (1950). In Chromatographic and Electrophoretic Techniques. Vol. 1: 96, 97. Med. Books Ltd. Wlm. Heinemann, London.

Smithies, O. (1967a). Cold Spring Harbor Symp. Quant. Biol. 32: 161.

Smithies, O. (1967b). Science 157: 267.

Spackman, D. H., Moore, S. & Stein, W. H. (1958). Analyt. Chem. 30: 1190.

Steinberg, A. G. (1969). Ann. Rev. Genetics 3: 25.

Steiner, L. A. & Porter, R. R. (1967). Biochemistry 6: 3957.

Stone, M. J. & Metzger, H. (1968). J. Biol. Chem. 243: 5977.

Suran, A. A. & Papermaster, B. W. (1967). Proc. Nat. Acad. Sci. 58: 1619.

Suter, L., Barnikol, H., Watanabe, S. & Hilschmann, N. (1969). Hoppe-Seyler's Z. Physiol. Chem. 350: 275.

Suzuki, T. (1969). Immunochemistry 6: 587.

Suzuki, T. & Deutsch, H. F. (1967). J. Biol. Chem. 242: 2725.

Svehag, S. E., Bloth, B. & Seligmann, M. (1969). J. Exp. Med. 130: 691.

Talmage, D. W. (1965). Science 150: 1484.

Terry, W. D. (1965). J. Immunol. 95: 1041.

Terry, W. D. & Fahey, J. L. (1964). Science 146: 400.

Terry, W. D., Fahey, J. L. & Steinberg, A. G. (1965). J. Exp. Med. 122: 1087.

Terry, W. D., Matthews, B. W. & Davies, R. R. (1968). Nature 220: 239.

Thomas, C. A. (1966). Progr. Nuc. Acid Res. Mol. Biol. 5: 315.

Thompson, A. R. & Miles, B. J. (1964). Nature 203, 483.

Thorpe, N. D. & Deutsch, H. F. (1966). Immunochemistry 3; 329.

Tischendorf, F. W. & Osserman, E. F. (1969). J. Immunol. 102: 172.

Titani, K., Whitley, E., Avogardo, L. & Putnam, F. W. (1965). Science 149: 1091.

Titani, K., Whitley, E. J. & Putnam, F. W. (1969). J. Biol. Chem. 244: 3521.

Todd, C. W. (1963). Biochem. Biophys. Res. Comm. 11: 170.

Todd, C. W. & Inman, F. P. (1967). Immunochemistry 4: 407.

Tomasi, T. B. (1965). J. Exp. Med. 121: 101.

Tourville, D. R., Adler, R. H., Bienenstock, J. & Tomasi, T. B. (1969). J. Exp. Med. 129: 411.

Vaerman, J. -P. & Heremans, J. F. (1966). Science 153: 647.

Vaerman, J. -P., Heremans, J. F. & Laurell, C. B. (1968). Immunology 14: 425.

Valentine, R. C. & Green, N. M. (1967). J. Mol. Biol. 27: 615.

Waldenstrom, J., Winblad, S., Hallen, J. & Liungman, S. (1964). Acta Med. Scand. 176: 619.

Wang, H. C. & Fudenberg, H. H. (1969). J. Mol. Biol. 44: 493.

Wang, H. C., Shuster, J. & Fudenberg, H. H. (1969). J. Mol. Biol. 41: 83.

Warner, N. L. & Herzenberg, L. A. (1966). J. Immunol. 97: 525.

Wikler, M., Kohler, H., Shinoda, T. & Putnam, F. W. (1969). Science 163: 75.

Wilkinson, J. M. (1969a). Biochem. J. 112: 173.

Wilkinson, J. M. (1969b). Nature, 223: 616.

Witney, P. L. & Tanford, C. (1965). J. Biol. Chem. 240: 4271.

Woods, K. R. & Wang, K. T. (1967). Biochim. Biophys. Acta 133: 369.

