

# THE GENETICS, EVOLUTION AND EXPRESSION OF ANTIBODY MOLECULES

L. HOOD

Division of Biology  
California Institute of Technology  
Pasadena, California 91109

## CONTENTS

SUMMARY . . . . .	74
1. INTRODUCTION . . . . .	74
1.1 INFORMATION STORAGE . . . . .	74
1.2 INFORMATION EXPRESSION . . . . .	75
1.3 INFORMATION EVOLUTION . . . . .	76
2. THE BIOLOGY OF ANTIBODIES . . . . .	76
The immune response . . . . .	76
The immune system . . . . .	76
Antigen . . . . .	76
Antibody heterogeneity . . . . .	76
The clonal selection hypothesis . . . . .	77
3. SYSTEMS FOR THE STUDY OF ANTIBODY MOLECULES . . . . .	78
Myeloma immunoglobulins . . . . .	78
Homogeneous antibodies . . . . .	79
A caution about phenotypic selection . . . . .	79
4. THE CHEMISTRY AND GENETICS OF ANTIBODY MOLECULES . . . . .	80
4.1 THE CHEMICAL ANALYSIS OF IG POLYPEPTIDES . . . . .	80
Light and heavy chains . . . . .	80
Three antibody families . . . . .	81
Variable and constant regions . . . . .	81
4.2 THE IMMUNOGENETIC ANALYSIS OF IG POLYPEPTIDES . . . . .	83
4.2.1 Human genetic markers . . . . .	84
4.2.2 Rabbit genetic markers . . . . .	86
5. ANTIBODY DIVERSITY AND INFORMATION STORAGE . . . . .	88
5.1 THE GENEALOGIC ANALYSIS OF V REGIONS . . . . .	90
5.1.1 The significance of a genealogic tree . . . . .	90
5.1.2 A genealogic tree for the human antibody families . . . . .	93
5.1.3 A genealogic analysis of mouse Ig chains . . . . .	94
The kappa family . . . . .	94
The lambda family . . . . .	98
Hypervariable regions . . . . .	98
5.2 MESSENGER RNA-DNA HYBRIDIZATION EXPERIMENTS . . . . .	98
5.3 AN ESTIMATE OF ANTIBODY DIVERSITY BY ELECTROFOCUSING . . . . .	99
5.4 THE PROBLEM OF ANTIBODY DIVERSITY REVISITED . . . . .	100
5.4.1 Somatic hypermutation . . . . .	100
5.4.2 Normal somatic mutation . . . . .	100
5.4.3 Somatic recombination . . . . .	101
5.4.4 Germ line . . . . .	101
6. THE ORGANIZATION OF IG GENES IN ANTIBODY FAMILIES . . . . .	102
6.1 SEPARATE V AND C GENES . . . . .	102
Biological data . . . . .	102
Serological data . . . . .	103
Sequence analysis . . . . .	103

6.2	THE ALTERNATIVE - THE ONE V-C GENE MODEL . . . . .	104
6.3	A HYBRID IMMUNOGLOBULIN MOLECULE . . . . .	105
7. A	JOINING MECHANISM AND ANTIBODY EXPRESSION . . . . .	105
7.1	JOINING MODELS . . . . .	105
7.2	RANDOM VS. PROGRAMMED EXPRESSION OF GERM LINE INFORMATION . . . . .	107
7.3	V AND C JOINING - A MECHANISM FOR DIFFERENTIATION . . . . .	111
8. THE	EVOLUTION OF ANTIBODY MOLECULES . . . . .	112
8.1	HOMOLOGY UNITS AND GENE DUPLICATION . . . . .	112
8.2	FUNCTIONAL DOMAINS . . . . .	113
8.3	THE ANTIBODY COMBINING SITE AND HYPERVARIABLE REGIONS . . . . .	115
8.4	THE EVOLUTION OF C GENES . . . . .	116
8.5	THE EVOLUTION OF V GENES . . . . .	117
8.6	MECHANISMS FOR COINCIDENTAL EVOLUTION IN MULTIGENE FAMILIES . . . . .	120
8.7	A SPECIAL EVOLUTIONARY PROBLEM - RABBIT V REGION GENETIC MARKERS . . . . .	126
8.8.	THE EVOLUTION OF THE IMMUNE SYSTEM FROM CELL SURFACE RECEPTOR MOLECULES . . . . .	128
9. SYNOPSIS	. . . . .	129
LITERATURE CITED	. . . . .	131

## SUMMARY

*The vertebrate immune system shares certain general properties with other complex systems in higher organisms and, accordingly, may serve as an ideal model for studying various aspects of differentiation and evolution in eukaryotes. I will discuss the general patterns which have emerged from genetic, molecular and cellular studies of this system. These observations will be related to three aspects of the problem of "information handling" in higher organisms; namely, how the information from a complex system is stored in the genome, how it is expressed during the differentiation of individual organisms, and how it changes during the evolution of eukaryotes.*

## 1. INTRODUCTION

The vertebrate immune system is a fruitful model for studying various aspects of differentiation and evolution in higher organisms. Indeed, it serves as an ideal model for considering questions relating to information storage, information evolution and information expression in a genetically complex and multigenic system. A series of provocative suppositions have emerged from genetic, molecular and cellular studies of the immune system which relate to each of these areas of information handling. These suppositions will be outlined briefly before turning to a detailed analysis of the data and arguments on which they are based.

## 1.1 INFORMATION STORAGE

I propose to call the fundamental genetic unit which underlies antibody (Ab) production a family (Figure 1). A family is a chromosomal segment, transmitted from one generation to the next, that contains all of the structural information required for the synthesis of a particular type of Ab polypeptide chain. The antibody family consists of a series

of nucleotide sequences coding for the amino terminal region of Ab chains (V region) and separately a series of nucleotide sequences coding for the carboxy terminal region of Ab chains (C region) (Figure 1). The mammalian germ cell appears to contain at least three Ab families, each located on a different nonhomologous autosome. These families code the distinct polypeptides that comprise the antibody molecules.



FIGURE 1. A general model of an antibody family.

A detailed analysis of Ab V and C regions has led to the supposition that each is generally encoded by multiple but separate germ line genes (Figure 1). Thus the antibody families constitute three distinct multigene systems. The number and arrangement of V genes within a family is a matter of considerable controversy. Some immunologists believe that the information content of the Ab family can be expanded during the differentiation of each individual by somatic mutation or recombination of individual Ab genes. This novel view, appropriately termed the *somatic theory* of antibody diversity, suggests that all of the information necessary for constructing a creature is not present in the germ line, but rather some is generated during somatic differentiation. In contrast, other immunologists feel that all or most of the information necessary for Ab synthesis is contained in the germ line (i.e. the *germ line theory*) and that these "germ line genes" arise through ordinary chemical evolution--that is, gene duplication and mutation followed by selection. More information is available concerning the number and arrangement of C genes within a family.

## 1.2 INFORMATION EXPRESSION

Perhaps the most provocative supposition that has emerged from the study of the Ab system is the suggestion that two separate germ line genes (a V gene and a C gene - Figure 1) encode each Ab polypeptide chain and that these genes (or their products) must be joined at some level of protein synthesis. This proposition has implications that extend far beyond the mere detailed understanding of the genetic mechanisms by which antibody synthesis occurs. If this joining mechanism operates at the DNA level, it may be the means by which an antibody producing cell is committed (differentiated) to the synthesis of a single molecular species of antibody (e.g. a mechanism of information expression). It is attractive to postulate, based on evolutionary arguments, that this joining mechanism may be employed by other genetically complex systems for the determination (commitment) of their effector cells.

### 1.3 INFORMATION EVOLUTION

The chromosomal units I have termed antibody families also have unusual evolutionary properties. Certain of these evolutionary features are shared with other well documented multigene systems. This leads me to believe that a variety of distinct multigene systems will in fact share a series of common evolutionary and genetic mechanisms. The study of multigene systems and their evolutionary mechanisms is, I believe, one of the exciting new frontiers in modern genetic research.

From the above summary of the properties of Ab families, it is evident that this hypothesis entails many unusual features. I will attempt below to provide a detailed and critical analysis of the data and the arguments which have led to the formulation of the model of the Ab family. To begin, let us briefly review salient features of the biology of Ab molecules

## 2. THE BIOLOGY OF ANTIBODIES

### The immune response

All vertebrates can respond to the invasion of foreign microorganisms through the elaboration of Ab molecules. These Ab molecules are specific in that they exhibit a molecular complementarity for the particular foreign organism which initiated the response. The chemical basis for this molecular complementarity is analogous to that exhibited by enzyme molecules for their corresponding substrates. Thus each Ab molecule exhibits a molecular complementarity for a particular three-dimensional configuration (antigenic determinant) and is accordingly, said to be specific for that particular structure. The combination of antibody and microorganism can trigger one or more mechanisms, termed effector functions, which lead to the destruction or elimination of the foreign organism.

### The immune system

That system in vertebrates which is concerned with the synthesis of Ab molecules is designated the immune system. This system is comprised of a variety of organs (e.g. thymus, lymph nodes, spleen) and cells (e.g. plasma cells and lymphocytes) which migrate freely throughout the circulatory system and most tissues.

### Antigen

Any foreign substance which evokes an immune response is termed an antigen. Essentially any macromolecule which is foreign to a particular vertebrate host can be antigenic and elicit the formation of complementary antibody.

### Antibody heterogeneity

The antibody response to the simplest of antigens is extremely heterogeneous; that is, many different molecular

species of antibody, perhaps hundreds, are generally produced in response to even a single antigen. Accordingly, each vertebrate organism appears to be capable of producing a large library of Ab molecules for most antigens. In addition, it appears that vertebrates can respond to an apparently unlimited variety of antigens and, accordingly, they must have the capacity to synthesize a large library of Ab molecules with distinct specificities. Let us now consider a hypothesis, the clonal selection theory, which is the central dogma of modern immunology.

The clonal selection hypothesis (BURNETT 1959)

This hypothesis suggests that at some time in ontogeny and independently of antigen, individual lymphocytes become committed to responding to a single type of antigenic determinant (Figure 2). This commitment is expressed through antigen-specific receptors, presumably antibody, on the lymphocyte cell surface. Thus, when antigen is introduced into the body it selects out those lymphocytes which already have receptors for the antigen on their surface. Presumably the interaction

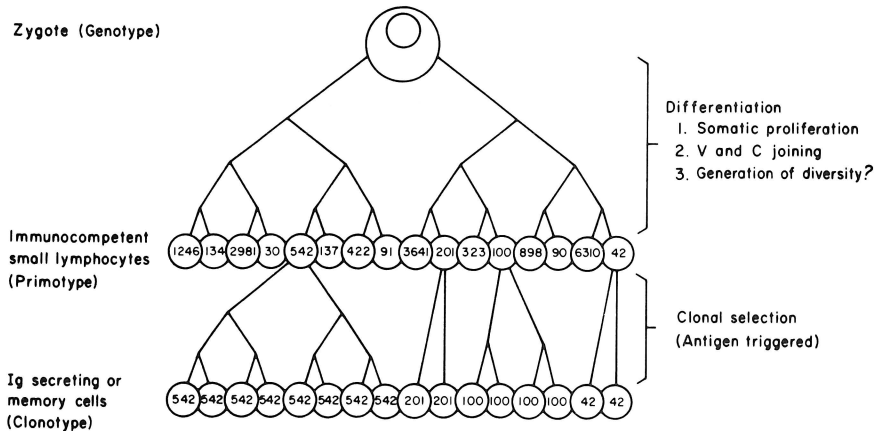


FIGURE 2. A model of somatic differentiation of antibody producing cells according to the clonal selection theory. This model suggests that there are two stages of development--differentiation which is independent of antigen and clonal expansion which is triggered by antigen. The different numbers signify the commitment of distinct lymphocytes to a different molecular species of antibody (see text).

of surface receptor molecules and antigen activates the specific lymphocytes and leads to clonal proliferation and the secretion of specific antibody (Figure 2). Daughter cells derived from clonal expansion probably are restricted

(committed) to the production of that same antibody. Thus differentiation of lymphocytes occurs in the absence of antigen, whereas specific activation (clonal expansion and Ab synthesis) is an antigen initiated process (Figure 2). The clonal selection hypothesis is now firmly based on experimental evidence (see RAFF 1973, GALLY and EDELMAN 1972).

The amount of Ab produced in a given immune response is proportional to the number of cells synthesizing that Ab. Generally, a single antigen can stimulate many distinct antibody producing clones leading to the synthesis of specific Ab molecules with a great deal of chemical heterogeneity. Thus the chemical heterogeneity of antibodies has its origin in the underlying cellular heterogeneity of the normal lymphocyte population. Let us consider systems in which this molecular heterogeneity can be studied.

### 3. SYSTEMS FOR THE STUDY OF ANTIBODY MOLECULES

Ab molecules whose functions are unknown are termed immunoglobulins (Ig). Most genetic and chemical studies have been carried out on immunoglobulins rather than antibodies. The animals whose Ig molecules have been most thoroughly studied are man, mouse and rabbit.

For the protein chemist the heterogeneity of ordinary Ab molecules presents a problem of immense proportion. For example, the serum of any vertebrate has an extremely heterogeneous collection of Ab molecules whose specificities are unknown. They presumably reflect the past antigenic history of that particular organism and the isolation of a single molecular species of antibody molecule from such a heterogeneous mixture has proved essentially impossible. Thus most protein chemists have turned to one of two systems for producing homogeneous Ig molecules.

#### Myeloma immunoglobulins

Myeloma globulins are Ig molecules that appear in the plasma of patients with multiple myeloma, a cancer of Ab producing cells (SNAPPER and KOHN 1971). Generally the neoplastic conversion occurs in a single plasma cell in each afflicted individual. This transformed cell divides repeatedly to become the predominant plasma cell population in the individual. Frequently the homogeneous Ig derived from such a neoplastic clone can comprise 95% of the serum Ig and, accordingly, this homogeneous molecule can be readily purified for chemical analysis. It is now generally accepted that myeloma globulins are not abnormal proteins, that is, proteins whose intrinsic structure differs from those of normal antibody molecules. Rather, these proteins occur as a result of the clonal expansion of one Ab producing cell in the normal population (KUNKEL 1965, GRANT and HOOD 1971). The myeloma globulin from one patient is generally different from the myeloma globulins of all other patients and indeed myeloma proteins can be found

which represent all the known classes and types of antibody molecules (see COHEN and MILSTEIN 1967). In addition to myeloma globulins, myeloma patients frequently synthesize a homogeneous immunoglobulin light chain (termed a Bence Jones protein) which, because of its low molecular weight, passes through the kidney and is excreted in the urine. Some myeloma tumors synthesize only the Bence Jones protein, whereas others synthesize, in addition, an Ig molecule whose light chain is identical to the Bence Jones protein. Thus myeloma proteins are an ideal model system for studying the structure of human Ig.

The discovery that myeloma can be induced in a laboratory strain of mouse (BALB/c) by the injection of mineral oil into the peritoneal cavity has provided an important new system for studying Ig structure and genetics (POTTER 1967). Since this particular strain of mouse is highly inbred, the Ig genes of all individual mice from this strain should be identical. Furthermore, induced tumors can be transplanted from one mouse to many other individuals of this strain and, accordingly, the production of myeloma globulin amplified. In addition, tumors can be frozen away for subsequent use or analysis. A large bank of mouse myeloma tumors, numbering in the thousands, has been generated in a number of laboratories (COHN 1967, POTTER 1967). Most of the amino acid sequence analysis work on Ig molecules to be discussed subsequently has been carried out on myeloma globulins.

#### Homogeneous antibodies

It should also be noted that particular immunization schedules with a few antigens in certain animals can, infrequently, lead to the production of homogeneous Ab (KRAUSE 1970). Various procedures for producing homogeneous antibody have been particularly useful in the rabbit where myeloma proteins have not been observed. In addition, homogeneous Ab offers an important tool for structure-function correlations on the nature of the antigen binding site.

#### A caution about phenotypic selection

The only Ig molecules which have been thoroughly examined by chemical analysis are those which have been clonally expanded--either by the myeloma process or by antigen. Thus Ig molecules whose properties the immunologist can examine may represent only a small fraction of those which the vertebrate organism is capable of generating. Accordingly, distinctions must be made for the various levels of expression of Ig genes and these are depicted in Figure 2 (GALLY and EDELMAN 1972). First, the "genotype" represents the genetic information stored in the egg and sperm which can be transmitted to progeny. This includes the structural genes for the three antibody families and their regulatory genes. Second, the "primotype" depicts the totality of different Ig molecules that can be generated by an organism during its lifetime. This number is larger than the genotype because two separate polypeptide chains comprise the Ab molecule. Accordingly, assume the number of germ line

genes for one type of chain is  $p$  and for the second is  $q$ . If the chains can associate randomly,  $p \times q$  antibody molecules could be generated from  $p + q$  genes (see EDELMAN 1971). In addition, somatic mutation may amplify the number of Ab genes. Presumably many of the Ab molecules (e.g. antibody producing cells) an organism expresses never come in contact with their respective antigens and, accordingly, are only represented in the total lymphocyte population by one or a few cells. Finally, the "clonotype" designates Ig molecules which have been produced by clonal expansion in quantities sufficient for various types of analysis. Thus the clonotype is a subset of the primotype and the primotype may or may not be a subset of the genotype, depending on whether somatic mutation can usefully amplify the information content of the antibody families.

The critical question is, of course, how faithfully the clonotype reflects the primotype and, indeed, the genotype. Does the myeloma process introduce a bias so that only a small subset of the primotype is expanded? This question can not be answered unequivocally, but there are definite indications of bias (GRANT and HOOD 1971, HOOD *et al.*, 1970c). Obviously one must be extremely cautious in drawing conclusions about genotype from studies of the clonotype. This issue will be considered in more detail as we turn now to the chemistry and genetics of antibody molecules.

#### 4. THE CHEMISTRY AND GENETICS OF ANTIBODY MOLECULES

Two separate experimental approaches have been important in analyses of the genetic control of Ig synthesis. The first, the amino acid sequence analysis of homogeneous Ig molecules, uses the genetic code dictionary to translate from protein to nucleic acid sequences. Detailed comparisons of many Ig sequences are carried out to elicit and analyze the variety of patterns which have provided insights into the organization and evolution of the Ab family. The second approach consists of the classic genetic analysis of antigenic markers present on Ig polypeptide chains and the transmission of these markers between generations. Although these approaches were generally initiated independently of one another, their results have converged and intertwined in such a fashion that a coherent picture of the chromosomal organization of Ab families is emerging. Let us consider the major patterns which have emerged from chemical studies and the complementary results from genetic analyses.

##### 4.1 THE CHEMICAL ANALYSIS OF IG POLYPEPTIDES

###### Light and heavy chains

The basic subunit of the Ab molecule is two identical light (L) polypeptide chains and two identical heavy (H) chains linked by disulfide bonds and non-covalent interaction (Figure 3). The smaller or light chain is about 23,000 daltons; the larger, or heavy chain ranges between 53,000 and 70,000 daltons (EDELMAN



and GALL 1969). The Ab molecule can be divided into three fragments by proteolytic enzyme cleavage at the middle of the heavy chain (the "hinge region"). Two of these, the Fab fragments, are identical and contain the antigen combining site, whereas the third, the  $F_C$  fragment, appears to be responsible for certain effector functions carried out by antibody molecules.

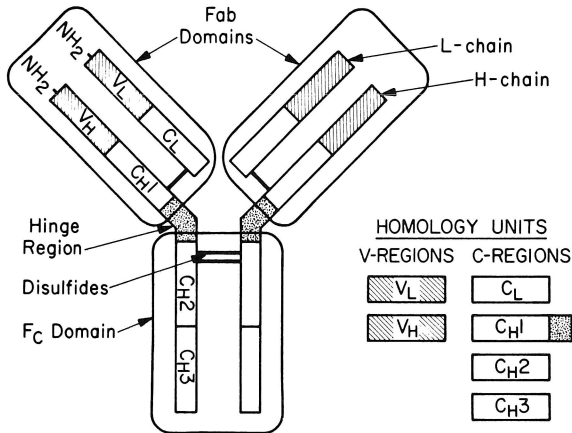


FIGURE 3. A model depicting the basic structure of the immunoglobulin (IgG) molecule. The hinge region of the heavy chain is accessible to proteolytic enzymes which can be used to cleave the molecule into its respective Fab and  $F_C$  domains. The homology units are  $\sim 110$  residues in length with a centrally placed disulfide bridge of about 60 residues (from SMITH *et al.* 1971).

### Three antibody families

Three major families of Ab polypeptides--the heavy chain family (H) and two light chain families, lambda ( $\lambda$ ), and kappa ( $\kappa$ )--are defined by their corresponding amino acid sequence homologies (SMITH *et al.* 1971). For example, kappa chains from two distinct myeloma tumors are more similar to one another than they are to any lambda or heavy chain (Table 1).

### Variable and constant regions

All immunoglobulin polypeptide chains can be divided into an amino terminal portion, the variable (V) region, and a carboxy terminal portion, the constant (C) region (LENNOX and COHN 1967). The V regions for both light (V<sub>L</sub>) and heavy (V<sub>H</sub>) chains are generally about 110-120 amino acid residues in length, whereas the C regions appear to be about 110 residues in length for light (C<sub>L</sub>) chains and 3-4 times this size for various heavy (C<sub>H</sub>) chains (EDELMAN *et al.* 1969, PUTNAM *et al.* 1971). The V and C regions were initially defined by sequence

variation (or the absence of variation). For example, if the amino acid sequences from ten IgG molecules (see Figure 3) were aligned to maximize sequence homology, the  $V_L$  regions would show marked sequence variations as would the  $V_H$  regions. The  $C_L$  regions for a given chain type (or class) would be identical as would the  $C_H$  regions, apart from minor genetic variants. The  $V_L$  regions generally differ from one another by 1 to 45 amino acid residues where the  $V_H$  regions can differ even more (SMITH *et al.* 1971, MILSTEIN and PINK 1970). Sequence diversity in the V region presumably reflects the diversity of antigen binding sites for various antibodies, since the primary amino acid sequence of a polypeptide determines its corresponding three-dimensional structure. The  $V_L$  and  $V_H$  regions show striking amino acid sequence homologies with one another as do the  $C_L$ ,  $C_{H1}$ ,  $C_{H2}$ , and  $C_{H3}$  homology units depicted in Figure 3 (HILL *et al.* 1966, EDELMAN *et al.* 1969).