Zottervall, O., Sjoquist, J., Waldenstrom, J. & Winblad, S. (1966). Clin. Exp. Immunol. 1: 213.

Zullo, D. M., Todd, C. W. & Mandy, W. J. (1968). Proc. Can. Fed. Biol. Soc. 11: 111.

(Reprinted from Nature, Vol. 223, No. 5208, pp. 807-810, August 23 1969)

Comparative Study of Two Immunoglobulin G Fd-Fragments

by
E. M. PRESS
N. M. HOGG
MRC Immunochemistry Unit,
Department of Biochemistry,
University of Oxford

The sequence of two γ1 heavy chains has revealed the extent of the variable region and the existence of subgroups.

THE immunoglobulins all have a four-chain structure, consisting of two heavy chains and two light chains. Immunoglobulin G is the predominant class and antibody activity is associated with a fragment, the Fab-fragment, consisting of the light chain and the N-terminal half of the heavy chain, the Fd-fragment[1]. Antibodies are chemically very heterogeneous and it is assumed that part of this heterogeneity is related to their antigen binding specificity. In order to find out the relationship between specificity and chemical structure, attempts have been made to determine the amino-acid sequence of the constituent light and heavy chains. All antibodies are, however, mixtures of molecules, and sequence studies are therefore very difficult, but considerable information on the extent of the variations in sequence has been obtained by studying the homogeneous proteins produced in large amounts by patients having myelomatosis or related disorders. It has been shown that the sequence of the N-terminal half of the light chain varies from one chain to another, but that, within a class, the C-terminal sequence is the same[2,3].

Comparison between DAW and COR Fd-fragments

In order to locate the extent of the variable and constant amino-acid sequences in heavy chains, a partial sequence of two pathological γ1 chains, both Gm a^+z^+, has been determined. The sequence of DAW heavy chain from 1–84 has already been reported[4] and the sequence of this heavy chain has now been extended to residue 225 (the C-terminal residue of the Fd-fragment). Preliminary work on COR γ1 chain has also been reported[4] and this γ1

DAW PCA- Val- Thr- Leu- Arg- Glu- Ser- Gly- Pro- (10) Ala- Leu- Val- Arg- Pro- Thr- Gln- Thr- Leu- Thr- (20) Leu- Thr- Cys- Thr- Phe- Ser- Gly-
COR PCA- Val- Thr- (Lys, Arg, Thr_2, Ser_5, Glx_2, Pro_2, Gly_3, Ala, Val, Cys, Leu_2, Phe_2)

DAW Phe- Ser- Leu- (30) Ser- Gly- Glu- Thr- Met- Cys- Val- Ala- Trp- Ile- (40) Arg- Gln- Pro- Pro- Gly- Glu- Ala- Leu- Glu- Trp- (50) Leu- Ala- Trp-
COR Met- Cys- Val- Gly- Trp- Ile- Arg- Gln- Pro- Pro- Gly- Lys- Gly- Leu- Glu- Trp- Leu- Ala- Arg-

DAW Asp- Ile- Leu- Asn- Asp- Asp- Lys- (60) Tyr- Tyr- Gly- Ala- Ser- Leu- Glu- Thr- Arg- Leu- (70) Ala- Val- Ser- Lys- Asp- Thr- Ser- Lys- Asn-
COR Ile- Asx- Trp- Asp- Asp- Asp- Lys- Tyr- Tyr- Asn- Thr- Ser- Leu- Glu- Thr- Arg- Leu- Thr- Ile- Ser- Lys- Asp- Thr- Ser- Arg- Asn-

DAW Gln- (80) Val- Val- Leu- Ser- Met- Asn- Thr- Val- Gly- Pro- (90) Gly- Asp- Thr- Ala- Thr- Tyr- Tyr- Cys- Ala- Arg- (100) Ser- Cys- Gly- Ser- Gln
COR Gln- Val- Val- Leu- Thr- Met- Asp- Pro- Val- — — — Asp- Thr- Ala- Thr- Tyr- Tyr- Cys- Ala- Arg- Ile- Thr- Val- Ile- Pro-