Table 1. Amino terminal sequences of immunoglobulin chains.\*

Anti-body family	V region subgroup	Protein	Position											
			1	5	10	15	20	25	30	35	40			
κ	I	AG	-	D I Q M T Q S P	S S L S A S V G	D R V T I T C	Q A S Q	-	-	-	-	-	-	D I N H Y L
κ	I	EU	-	D I Q M T Q S P	S T L S A S V G	D R V T I T C	R A S Q	-	-	-	-	-	-	S I N T W L
κ	I	HAU	-	D I Q M T Q S P	S S L S A S V G	D R V T I T C	R A S Q	-	-	-	-	-	-	S I S S Y L
κ	I	OU	-	D I Q M T Z S P	S S L S A S V G	B R V T I T C	R A S Z	-	-	-	-	-	-	T I S S Y L
κ	I	ROY	-	D I Q M T Q S P	S S L S A S V G	D R V T I T C	Q A S Q	-	-	-	-	-	-	D (I.S) I F L
κ	II	CUM	E	D I V M T Q T P	L S L P V T P	G E P A S I	S C R S S Q	S L	L D S G	D G N T Y L				
κ	II	MIL	-	D I V L T Q S P	L S L P V T P	G E P A S I	S C R S S Q	N L	L L Z	-	-	-	-	S B G B Y L
κ	II	TEW	-	D I V M T Q S P	L S L P V T P	G E P A S I	S C R S S Q	-	-	-	-	-	-	H (G,B) S F L
κ	III	B6	-	(Z.I.V.L.T.Z.S.P.G.T.L.S.L.S.P.G.Z)	R.A.A.L.S	C.R.A.S.Q.S	-	-	-	-	-	-	-	L.S.G.N.Y.L
κ	III	TI	-	E I V L T Q S P	G T L S L S P	G E R A T L S	C R A S Q S	-	-	-	-	-	-	V S N S F L
λ	I	HA	Z	S V L T Q P P S	V S V S G T P G	Q R V T I S C	S G G S S N G	T G N N Y	V Y W Y Q Q					
λ	I	NEW	Z	S V L T Q P P S	V S A A P F G	Q K V T I S C	S G G S T N	-	I G N N Y	V S W H Q H				
λ	II	VIL	H	S A L T Q P A S	V S G S L G Q	S I T I S C	T G T S S D	V	G G Y N Y	S W F Q Q				
λ	II	BO	Z	S A L T Q P P S	A S G S P G Q	S V T I S C	T G T S S D	V	G D N K Y	V S W Y Q Q				
λ	III	SH	-	S E L T Q D P A	V S V A L G Q	T V R I T C	G G D S	-	-	L R G Y D A A	W Y Q Q			
λ	IV	BAU	-	Y G L T Q P P S	L S V S P G Q	T A S I T C	S G D K	-	-	L G E Q Y V C	W Y Q Q			
λ	IV	KERN	-	Y A L T Q P P S	V S V S P G Q	T A V I T C	S G D N	-	-	L E K T F V S	W F Q Q			
λ	IV	X	-	Y D L T Q P P S	V S V S P G Q	T A S I T C	S G D K	-	-	L G D K D V C	W Y Q Q			
Hyl	I	EU	Z	V Q L V Q S G A	E V K K P G S	S V K V S C	K A S G G T F S	-	-	R S A I I	I W V R			
Hμ	II	OU	Z	V T L T E S G P A	L V K P K Q	P L T L T C T	F S G F S L S T S R M R V S W I R							
Hyl	II	HE	Z	V T L K E N G P T	L V K P T E T	L T L T C T	L S G L S L T T D G V A V G W I R							
Hyl	II	DAW	Z	V T L R E S G P A	L V R P T Q T	L T L T C T	F S G F S L S T S G E T M C V A W I R							
Hyl	II	COR	Z	V T L R E S G P A	L V K P T Q T	L T L T C T	F S G F S L S S T G M C V G W I R							
Hyl	III	NIE	Z	V Q L V Q S G G	G V V Q P G R	S L R L S C A	A S G F T F S	-	-	R Y T I H	H V R			

\* One letter amino acid code: A, Ala; B, Asx; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, ?; Y, Tyr; Z, Glx; -, Gap. This data is taken from Dayhoff (1972).

There are three families of variable regions which correspond to the three antibody families described above ( $V_\lambda$ ,  $V_\kappa$ ,  $V_H$ ). Once again the V regions from a given family are much

more closely related to one another than they are to the V regions of other families (Table 1). In certain Ig families (e.g. mouse  $\kappa$  and human  $\lambda$ ) enormous heterogeneity is seen among the V region sequences examined to date (HOOD *et al.* 1973, SMITH *et al.* 1971). In contrast, certain other Ig families (e.g. mouse  $\lambda$ ) seem to exhibit an extremely limited degree of V region heterogeneity (WEIGERT *et al.* 1970).

Antibody families may have one or more constant regions (see GALLY and EDELMAN 1972). For example, in man the heavy chain family has at least 10 distinct C regions, the lambda has three and the kappa has a single C region (Figure 4).

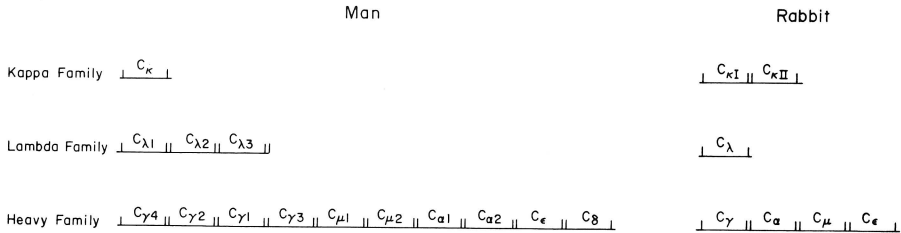


Figure 4. A model of the constant genes for the three Ab families of man and rabbit.

Some of these C regions differ by a single amino acid residue (e.g. C $\lambda$ 2, C $\lambda$ 3) whereas others differ more extensively in sequence (C $\mu$ , C $\gamma$ , C $\alpha$ , C $\delta$ , and C $\epsilon$ ) (DAYHOFF 1972). The class or subclass of an Ig molecule is determined by its C<sub>H</sub> region. Ig molecules with C<sub>H</sub> regions differing extensively are said to belong to different classes of Ab (e.g., IgM [ $\mu$  chain], IgG [ $\gamma$  chain], IgA [ $\alpha$  chain], IgD [ $\delta$  chain], IgE [ $\epsilon$  chain]). Ig molecules with more similar C<sub>H</sub> regions are said to belong to different subclasses of Ab (e.g. IgG<sub>1</sub>, IgG<sub>2</sub>, IgG<sub>3</sub> and IgG<sub>4</sub>). The heavy chains from different classes vary in molecular weight from 53,000 to 70,000 daltons (PUTNAM *et al.* 1971) and presumably carry out distinct physiologic or effector functions (Table 2).

Other important patterns have emerged from a detailed amino acid sequence analysis of Ig polypeptides. These relate to topics which will be covered later and will accordingly be discussed at that time.

#### 4.2 THE IMMUNOGENETIC ANALYSIS OF IG POLYPEPTIDES

The Ig molecules from individual species of vertebrates represent a polymorphic system with genetically inherited variations in their chemical structures. Many of these variations

can be detected by serologic techniques and, accordingly, a simple assay system is available for examination of the distribution of these markers in family and population studies (GRUBB 1970). These genetic analyses have provided important insights

Table 2. Human immunoglobulin classes.

Immunoglobulin Class	IgG or $\gamma$ G	IgA or $\gamma$ A	IgM or $\gamma$ M	IgD or $\gamma$ D	IgE or $\gamma$ E
Usual Molecular Formula	$L_2\gamma_2$	$L_2\alpha_2$ or $(L_2\alpha_2)_2SC$ , **J*	$(L_2\mu_2)J^*_5$	$L_2\delta_2$	$L_2\epsilon_2$
Subclasses	$\gamma_1, \gamma_2, \gamma_3, \gamma_4$	$\alpha_1, \alpha_2$	$\mu_1, \mu_2$	None	None
Molecular Weight	150,000	152,000 or 385,000	900,000	175,000	190,000
Sedimentation Constant	6.6 S	6.8-11.4 S	19.0 S	7.0 S	8.0 S
Carbohydrate Content	2.5%	5-10%	5-10%	-	11.5%
Serum Levels (mean, adult, mg/ml)	12	1.8	1.0	0.03	0.0003
Half Life (days)	25	6	5	2.8	2
Synthetic Rate (gm/day/70 kg)	2.3	2.7	0.5	0.03	-
Physiologic (effector) Function	Complement fixation, Placental transfer	External secretions	Complement fixation, Early response, High agglutination efficiency	??	Mast cell fixation, Reagin activity

\* J = J chain, a polypeptide which seems to join the higher multimeric forms of Ig.

\*\* SC = secretory component, a polypeptide found on IgA which is released into external secretions.

L = light chain ( $\kappa$  or  $\lambda$ ).

as to the chromosomal organization of the Ab families. Studies in two systems, man and rabbit, will be described below because the nature and the distribution of their genetic markers is representative of most other systems examined.

#### 4.2.1 Human genetic markers

Serologic reagents for subtle chemical differences on molecules coded by allelic genes (e.g. single residue substitutions) usually must be generated in the same or closely related species so that the only difference between host and donor polypeptides is the genetic marker (amino acid residue) of interest. For example, in man some of the most useful serologic reagents for genetic markers have come from multiparous mothers (frequently immunized by the genetically distinct fetus) and from individuals receiving multiple blood transfusions

(STEINBERG 1962). The best characterized genetic markers from the Ig chains of man are indicated in Table 3.

Table 3. Genetic markers on human immunoglobulin molecules.

Ig polypeptide	Marker	Location*	Residue(s)
κ	InV(1)	191	Leu
κ	InV(3)	191	Val
κ	InV(1,2)	unknown	
γ1	Gm z(17)	218	Lys
γ1	Gm f(4)	218	Arg
γ1	Gm a(1)	356-358	Asp-Glu-Leu
γ1	Gm y(22)	356-358	Glu-Glu-Met
γ3	Gm g(21)	400	Tyr
γ3	Gm n(5)	400	Phe

\* The residue positions on human κ chains use the Bence Jones protein Ag as a reference (DAYHOFF 1972); on H chains Eu is used as a reference (EDELMAAN *et al.* 1969).

The human κ chain has a genetic marker termed InV which correlates with a single amino acid substitution, valine to leucine, at position 191 (TERRY *et al.* 1969) (Table 3). These markers are inherited in a simple autosomal codominant fashion (STEINBERG 1962) and presumably arose from a nucleotide substitution in the single structural gene coding for the human C<sub>κ</sub> region. The fact that serologic reagents can detect this subtle difference of one in more than 200 amino acid residues is an impressive display of the exquisite specificity of the immune system.

Genetic markers have not been observed in the human λ family. Three very similar C<sub>λ</sub> sequences appear to be present in all normal individuals (EIN 1968, GIBSON *et al.* 1971, HESS *et al.* 1971) and, accordingly, they are probably encoded by three distinct structural genes which arose by recent gene duplication. The absence of segregating genetic markers prevents one from drawing any general conclusions about the linkage (or lack of linkage) of the human λ family to other immunoglobulin families.

Genetic polymorphisms termed Gm markers have been found on three of the four C<sub>γ</sub> regions (Figure 4) (EDELMAAN *et al.* 1969, FRANGIONE *et al.* 1969, PRESS and HOGG 1969, PRAHL 1967, GREY and ABEL 1970). Four of the Gm markers appear to correspond to a simple amino acid substitution (Table 3). Family and population studies indicate that the Gm markers are closely linked to one another (NATVIG *et al.* 1967). Rare interallelic recombinants among each of the C<sub>γ</sub> genes permit the linkage

order for the  $C_\gamma$  genes given in Figure 4 to be proposed (KUNKEL *et al.* 1969a).

A serologic marker on one of the two  $C_\alpha$  genes indicates that it is also closely linked to the  $C_\gamma$  genes described above (KUNKEL *et al.* 1969b). The cluster of  $C_H$  genes is not linked to the  $C_\kappa$  gene (STEINBERG 1962). Genetic markers have not been found for the remainder of the human  $C_H$  genes.

4.2.2 Rabbit genetic markers

Ig molecules from one rabbit can be injected into a second rabbit and the recipient will respond with the synthesis of specific antibodies against those antigenic determinants (markers) which are not present on its own Ig (OUDIN 1960). Thus specific antiserum can be raised against allelic forms of Ig present in one rabbit but not in a second. Such serologic (genetic) markers are termed allotypes. In mating experiments most allotypes segregate in a classic Mendelian fashion (OUDIN 1960). There are eight groups of genetic markers in the rabbit: group a -  $V_H$ ; group b -  $\kappa$ ; group c -  $\lambda$ ; group d -  $C_\gamma$ ; group e -  $C_\gamma$ ; group f -  $\alpha$ ; group g -  $\alpha$ ; group n -  $\mu$  (Figure 5). The a, b and c markers may appear in all classes of Ig, whereas the others are found only in the class which has their respective  $C_H$  region.

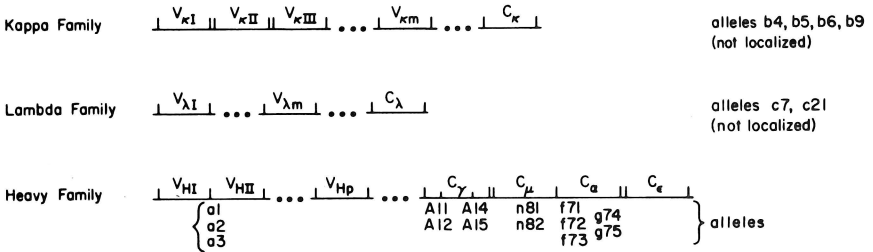


FIGURE 5. A model of the three families of antibodies genes in the rabbit depicting the various allotype markers under their respective C genes. The groups b and c alleles have not been localized.

Rabbit  $\kappa$  chains have four allotypes designated b4, b5, b6 and b9 (Table 4) (DRAY *et al.* 1963, MAGE 1971). The different antigenic forms behave as codominant alleles at a single genetic locus (DRAY *et al.* 1963). Multiple amino acid substitutions do exist among the four C regions (APPELLA *et al.* 1969). It appears that certain types of V sequences may also correlate with the b markers (WATERFIELD *et al.* 1973). Accordingly, these markers have not been localized to a particular region of the  $\kappa$  chain.

Two markers (c7, c21) have been observed in rabbit  $\lambda$  chains which initially appeared to be inherited as simple

Mendelian alleles (MAGE *et al.* 1968, GILMAN-SACHS *et al.* 1969). More recent studies have indicated that in some pedigrees these markers behave as duplicated genes (MAGE 1971). Thus the status of these markers is uncertain. The  $\lambda$  and  $\kappa$  markers appear to be unlinked (DRAY *et al.* 1963).

Table 4. Genetic markers on rabbit immunoglobulin molecules.

Ig polypeptide	Marker	Location*	Residue
$\kappa$	b4	uncertain <sup>†</sup>	? multiple residues
$\kappa$	b5	"	"
$\kappa$	b6	"	"
$\kappa$	b9	"	"
$\lambda$	c7	uncertain	
$\lambda$	c21	"	
$\gamma$	d11	225	Met
$\gamma$	d12	225	Thr
$\gamma$	e14	309	Thr
$\gamma$	e15	309	Ala
$\alpha$	f71	somewhere on Ca	
$\alpha$	f72	"	
$\alpha$	f73	"	
$\alpha$	g74	"	
$\alpha$	g75	"	
$\mu$	n81	somewhere on Cu	
$\mu$	n82	"	
V <sub>H</sub>	a1	see text	? multiple residues
V <sub>H</sub>	a2	"	"
V <sub>H</sub>	a3	"	"

\* See legend to Table 2.

<sup>†</sup>The position(s) of these markers is uncertain, although certain differences in the C<sub>K</sub> regions have been observed (Appella *et al.* 1969)

Six allotype markers (A8, A10, d11, d12, e14, e15) exist for the rabbit  $\gamma$  chains and four of these have been localized (Table 4) (HAMERS *et al.* 1966, PRAHL *et al.* 1969, FLORENT *et al.* 1970, APPELLA *et al.* 1971, DUBISKI 1969, MANDY and TODD 1970). Five allotype markers are present on rabbit  $\alpha$  chains (HANLEY *et al.* 1973) and at least two are present on rabbit  $\mu$  chains (GILMAN-SACHS and DRAY 1972) (Table 4). Each of these groups of antigenic markers behave as codominant alleles at a single locus. These H chain allotype markers are closely linked to one another in the rabbit genome but they are not linked to the  $\kappa$  or  $\lambda$  allotype markers. There is no evidence

for multiple  $C_H$  genes in any of the major classes of Ig (e.g.,  $\gamma$ ,  $\alpha$ , or  $\mu$ ), in direct contrast to the situation for the human H chain family (Figure 4). There is presumptive evidence for an  $\epsilon$  chain in rabbits (KINDT and TODD 1969), but genetic markers are not available for this chain.

The group a allotype markers appear to be the only bona-fide V region genetic markers known in any species (OUDIN 1960). They have been localized to the amino terminal half of the H chain (KOSHLAND 1967, PRAHL and PORTER 1968, WILKINSON 1969a). Because the group a allotypes are associated with all classes of heavy chain ( $\gamma$ ,  $\alpha$ ,  $\mu$  and  $\epsilon$ ) (PRAHL *et al.* 1970), it appears that these markers are present on  $V_H$  regions, which, apparently can be associated with any of the  $C_H$  regions. Multiple amino acid substitutions in the  $V_H$  regions appear to correlate with the group a allotypes (KOSHLAND 1967, WILKINSON 1969), although recent experimental data suggest that many of these apparent correlations may be misleading (JATON *et al.* 1973). The group a markers are closely linked to markers from groups d, e, f, g and h and are unlinked to the group b and c markers. Recombination between the group a and group d or e markers occurs with a frequency of about 1% (MAGE *et al.* 1971, KINDT and MANDY 1972). Thus the genes of the heavy chain family appear to be linked to one another but are unlinked to those of the light chain families.

Genetic and chemical studies on human and rabbit Ig molecules permit four important conclusions to be drawn. First, most of these genetic markers are or will be correlated with simple amino acid substitutions (Tables 3 and 4). An almost certain exception is the rabbit group a markers (and possibly the group b) which will be discussed subsequently. Second, there are three Ab families which are unlinked in the mammalian genome. Clearly the  $\kappa$  and H families are unlinked, and if the c markers in the rabbit are alleles, then the  $\lambda$  family is unlinked to either of the others (DRAY *et al.* 1963, GILMAN-SACHS *et al.* 1969). Third, the  $C_H$  genes appear to be closely linked to one another as well as closely linked to the  $V_H$  genes. Genes for  $\gamma_1$ ,  $\gamma_2$ ,  $\gamma_3$  and  $\alpha_2$  are closely linked in man and genes for  $\gamma$ ,  $\alpha$ , and  $\mu$  are closely linked in rabbit. Thus it appears that the  $C_H$  genes of mammals will be closely linked to one another and to their corresponding  $V_H$  genes. The V regions of one Ab family appear to be capable of associating with any of the C regions of that same family, but never with the C regions of other Ab families. Finally, enormous sequence diversity generally exists among the V regions of each Ab family (Table 1). Let us turn now to the question of how this V region diversity is stored in the mammalian genome.

## 5. ANTIBODY DIVERSITY AND INFORMATION STORAGE

Information can be stored in the mammalian genome by two general mechanisms. First, each unit of information (V gene) may be stored in the zygote as a germ line gene which arose



during the evolution of the species (germ line theory). Alternatively, information may be stored as a limited number of germ line genes which are expanded by a mutational or recombinational process during somatic differentiation of the individual. Indeed, the controversy between these two alternative points of view has been one of the major driving forces in molecular immunology for the past ten years.

Prior to the availability of detailed V region sequence data, the somatic theories were appealing to some because molecular immunologists, fresh from studies of bacterial and viral genetics, were attracted to simple genetic explanations (COHN 1968, EDELMAN and GALLY 1967, JERNE 1967, SMITHIES 1967). A single V gene for each Ab family made it simple to explain the existence of V region genetic markers (the group a allotypes of the rabbit). It was also easy to imagine how natural selection might operate on a single V gene but rather difficult to imagine how it might operate in a germ line model where the "silent" or unused Ab genes of each generation could mutate, unchecked by selection pressures, and lose their respective functions (COHN 1968). Other immunologists contended that the germ line model was attractive because it did not require an *ad hoc* mechanism (mutational or selectional) for explaining V region diversity (DREYER *et al.* 1967). Phylogenetic evolution through gene duplication, mutation, and selection in the germ line had produced gene diversity in other systems such as hemoglobins and cytochromes and could, accordingly, it was suggested, also generate antibody diversity.

Four distinct models for generating antibody diversity had been proposed by 1968.

1. Somatic hypermutation. This model suggests that a special mutational mechanism operates on antibody genes to generate sequence diversity (LEDERBERG 1959, BRENNER and MILSTEIN 1966).
2. Normal somatic mutation. This theory postulates that ordinary somatic mutation can produce V gene variants (COHN 1968).
3. Somatic recombination. This hypothesis suggests that two or more antibody V genes can undergo somatic recombination to generate variants (GALLY and EDELMAN 1967, SMITHIES 1967).
4. Germ line genes. This model postulates that most, if not all, antibody V genes are coded in the zygote or germ line (DREYER *et al.* 1967, HOOD *et al.* 1967).

The availability of extensive V region sequence data caused a remarkable evolution of the various theories of antibody diversity and has in a very real sense made them much more difficult to distinguish one from another as we shall subsequently see (HOOD and PRAHL 1971, GALLY and EDELMAN 1972, COHN 1971, MILSTEIN and PINK 1970, JERNE 1971). The accumulation of more than 150 partial or complete V region sequences (DAYHOFF 1972, HOOD *et al.* 1973) has also created a problem of data

analysis. How can patterns be abstracted from the extensive amino acid sequence data which may tell immunologists more about the genetic organization of antibody families? The genealogic analysis, discussed below, permits sequence data to be compressed and visualized in a form that is suitable for abstracting patterns. Perhaps in no other system has genealogic analysis been so useful as in delineating the genetic organization of antibody families. Indeed, in the analysis of complex and multigenic systems, the genealogic analysis of protein or nucleic acid sequences is, perhaps, the most powerful tool available to the geneticist.

## 5.1 THE GENEALOGIC ANALYSIS OF V REGIONS

### 5.1.1 The nature and significance of a genealogic tree

All immunologists agree that V region diversity should be explained by the simplest possible genetic mechanism that is compatible with the experimental data. The genealogic approach assumes that contemporary genes are related to one another by divergent evolution and that their relationships can be depicted by a genealogic tree which permits the determination of both immediate and distant ancestral (nodal) genes (Figure 6). Accordingly, genealogic analysis is an attempt to retrace the events of divergent evolution and determine how a given set of sequences (V genes) can be derived from a single ancestral gene using the minimum number of genetic events. There are three types of genetic events--gene duplication, single base substitution, and the insertion or deletion of one or more codons (sequence gaps).

Let us consider the events required to generate a hypothetical genealogic tree for human  $V_L$  genes (Figure 6). A single primordial L gene (level A) is the ancestor to all contemporary  $V_L$  genes (level F). This ancestral gene undergoes a gene duplication to produce two daughter genes each of which will diverge in nucleotide sequence from the other to become, respectively, the primordial  $V_K$  and  $V_\lambda$  genes (level B). The number 5 on the branches joining levels A and B indicates nucleotide substitutions at certain codon positions. In addition, the  $V_\lambda$  gene had codon deletions at positions 9 and 94-95 which are indicated by brackets. All the  $V_\lambda$  genes above level B on the genealogic tree will share these genetic events (deletions and base substitutions) with the primordial  $V_\lambda$  gene. It is these shared genetic events in contemporary proteins (genes) which allow the ancestry to be traced. Gene duplication followed by mutation and occasional codon deletions occurs at each of the successively higher levels of the genealogic tree (C, D, E) until the diversity of contemporary genes (level F) is produced. In actual practice, genealogic trees are generated by the reverse process. The gene (protein) sequences at level F are given. With the aid of a computer program, genealogic trees similar to that of Figure 6 have been derived from amino acid sequence data on a variety of systems such as cytochrome c and the globins (FITCH AND MARGOLIASH 1967, DAYHOFF 1972). It should be stressed that the

genealogic tree determines the number and types of genetic events that are required to generate a given set of genes irrespective of whether these mutational events occurred in the germ line or in the soma.

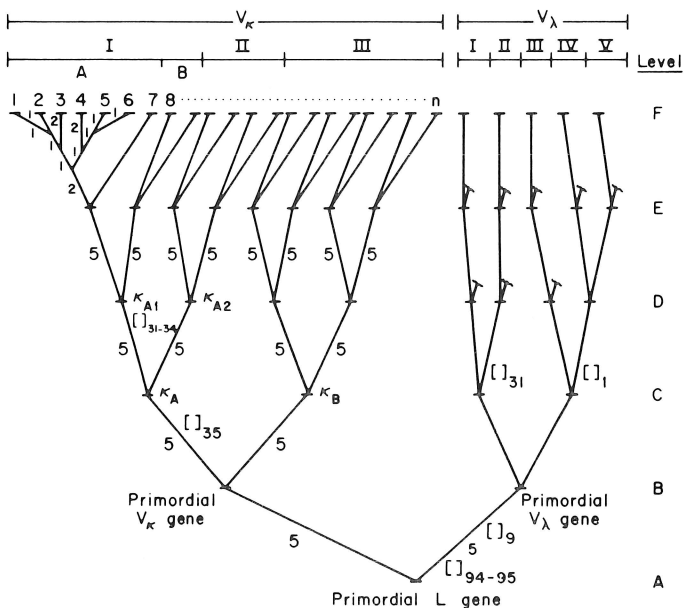


FIGURE 6. A hypothetical genealogic tree for human  $V_L$  regions. The tree is constructed from a set of proteins, such as human  $V_L$  regions, by generating a series of ancestral or nodal sequences (levels E, D, C, and so on) using the minimum possible number of "genetic events" (base substitutions, sequence insertions or deletions, and gene duplications). The genetic events responsible for generating this genealogic pattern could occur, in part, during somatic differentiation (somatic theory) or entirely during the evolution of the species (germ line theory). The  $V_K$  branch is divided into three sub-branches (subgroups) designated I, II and III. The fact that each of these subbranches can be further subdivided is indicated by IA and IB. The fine details of one region of the genealogic tree are represented by  $V_K$  regions 1-8. (From HOOD 1972).

What would be the general amino acid sequence characteristics of this set of hypothetical V regions (level F)? Amino acid sequence analysis would show that proteins from common genealogic branches would form related groups or clusters of sequences (termed subgroups by immunologists) (see Table 1). For example, the  $V_K$  regions 1-7 in Figure 6, derived from a common nodal sequence at level E, would differ from one

another by no more than 10 amino acid residues. This set of closely related  $V_K$  regions, a "subgroup," is indicated as IA in Figure 6. Likewise the  $V_K$  regions in subgroup II (derived from nodal sequence  $\kappa_{A2}$ ) would differ from one another by no more than 20 amino acid substitutions. In contrast, each  $V_{KIA}$  region would differ from each  $V_{KII}$  region by about 30 residue substitutions and one sequence gap at positions 31-34. Thus subgroups are no more than related sets of V regions which share a common nodal sequence on the V region genealogic tree. The concept of V region subgroups is useful because of its descriptive nature and because all immunologists agree that each distinct V region subgroup must represent one or more germ line V genes. Immunologists can not, however, agree on the precise definition of a subgroup as we shall see subsequently.

How is the genealogic pattern related to theories of antibody diversity? It should be stressed that the genealogic tree presented in Figure 6 is generated by gene duplication, mutation and selection. In theory, these events could occur in the soma, or they could occur in the germ line. The germ line theory contends that the terminal most twigs of the genealogic tree (level F - Figure 6) represent germ line genes, whereas somatic theories would place the boundary between germ line genes and somatic mutational events at an earlier level (somewhere below F). It should be noted that genealogic analysis also permits us to derive the most probable sequences for those genes which are germ line under a somatic model (these are the nodal sequences at an appropriate level).

Obviously the critical question with regard to theories of antibody diversity is, "What level of the genealogical tree represents germ line genes?" This question cannot be answered precisely, but two features of any somatic theory should be considered. First, how many mutational events are required during somatic differentiation to generate most antibody V genes? To take an extreme case, for example, suppose only the primordial  $V_L$  gene is in the germ line (level A - Figure 6). Then 25 somatic nucleotide substitutions plus various deletions must occur in each immunocyte line to generate serum light chains with the hypothetical  $V_\lambda$  and  $V_K$  regions given in Figure 6. A second but related question is how much parallel mutation is likely to occur with a somatic theory? Let us illustrate this point by assuming that proteins 1 and 2 (Figure 6) had come from separate individuals. Once again, if we assume that only the primordial gene (level A) is in the germ line, then 25 mutations are required to generate genes 1 and 2. However in this case 24 mutations as well as two deletions must be identical (e.g. parallel mutations) in these two separate individuals (hence in two distinct lines of immunocytes). The number of mutations required (parallel or not) is obviously reduced by moving the germ line level up the genealogic tree (e.g. if the germ line level is at E only 5 mutations [4 identical] are required to generate genes 1 and 2). These two issues will be considered with regard to individual theories of antibody formation after an examination of actual V region data.

5.1.2 A genealogic tree for the human antibody families

The genealogic tree for the complete human V region sequences from Table 1 is depicted in Figure 7. Two general conclusions can be drawn. First, there are three major branches on this tree which correspond to the three antibody

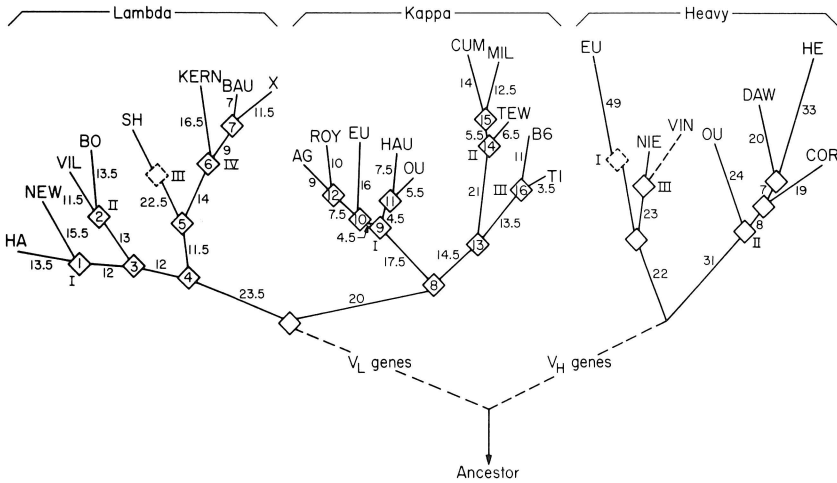


FIGURE 7. A genealogic tree for V regions from the three Ab families of man. A correction factor has been added to account for the possibility of multiple nucleotide substitutions at a single codon position. Thus for example,  $V_K$  region BAU has diverged from nodal sequence "7" by seven mutational events and from the next most similar  $V_K$  region, X, by 18.5 mutational events. (Adapted from DAYHOFF 1972).

families-- $\lambda$ ,  $\kappa$  and H. For example, 43.5 residues separate the nodal sequences from the  $\lambda$  and  $\kappa$  families (nodal sequences 4 and 8 respectively). Thus all immunologists agree that these families are coded by distinct germ line genes. Second, each major branch (antibody family) has distinct subbranches. The set of V regions which constitute a subbranch of the genealogic tree for one Ab family is, as indicated above, termed a subgroup (HOOD *et al.* 1967, MILSTEIN 1967). In Figure 7 the nodal sequence for each proposed subgroup is designated by a Roman numeral (e.g. the  $\kappa$  family has three subgroups--I, II and III). The assignment of V regions to a particular branch is easily carried out for some Ab families (human  $\kappa$ ) by visual inspection based on linked amino acid residues and sequence gaps (Table 1). For other families the assignment is more difficult. Human V regions within the individual subgroups of the  $\kappa$  family differ by about 3.3-13 residues; those from the  $\lambda$  family differ by

about 37-18 residues; and those from the H family differ by about 72-53 residues (see legend to Figure 7). All immunologists agree that the major branches for each antibody family (e.g. I, II and III for  $\kappa$ ) are coded by distinct germ line genes. The controversial point at issue is what level on the actual genealogic tree of human Ig chains represents germ line genes. Some immunologists have suggested that genetic polymorphism in the human population may obscure the actual mutational events produced by a somatic mechanism of antibody diversity (COHN 1968). Thus interest has turned to an examination of Ig polypeptides from an inbred strain of mice (BALB/c) in which myeloma tumors can readily be induced and transplanted (POTTER 1967).

### 5.1.3 A genealogic analysis of mouse Ig chains

The BALB/c mouse strain is highly inbred and, accordingly, genetic polymorphisms which might obscure sequence patterns pertinent to the genetic mechanisms of antibody diversity should be minimized, if not eliminated.

#### The kappa family

The amino terminal sequence of 44 BALB/c kappa chains has been examined over their amino terminal 23 residues (HOOD *et al.* 1973) (Table 5). Thirty-two of the 44 proteins have distinct sequences and indeed a majority of the proteins clearly differ from one another by multiple nucleotide substitutions. A genealogic tree for the amino terminal portion of certain mouse and human  $V_{\kappa}$  sequences is present in Figure 8. A more complete genealogic analysis reveals that more than 20 branches exist on the mouse  $V_{\kappa}$  genealogic tree which are about as distinct as the three major branches seen on the human  $V_{\kappa}$  genealogic tree (Figure 8). Accordingly, there must be at least 20 subgroups (V germ line genes) in the mouse library. It is important to stress that the BALB/c  $V_{\kappa}$  sequences have demonstrated no saturation of diversity; that is, the first 20  $V_{\kappa}$  sequences examined yielded approximately as many new genealogic branches or subgroups as did the second 20 examined. Presumably if a third set of 20  $V_{\kappa}$  sequence were examined, many additional subgroups would be found. Thus in the highly inbred BALB/c mouse there appears to be a large number of  $V_{\kappa}$  genes, certainly many more than 20 even by conservative standards (note that about one  $V_{\kappa}$  subgroup [germ line gene] has been observed per 2  $V_{\kappa}$  regions examined).

It is possible to demonstrate that sequence diversity exists among those  $V_{\kappa}$  regions which appeared identical at their amino terminus. Four such proteins were examined (M70, M321, M63, and T124 - See Table 5) (MCKEAN *et al.* 1973a, 1973b). M70 differed from the other three  $V_{\kappa}$  regions by 20 to 21 residues; M63 differed from the other three  $V_{\kappa}$  regions by about eight residues; whereas M321 and T124 differed from one another by just three residues. A genealogic tree of these  $V_{\kappa}$  regions is present in Figure 9. Thus the consequences of postulating a germ line gene at each of the levels (A, B, or C - Figure 9) of this genealogic tree can readily be visualized. If the germ

Table 5. Amino terminal sequences of BALB/c  $\kappa$  chains.\*

Tumor	Class	1	5	10	15	20	23																		
M70	IgF	Asp	Ile	Val	Leu	Thr	Gln	Ser	Pro	Ala	Ser	Leu	Ala	Val	Ser	Leu	Gly	Gln	Arg	Ala	Thr	Ile	Ser	Cys	
M321†	BJ	-----																							
T124	BJ	-----																							
M63	BJ	Asn	-----																						
B32	BJ	-----																							
B61	BJ	-----																			Glx	Lys	-----		
M46	BJ	Asp	Ile	Val	Leu	Thr	Gln	Ser	Pro	Ala	Thr	Leu	Ser	Val	Thr	Pro	Gly	Asp	Ser	Val	Ser	Leu	Ser	Cys	
M172	IgH	-----																							
M316†	BJ	-----																							
M603	IgA	Asp	Ile	Val	Met	Thr	Gln	Ser	Pro	Ser	Ser	Leu	Ser	Val	Ser	Ala	Gly	Glu	( )	Val	Thr	Met	Ser	Cys	
M870	IgA	-----																			Lys	-----			
M384	IgA	-----																			Lys	-----			
M167	IgA	Asp	Ile	Val	Ile	Thr	Gln	Asx	Glu	Leu	Ser	Asp	Pro	Val	Thr	Ser	Gly	Glu	Ser	Val	Ser	Ile	Thr	Cys	
M511	IgA	-----																							
T15	IgA	Asp	Ile	Val	Met	Thr	Gln	Ser	Pro	Thr	Phe	Leu	Ala	Val	Thr	Ala	Ser	Lys	Lys	Val	Thr	Ile	Ser	Cys	
H8	IgA	-----																							
M467	IgA	Asp	Val	Leu	Met	Thr	Gln	Thr	Pro	Leu	Ser	Leu	Pro	Val(Ser)	Leu	Gly	Asp	Glu	Ala	( )	Ile(Ser)	Cys	-----		
M37	IgH	-----																							
T173	BJ	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ala	Ser	Leu	Ser	Val	Ser	Val	Gly	Glx	Thr	Val	Thr	Ile	Thr	Cys	
R23	IgF	-----																			Glu	( )	-----		
M31c	IgF	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ala	Ser	Leu	Ser	Ala	Ser	Val	Gly	Glu	Arg	Val	Thr	Ile	Thr	Cys	
MX	-----																			Val	-----				
M674	BJ	Asp	Val	Val	Met	Thr	Gln	Thr	Pro	Leu	Thr	Leu	Ser	Val	Thr	Ile	Gly	Glu	Pro	Ala	Ser	Leu	Ser	Cys	
M843	BJ	-----																			( )	Ile	-----		
M265	BJ	Glu	Thr	Thr	Val	Thr	Gln	Ser	Pro	Ala	Ser	Leu	Ser	Met	Ala	Ile	Gly	Glu	Lys	Val	Thr	Ile	[+]	Cys	
M773	BJ	-----																							
S10	IgA	Glu	Ile	Val	Leu	Thr	Glx	Ser	Pro	Ala	Ile	Thr	Ala	Ala	Ser	Leu	Gly	Glx	Arg	Val	Thr	Ile	Thr	Cys	
T191	-----																								
M29	BJ	Glu	Asn	Val	Leu	Thr	Glx	Ser	Pro	Ala	Ile	Met	Ser	Ala	Ser	Pro	Gly	Glu	Arg	Val	Thr	Met	Thr	Cys	
T153	BJ	-----																			( )	Glx	Asx	-----	
L1	IgG	Asp	Ile	Val	Met	Thr	Gln	Ser	Pro	Ser	Ser	Met	Gln	Ala	Ser	Ile	Gly	Glu	Lys	Val	Thr	Ile	Ser	Cys	
M35	BJ	Asp	Ile	Val	Met	Thr	Gln	Thr	Pro(Asx)	Phe	Leu	Leu	Val	Ser	Ala	Gly	( )	Arg	Val	Thr	Ile	Thr	Cys	-----	
T157	BJ	Asp	Ile	Val	Met	Thr	Gln	Ser	Gln	Ser	Phe	Met	Ser	Thr	Ser	Val	Gly	Asp	Arg	Val	Ser	Val	Thr	Cys	
M379	BJ	Asp	Ile	Val	Met	Thr	Gln	Ser	Pro	Ser	Phe	Met	Val(Thr)	Ser	Val	Gly(Glx)	( )	Val	-----						
H5	IgA	Asp	Ile	Val	Met	Thr	Gln	Ser	Thr	Lys	Phe	Met	Ser	Thr	Ser	Val	Gly(Lys)	Arg	Val	Ser	Ile	Thr	Cys	-----	
M21	IgF	Asn	Ile	Val	Met	Thr	Gln	Ser	Pro	Lys	Ser	Met	Ser	Met	Ser	Val	Gly	Glu	Arg	Val	Thr	Leu	Thr	Cys	
M173	IgG	Asp	Ile	Gln	Met	Thr	Gln	Thr	Thr	Ser	Ser	Leu	Ser	Ala	Ser	Leu	Gly	Asp	( )	Val	Thr	Ile	Thr	Cys	
M41	BJ	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ser	Ser	Leu	Ser	Ala	Ser	Leu	Gly	Glu	Arg	Val	Ser	Leu	Thr	Cys	
J606	IgA	Asp	Val	Gln	Met	Ile	Gln	Ser	Pro	Ser	Ser	Leu	Ser	Ala	Ser	Leu	Gly	Asp	Arg	Val	Thr	-----			
M149	BJ	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Asx	Tyr	Leu	Ser	Ala	Ser	Val	Gly	Glu	Thr	Val	Thr	Ile	Thr	( )	
M600	IgF	Asp	Ile	Gln	Met	Ile	Glx	Ser	Pro	Ser	Ser	Met	Phe	Ala	Ser	Ile	Gly	Asp	Glx	Val	Ser	Ile	Ser	Cys	
M460	IgA	Asp	Val	Val	Met	Thr	Gln	Thr	Pro	Leu	Ser	Leu	Thr	Val	Ser	Leu	Gly	Asp	Arg	Ala	Ser	Ile	Ser	Cys	
T29	BJ	Glu	Val	Val	Leu	Thr	Glx	Ser	Pro	Ala	Ile	Met	Ser	Ala	Ser	Leu	Gly	Leu	Arg	Val	Ser	Met	Ser	Cys	
M47	IgF	Glu	Val	Val	Met	Thr	Gln	Thr	Pro	Leu	Ser	Leu	Ala	Val(Ser)	Leu	Gly	( )	Glx	Ala(Ser)	-----					

\*References are given in Hood et al. (1973).

†Sequences identical or nearly identical to a sequence directly above are indicated by lines.

‡Unknown or uncertain residue positions are indicated by brackets.