DAW — — — — Tyr- Phe- Asp- Tyr- Trp- (110) Gly- Gln- Gly- Ile- Leu (115) Val- Thr- Val- Ser- Ser- (120) Ala- Ser- Thr- Lys- Gly- Pro- Ser-
COR Ala- Pro- Ala- Gly- Tyr- Met- Asp- Val- Trp- Gly- Arg- Gly- Thr- Pro

DAW/COR Val- Phe- Pro- (130) Leu- Ala- Pro- Ser- Ser- Lys- Ser- Thr- Ser- Gly- (140) Gly- Thr- Ala- Ala- Leu- Gly- Cys- Leu- Val- Lys- (150) Asp- Tyr- Phe-

DAW/COR Pro- Glu- Pro- Val- Thr- Val- Ser- (160) Trp. Asn- Ser- Gly- Ala- Leu. Thr- Ser- Gly- Val- (170) His. Thr- Phe- Pro- Ala- Val- Leu. Gln- Ser-

DAW/COR Ser- (180) Gly- Leu- Tyr. Ser- Leu- Ser- Ser- Val- Val- Thr- (190) Val- Pro- Ser- Ser- Leu- Gly- Thr- Gln- Thr- Tyr. (200) Ile- Cys- Asn- Val- Asn-

DAW/COR His (Lys, Pro, Ser) Asn- (210) Thr- Lys- Val- Asp- Lys- Lys- Val- Glu- Pro- Lys- (220) Ser- Cys- Asp- Lys- Thr- (225) His

Fig. 1. Comparison of amino-acid sequence of two γ1 Fd-fragments, DAW and COR. The long dashes indicate deletions; the full stops indicate that overlapping peptides at these points have not been isolated. Identical residues are in bold type. PCA stands for pyrrolid-2-one-5-carboxylic acid.

chain has now been partially sequenced from residue 1 to 225. These two sequences are compared in Fig. 1. There are variations in about 30 per cent of the positions up to residue 114; thereafter the sequence is identical for the two Fd-fragments. The stops after residues 160, 165, 170, 176, 182 and 199, indicate that these five chymotryptic peptides have not been overlapped but they have been arranged as shown, because of the similarity in sequence to rabbit γ1 chain in this region (unpublished results of R. G. Fruchter and S. Jackson). This arrangement is in agreement with the complete sequence of another γ1 heavy chain, EU[5].

It had been deduced, from a comparison of DAW and COR Fd-fragments, that the constant region starts at residue 115, but another γ1 heavy chain, HE, has an alanine at residue 116 (ref. 5) instead of threonine. The sequences near the beginning of the constant section in light chains and the γ1 heavy chains are very similar (see Fig. 2; identical residues in box), and when more γ1 sequences are available it may be found that the constant region actually starts at residue 120 (DAW numbering) in a position analogous to that in the light chains. In fact, the incidence of variable residues is very low just before the constant region in the light chains, particularly within subgroups of λ chains[6].

The number of positions between 34 and 99 at which COR and DAW heavy chains have different residues is 19 out of 65; that is, 71 per cent of positions have identical residues. In contrast, the number differing from 100 to 114 is 10 out of 15: only 33 per cent of positions have the same residue. It has been noted that the corresponding region of light chain is more highly variable than the N-terminal region within subgroups of κ and λ chains[6–8].

The sequence of the residues 4 to 33 of protein COR has not been determined, but the composition is so similar to that of DAW (Fig. 1) that it is most likely that the sequence of COR will show very few differences. The order of variability in sequence between DAW and COR heavy chain (71 per cent of positions identical between 34–99) is similar to that found within subgroups of κ and λ light chains, and it is highly likely that these two heavy chains belong to the same subgroup of Fd-fragment. The two chains are not, however, precisely the same length: COR has a deletion of three residues 88–90, and an insertion of four residues after 104. Deletions and insertions have also been found in light-chain sequences within subgroups, for example, in κ subgroup II CUM has a two residue insertion relative to MIL[9,10]. Also in λ chain subgroup I, HA has an insertion after 28 relative to NEW[6].