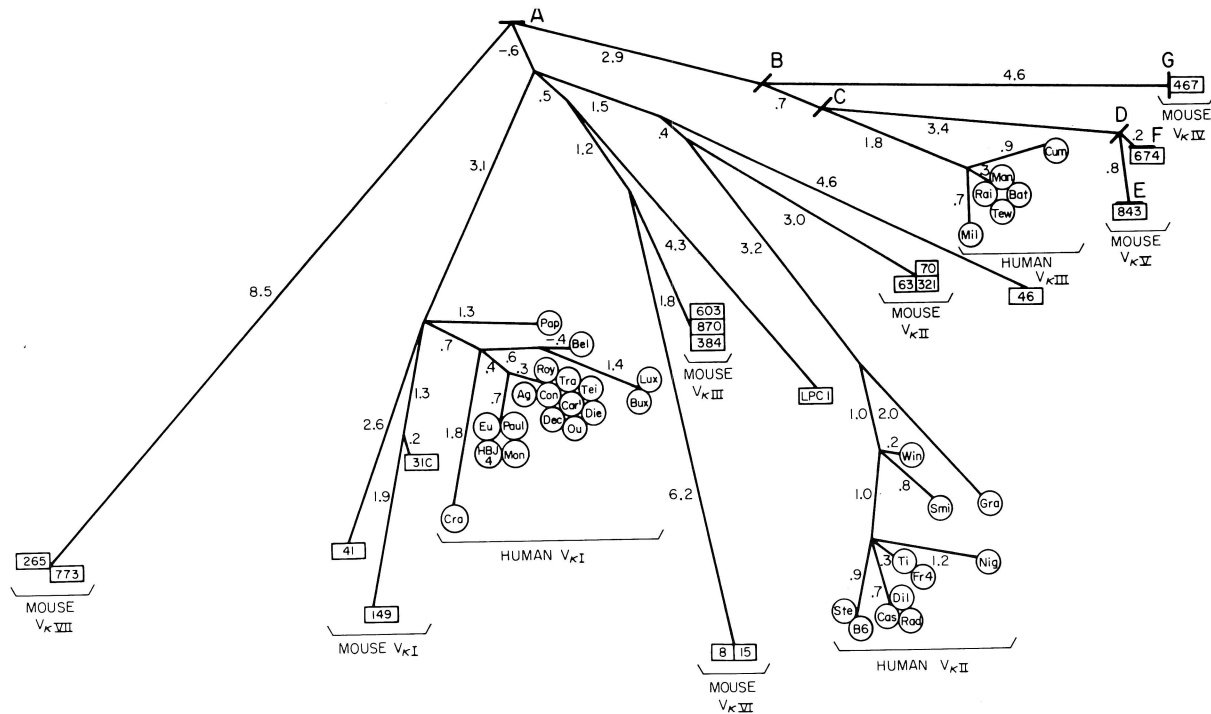


FIGURE 8. A genealogic tree for positions 1-22 of  $V_{\kappa}$  regions from man and mouse. The major branches on this genealogic tree are indicated by species name and subscripted Roman numerals (e.g. mouse  $V_{\kappa I}$ ). In this tree no correction has been made for superimposed mutations (FITCH and MARGOLIASH 1967) (Adapted from SMITH *et al.* 1971.).



line level exists at A, 12.3 mutational events are required to generate M70, 8.3 to produce M63 and 9.7 to produce T124 and M321. Furthermore, 5.3 parallel mutational events would be required for M63 and 8.2 parallel mutational events for T124 and M321. This number of parallel mutations is unacceptable to most theories of Ab diversity. If the germ line levels

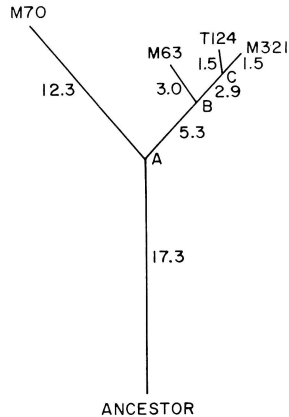


FIGURE 9. A genealogic tree of four mouse  $V_{\kappa}$  regions with "limited" sequence differences. Kindly prepared by W. BARKER and M. DAYHOFF according to DAYHOFF, 1972.

exists at B, 3.0, 4.4, and 4.4 mutations would be required for M63, T124 and M321 respectively. In addition, 2.9 parallel mutations would still be required for proteins T124 and M321. Only by moving the germ line level to C can parallel mutations be avoided. Even from the viewpoint of somatic theories, the four mouse  $\kappa$  chains which appeared nearly identical at their amino terminus apparently are coded by two (one for M70 and B) or even three (one for M70, one for M63 and C) germ line genes.

One additional line of evidence also suggests that the  $V_{\kappa}$  regions of the BALB/c mouse are coded by many germ line genes. In an examination of about 50 consecutive Bence Jones proteins, one pair was observed which appears to be identical by immunologic (PERIMAN, personal communication), peptide map (HOOD, unpublished data), and electrophoretic criteria (PERIMAN, personal communication). The sequence analysis of both of these proteins is nearly complete and it appears that they will indeed be identical (BARSTAD, PERIMAN and HOOD, unpublished data). A statistical calculation can be carried out to determine how large the pool size of different  $V_{\kappa}$  regions must be such that a sample of 50 yields one identical pair (see COHEN and MILLSTEIN 1967). At the 90% confidence level the pool size ranges between 700 and 10,000 germ line  $V_{\kappa}$  genes.

Thus the extensive screen of amino terminal sequences, the examination of four similar  $\kappa$  chains and even the examination of two apparently identical  $\kappa$  chains suggest that the  $V_{\kappa}$

family in the BALB/c mouse has a large family of distinct germ line genes. The mouse  $\lambda$  family presents a striking contrast.

#### The lambda family

Fifteen nearly complete mouse lambda sequences have been examined to date (WEIGERT and COHN, personal communication; WEIGERT *et al.* 1970) Nine of these fifteen sequences are identical and the others differ from the common sequence by one to four base substitutions. Accordingly, it has been postulated that the variant  $V_\lambda$  genes are derived from a single ancestral gene through random somatic mutation and selection (WEIGERT *et al.* 1970, COHN 1970). Whether the variant  $V_\lambda$  sequences are due to somatic mutation or in fact are coded by germ line genes, it appears that the sequence diversity and presumably the number of germ line  $V_\lambda$  genes is highly restricted in the BALB/c  $\lambda$  as compared to the  $\kappa$  family.

#### Hypervariable regions

An important sequence pattern, the hypervariable regions, should be considered at this point. When Ig light chains from man and mouse are compared to one another it appears that three distinct regions are far more variable than the remainder of the V region (WU and KABAT 1970). These regions are termed the hypervariable regions and comprise about 25% of the V region (approximately residues 25-34, 50-56, and 89-97). These regions are probably related to the antigen binding site as will be discussed subsequently. It is remarkable that all of the mouse  $\lambda$  variants and most of the substitutions among the three closely related  $\kappa$  proteins, T124, M63 and M321, fall into these regions (Figure 10). Thus selection must favor for clonal expansion those lymphocytes which have substitutions in the hypervariable regions, whether these mutations occur in the germ line or in the soma. A question of considerable interest is whether these hypervariable regions represent hot spots of mutation at the DNA level.

### 5.2 MESSENGER RNA-DNA HYBRIDIZATION EXPERIMENTS

An additional approach to the question of antibody diversity, mRNA-DNA hybridization, is in the early stages of study (STAVNEZER AND HUANG 1971, SWAN *et al.* 1972). Putative messenger RNA has been isolated from a number of myeloma light chain tumors (BALB/c  $\kappa$  chains) and hybridized to DNA from various somatic cells (STORB 1972, DELOVITCH and BAGLIONI 1973). With appropriately purified mRNA and under appropriate hybridization conditions, the rate of the bimolecular reaction which leads to RNA-DNA duplexes gives an estimate of the number of complementary gene copies in the somatic cell genome. All of the technical complexities of this experimental procedure lead to low estimates of the V gene number, nevertheless in two laboratories the numbers obtained ranged between 40 on the one hand to thousands on the other hand (DELOVITCH and BAGLIONI 1973, STORB, personal communication). Since the fraction of the total number of V genes with which any myeloma mRNA can hybridize is uncertain, any V gene

estimate which comes from these experiments must be multiplied by an unknown factor to obtain the true V gene number. Perhaps competition experiments with a variety of mRNAs will allow some estimate of this unknown factor. In any case, the  $V_K$  genes of the mouse also appear by this experimental approach to be coded by multiple germ line genes. The major reservation with regard to these studies is purity of the mRNA, for there are indications that even the "purest" of light chain mRNAs may be cross contaminated with a significant amount of other mRNAs (AVIV *et al.* 1973).

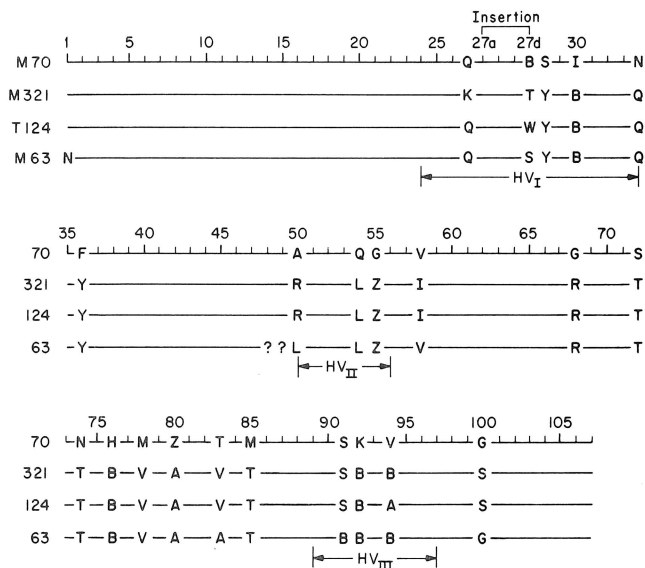


FIGURE 10. Amino acid sequence comparison of M70, M321, T124, and M63 indicating the three hypervariable regions, HV<sub>I</sub>, HV<sub>II</sub> and HV<sub>III</sub>, as defined by WU and KABAT (1970). Where proteins are identical they are represented by a horizontal line; where different they are represented by the one letter amino acid code of DAYHOFF (1972) (see Table 5). (From MCKEAN *et al.* 1973b).

### 5.3 AN ESTIMATE OF ANTIBODY DIVERSITY BY ELECTROFOCUSING

Polyacrylamide gel isoelectric focusing with the use of radioactive antigen overlay and autoradiography for the detection of specific isoelectric spectra has been employed to estimate the number of different anti-NIP (3-nitro, 4-hydroxy-5-iodophenylacetyl) antibodies in inbred CBA/H mice (KRETH and WILLIAMSON 1973). Purified myeloma proteins and monoclonal antibodies show a characteristic isoelectric spectrum consisting of a major band and a small number of minor bands. The Ig produced by a single clone of cells is biosynthetically homogeneous, but postsynthetic changes, such as deamidation,

generate microheterogeneity. This is useful because it theoretically permits up to  $5 \times 10^4$  different isoelectric spectra (Ab clones) to be detected by this method. Cell transfer (dilution) experiments were carried out from immunized donors to irradiate hosts so as to see individual clones of anti-NIP producing cells in isolation. The independent occurrence of anti-NIP molecules with identical isoelectric spectra (presumably identical sequences) was observed five times in a sample of 337 antibody molecules. This frequency permitted a statistical estimate of the minimum number of different anti-NIP molecules potentially capable of being elicited in CBA/H mice. The most likely value of this pool is 8,000 with a 98% probability that the pool size exceeds 3,000 and a 95% probability that the pool size is less than 30,000. Taking into account the fact that  $V_H$  and  $V_L$  regions contribute to antibody specificity and the fact that two  $C_H$  regions were seen in these antibodies, one may estimate that to generate a pool of 8,000 antibodies about 70  $V_H$  and 70  $V_L$  genes are required. As will be discussed subsequently, it appears unlikely that random somatic mutation could generate antibodies with identical isofocusing patterns because of the unattractive requirement of parallel mutation (see next section). Hence the repeat patterns seen here do not fit a somatic model of diversity, rather these repeats suggest that the corresponding genes are carried in the germ line. Accordingly, by amino acid sequence analysis, by mRNA-DNA hybridization studies, and by isoelectric focusing analyses the number of germ line  $V_K$  genes in the mouse is large. In each case the numbers obtained are minimum estimates.

#### 5.4 THE PROBLEM OF ANTIBODY DIVERSITY REVISITED

Let us consider once again each of the theories of antibody diversity in terms of the V region sequence data.

##### 5.4.1 Somatic hypermutation

This model requires the *ad hoc* assumption that a special mutational mechanism operates on V genes, but not on C genes. It can explain the fixation of relatively large numbers of mutations in a given lymphocyte line, but it cannot account for parallel mutation unless somatic mutation is coupled with some type of intense selective pressure. In addition, the types of substitutions seen in V regions of antibodies are similar to those seen in evolutionary related sets of proteins such as the globins and cytochrome c (HOOD and TALMAGE 1970). Accordingly, there is no evidence for a special hypermutational mechanism. This hypothesis appears to have been abandoned by most immunologists (MILSTEIN and PINK 1970).

##### 5.4.2 Normal somatic mutation

This hypothesis implies that ordinary somatic mutation generates occasional useful V region variants which are clonally expanded by some type of selection. It must be stressed that after a single useful mutation has occurred, the

the variant lymphocyte must be clonally expanded before a second useful mutation can be fixed in this same variant line. For example, assume that the ordinary rate of somatic mutation is  $10^{-5}$  per cell division per genetic locus. Thus  $10^5$  cells of variant 1 must be produced before a second mutation will occur in this same V gene. The number of variant lymphocytes required will be much larger than  $10^5$  in that most mutations are not useful (perhaps 1 in  $10^6$ ?). Thus most immunologists with a leaning toward this theory would hesitate to require more than three or four mutations in each lymphocyte line (COHN, personal communication). Once again parallel mutation is difficult to explain by this theory. For example, certain myeloma globulins of the BALB/c mouse that were derived from independent tumors appear to be identical (HOOD *et al.* 1973; also see CAPRA and KUNKEL 1970). To avoid parallel mutation identical pairs of V regions must be coded by a distinct germ line gene (COHN 1971). In this particular case the level of germ line genes has been moved to the terminal most twigs of the genealogic tree and, accordingly, the germ line and somatic theories are in agreement. A second necessary component of this theory is some type of intense selective pressure to expand the clone size of the useful variants (see COHN 1970, COHN 1971, JERNE 1971). This theory postulates that as a lower limit 20 or more V genes code mouse  $\kappa$  chains and a single V gene codes mouse  $\lambda$  chains.

#### 5.4.3 Somatic recombination

This model requires that two or more V genes undergo recombination events to generate variants. The recombining genes can be highly selected so that many of the variants are useful. Thus a single recombination event could generate a variant V gene quite different from that of either parent V gene. However, there is no evidence for somatic recombination when the V region sequence data are carefully analyzed (SMITH *et al.* 1970). This difficulty can be circumvented by postulating an even larger family of V genes that can undergo recombination. Indeed, it has been postulated that each V region subgroup has 10-100 genes which can recombine (GALLY and EDELMAN 1970). Thus the mouse  $\kappa$  family with more than 20 subgroups apparently is coded by 200-2000 genes according to this theory.

#### 5.4.4 Germ line

This theory suggests that the level of germ line genes on the genealogic tree is at the terminal most twigs of the genealogic tree (level F - Figure 6). Clearly some Ab families (mouse  $\kappa$ ) have a much larger number of germ line genes than others (mouse  $\lambda$ ).

All contemporary theories of antibody diversity appear to be multigenic. Once those who favor a somatic point of view concede 20-200 V genes for the mouse  $\kappa$  family (COHN 1971, GALLY and EDELMAN 1970), it becomes extremely difficult to distinguish among the alternative theories. Additional V region sequence data or mRNA-DNA hybridization experiments can increase the required number of germ line V genes, but it can never

resolve the fundamental issue which distinguishes these two points of view. Namely, it is apparent that in any multigenic system somatic mutation and recombination will occur. The real issue is whether the vertebrate organism can effectively select from among these random somatic variants, lymphocytes which produce higher affinity antibody so as to enlarge the useful library of antibody diversity. At this point in time it is not obvious how the problem of Ab diversity can be approached experimentally. Now let us consider the organization of antibody genes.

## 6. THE ORGANIZATION OF IG GENES IN ANTIBODY FAMILIES.

The data described in the preceding section can most easily be explained by assuming that there are three families antibody genes located on three nonhomologous autosomes (Figure 11). Each family contains separate V and C genes. The V and C genes probably are located in different areas of that chromosomal region coding the antibody family because infrequent recombination events allow, for example, all of the

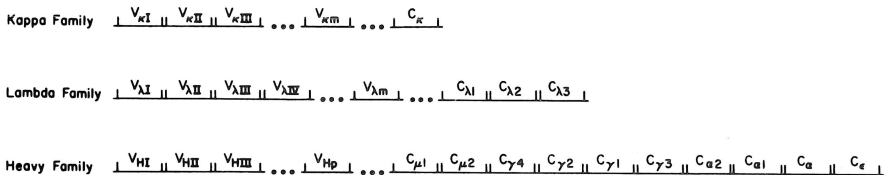


FIGURE 11. A model of the V and C genes present in the three antibody families of the human germ cell.

rabbit V $H$  genes of one group a allotype to be joined to a new C $\gamma$  allotype (group e) (MAGE 1971). This would not be possible by a single recombination event if the V and C genes were intermingled. Let us consider the evidence which favors separate V and C genes.

### 6.1 SEPARATE V AND C GENES

#### Biological data

During the maturation of the immune response there is, over a period of a few weeks, normally a shift from one class of antibody (IgM) to a second (IgG). Despite this shift in Ig class, the specificity of the antibody is maintained. During the transition period (IgM  $\rightarrow$  IgG), single antibody producing cells appear to have specific IgM and IgG antibodies (NOSSAL *et al.* 1964). Thus a single cell appears to change classes of antibody while maintaining specificity.

Serologic data

An explanation for the observations noted above was afforded by three serologic observations. First, specific antiserum (termed "idiotypic" antiserum) can be produced against the V regions of specific antibody. Idiotypic determinants are maintained during the shift from cells producing IgM Ab to those producing IgG Ab during the maturation of the immune response (OUDIN and MICHEL 1969). This implies that similar (or identical) specific V regions are attached to the C<sub>μ</sub> and to the C<sub>γ</sub> regions. Second, the rabbit group a allotypes are found on all classes of H chain (PRAHL *et al.* 1970). Three rabbit C<sub>H</sub> genes have distinct allotypes (C<sub>γ</sub>, C<sub>μ</sub> and C<sub>α</sub> - see Table 4). Thus the association of the V<sub>H</sub> allotypes with each of three distinct C<sub>H</sub> genes suggests that the V and C genes are indeed separate. Finally, a single myeloma patient has been observed with two myeloma proteins (an IgG and an IgM) produced by two separate clones of plasma cells (LEVIN *et al.* 1971). Anti-idiotypic antiserum directed against either protein reacts equally well with the other. Thus it appears that V<sub>H</sub> and V<sub>L</sub> regions for these two proteins are serologically identical. This supposition has been supported by amino acid sequence analysis.

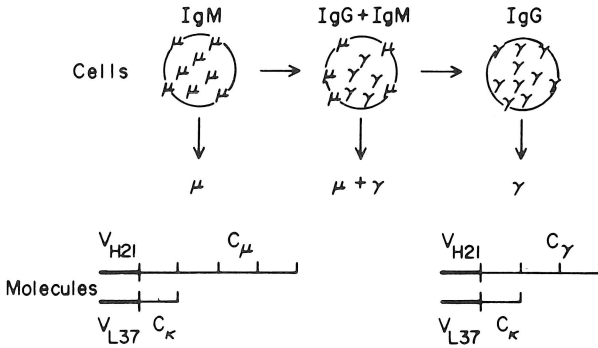


FIGURE 12. A model that depicts the events which occur during the maturation of the immune response as reflected in the shift from the production of IgM to IgG antibody. It appears likely that a single cell (upper figure) goes through three states of antibody synthesis: (1) μ synthesis; (2) γ and/or μ synthesis; and (3) γ synthesis. During this shift the antibody molecule (lower figure) may remain invariant apart from a shift from the C<sub>μ</sub> to C<sub>γ</sub> gene. (from HOOD 1972).

Sequence analysis

The biconal myeloma proteins described above appear to have identical light chains (WANG *et al.* 1969) and identical V<sub>H</sub> regions (WANG *et al.* 1970, 1971, 1973) insofar as these chains have been analyzed. Apparently, only the C<sub>H</sub> regions differ in these two molecules (Figure 12). Thus the same V<sub>H</sub>

region appears to be joined to distinct  $C_H$  regions (also see PERNIS *et al.* 1971). In addition, the existence of multiple  $V_K$  regions joined to a single  $C_K$  region in man and in mouse suggests that the corresponding genes are separate (HOOD *et al.* 1967, HOOD and EIN 1968, HOOD 1972). In summary, for certain cases multiple  $C_H$  regions can be associated with a single  $V_H$  sequence, whereas in other cases multiple V regions appear to be associated with a single C region. In each case the simplest explanation appears to be separate V and C genes (Figure 11). However there is an alternative to the two gene: one polypeptide model (Figure 1).

6.2 THE ALTERNATIVE - THE ONE V-C GENE MODEL

This model suggests that for each germ line V gene there is a contiguous C gene (Figure 13). This model requires that

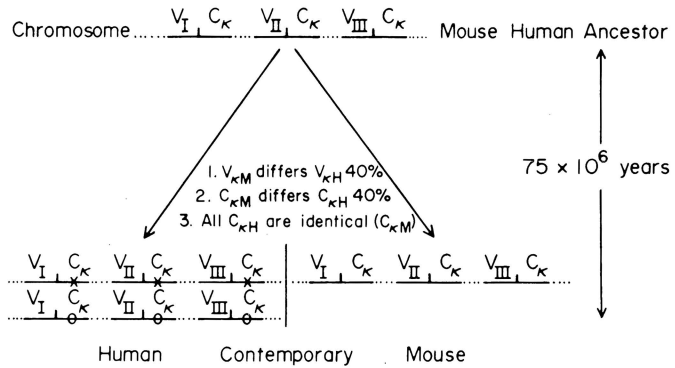


FIGURE 13. A model that depicts the evolution of three complete  $V_K$ - $C_K$  genes in the mouse and human evolutionary lines.  $V_I$ ,  $V_{II}$ , and  $V_{III}$  designate genes encoding major branches of the V region genealogic tree (V region subgroups).  $V_{KM}$  designates a mouse kappa V region and  $V_{KH}$  a human kappa V region. Ancestor designates the ancestral genes which the mouse and human evolutionary lines diverged 75 million years ago. Contemporary indicates contemporary mice and men. The amino acid sequence differences (or identities) are indicated by notes 1, 2 and 3. The x and o in the contemporary human  $C_K$  genes represent a single amino acid substitution which segregates in a Mendelian fashion. Accordingly, for this model one must envision that (1) multiple  $C_K$  genes within a species remain constant in spite of the tremendous divergence which occurs among their V regions, (2) allelic substitutions can arise in multiple human  $C_K$  genes and (3) meiotic recombination does not occur to scramble the human genetic markers. (From HOOD 1972).

each antibody family evolve multiple, identical C genes. This model will not be considered in detail as the data presented in



the next section suggest that it is very unlikely (see legend to Figure 13).

### 6.3 A HYBRID IMMUNOGLOBULIN MOLECULE

Serologic studies with certain Gm markers revealed a human family in which individuals were found with hybrid immunoglobulin molecules (KUNKEL *et al.* 1969a). These molecules were comprised of the amino terminal portion of the C $\gamma$ 3 gene and the carboxy terminal portion of the C $\gamma$ 1 gene. One individual appeared to be homozygous for the gene which codes this hybrid molecule; that is, he was completely lacking normal  $\gamma$ 1 or  $\gamma$ 3 molecules and in their place he had hybrid  $\gamma$ 1- $\gamma$ 3 molecules. These observations are critical because they suggest that the hybrid gene arose by homologous but unequal crossing over in a fashion analogous to that observed between the  $\beta$  and  $\delta$  globin genes to generate a hybrid globin gene--hemoglobin Lepore (BAGLIONI 1962). More important, these observations suggest that there is a single germ line copy of the C $\gamma$ 1 and C $\gamma$ 3 genes. If there were multiple copies of these C genes, each pair would have to simultaneously undergo a similar recombinational event in order to generate an individual homozygous for the gene(s) coding the hybrid immunoglobulin. This appears a very unlikely possibility. Accordingly, the presence of a hybrid immunoglobulin chain in an individual homozygous for this trait strongly supports the existence of separate V and C genes (Figure 11). Now we must consider the level in protein synthesis at which the V and C genes are joined.

## 7. A JOINING MECHANISM - ANTIBODY EXPRESSION

Accepting for the moment that separate V and C genes do exist, these genes (or their products) can be joined at the DNA, the RNA, or the protein level. Although no unequivocal experiments rule out union at the protein level, two observations render joining at this level unlikely. First, double label experiments have been carried out to ask how many growing (initiation) points does the immunoglobulin polypeptide chain have? It appears that there is a single growing point for the translation of the immunoglobulin light chain and, accordingly, this suggests that the light chain is synthesized from a single messenger (KNOPF *et al.* 1967, also see FLEISCHMAN 1967). Second, a 12S messenger RNA has been isolated from myeloma cells which is capable of synthesizing intact light chains in a heterologous cell-free system (STAVNEZER and HUANG 1971, RALPH and RICH 1971, MILSTEIN *et al.* 1972). This messenger RNA is about twice the size expected for the light chain mRNA, and, as with other mRNAs, appears to have a string of poly A nucleotide residues at the 3' terminus (SWAN *et al.* 1972). Thus it appears that antibody polypeptides are probably synthesized from a conventional messenger RNA which codes the entire Ig chain. Experimental evidence does not permit us to distinguish between DNA level and RNA level joining of V and C genes, although a variety of models have been postulated for joining at the DNA level.

### 7.1 JOINING MODELS

Although no compelling evidence exists for joining at the DNA level, the general implications of such a mechanism, as we

shall see in just a moment, are fascinating and have led to three models for somatic, intrachromosomal, DNA level joining of V and C genes. There are two joining models which propose that one of a tandem array of V genes is moved next to a C gene in another part of the chromosome by a crossing over event similar to that which integrates the phage lambda genome into a specific part of the *E. coli* genome. In the first model, diagrammed in Figure 14, a random translocation and integration

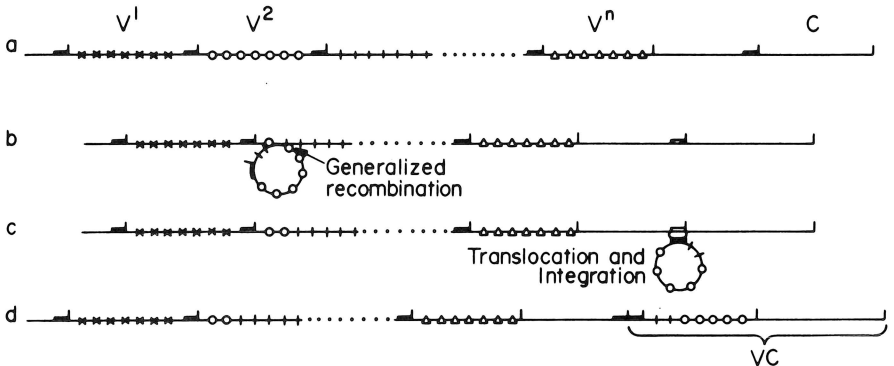


FIGURE 14. Hypothetical mechanism for V and C gene joining. The tandem V genes have base substitutions which result in variations in sequence and are indicated along each V gene by the symbols x, o, and Δ. Intrachromosomal recombination between adjacent V genes would lead to the formation of a V gene episome, b. This process would be analogous to generalized recombination seen in bacterial systems i.e., recombination would occur wherever homologous DNA strands pair to form hybrids, not necessarily at any specific nucleotide sequences. The episome formed could therefore contain new sequence variants composed of sequences from two adjacent V genes. Integration of the episome, c, is assumed to require an enzyme capable of recognizing specific nucleotide sequences in both the episome and the C gene as a first step in the formation of a complete VC gene, d. The enzyme is shown bound to the attachment regions of the episome and the C gene. (From GALLY and EDELMAN 1970).

of V into C genes is proposed (GALLY and EDELMAN 1970). The second model, diagrammed in Figure 15 suggests that joining is effected through a special organelle that operates in a programmed fashion to methodically express all of the germ line V gene information in differentiated immunocytes. The third model, the lateral array model, is depicted in Figure 16 and envisions a branched network of DNA forming a lateral array of V genes, all of which are simultaneously adjacent to a similar lateral but less extensive array of C genes. This last model obviates the problems of translocating genes by introducing new rules for replication and transcription. The details of each

of these models are discussed in the legends of the figures. It will be difficult to distinguish experimentally among these models and perhaps even more difficult to delineate the correct joining mechanism. The availability of purified Ig mRNA and reverse transcriptase and various techniques such as DNA-RNA hybridization (see DAVIDSON *et al.* 1973), *in situ* hybridization (PARDUE and GALL 1970) and gene mapping by electron microscopy (DAVIS *et al.* 1971) offer some promise from distinguishing among the tandem and the lateral array models. One possible difference between the two tandem models is whether the germ line information is expressed in an orderly (programmed) or in a random fashion.

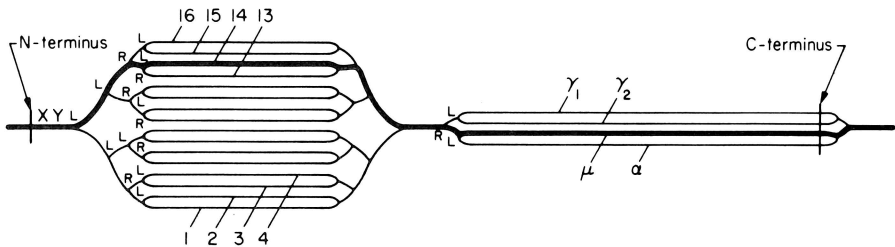
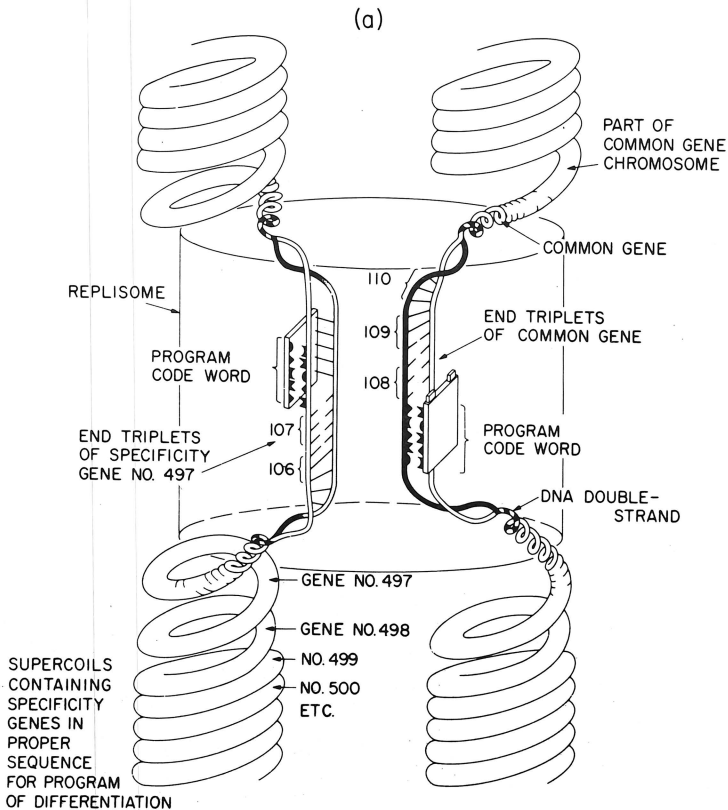


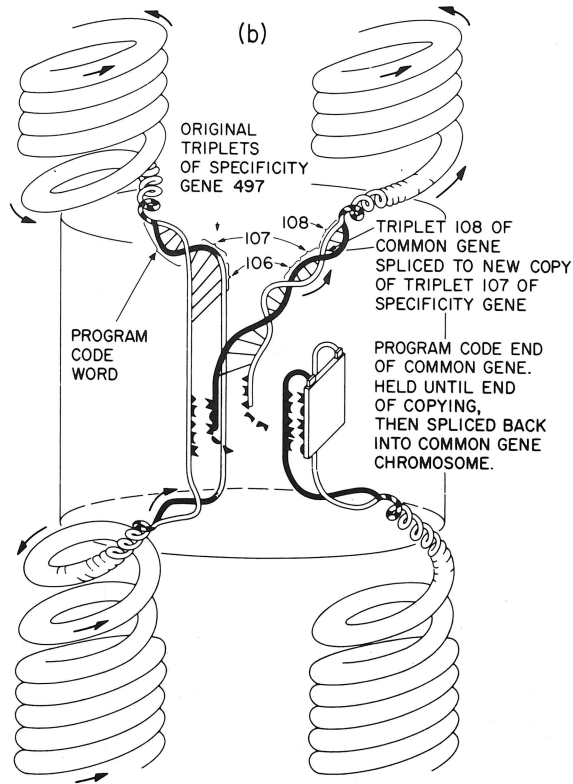
FIGURE 16. A hypothetical locus for an immunoglobulin heavy chain with 16 lateral genes for variable regions ( $V_1$  to  $V_{16}$ ) and four lateral genes for constant regions ( $C_{\gamma_1}$ ,  $C_{\gamma_2}$ ,  $C_{\mu}$ , and  $C_{\alpha}$ ). Each line represents a DNA double helix. Commitment to one of the possible proteins coded by the locus is determined by the random setting of its DNA forks in left (L) or right (R) configurations; x represents a position at which allotypic variants could be found easily. The path of an RNA polymerase molecule is indicated by the heavy line; it would transcribe messenger RNA corresponding to  $V_{14}-C_{\mu}$ . (From SMITHIES 1970).

## 7.2 RANDOM VS. PROGRAMMED EXPRESSION OF GERM LINE INFORMATION

A programmed model for gene expression would read out the information from tandemly linked genes in an orderly and sequential fashion much as instructions are read from a punched tape into a computer (see legend to Figure 15). Presumably one precursor lymphocyte would produce a successive series of differentiated daughter cells, each expressing one successive V genes from the tandemly linked series (Figure 15). Thus this cell division is asymmetric in that the precursor cell would remain undifferentiated while producing a series of differentiated daughter cells. Two implications could follow, depending on the length of time required for the genotype to be read into primotype (see last section of 3, page 79). First, there may be an orderly and temporally reproducible maturation of specificities in the immune response which correspond to the successive



Base recognition at splice region



Copying of specificity gene to commit progeny cell to a particular kappa chain

© CIT 1967

FIGURE 15. Proposed "copy-splice" mechanisms for production of complete V-C antibody genes. The process is depicted as a conservative copying, involving two different stretches of double-stranded DNA. The specificity (variable gene) strand (left in a and b) contains many individual specificity (variable) genes separated by "program code" sequences of DNA; the other strand (right in a and b) contains the unique common region gene, plus its particular program code DNA. a: Base recognition at splice region. The replisome is shown as containing two adapters containing anticodons which are able to recognize and pair with the program code words on the DNA strands, thus bringing the specificity and common genes into the correct alignment. These adapters might be considered somewhat analogous to ribosomal transfer RNA molecules. The specificity (variable) gene adapter is positioned at the start of a particular gene (no. 497), after having been involved in the copy-splicing of gene no. 496 in the previous round of cell division. It thus acts as a program reader, enabling the various specificity genes to be expressed in an orderly sequence. The recognition process is shown as being exclusively between genes and adapters; this is certainly not obligatory, and there may be direct pairing between common and specificity strands in the splice region or elsewhere. b: Copying of specificity (variable) gene to commit progeny cell to a particular receptor. The strand containing the common gene has been severed immediately after triplet 108, and the distal ends have been temporarily attached to the replisome. For clarity, we have shown the remainder of this strand as being stationary during the copying process--as would be probable if the replisome were membrane bound. The free ends of the common gene then act as growing points for the addition of nucleotides, and a copy of the specificity (variable) gene is built up onto it, working back toward the first triplet. Some of the free energy of reaction is used to drive the DNA strands through the replisome. As the strands emerge they take up the characteristic double-helix structure, producing coils and super-coils and turning the DNA in the direction shown. When the copy is complete, the ends of the common gene strand are joined (spliced), remaking a continuous stretch of chromosomal DNA. The net result of these processes is that the common gene chromosome contains a stretch of DNA coding for a specific, intact light chain subunit, and the program for antibody production has progressed one more step. The mechanism illustrated resembles the "copy-choice" mechanism proposed by some molecular geneticists to explain certain types of recombination. The "breakage-reunion" mechanism might well be used to accomplish essentially the same process. It is not illustrated here due to the complexity of the drawing required. In either case an asymmetric cell division is assumed as the committed daughter cell is produced from the stem cell. (From DREYER 1970).

expression of tandem V genes. Second, if the committed descendants of a single undifferentiated lymphocyte could be clonally expanded and examined, one might expect them to express distinct but adjacent (and possibly similar) tandem genes (see legend to Figure 15). There are two lines of evidence that are consistent with these suppositions, although each is subject to alternative interpretations. The observations are, however, sufficiently provocative to warrant presenting here in some detail.

With regard to the temporal maturation of antibody specificities, studies carried out in the fetal lamb demonstrate that immunologic competence to all antigens does not arise simultaneously in the developing fetus--rather, there is a stepwise maturation of immunologic competence to different antigens at different stages of development (SILVERSTEIN *et al.* 1963, STERZL and SILVERSTEIN 1967). Among the antigens tested, antibody formation occurs first to bacteriophage ØX174, later in gestation to ferritin and hemocyanin, and even later to ovalbumin. The immune response continues to mature even after term. The precision with which the fetus develops competence to a given antigen at a given stage of gestation is remarkable. For example, the fetal lamb developed competence to ovalbumin at 120 to 125 days of gestation (150 days gestation period) in some 60 animals tested (SILVERSTEIN and PRENDERGAST 1970). Thus the ability to respond to a series of antigens appears to be a carefully controlled and temporally reproducible sequence within a given species. One explanation for this temporally reproducible expression of antibody activity is that sequential tandem germ line genes are being expressed in immunocytes in a reproducible temporal fashion. An alternative explanation is that these observations relate to the maturation of the antigen processing system.

An observation in one patient with Waldenstrom's macroglobulinemia, a clonal proliferation of IgM producing immunocytes, is consistent with the sequential expression of distinct but related  $V_{\kappa}$  genes. Certain afflicted individuals can produce two, three, or even four molecular species of immunoglobulins from these cancer-like cells (HARBOE *et al.* 1972). In one case the light chains of patient, Tö, with four distinct IgM molecules were examined (Table 6) (HANNESSTAD and SLETTEN 1971). Although the sequence data is limited, these four  $\kappa$  chains have two interesting properties. First, they probably belong to an extremely infrequent V region subgroup in that the sequence Val-Val at positions 2 and 3 has not been observed in more than 100 human  $V_{\kappa}$  regions previously characterized (DAYHOFF 1972, HOOD and TALMAGE 1970, HOOD, unpublished observations). Thus the probability of four unrelated myeloma proteins expressing this new sequence is very unlikely. Second, three of the four  $V_{\kappa}$  regions differ in their amino terminal four residues by one or two substitutions. Mechanisms to be discussed subsequently for the evolution of antibody families suggest that similar V genes (e.g. those of a particular subgroup) may be adjacent in the mammalian genome. Thus the Waldenstrom's transformation may have occurred in an undifferentiated precursor lymphocyte. Subsequent to the transformation, differentiated daughter cells were produced--each transformed and each committed to the synthesis of a

distinct but closely related and possibly adjacent  $V_{\kappa}$  gene (Figure 15). An alternative explanation would be that each of these four clones is derived from a single ancestor by somatic mutation. This seems unlikely because too many mutational

Table 6. Amino terminal sequence of  $\mu$  and  $\kappa$  chains from the polyclonal sera Næ and Tö.\*

Component	1	2	3	4	5	6	7	8	9	10	11	12	13
$\kappa$ chain Næ 1	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ser	Ser	Leu	Ser	Ala
$\kappa$ chain Næ 2 <sup>†</sup>	-----												
$\mu$ chain Næ 1	Glu	Val	Gln	Leu	Glu	Gln							
$\mu$ chain Næ 2	-----												
$\mu$ chain Tö 1	Glx	Val	Glx	Leu	Val	Glx							
$\mu$ chain Tö 2-4	-----												
$\kappa$ chain Tö 1	Asp	Val	Val	Leu	Thr	Gln	Ser	Pro	Leu (Thr/Ser)	Leu	Pro	Val	
$\kappa$ chain Tö 2	Asp	Val	Val	Met									
$\kappa$ chain Tö 3	Glu	Val	Val	Met									
$\kappa$ chain Tö 4	Asp	Val	Val	Leu									

\* Data from Hannestad & Sletten (1971).

<sup>†</sup>Solid line indicates sequence identical to the one immediately above.

events seem to be required (e.g. 1 substitution in 4 residues extrapolates to about 25 substitutions in 100 residues--the length of the V region). Thus these preliminary observations are consistent with the sequential read-out of similar (adjacent) germ line genes or at least with a preliminary commitment or "focusing down" on an unusual group of V genes in one lymphocyte line.

Other features of control mechanisms can also be studied in these patients. Table 6 shows that the four  $\mu$  chains from Tö are identical for six residues. Are these heavy chains identical throughout their  $V_H$  regions? If so, commitment to a  $V_H$  gene may in this case have occurred prior to that for the  $V_L$  gene. The Ig chains from a second patient, Næ, show the inverse relationship, namely  $\kappa$  chain identity with H chain variation. Perhaps in this case the commitment to a  $V_L$  gene occurred prior to that for the  $V_H$  genes. Clearly the multi-clonal myeloma and Waldenstrom's Ig molecules afford some exciting possibilities for dissecting various levels of control in the synthesis of Ig.

### 7.3 V AND C JOINING - A MECHANISM FOR DIFFERENTIATION

It is important to stress that a single antibody producing cell generally appears to synthesize just a single molecular species of antibody. This supposition is supported by a variety of experimental data. First, myeloma tumors generally synthesize a single molecular species of Ig (KUNKEL 1965,

POTTER 1967). In those rare instances where two or more myeloma proteins are present, each appears to be synthesized by a separate clone of plasma cells (SNAPPER and KOHN 1971). Second, fluorescent antibody techniques have demonstrated that normal Ig producing cells are generally restricted to the synthesis of a single  $C_H$  region and a single  $C_L$  region (CEBRA *et al.* 1966, PERNIS *et al.* 1965, MAKELA and CROSS 1970). The question of whether Ig producing cells are synthesizing a single  $V_L$  and  $V_H$  regions is obviously more difficult to answer. The only reasonable criterion available for the study of heterogeneity in the V regions of single cells is that of specificity. The consensus is, however, that generally one cell makes antibody of a single specificity (GREEN *et al.* 1967a, GREEN *et al.* 1967b, MAKELA and CROSS 1970). Finally, the individual Ig producing cell also seems to be restricted to the synthesis of one of two allelic products in animals heterozygous for a given genetic marker (CEBRA *et al.* 1966). This phenomenon is termed "allelic exclusion" and superficially resembles that observed for genes on the X chromosome of female mammals. In female somatic cells, one of the two X chromosomes is randomly condensed into a Barr body (heterochromatin) and, accordingly, all the genes contained in the condensed linkage group are not expressed. Allelic exclusion in the immune system also appears to randomly repress either allele, although there is no evidence to suggest an entire autosome is inactivated. It is, therefore, appropriate to consider how an antibody producing cell may be restricted by the synthesis of one antibody molecule.

Each of the joining models proposes a mechanism whereby a single cell might become committed to the synthesis of a single molecular species of immunoglobulin. In each case specific molecular events at the DNA level commit the immunocyte to the synthesis of a single joined VC pair for each antibody polypeptide chain. Thus differentiation in the Ab system may occur through differential gene alteration (joining) rather than by the more conventional model of differential control of transcription. Evolutionary considerations suggest that such a joining mechanism may play a general role in the differentiation of other complex systems in that it may provide the means whereby their effector cells can also become committed to a single "unit" of information (see section 8.8, page 128). Let us now turn to these evolutionary considerations.

## 8. THE EVOLUTION OF ANTIBODY MOLECULES

The evolution of antibody molecules employs essentially all of the known mechanisms of molecular evolution. This system also demonstrates the power of using the amino acid sequence information from contemporary proteins to deduce a detailed history of the evolution of a given gene family.

### 8.1 HOMOLOGY UNITS AND GENE DUPLICATION

All antibody polypeptide chains are comprised of sequence units about 110 amino acid residues in length with a centrally



placed disulfide bridge spanning about 60 residues (Figure 17). These sequence units are termed homology units (Figure 3). The IgG heavy chain has four such homology units ( $V_H$ ,  $C_H1$ ,  $C_H2$ ,

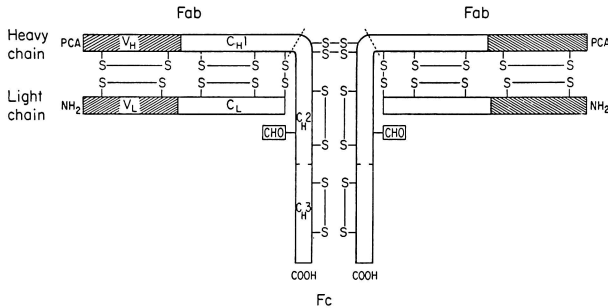


FIGURE 17. A model of the antibody molecule (IgG) depicting the linear arrangement of intrachain disulfide bonds which are centrally placed in homology units. Interchain disulfide bridges are also depicted. The boundary between the Fab and Fc is indicated by a dotted line. CHO represents a polysaccharide chain. (From GALLY and EDELMAN 1970).