The intrachain disulphide bridges of several myeloma heavy chains have been studied and a constant feature is two intrachain bridges in the Fd-fragment[11] and two in the Fc-fragment[12]. DAW Fd-fragment, however, has three intrachain disulphide bridges. By comparing the sequences around the cysteine residues in DAW with those reported for other myeloma γ1 chains[11], we may assume that cysteine-22 is linked to 97, and 146 to 201 (Fig. 3). Presumably, therefore, cysteine-35 in DAW heavy chain is linked to cysteine-101, and this is in agreement with the observation that the fragments produced by cleavage of DAW Fd with cyanogen bromide: 2b (1–34) and 4 (35–84) are both linked by disulphide

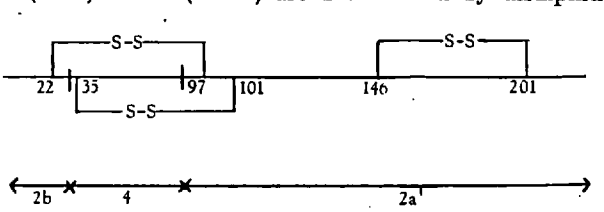


Fig. 3. The intrachain disulphide bonds of DAW Fd-fragment. The vertical bars represent the points of cleavage by cyanogen bromide, giving rise to fragments 2b, 4 and 2a¹.

bonds to fragment 2a¹ (85–225) (ref. 13). The Fd-fragment of COR has only two intrachain disulphide bonds: cysteine-22 is linked to cysteine-97 and 146 to 201. There is, however, a cysteine residue at 35, as in DAW, but not at position 101. Because no other cysteine residues were found in COR Fd-fragment, apart from those at 22, 97, 146, 201 and 221 (the interchain linkage with light chain), it must be assumed that the cysteine residue in COR at position 35 has a free SH group or is linked to a free cysteine residue, as found in a light chain[14].

Comparison with other Heavy Chains

The sequences of two other human heavy chains have been reported. A μ chain, OU, has been sequenced from 1–105 (ref. 15) and another γ1 chain, EU, from 1–446 (refs. 5, 16). The μ chain sequence is remarkably similar

							constant								
Human light chain κ							108 Arg-	Thr-	Val-	Ala-	Ala-	Pro-	Ser-	Val-	Phe
Human light chain λ							109 Gln-	Pro-	Lys-	Ala-	Ala-	Pro-	Ser-	Val-	Thr
						constant									
Human γ1 heavy chains	DAW and COR	115 Val-	Thr-	117 Val-	Ser-	Ser-	120 Ala-	Ser-	Thr-	Lys-	Gly-	Pro-	Ser-	Val-	Phe
	EU	Val-	Thr-	115 Val-	Ser-	Ser-	Ala-	Ser-	Thr-	Lys-	Gly-	Pro-	Ser-	Val-	Phe
	HE	Val-	Ala-	Val-	Ser-	Ser-	Ala-	Ser-	Thr-	Lys-	Gly-	Pro-	Ser-	Val-	Phe

Fig. 2. The N-terminal sequences of the constant regions of light and heavy chains.

to that of DAW γ1 chain, having 70 per cent of the positions between 1 and 99 occupied by the same residue, in contrast to the sequence of EU γ1 chain, which has only 29 per cent of the positions from 1–99 occupied by the same residue as in DAW. It seems probable that there are Fd-fragment subgroups as has been postulated for the variable part of the light chains. The two γ1 chains—COR and DAW and the μ chain OU—would belong to the same subgroup, and EU γ1 chain to another subgroup, these subgroups being unrelated to the classes of heavy chain (γ, μ and α) which are characterized by determinants in the Fc-fragments of the heavy chains. Another γ1 chain, STE (Gm a-f^+), has been partially sequenced in this laboratory and from its similarity to EU must belong to this second subgroup (unpublished work of C. E. Fisher). The subgroups may also be unrelated to the Gm groups: although COR and DAW were chosen for study because they belonged to the same Gm group (a^+z^+) and EU and STE heavy chains are Gm a-f^+, the μ chain does not express Gm determinants. Bennett[17] has reported the N-terminal pentapeptide sequence of 4 μ chains, which may all belong to yet a third subgroup.