$C_H3$ ), whereas the light chain has two ( $V_L$  and  $C_L$ ) (Figure 17). The homology units of the  $C_L$  and the  $C_H$  regions exhibit marked homology for one another as do those of the  $V_L$  and the  $V_H$  regions (EDELMAN *et al.* 1969, DAYHOFF 1972). No obvious homology exists between the V and the C homology units apart from a similarity in size and disulfide bridge placement, although a putative evolutionary relationship may have been obscured by a very ancient divergence. Accordingly, all antibody genes may have evolved from a precursor gene about 330 nucleotide pairs in length through the scheme depicted in Figure 18. Initially, the early precursor gene may have duplicated to form separate ancestral V and C genes. Then chromosomal translocation and/or polyploidization may have generated the three separate families of immunoglobulin genes. In each Ab family an expansion of the number of V and in some cases the C genes occurred by discrete gene duplication (Figure 18). The  $C_H$  genes also increased in size by contiguous gene duplication. Contiguous duplication generates a single gene approximately twice the size of the original ancestor by gene fusion and loss of the internal initiation and termination sites. Thus all antibody polypeptides are comprised of homology units which reflect their past evolutionary history (Figure 18). The homology units also appear to have functional significance.

## 8.2 FUNCTIONAL DOMAINS

The antibody molecule appears to be comprised of at least three distinct domains as depicted in Figure 3, two identical Fab domains and an Fc domain. Fab fragments contain fully

functional antigen binding sites, whereas the Fc fragment carries out a variety of physiologic or effector functions such as complement fixation, skin fixation, and placental

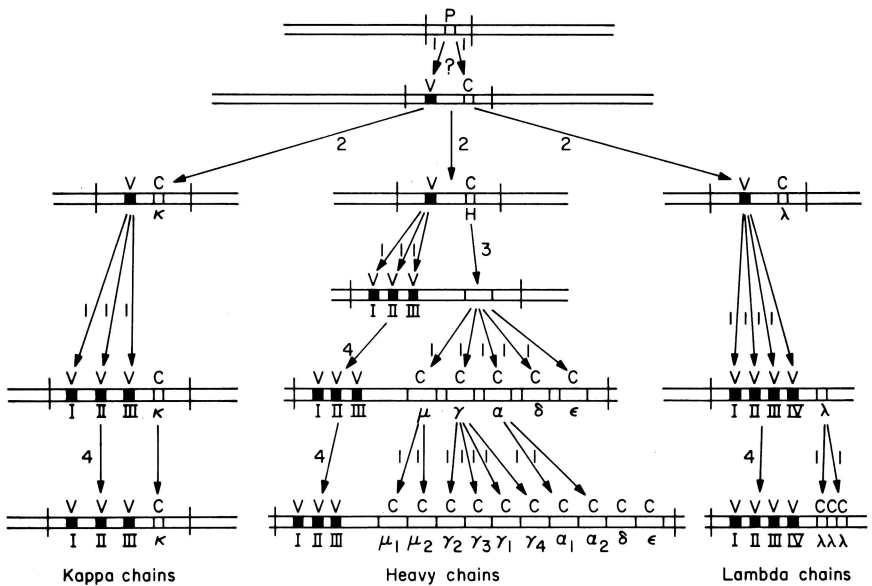


FIGURE 18. A hypothetical scheme for the evolution of antibody families. The order of gene duplication events is unknown. For example, a library of V genes probably evolved prior to the divergence of the three antibody families (see section 8.8). The shaded V genes may each represent a single locus undergoing somatic mutation and selection, a set of tandemly duplicated sequence undergoing frequent recombination, or a set of germ line genes. A number of genetic mechanisms seem to be employed in the evolution of these families as indicated by numbers adjacent to arrows. These are:

1. discrete gene duplication.
2. gene duplication by polyploidization or chromosomal translocation.
3. contiguous gene duplication.
4. coincidental evolution of multiple genes.

Mechanisms 1 and 4 are probably identical (see sections 8.5 and 8.6). (Adapted from GALLY and EDELMAN 1972).

transfer that are not directly related to the antigen binding function (see GOOD and FISHER 1971). The different classes of antibodies (e.g. IgM, IgG, IgE, etc.) appear to carry out

different kinds of effector functions (Table 2). The domain hypothesis suggests that each of the homology units of the antibody molecule folds into a compact structure and that each carries out distinct functions (EDELMAN 1970). Accordingly, each Ab domain has evolved distinct functions. X-ray crystallographic studies on Fab crystals (POLJAK *et al.* 1972) and on crystals of the entire IgG molecule (DAVIES *et al.* 1971, SARMA *et al.* 1971) are in accord with the generalizations drawn above. Thus the antibody molecule fuses together two separate and discrete kinds of functions--those related to antigen binding and those related to biologic effector functions--to generate a single molecular entity of enormous biological sophistication. For a more detailed discussion of the general structure of the antibody molecule see EDELMAN (1971).

### 8.3 THE ANTIBODY COMBINING SITE AND HYPERVARIABLE REGIONS

The antibody combining site appears to have a number of interesting features (PORTER 1971). First, biological studies have generally demonstrated that both the L and the H chains are necessary for effective antigen binding (see EDELMAN 1971). Second, sequence correlates of the active site seem to exist in that, as previously described, there are three regions of extreme sequence diversity in the comparisons of myeloma light chains (Figure 10) (WU and KABAT 1970). Three to four hypervariable regions exist for a comparable comparison of the V<sub>H</sub> regions (CAPRA 1971, KEHOE and CAPRA 1971). Two of these hypervariable regions for both the L and for the H chains are linearly juxtaposed by a disulfide bridge. Third, normal immunoglobulin H chains from inbred guinea pigs give a single major sequence in the V<sub>H</sub> region apart from three areas of extreme amino acid sequence diversity which correspond to the hypervariable regions observed in the myeloma system (BIRSHTEN and CEBRA 1971). When specific antibody is induced in guinea pigs to a variety of antigens, one major sequence is observed in two of the hypervariable regions for each specific antibody preparation and these major sequences are characteristic for each type of antibody (RAY and CEBRA 1971, CEBRA *et al.* 1971). Fourth, special antigens, termed "affinity labels," can be synthesized which have two components: (i) an antigenic determinant which positions the antigen in the binding site and (ii) a chemical group that is capable of covalently coupling to amino acid residues in or near the combining site. A variety of affinity labels have been constructed and they appear to link to antibody polypeptide chains only at or near the hypervariable regions (SINGER *et al.* 1971, HAIMOVICH *et al.* 1971, EISEN 1971, RAY and CEBRA 1971). Finally, X-ray crystallographic studies show that the antibody combining site is a large cleft bounded by both the light and heavy chains (SARMA, personal communication).

In summary, the hypervariable regions probably do comprise at least a portion of the active site cleft. It appears likely, however, that sequence substitutions outside the hypervariable regions, even if not located directly in the active site, could modulate the nature of the antigen binding site. The sequence diversity of the hypervariable regions presumably reflects

selection at the soma or germ line level for a corresponding diversity of functional antibody molecules.

#### 8.4 THE EVOLUTION OF C GENES

Most mammals appear to have two or more  $C_\gamma$  genes (Figure 11) (MILSTEIN and PINK 1970, DAYHOFF 1972). By immunologic and limited sequence criteria,  $C_\gamma$  regions in various mammals appear to be much more closely related to other  $C_\gamma$  regions in the same species than they are to  $C_\gamma$  regions in other species; that is, they have evolved as species specific clusters of sequences (MILSTEIN and PINK 1970, DAYHOFF 1972, CEBRA, personal communication). This observation implies that the  $C_\gamma$  genes of individual mammalian species have undergone a series of discrete gene duplications after the mammalian radiation which occurred about 75 million years ago (Figure 19). Accordingly, in recent evolutionary history the  $C_\gamma$  genes have

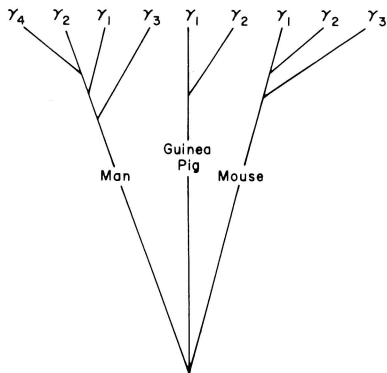


FIGURE 19. A genealogic tree of the  $C_\gamma$  genes of man, mouse and guinea pig based on serological and chemical comparisons (see text). These mammals diverged from one another about 75 million years ago.

expanded independently in various mammalian evolutionary lines by a process of repeated gene duplication. It will be interesting to determine whether the mammalian ancestor had multiple  $C_\gamma$  genes. If so, this would imply that  $C_\gamma$  genes of mammals are undergoing a process of gene duplication and deletion (gene expansion and contraction) similar to that to be described for the V genes in the next section. If, for example, birds, amphibia and fish all have multiple  $C_\gamma$  genes, then the simplest supposition is that the ancestor to these three groups and mammals did also. I suspect that a similar picture of multiple  $C_H$  genes will emerge for the IgA and IgM classes of antibody also (see GITLIN *et al.* 1973). Once again, if this is so, it would suggest that the entire family of  $C_H$  genes is undergoing gene expansion and contraction.

## 8.5 THE EVOLUTION OF V GENES

Each Ig family of V genes appears to have two unusual evolutionary features. The first is that the number of V genes contained in a given antibody family differs markedly among the mammalian evolutionary lines. This supposition is based on the observation that serum  $\lambda$  to  $\kappa$  ratios in different mammals can range from essentially all  $\kappa$  (mouse) to essentially all  $\lambda$  (horse) (Figure 20) (HOOD *et al.* 1967, 1970b). The inference is that the ratio of serum light chains reflects a corresponding ratio of diversity in the germ line families for  $V_\lambda$  and

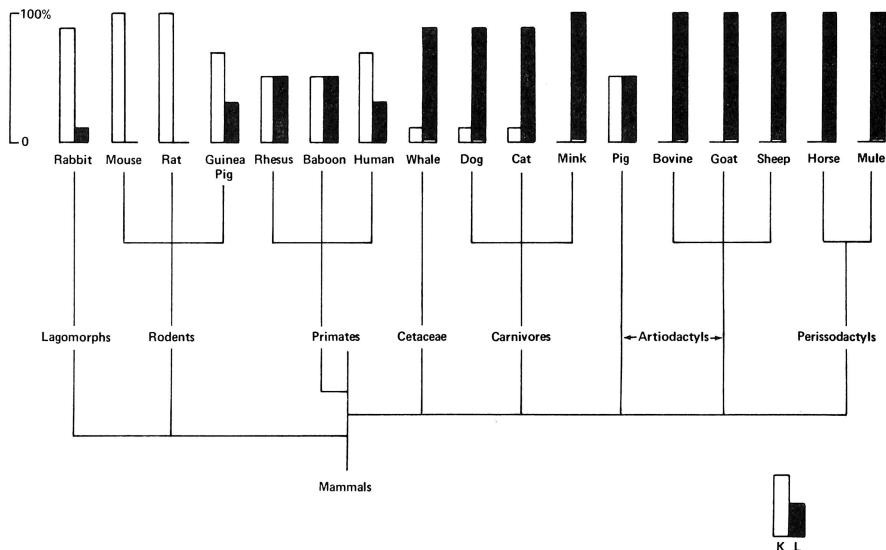


FIGURE 20. Distribution of  $\lambda$  and  $\kappa$  light chains among various mammals, based on data of HOOD *et al.* 1967 and 1970b. The vertical bar on the left above each species indicates proportion of light chains in  $\kappa$  class (K); right-hand (shaded) bar indicates  $\lambda$  chains (L). (from HOOD *et al.* 1970b).

$V_\kappa$  genes. This inference is supported by the previously cited sequence studies on the myeloma  $V_L$  regions of the BALB/c mouse. The serum ratio of  $\lambda$  to  $\kappa$  chains in the mouse is 3/97 (MCINTIRE and ROUSE 1970). There is extreme sequence diversity among the  $V_\kappa$  regions (HOOD *et al.* 1973) and extremely limited sequence diversity among the  $V_\lambda$  regions (WEIGERT *et al.* 1970). Accordingly, in the mouse there appear to be many  $V_\kappa$  genes and a very few  $V_\lambda$  genes. This ratio of diversity correlates with  $\lambda$  to  $\kappa$  ratio in the serum of the mouse. By similar arguments, for example, the cow appears to have many  $V_\lambda$  genes and only a few  $V_\kappa$  genes (Figure 20). Since all mammals shared a common

ancestor less than 75 million years ago, the  $V_{\lambda}$  and  $V_{\kappa}$  families of germ line genes must have changed markedly in gene number for various mammalian evolutionary lines.

A second unusual evolutionary feature of the Ab families of V genes is that the V genes of a given evolutionary line tend to evolve as similar clusters of similar sequences. A cluster is defined as a group of V regions (e.g., a branch on the genealogic tree) which are more closely related to one another than to the V regions of other species. In Figure 21, which presents a genealogic tree for most of the complete  $V_{\kappa}$  sequences of mouse and man, the mouse proteins M70, M321, M63 and T124 are one cluster and the human proteins diverging from ancestral sequences 11, 17, and 19 constitute three additional

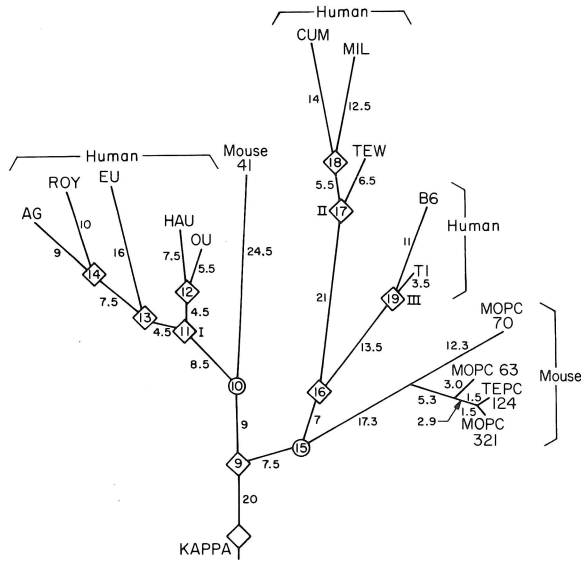


FIGURE 21. A genealogic tree of the  $V_{\kappa}$  regions from man and mouse. The nodal sequences for the three major branches on the human  $V_{\kappa}$  tree are indicated by I, II and III. (Adapted from DAYHOFF 1972 and BARKER AND DAYHOFF, personal communication).

clusters. Although complete sequences for additional mouse  $\kappa$  chains are not yet complete, it appears that there are additional  $V_{\kappa}$  regions. When the amino terminal sequence data of human and mouse  $V_{\kappa}$  regions are examined, it is apparent that there are many additional clusters of mouse  $V_{\kappa}$  regions (Figure 8). Thus all known clusters of mouse  $V_{\kappa}$  regions are distinct from their counterparts in man. I shall term the tendency for

V genes to evolve in clusters which are distinct for each species "coincidental evolution" (e.g. the genes for a given cluster coincide or correspond to one another).

Those who believe in a somatic theory of antibody diversity might contend that coincidental evolution is easily explained by postulating that just a single germ line V gene exists for each cluster of sequences and that these germ line V genes undergo ordinary divergent evolution. This explanation is not, however, satisfactory. Consider the cluster of similar mouse  $V_K$  sequences T124, M63, M312 and M70 (Figure 21). An earlier consideration of the M63, T124 and M321 V regions suggested that at least two  $V_K$  germ line genes must be postulated to avoid (i) multiple parallel mutations or (ii) more than three or four somatic mutations per lymphocyte line (see section 5.1.3, page 94). Indeed, even if a single  $V_K$  gene is postulated to code this trio, a second  $V_K$  gene must be postulated for the M70 protein to avoid the requirement for excessive somatic mutations. M70 is more closely related to the other three mouse  $V_K$  regions than it is to any of the eleven completely sequenced human  $V_K$  regions (Figure 21). Thus this cluster of mouse V regions is coded by at least two or three germ line  $V_K$  genes. Although the data is admittedly limited, these comparisons suggest that at least two mouse  $V_K$  genes are evolving in a coincidental fashion. This is an important example because it can no longer be argued that coincidental evolution can be explained by a single gene for each cluster of V regions. Additional examples will certainly be found as more sequence data accumulate (MCKEAN, personal communication; BARSTAD and HOOD, unpublished data).

One additional sequence pattern is relevant to a consideration of coincidental evolution in a multigenic family. The presence of "species associated residues" (also termed species specific residues or phylogenetically associated residues) has been reported in a number of antibody families from various species (HOOD *et al.* 1970, KEHOE and CAPRA 1972, NOVOTNY *et al.* 1972). The phenomenon of species associated residues is illustrated in Table 7. For example, most of the  $V_K$  regions in normal and antibody light chains of the rabbit appear to have a valine at position 11 and a glycine at position 17, whereas their human counterparts have leucine and aspartic acid or glutamic acid at the corresponding positions (Table 7). Thus valine and glycine are residues associated with most rabbit  $V_K$  regions at certain positions which are not found in the homologous positions of a different species (e.g. species associated residues). The evolution of species associated residues is, accordingly, a special case of coincidental evolution in which a particular gene family apparently expresses just a single cluster of similar V genes. The extreme example of species associated residues is the mouse  $\lambda$  chain in which, because of the limited heterogeneity expressed, a single residue is present at most positions. Accordingly, species associated residues occur when residue positions in mouse  $\lambda$  chains differ from the alternatives expressed at homologous positions in the  $\lambda$  chains of other species. The general implication of species associated residues is that the information content of

the Ig gene family is small. For example, species associated residues are not seen in comparing the  $\kappa$  families of man and mouse, which is consistent with the belief that they have

Table 7. Species associated residues in the  $V_{\kappa}$  regions of rabbit and man<sup>a</sup>

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17 <sup>b</sup>
Rabbit $\kappa$ (9) <sup>c</sup>	Ala	Ala	Ile	Val	Met	Thr	Glx	Thr	Pro	Ala	Ser	<u>Val</u>	Ser	Glx	Pro	Val	Gly	<u>Gly</u>
		Asp	Val	Glx	Val					Ser			Thr	Ala	Ala			
			Phe		Leu								Glx	Val				
Human $\kappa$ (41) <sup>c</sup>		Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ser	Ser	<u>Leu</u>	Ser	Ala	Ser	Val	Gly	Asp
	Glu	Glu	Val	Val	Leu			Thr		Ala	Thr		Pro	Leu	Thr	Pro	Arg	Glu
		Lys	Met	Leu						Leu	Phe			Val		Leu		
				Ile						Gly				Met				
										Thr								
										Asx								

<sup>a</sup>These data are taken from Hood et al. 1970a, Hood and Talmage 1970. <sup>b</sup>Numbers indicate residue position from the N-terminus of mouse and human  $\kappa$  chains. <sup>c</sup>These numbers indicate the number of homogeneous chains from which the comparisons have been made. <sup>d</sup>Boxes indicate species associated residues.

large families of  $V_{\kappa}$  genes. One important qualification must be made in this regard. One must always ask whether the clone-type examined has in some fashion been selected so that only a limited portion of the V gene family is amplified. If so, no conclusions can be drawn about genotype. This reservation may hold for normal serum Ig chains as well as the population of myeloma proteins (GRANT and HOOD 1971).

The evolution of species specific clusters of genes poses an interesting evolutionary problem. Since many mammalian species appear to have multiple  $V_{\kappa}$  genes, the ancestor common to all contemporary mammalian lines probably also had multiple  $V_{\kappa}$  genes. If so, how is one to explain the observation that individual V genes of a given evolutionary line have evolved in clusters whose individual members are more closely related to one another than to the V genes of a second species. As previously indicated, I have chosen to term this type of evolution in a multigene family "coincidental evolution" so that the phenomenon has a descriptive designation which does not imply a genetic or evolutionary mechanism. Let us now consider some possible mechanisms.

#### 8.6 MECHANISMS FOR COINCIDENTAL EVOLUTION IN MULTIGENE FAMILIES

Two general mechanisms might explain coincidental evolution. First, the genes of the ancestor may be directly passed to contemporary species and coincidental evolution could occur



by parallel evolution, either through natural selection or by a gene correction process (Figure 22). Second, coincidental evolution could occur by a process of gene expansion and contraction in which different genes are expanded (or contracted) in distinct evolutionary lines (Figure 23). Let us consider in detail each of these possibilities.

Parallel evolution by natural selection. Parallel evolution suggests that a cluster of V genes in one evolutionary line can evolve together such that the members of the cluster are much more closely related to one another than to other V genes in the different species (or even different clusters of same species) (Figure 22). Natural selection must operate on the random single base substitutions that occur independently

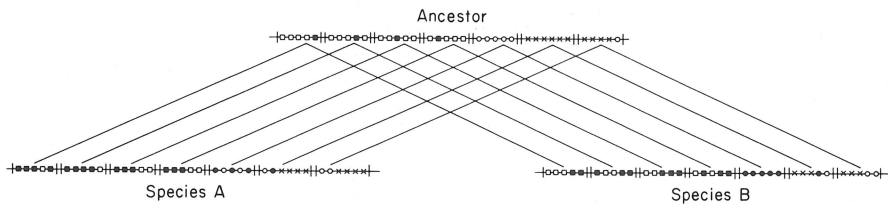


FIGURE 22. A model of parallel evolution in a multigene system. The genes are transferred directly from ancestor to contemporary species A and B. The various combinations of  $\blacksquare$ ,  $\square$ ,  $\circ$  and  $\times$  indicate nucleotide sequences (see text).

in each gene. Intense selective pressure must then be generated to restrict the range of changes that can be passed on to progeny for each cluster of V genes. This possibility seems very unlikely to me for a number of reasons. First, the function of a distinct antibody V gene appears to overlap with those of many other antibody V genes. That is, V regions present in the ordinary heterogeneous antibody response can be drawn from many different branches (different V genes) of the V region genealogic tree (GRANT *et al.* 1969, HOOD *et al.* 1973). Accordingly, it is difficult to imagine how selection can operate on individual genes if many other V genes can carry out similar functions. An extremely important generalization emerges from these considerations, namely that natural selection probably can not act on individual Ab V genes. Rather, natural selection must survey the functions of the antibody system taken as a whole. For example, natural selection is indifferent as to which specific V genes are used to mount an antibody response to a particular bacterial infection as long as some combination of the V genes available is effective. Accordingly, because of the overlapping functions of individual V genes, natural selection appears to operate on the families of V genes taken as a whole and not upon individual V gene members of that family. Thus it is difficult to see how

natural selection could cause parallel evolution in clusters of V genes. Second, in a multigene system natural selection would impose a large genetic load (e.g. need to get rid of bad or ineffective antibody genes by wasting individuals) on vertebrate organisms (OHNO 1972), particularly since coincidental evolution appears to occur also in other multigene systems (HOOD 1972, CAMPBELL *et al.* 1973). Accordingly, another evolutionary mechanism must be postulated to explain coincidental evolution.

Parallel evolution by gene correction. Gene correction can be defined as a superposition of information from one piece of DNA onto another. This model implies once again that the genes in an Ab family are transferred from the ancestor to contemporary creatures in a linear fashion and that coincidental evolution occurs by some process of gene correction (Figure 22). The process of gene correction in yeast, in which one allele is corrected to the sequence of the second allele, seems to be an example of this process (FOGEL and MORTIMER 1969). The correction process is envisioned as a matching of homologous sequences and a copy repair mechanism which extends over 150 to 300 nucleotides. The correction process in multigene families, however, requires that correction occurs not in just a single allele but that it occurs for each of the members of a given gene cluster (Figure 22). Several specific correction mechanisms have been proposed. The master-slave hypothesis suggests that one gene in a multigene family (the master) corrects the remainder of the genes (slaves) at each DNA division (CALLAN 1967, THOMAS 1970). In the master-slave model the entire multigene family evolves in parallel with the master gene. This model was postulated to explain the evolution of multigene families whose individual members are identical in sequence. It is not clear how this model could be adapted to a nonidentical multigene family such as antibodies although some have tried (WHITEHOUSE 1967).

In this context, a gene correction mechanism for coincidental evolution in a nonidentical multigene family has been proposed (EDELMAN and GALLY 1970) and is termed "democratic gene conversion." This theory suggests that any member of a given multigene family can correct the DNA sequence of any other member. Presumably those genes which experience selectively favorable mutations can then spread these mutations through the multigene family by gene correction. Thus the multigene family could serve as a network to trap and spread favorable mutations among all members of the set.

I find this model, as applied to antibodies, unattractive for several reasons (many of these objections also apply to parallel evolution by natural selection). First, it does not explain the change in the size of a given antibody family that occurs among various evolutionary lines (Figure 20). Second, if the gene correction process does operate over regions 150-300 nucleotides in length, then the process of spreading favorable mutations to other members of the multigene family would invariably eliminate the diversity of the hypervariable regions and, accordingly, cause a loss of that sequence diversity which

seems to be associated with specificity in the antigen binding site. Third, how can a gene correction process lead to the coincidental evolution of V region clusters? It has been argued that gene correction occurs much more frequently among very closely related V genes--thus clusters of genes would tend to evolve independently (EDELMAN and GALLY 1970). It is not clear with a gene correction process why the resulting members of a cluster should exhibit the genealogic structure typical of genes which have evolved by divergent evolution (FITCH and MARGOLIASH 1967, NOLAN and MARGOLIASH 1968). Finally, it is difficult to explain the phenomenon of species associated residues seen in comparing rabbit and human  $V_K$  regions (Table 7). The gene correction model must postulate that subsequent to the divergence of the rabbit and human evolutionary lines, most of the rabbit  $V_K$  genes were corrected to express the species associated residues characteristic of positions 11 and 17. Because there is extensive  $V_K$  sequence diversity seen among  $V_K$  regions from different subgroups in mouse and man (Figure 21), it appears reasonable to assume that the ancestor to mammals had extensive sequence diversity in its  $V_K$  regions (genes). Thus the correction process must be imposed on V genes which in the ancestor were quite divergent from one another. Yet the correction of dissimilar sequences appears unlikely from what we know about the process of gene conversion. Furthermore, it was argued that clusters of genes evolve because the correction process generally occurs among similar genes. Perhaps a more satisfactory solution to these evolutionary problems is a process of gene expansion and contraction.

Gene expansion and contraction (HOOD *et al.* 1970a, MILSTEIN and SVASTI 1971, HOOD and PRAHL 1971). The gene expansion and contraction model suggests that the number of germ line genes in a given multigene family can be expanded by duplication or contracted by deletion of individual genes (Figure 23). Infrequent expansions or deletions, like rare

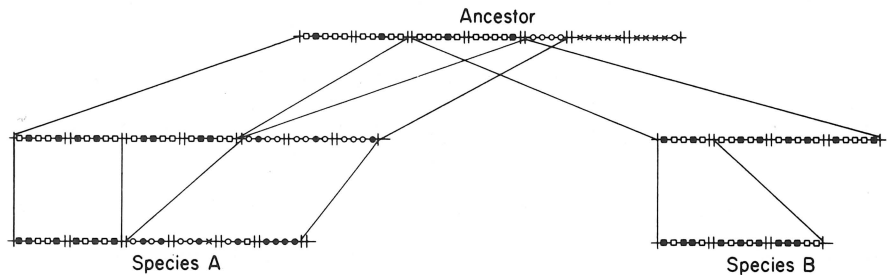


FIGURE 23. A model of gene duplication and deletion (gene expansion and contraction) in a multigene system.

mutations, could confer a selectively favorable phenotype and become fixed in the population by natural selection. The antibody families in different evolutionary lines could then evolve

in a coincidental fashion by replicating some and deleting different members of a given multigene family. Thus, as illustrated in Figure 23, contemporary species (A and B) could have family members which diverged from different V genes in the ancestor. The expansion events after speciation would lead to clusters of related genes. At least two general mechanisms have been proposed which lead to the expansion of multigene families--homologous but unequal crossing over and saltatory replication.

Homologous but unequal crossing over. Homologous but unequal crossing over in a family of duplicated genes leads to one chromosome with an increased number of genes (expanded family) and a second with a decreased number of genes (contracted family) (Figure 24). The same mechanism can explain the decrease in size of a gene family as well as an increase in size. Homologous but unequal cross over might occur interchromosomally by homologue exchange during meiosis or intrachromosomally by sister chromatid exchange during meiosis or intrachromosomally by sister chromatid exchange during meiosis or germ line mitosis. The repeated genes need not be identical but only similar enough to promote unequal exchange.

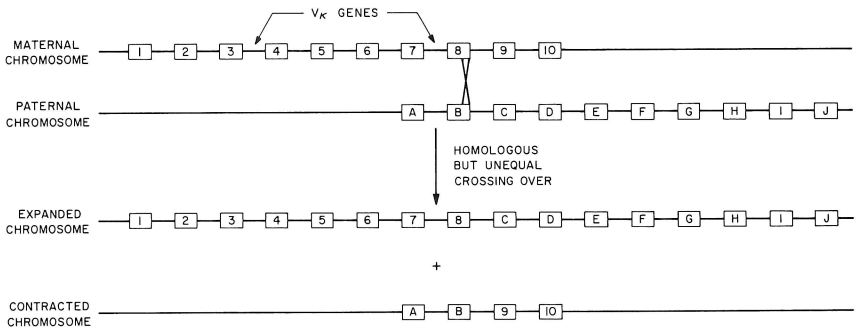


FIGURE 24. A model of gene duplication and deletion (expansion and contraction) of homologous genes on chromosomes by homologous but unequal crossing over. (From SMITH *et al.* 1971).

This mechanism can also explain coincidental evolution in multigene families. For example, a computer simulation of homologous but unequal crossing over in a family of 500 genes demonstrated that within 10,000 cross over events, all the genes in a given family could have descended from a single ancestor (SMITH, personal communication). The gene family in this exercise fluctuated between 450 and 550 members ( $\pm 10\%$ ) and, accordingly, radical reductions in family size are not necessary to explain coincidental evolution in multigene families (or even the emergence of species associated residues). This same study recorded at each step the gene order on the

chromosome and demonstrated that although genes in the same cluster (subgroup) are not necessarily contiguous, they tend to be near one another on the chromosome. Thus homologous but unequal crossing over can explain the change in size of multigene families as well as their tendency to evolve in a coincidental fashion. Let us consider a second mechanism for expansion.

**Saltatory replication.** This theory proposes that one gene is copied many times over and incorporated into the genome. This mechanism can be viewed as a chromosomal event in which a single sequence is reiterated during normal DNA replication or as an extra chromosomal event in which multiple copies of a chromosomal element are generated and subsequently reintegrated into the genome (Figure 25). The supposition that saltatory replication can occur in eukaryotes as a chromosomal event is based on the existence of DNA satellites. For example, the satellite DNA of the house mouse (*Mus musculus*) is a

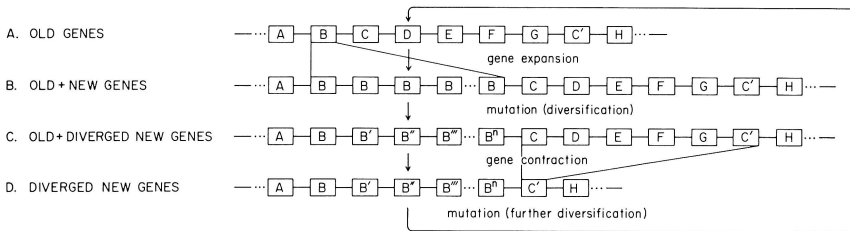


FIGURE 25. A model of gene expansion by saltatory replication. Gene contraction presumably occurs by homologous but unequal crossing over (Figure 24). Mutation of the duplicated gene (B) leads to diversification. (From CAMPBELL *et al.* 1973).

repeating sequence of about 300 nucleotides which is present as  $10^6$  copies (WALKER *et al.* 1969). The satellite sequences are homologous to one another at the 95% level and are not found in the genomes of other animals closely related to *Mus musculus* (e.g. the rat and the British wood mouse). Since these sequences are so similar and are not found in closely related evolutionary lines--it has been suggested that they arose by a sudden saltatory event, that is massive gene duplication. Thus saltatory replication could expand out one or a few V genes in a given evolutionary line and subsequent divergence could lead to the diversity seen in the antibody V families (Figure 25). If different V genes are expanded in different evolutionary lines, coincidental evolution will follow naturally (Figure 23). A variety of specific mechanisms have been proposed for saltatory replication (BRITTEN and DAVIDSON 1971). Saltatory replication is theoretically interesting because it predicts that at certain stages in the evolution of antibody families there will be multiple copies of identical or nearly identical V genes (Figure 25). Thus

families, such as mouse  $\lambda$ , may not be coded by a few V genes, rather they may be coded by a larger number of identical or nearly identical V genes. The mouse  $V_\lambda$  family seems to be an ideal system for gene counting by mRNA-DNA hybridization techniques because the multiple copies, if they exist, should be similar in sequence and, accordingly, obviate some of the technical difficulties in the RNA-DNA hybridization procedure.

The gene expansion-contraction model for evolution in a multigene family is attractive for a variety of reasons. First, gene duplication of various types accounts for the general features of antibody evolution as discussed in section 8.1 (Figure 18). Second the family of  $C_H$  genes in mammals probably expands (and contracts) as discussed in section 8.4 (Figure 19). Thus expansion and contraction occurs, presumably by the same mechanism, throughout the genes of the antibody family--with C as well as V genes. Third, the two unusual evolutionary features of antibody genes, coincidental evolution and a change in family size, can be explained by a single mechanism. Fourth, both mechanisms for gene expansion and contraction place similar genes near one another in the genome and accordingly, are consistent with the observations discussed earlier from the patient (TÖ) with a tetraclonal Waldenstrom's macroglobulemia (see section 7.2, page 109).

#### 8.7 A SPECIAL EVOLUTIONARY PROBLEM - RABBIT V REGION GENETIC MARKERS

The rabbit allotype markers a1, a2, and a3 are localized to the amino terminal half of all classes of heavy chain which suggests these markers are on  $V_H$  regions which are shared by all H chain classes. Studies on heterogeneous normal H chains from several Ig classes suggests that (i) several different major  $V_H$  sequences (? V subgroups) are present in various proportions in each of the three homozygous rabbits (a1, a2, a3) and (ii) multiple V region sequence differences in these three populations of H chains seem to correlate with the three allotypic markers (WILKINSON 1969a, 1969b, 1970, KOSHLAND *et al.* 1969, MOLE *et al.* 1971). More recent studies using restricted or homogeneous rabbit antibodies suggest that certain of the sequence difference originally observed do not in fact correlate with a particular allotype (JATON *et al.* 1973). Thus it seems likely that rabbit heavy chains of a particular group a allotype are coded by multiple  $V_H$  genes and that the chains of one allotype share some general structural feature(s) which is detected by the allotype antisera. For example, antisera can be made which distinguish V regions from each of the three major human  $\kappa$  subgroups (see Figure 21) (SOLOMON and MCLAUGHLIN 1971).  $\kappa$  chains within a subgroup can differ by as many as 33 residues, yet they still share one or more common antigenic features. Perhaps the same is true for the allotype differences among rabbit  $V_H$  regions. If so, perhaps the differences among the  $V_H$  genes of differing allotypes may have arisen by the same mechanism which explains the coincidental evolution of clusters of genes in antibody families. The mechanism I favor in this regard is gene expansion and contraction. In a

multigene model, however, this process would have to involve multiple crossing-over events, which would inevitably scramble the differences as fast as they arose unless crossing over between the three types of chromosomes was in some way prevented. This might have occurred in two ways. First, three geographically isolated populations of rabbits may have evolved the three chromosomes independently as if they were separate species (these populations may have been remixed too recently for detectable recombination to have occurred). These features might be explained if the a1 and a3 sequences diverged not as alleles but as gene duplications just as with the V region subgroups. Different rabbit chromosomes, however, might have undergone different subsequent gene deletions and/or duplications, so that they now contain genes for different subgroups. In this way, linked genes which diverged long before the rabbits became a species, and thus have accumulated a large number of differences, might have become alleles (or pseudo-alleles) in contemporary rabbits. If the number of genes is very limited only a few deletions and duplications could be required. Second, interchromosomal crossing over may have for some reason been prevented (the proposed gene expansion-contraction could be due to intrachromosomal crossing over). The assumption that intrachromosomal crossing over occurs frequently while interchromosomal crossing over is undetectable must be regarded as somewhat contrived. Interchromosomal crossing over might have been inhibited in one of two ways. First, small chromosomal inversions and/or translocations involving the V<sub>H</sub> gene region might render the V<sub>H</sub> genes on different chromosomes unable to pair and cross over. Secondly, meiotic crossing over might not occur at all in the V<sub>H</sub> locus (perhaps because of some deficiency analogous to that which prevents crossing over in the male *Drosophila* with intrachromosomal crossing over occurring during the many mitotic divisions in the germline. Unequal intrachromosomal crossing over has been demonstrated at the *bar* locus of the female *Drosophila* (PETERSON and LAUGHAN 1963), but it is not known whether intrachromosomal crossing over occurs at a comparable rate in this locus in the male *Drosophila*, in which all interchromosomal crossing over is inhibited).

Two additional possible explanations of rabbit allotypes should be mentioned. First, the differences might have arisen by a gene correction mechanism (see section 8.6, page 122 and GALLY and EDELMAN 1970). Second, it is possible that a1, a2 and a3 chromosomes have indistinguishable complements of multiple V<sub>H</sub> genes, their differences lying instead in the gene(s) coding for the apparatus by which V and C genes are joined; the three proposed allelic forms of this joining apparatus might preferentially join three different subgroups (a1, a2 a3) of V<sub>H</sub> genes to the C<sub>H</sub> genes. Thus, the subgroups themselves would appear on genetic analysis to be allelic.

It should be stressed that there is not entirely straightforward interpretation for this phenomenon, and it may be that additional evidence will suggest an entirely unsuspected explanation. There are three promising lines of approach: first,

serological and structural analysis of  $V_H$  regions in species related to the rabbit (hares and pikas) may give some important clues about how these allelic differences arose; second, homogeneous rabbit antibodies might allow this extremely complex system to be analyzed in exactly the same way that the homogeneous human myeloma proteins allowed the Gm allotypes to be organized into a coherent genetic scheme; finally, the observation that certain  $V_K$  sequences appear to correlate with the b allotypes suggests that a similar phenomenon may occur in rabbit light chains which are easier to study for a variety of technical reasons (e.g., light chains are half the size of heavy chains and have a free  $\alpha$  amino group) (WATERFIELD *et al.* 1973). The existence of V region genetic markers is difficult to explain by all theories of antibody diversity which postulate multiple  $V_H$  genes and, accordingly, cannot be used in a very convincing fashion as an argument for or against any theory. Let us now consider the evolution of the immune system in more general terms.

#### 8.8 THE EVOLUTION OF THE IMMUNE SYSTEM FROM CELL SURFACE RECEPTOR MOLECULES

The evolution of the immune system has required the development of at least five separate entities which are listed in Table 8. The possibility that the library (family)

Table 8. Requirements for the evolution of the immune system

- 
1. Evolution of a library of V (and C) genes
  2. Evolution of a joining mechanism
  3. Evolution of a means for initiating cell proliferation once the membrane receptor molecule has combined with antigen (amplification of the response)
  4. Evolution of cellular effector functions (cellular immunity)
  5. Evolution of a mechanism for the synthesis and secretion of antibody molecules upon antigenic stimulation (humoral immunity)

of V genes may have evolved from membrane placed receptor molecules is attractive (DREYER *et al.* 1967). As one ascends the evolutionary scale from single celled organisms to complex metazoa, more and more cell receptor molecules are required for a variety of different functions such as cell recognition, scavaging of debris, and hormonal triggering. Thus libraries of V-like genes may have evolved to carry out diverse receptor functions. Presumably a joining mechanism evolved to unite these V-like cell surface receptor genes to their corresponding C genes. Accordingly, a single C gene could serve as a handle whereby receptor molecules of differing specificities could be



appropriately positioned in the cell membrane. Perhaps an event such as chromosomal doubling provided a new library of V genes to initiate the eventual evolution of the antibody families. Initially there may have been a very limited number of C genes which functioned to position the putative antibody V genes on the cell membrane. Certainly a critical change which differentiated the early immune response from other cell surface receptor functions was the ability to initiate cell division after the antibody receptor molecule had combined with its corresponding antigen--thus producing more cells with identical antibody receptor molecules. Presumably this mechanism for amplification of the immune response occurred just prior to or at the divergence of vertebrates from invertebrates as it is uncertain that bona fide immune responses exist in invertebrates. Later certain putative lymphocytes, upon antigenic stimulation, acquired the respective properties of the cellular or the humoral immune system (Table 8). This view of the evolution of the antibody families suggests that the existence of V gene libraries and the joining mechanism antedated the emergence of the immune system and that both of these strategies may be general mechanisms in the differentiation of other complex systems in higher organisms. The existence of repeating families of DNA sequences has been established in all eukaryotes by DNA-DNA hybridization techniques (BRITTEN and KOHNE 1968). Certain of these repeating DNA sequences appear to be tandemly

Table 9. Possible multigene systems

- 
1. Cell surface molecules
    - a. Transplantation antigens
    - b. Membrane receptors (hormonal, cAMP)
    - c. Membrane transport systems
    - d. Embryonic (cancer) cell surface antigens
    - e. Immune response genes
  2. Developmental control systems
    - a. The T allele system of mice
    - b. Transcriptional signals
    - c. Translational signals
  3. Nervous system
    - a. Wiring
    - b. Information storage
-

linked families of multiple genes with well defined functions (e.g., histone genes, tRNA genes, 18 S RNA genes, 28 S RNA genes and 5 S RNA genes--see BROWN 1971, KEDES and BIRNSTIEL 1970). These known multigene families also share both of the unusual evolutionary features seen in the antibody families (e.g., the potential for rapid change in family size and coincidental evolution) (CAMPBELL *et al.* 1973). Thus similar evolutionary mechanisms may be shared by each of these multigene systems. Perhaps of greater interest are the potential multigene systems given in Table 9 which, like the antibody system, must code a large amount of information. It will be of great interest in future studies to determine which, if any, of the strategies and mechanisms seen in the antibody system are shared by these other potential "informational multigene systems."

## 9. SYNOPSIS

The antibody family is a provocative model for the study of differentiation in genetically complex systems in higher organisms. The analysis of general patterns that have emerged from genetic, molecular, and cellular studies have given us important insights into mechanisms for the storage, expression, and evolution of information in antibody families. Some of these patterns are summarized in Table 10. Information (antibody V genes) is stored in the form of multiple germ line genes. It is still an open question whether all of the information is present as germ line genes or whether somatic mutation or recombination amplifies the germ line information during somatic differentiation. It is attractive to postulate that information is expressed through a joining mechanism which operates at the DNA level during somatic differentiation. Although the nature of this joining mechanism is unclear, it does provide a general model for the specific commitment of cells in higher organisms. Finally, the evolution of information (antibody V genes) appears to require some unusual evolutionary mechanisms to explain the change in family size and coincidental evolution seen in mammalian antibody families.

The vertebrate antibody family is a unique system for the analysis of information storage, information expression, and information evolution as it is the only multigene system in which single gene products can be obtained for comparison and detailed analysis (see Table 9). In the future we can expect to gain a far more detailed understanding of the mechanisms of information expression and evolution. Perhaps the most provocative question which emerges from these studies is whether and to what extent other multigene systems will share the strategies and mechanisms of the antibody family.

Table 10. Patterns derived from amino acid sequence and serological studies of Ig polypeptides.