Comparison of the sequences of DAW, COR and OU (Fig. 4) shows that the positions which are occupied by different residues in DAW and COR are the same ones that vary in OU, so that the invariant residues are staggered in blocks of up to nine residues, for example, 47–51 and 91–99. Although the γ1 chain, EU has only 29 per cent of positions occupied by the same residues as

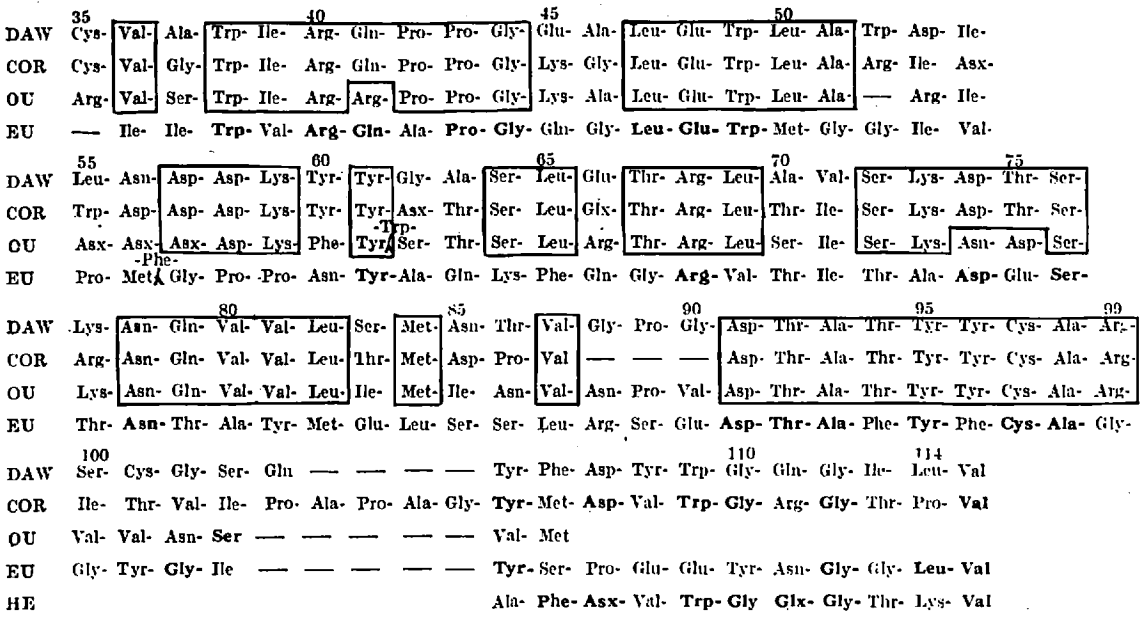


Fig. 4. A comparison between the variable region sequences of γ1 heavy chains DAW and COR and the μ chain OU (ref. 15) contrasted with the γ1 chain EU (refs. 5, 16).

in DAW, these also fall within the "invariant blocks", namely 43 and 44, and 47–49. As noted, the region just beyond cysteine-97 shows a higher incidence of variation; in fact, comparing DAW, COR, OU and EU from 100–106, the division into subgroups is not apparent (Fig. 4). The residues identical with DAW are in bold type and although COR and DAW sequences from 107–114 are more similar, γ1 chain HE is also very similar to DAW. Sequence data are, however, not available on HE from 1–99, so it is not known to which subgroup the heavy chain HE belongs. It has also been noted that the sequence at this point (100–106) in light chains, bears no relationship to subgroup[7].

The C-terminal half of the Fd-fragment 116–225 has an identical sequence in the γ1 chains DAW, COR and EU, except for a substitution of arginine for lysine at position 215 in EU compared with COR and DAW (DAW numbering). This latter difference could be related—as noted by Edelman *et al.*[5]—to the difference in Gm group, for the Gm f^+ (EU) and Gm z^+ (DAW and COR) antigenic determinants are in the Fab-fragments.

In conclusion, the comparative sequence studies on heavy chains reveal the existence of subgroups of variable regions, as has been noted for light chains. In contrast to the light chains, however, the same Fd subgroup may occur in heavy chains of different classes; whereas, the κ and λ chain subgroups are peculiar to the class of light chain. Thus the light chain classes, κ and λ, and the subgroups of their variable region, appear to have evolved in a different way from the heavy-chain classes and subgroups. The existence of subgroups of the variable part of the heavy chain, common to two γ1 chains and one μ chain, gives further support to the view that at least two genes are concerned in the synthesis of the heavy chain. It is possible that the gene controlling the variable sequence is able to translocate to either of the genes controlling synthesis of a γ chain or a μ chain[18].

The antigen binding site is presumably within the variable region of the Fab fragment, and it may involve the highly variable region between residues 100–106 referred to above, but other techniques, such as affinity labelling, will have to be used to solve the problem.

We thank Dr G. M. Edelman for a copy of the preprint of his paper before publication, Professor R. R. Porter for his enthusiastic support and guidance throughout the course of this work, Dr V. Wynn and Dr H. G. Kunkel for the generous supply of pathological sera, the Medical Research Council for their financial support, Miss C. Poole for technical assistance, and Mr T. Gascoyne for the operation of the amino-acid analysers. N. M. H. acknowledges a studentship from the Medical Research Council of Canada.

Received April 29, 1969.

[1] Fleischman, J. B., Porter, R. R., and Press, E. M., *Biochem. J.*, **88**, 220 (1963).
[2] Hilschmann, N., and Craig, L. C., *Proc. US Nat. Acad. Sci.*, **53**, 1403 (1965).
[3] Titani, K., Whitley, E., Avogardo, L., and Putnam, F. W., *Science*, **149**, 1091 (1965).
[4] Press, E. M., and Piggot, P. J., *Cold Spring Harbor Symp. Quant. Biol.*, **32**, 45 (1967).
[5] Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U., and Waxdal, M. J., *Proc. US Nat. Acad. Sci.* (In the press, 1969).
[6] Langer, Von B., Kayne, M. S., and Hilschmann, N., *Z. Physiol. Chem.*, **349**, 945 (1968).
[7] Milstein, C., *Nature*, **216**, 330 (1967).
[8] Putnam, F. W., *Science*, **163**, 633 (1969).
[9] Hilschmann, N., *Z. Physiol. Chem.*, **348**, 1077 (1967).
[10] Dreyer, W. J., Gray, W. R., and Hood, L., *Cold Spring Harbor Symp. Quant. Biol.*, **32**, 353 (1967).
[11] Frangione, B., and Milstein, C., *Nature*, **216**, 939 (1967).
[12] Frangione, B., Milstein, C., and Franklin, E. C., *Biochem. J.*, **106**, 15 (1968).
[13] Piggot, P. J., and Press, E. M., *Biochem. J.*, **104**, 616 (1967).
[14] Milstein, C., Clegg, J. B., and Jarvis, J. M., *Biochem. J.*, **110**, 631 (1968).
[15] Wikler, M., Köhler, H., Shinoda, T., and Putnam, F. W., *Science*, **163**, 75 (1969).
[16] Gottlieb, P. D., Cunningham, B. A., Waxdal, M. J., Konigsberg, W. H., and Edelman, G. M., *Proc. US Nat. Acad. Sci.*, **61**, 168 (1968).
[17] Bennett, J. C., *Biochemistry*, **7**, 3340 (1968).
[18] Lennox, E. S., and Cohn, M., *Ann. Rev. Biochem.*, **36**, 365 (1967).