Pattern	Implications
1. Antibody families.	1. Three unlinked families of genes.
2. V and C regions.	2. Coded for by separate genes which require a joining mechanism.
3. V region genealogic trees (subgroups).	3. V regions are coded for by multiple germ line V genes.
4. Distinct genealogic patterns (clusters) for the V genes of each species.	4. Coincidental evolution in a multigene family.
5. Mouse $V_{\lambda}$ regions exhibit highly restricted heterogeneity; mouse $V_{\kappa}$ are extremely heterogeneous. Difference in mammalian $\lambda/\kappa$ ratios.	5. The number of gene members in the same family can be very different in two species. Suggests that some gene expansion and contraction must occur.
6. Species associated residues.	6. A special case of coincidental evolution.
7. Rabbit group a allotypes.	7. Coincidental evolution plus a restriction of meiotic recombination by 1) geographic isolation, 2) pseudoallelic organization or 3) a special mechanism to inhibit meiotic recombination among $V_H$ genes.
8. Hypervariable regions.	8. Reflects diversity of the antigen binding sites.
9. The switch of $V_H$ from $C_{\mu}$ to $C_{\lambda}$ .	9. $V_H$ and $C_H$ genes are separate.
10. Hybrid immunoglobulin.	10. The $C_H$ genes are closely linked. The single VC gene model is unlikely.
11. Allelic exclusion.	11. An unknown control mechanism.

## LITERATURE CITED

- APPELLA, E., J. REJNEK and R. A. REISFELD 1969 Variations at the carboxyl-terminal amino acid sequence of rabbit light chains with b4, b5 and b6 allotypic specificities. *J. Mol. Biol.* 41:473.
- APPELLA, E., A. CHERSI, R. G. MAGE and S. DUBISKI 1971 Structural basis of the A<sub>14</sub> and A<sub>15</sub> allotypic specificities in rabbit IgG. *Proc. Nat. Acad. Sci. U.S.* 68:1341.
- AVIV, H., S. PACKMAN, D. SWAN, J. ROSS, P. LEDER 1973 In vitro synthesis of DNA complementary to mRNA derived from a light chain-producing myeloma. *Nature N.B.* 241:174.
- BAGLIONI, C. 1962 The fusion of two peptide chains in hemoglobin Lepore and its interpretation as a genetic deletion. *Proc. Nat. Acad. Sci. U.S.* 48:1880.

- BIRSHTEIN, B. K. and J. J. CEBRA 1971 Primary structure of C-1-a-a cyanogen bromide fragment of heavy chain from inbred guinea pig immunoglobulin G(2), which contains a markedly variable segment. *Biochemistry* 10:4930.
- BRENNER, S. and C. MILSTEIN 1966 Origin of antibody variation. *Nature* 211:242.
- BRITTEN, R. J. and D. E. KOHNE 1968 Repeated sequences in DNA. *Science* 161:529.
- BRITTEN, R. J. and E. H. DAVIDSON 1971 Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Quart. Rev. Biol.* 46:111.
- BROWN, D. D. 1971 The evolutionary solution to the antibody dilemma. In: *Molecular Genetics and Developmental Biology* (M. Sussman, Ed.) Society of General Physiologists, in press.
- BURNET, F. M. 1959 *The Clonal Selection Theory of Acquired Immunity*, Cambridge University Press, London and New York.
- CALLAN, H. G. 1967 The organization of genetic units in chromosomes. *J. Cell Sci.* 2:1.
- CAMPBELL, J., S. ELGIN, J. SMART and L. HOOD 1973 Evolution and multigene families. In preparation.
- CAPRA, J. D. 1971 A hypervariable region of human immunoglobulin heavy chains. *Nature* 230:62.
- CAPRA, J. D. and H. G. KUNKEL 1970 Amino acid sequence similarities in two human anti-gamma globulin antibodies. *Proc. Nat. Acad. Sci. U. S.* 67:87.
- CEBRA, J. J., J. E. COLBERG and S. DRAY 1966 Rabbit lymphoid cells differentiated with respect to  $\alpha$ -,  $\gamma$ -, and  $\mu$ - heavy polypeptide chains and to allotypic markers Aa1 and Aa2. *J. Exp. Med.* 123:547.
- CEBRA, J. J., A. RAY, D. BENJAMIN and B. BIRSHTEIN 1971 Localization of affinity label within the primary structure of  $\gamma_2$  chain from guinea pig IgG (2). Pp. 269-284. In: *Progress in Immunology* (B. Amos, Ed.) Academic Press, New York.
- COHEN, S. and C. MILSTEIN 1967 Structure and biological properties of immunoglobulins. *Adv. Immunol.* 7:1.
- COHN, M. 1967 Natural history of the myeloma. *Cold Spring Harbor Symp. Quant. Biol.* 32:211.
- COHN, M. 1968 Molecular biology of expectation. Pp. 671-715. In: *Nucleic Acids in Immunology* (O. J. Plecia and W. Braun, Eds.) Springer Verlag, New York.

- COHN, M. 1970 Selection under a somatic model. Cellular Immunol. 1:461.
- COHN, M. 1971 The take-home lesson - 1971. Ann. N.Y. Acad. Sci. 190:529.
- DAVIDSON, E. H., B. R. HOUGH, A. S. AMENSON and R. J. BRITTEN 1973 General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. J. Mol. Biol., in press.
- DAVIES, D. R., R. SARMA, L. LABAW, E. SILVERTON, D. SEGAL and W. D. TERRY 1971 X-ray diffraction and electron microscope studies on a crystalline human immunoglobulin. Ann. N. Y. Acad. Sci. 190:122.
- DAVIS, R. W., M. N. SIMON and N. DAVIDSON 1971 Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. Pp. 413. In: Methods in Enzymology, Nucleic Acids. Vol. 21D (L. Grossman and K. Moldave, Eds.).
- DAYHOFF, M. O., Ed. 1972 Atlas of Protein Sequence and Structure, Silver Spring, Nat. Biomed. Res. Found.
- DELOVITCH, T. L. and C. BAGLIONI 1973 Estimation of light-chain gene reiteration of mouse immunoglobulin by DNA-RNA hybridization. Proc. Nat. Acad. Sci. U.S. 70:173.
- DRAY, S. O., G. O. YOUNG and L. GERALD 1963 Immunochemical identification and genetics of rabbit  $\gamma$ -globulin allotypes. J. Immunol. 91:403.
- DREYER, W. J. 1970 A proposed new and general chromosomal control mechanism for commitment of specific cell lines during development. In: Development Aspects of Antibody Formation and Structure (J. Sterzl and I. Riha, Eds.) Academia Publishing House of the Czechoslovak Acad. of Sci. Prague Vol. 2:919.
- DREYER, W. J., W. R. GRAY and L. HOOD 1967 The genetic, molecular and cellular basis of antibody formation: Some facts and a unifying hypothesis. Cold Spring Harbor Symp. Quant. Biol. 32:353.
- DUBISKI, S. 1969 Immunochemistry and genetics of a "new" allotypic specificity Ae<sup>14</sup> of rabbit  $\gamma$ G immunoglobulins: Recombination in somatic cells. J. Immunol. 103:120.
- EDELMAN, G. M. 1970 The covalent structure of a human  $\gamma$ G immunoglobulin. XI. Functional implications. Biochemistry 9:3197.
- EDELMAN, G. M. 1971 Antibody structure and modern immunology. Ann. N. Y. Acad. Sci. 190:5.

- EDELMAN, G. M. and W. E. GALL 1969 The antibody problem. *Ann. Rev. Biochem.* 38:415.
- EDELMAN, G. M. and J. A. GALLY 1967 Somatic recombination of duplicated genes, an hypothesis on the origin of antibody diversity. *Proc. Nat. Acad. Sci. U.S.* 57:353.
- EDELMAN, G. M. and J. A. GALLY 1970 Arrangement and evolution of eukaryotic genes. Pp. 962. In: *The Neurosciences: Second Study Program*, (F. O. Schmitt, Ed.) Rockefeller Univ. Press., New York.
- EDELMAN, G. M., B. A. CUNNINGHAM, W. E. GALL, P. D. GOTTLIEB, U. RUTISHAUSER and M. WAXDAL 1969 The covalent structure of an entire  $\gamma$ G immunoglobulin molecule. *Proc. Nat. Acad. Sci. U.S.* 63:78.
- EIN, D. 1968 Nonallelic behavior of the OZ groups in human  $\lambda$  Ig chains. *Proc. Nat. Acad. Sci. U.S.* 60:982.
- EISEN, H. 1971 Combining sites of anti-2,4-dinitrophenyl antibodies. Pp. 243. In: *Progress in Immunology* (B. Amos, Ed.) Academic Press, New York and London.
- FITCH, W. M. and E. MARGOLIASH 1967 Construction of phylogenetic trees. *Science* 155:279.
- FLEISCHMAN, J. 1967 Synthesis of  $\gamma$ -globulin by rabbit lymph node cells. *Biochemistry* 6:1311.
- FLORENT, G., J. ROEFS-GRIPPA, C. HAMERS-COSTERMAN, W. VAN DER LOO and R. HAMERS 1970 A study of rabbit heavy chain allotypes. *Protides Biol. Fluids* 17:147.
- FOGEL S. and R. K. MORTIMER 1969 Informational transfer in meiotic gene conversion. *Proc. Nat. Acad. Sci. U.S.* 62:96.
- FRANGIONE, B., C. MILSTEIN and J. R. L. PINK 1969 Structural studies of immunoglobulin G. *Nature* 221:145.
- GALLY, J. A. and G. M. EDELMAN 1970 Somatic translocation of antibody genes. *Nature* 227:341.
- GALLY, J. A. and G. M. EDELMAN 1972 The genetic control of immunoglobulin synthesis. *Ann. Rev. Genetics* 6:1.
- GIBSON, D., M. LEVANSON and O. SMITHIES 1971 Heterogeneity of normal human Ig light chains. Non-allelic variation in the constant region of  $\lambda$  chains. *Biochemistry* 10:3114.
- GILMAN-SACHS, A. and S. DRAY 1972 Identification and genetic control of two rabbit IgM allotype specificities. *Eur. J. Immunol.* 2:505.

- GILMAN-SACHS, A., R. G. MAGE, G. O. YOUNG, C. ALEXANDER and S. DRAY 1969 Identification and genetic control of two rabbit immunoglobulin allotypes at a second light chain locus, the c locus. *J. Immunol.* 103:1159.
- GITLIN, D., A. PERRICELLI and J. GITLIN 1973 Multiple immunoglobulin classes among sharks and their evolution. *Comp. Biochem. Physiol.* 44b:225.
- GOOD, R. A. and D. W. FISHER, Eds. 1971 *Immunobiology*, Sinauer Associates Inc., Stamford, Conn.
- GRANT, J. A. and L. HOOD 1971 N-terminal analysis of normal immunoglobulin light chains. I. A study of thirteen individual humans. *Immunochemistry* 8:63.
- GRANT, J. A., M. E. LAMM and L. HOOD 1969 N-terminal sequence heterogeneity of guinea pig anti-DNP kappa chains. *Immunochemistry* 6:645.
- GREEN, I., P. VASSALI and B. BENACERRAF 1967a Cellular localization of anti-DNP-PLL and anticonveyor albumin antibodies in genetic nonresponder guinea pigs immunized with DNP-PLL albumin complexes. *J. Exp. Med.* 125:527.
- GREEN, I, P. VASSALI and B. BENACERRAF 1967b Specificity of the antibodies produced by single cells following immunization with antigens bearing two types of antigenic determinants. *J. Exp. Med.* 125:511.
- GREY, H. M. and C. A. ABEL 1970 Subclass and allotype differences in human G immunoglobulins reflected by the amino acid sequence of Fc glycopeptides. *Protides Biol. Fluids* 17:229.
- GRUBB, R. 1970 *The Genetic Markers of Human Immunoglobulins*, Springer-Verlag, New York.
- HAIMOVICH, J., H. EISEN and O. GIVOL 1971 Combining sites of immunoglobulins that bind the 2,4-dinitrophenyl (DNP) group specifically. *Ann. N.Y. Acad. Sci.* 190:352.
- HAMERS, R., C. HAMERS-COSTERMAN and S. LAGNARIX 1966 A new allotype in the rabbit linked with As 1 which may characterize a new class of IgG. *Immunology* 10:399.
- HANLY, W. C., F. A. LICHTER, S. DRAY and K. L. KNIGHT 1973 Rabbit immunoglobulin A allotypic specificities. Localization to two papain fragments, Fab<sub>2α</sub> and Fc<sub>2α</sub>, of secretory immunoglobulin A. *Biochemistry* 12:733.
- HANNESTAD, K. and K. SLETTEN 1971 Multiple M-components in a single individual. III. Heterogeneity of M-components in two macroglobulinemia sera with antipolysaccharide activity. *J. Biol. Chem.* 246:6982.

- HARBOE, M., K. HANNESTAD and K. SLETTEN 1972 Oligoclonal macroglobulinemia. *Scand. J. Immunol.* 1:13.
- HESS, M., N. HILSCHMANN, L. RIVAT and C. ROPARTZ 1971 Iso-types of human immunoglobulin  $\lambda$ - chains. *Nature N. B.* 234:58.
- HILL, R. L., R. DELANEY, R. E. FELLOWS, JR. and H. E. LEBOVITZ 1966 The evolutionary origins of the immunoglobulins. *Proc. Nat. Acad. Sci. U.S.* 56:1762.
- HOOD, L. 1972 Two genes, one polypeptide chain - fact or fiction? *Fed. Proc.* 31:177.
- HOOD, L. and D. EIN 1968 Immunoglobulin lambda chain structure: Two genes, one polypeptide chain. *Nature* 220:764.
- HOOD, L. and D. W. TALMAGE 1970 Mechanism of antibody diversity: Germ line basis for variability. *Science* 168:325.
- HOOD, L. and J. PRAHL 1971 The immune system: A model for differentiation in higher organisms. *Adv. Immunol.* 14: 291.
- HOOD, L., W. R. GRAY, B. G. SANDERS and W. J. DREYER 1967 Light chain evolution. *Cold Spring Harbor Symp. Quant. Biol.* 32:133.
- HOOD, L., K. EICHMAN, H. LACKLAND, R. M. KRAUSE and J. J. OHMS 1970a Rabbit Ab light chains and gene evolution. *Nature* 228:1040.
- HOOD, L., J. A. GRANT and H. C. SOX 1970b On the structure of normal light chains from mammals and birds: Evolutionary and genetic implications. In: *Developmental Aspects of Antibody Formation and Structure* (J. Sterzl and I. Riha, Eds.) Academia Publishing House of the Czechoslovak Acad. Sci. Prague Vol. 1:283.
- HOOD, L. E., D. MCKEAN and M. POTTER 1970c Immunoglobulin structure: comparison of amino terminal sequences of kappa chains derived from genetically similar mice (BALB/c). *Science* 170:1207.
- HOOD, L., D. MCKEAN, V. FARNSWORTH and M. POTTER 1973 Mouse immunoglobulin chains: A survey of the amino terminal sequence of kappa chains. *Biochemistry* 12:741.
- JATON, J. C., D. G. BRAUN, A. D. STROSBURG, J. E. MORRIS and E. HABER 1973 Homogeneous rabbit antibodies: Amino acid sequences of rabbit H chains of allotype a1, a2 and a3 in the region 80 to 94. *Biochemistry*, in press.
- JERNE, N. K. 1967 Waiting for the end. *Cold Spring Harbor Symp. Quant. Biol.* 32:591.



- JERNE, N. K. 1971 The somatic generation of immune recognition. *Europ. J. Immunol.* 1:1.
- KEDES, L. and M. BIRNSTIEL 1971 Reiteration and clustering of DNA sequences complementary to histone messenger RNA. *Nature N. B.* 203:165.
- KEHOE, J. M. and J. D. CAPRA 1971 Localization of two additional hypervariable regions in Ig heavy chains. *Proc. Nat. Acad. Sci. U.S.* 68:2019.
- KEHOE, J. M. and J. D. CAPRA 1972 Sequence relationships among the variable regions of immunoglobulin heavy chains from various mammalian species. *Proc. Nat. Acad. Sci. U.S.* 69:2052.
- KINDT, T. J. and W. J. MANDY 1972 Recombination of genes coding for constant and variable regions of immunoglobulin heavy chains. *J. Immunol.* 108:1110.
- KNOPF, P. M., R. M. E. PARKHOUSE and E. S. LENNOX 1967 Biosynthetic units of immunoglobulin heavy chains. *Proc. Nat. Acad. Sci. U.S.* 58:2288.
- KOSHLAND, M. E. 1967 Location of specificity and allotypic amino acid residues in antibody Fd fragments. *Cold Spring Harbor Symp. Quant. Biol.* 32:119.
- KOSHLAND, M. E., J. J. DAVIS and N. J. FUJITA 1969 Evidence for multiple gene control of a single polypeptide chain: The heavy chain of rabbit immunoglobulin. *Proc. Nat. Acad. Sci.* 63:1274.
- KRAUSE, R. M. 1970 The search for antibodies with molecular uniformity. *Adv. Immunol.* 12:1.
- KRETCH, H. W. and A. R. WILLIAMSON 1973 The extent of diversity of antihapten antibodies in inbred mice: Anti-NIP (3-nitro, 4-hydroxy, 5-iodo-phenylacetyl) antibodies in CBA/H mice. *Eur. J. Immunol.*, in press.
- KUNKEL, H. G. 1965 Myeloma proteins and antibodies. *Harvey Lect.* 59:219.
- KUNKEL, H. G., J. B. NATVIG and F. G. JOSLIN 1969a A "Lepore" type of hybrid  $\gamma$  globulin. *Proc. Nat. Acad. Sci. U.S.* 62:144.
- KUNKEL, H. G., W. K. SMITH, F. G. JOSLIN, J. B. NATVIG and S. D. LITWIN 1969b Genetic marker of the  $\gamma A_2$  subgroup of  $\gamma A$  immunoglobulins. *Nature* 233:1247.
- LEDERBERG, J. 1959 Genes and antibodies. *Science* 129:1649.
- LENNOX, E. S. and M. COHN 1967 Immunoglobulins. *Ann. Rev. Biochem.* 36: 365.

- LEVIN, A. S., H. H. FUDENBERG, J. E. HOPPER, S. K. WILSON and A. NISONOFF 1971 Immunofluorescence evidence for cellular control of synthesis of variable regions of light and heavy chains of immunoglobulin G and M by the same gene. Proc. Nat. Acad. Sci. U.S. 68:169.
- MAGE, R. G. 1971 Structural localization, allelic exclusion and linkage relationships of rabbit allotypes. Prog. Immunol. 1st Int. Congr. Immunol. (B. Amos, Ed.) Academic Press, New York. Pp. 47-59.
- MAGE, R. G., G. O. YOUNG and R. A. REISFELD 1968 The association of the c7 allotype of rabbits with some light polypeptide chains which lack b locus allotype. J. Immunol. 101:617.
- MAGE, R. G., G. O. YOUNG-COOPER and C. ALEXANDER 1971 Genetic control of variable and constant regions of immunoglobulin heavy chains. Nature 230:63.
- MÄKELÄ, O. and A. M. CROSS 1970 The diversity and specialization of immunocytes. Prog. Allergy 14:145.
- MANDY, W. J. and C. W. TODD 1970 Rabbit immunoglobulin allotypes A<sub>12</sub>: A new agglutinating specificity. Biochem. Genet. 4:59.
- MCINTIRE, R. and A. ROUSE 1970 Mouse immunoglobulin light chains: alteration of kappa:lambda ratio. Fed. Proc. 29:704.
- MCKEAN, D., M. POTTER and L. HOOD 1973a Mouse immunoglobulin chains. Partial amino acid sequence of a kappa chain. Biochemistry 12:749.
- MCKEAN, D., M. POTTER and L. HOOD 1973b Mouse immunoglobulin chains. Pattern of sequence variation among kappa chains with limited sequence differences. Biochemistry 12:760.
- MILSTEIN, C. 1967 Linked groups of residues in immunoglobulin kappa chains. Nature 216:330.
- MILSTEIN, C. and J. R. L. PINK 1970 Structure and evolution of immunoglobulins. Prog. Biophys. Mol Biol. 21:209.
- MILSTEIN, C., G. G. BROWNLEE, T. M. HARRISON and M. B. MATHEWS 1972 A possible precursor of immunoglobulin light chains. Nature 239:117.
- MILSTEIN, C. and SVASTI, J. 1971 Expansion and contraction in the evolution of immunoglobulin gene pools. Prog. Immunol. 1st Int. Congr. Immunol. (B. Amos, Ed.) Academic Press, N. Y. Pp. 33-46.
- MOLE, L., S. A. JACKSON, R. R. PORTER and J. M. WILKINSON 1971 Allotypically related sequences in the Fd fragment of rabbit Ig heavy chains. Biochem. J. 124:301.

- NATVIG, J. B., H. G. KUNKEL and S. D. LITWIN 1967 Genetic markers of the heavy chain subgroups of human  $\gamma$ G globulin. Cold Spring Harbor Symp. Quant. Biol. 32:173.
- NOLAN, C. and E. MARGOLIASH 1968 Comparative aspects of primary structures of proteins. Ann. Rev. Biochem. 37: 727.
- NOSSAL, G. J. V., A. SZENBERG, G. L. ADA and C. M. AUSTIN 1964 Single cell studies on 19S antibody formation. J. Exp. Med. 119:485.
- NOVOTNY, J., L. DOLEJS and F. FRANEK 1972 Uniformity and species-specific features of the N-terminal amino acid sequence of porcine immunoglobulin  $\lambda$  chains. Eur. J. Biochem. 31:277.
- OHNO, S. 1972 Simplicity of mammalian regulator systems. Develop. Biol. 27:131.
- LOUDIN, J. 1960 Allotypy of rabbit serum proteins II. J. Exp. Med. 112:125.
- LOUDIN, J. and M. MICHEL 1969 Idiotype of rabbit antibodies. II. Comparison of idiotype of various kinds of antibodies formed in the same rabbit against Salmonella typhi. J. Exp. Med. 130:619.
- PARDUE, M. L. and J. GALL 1970 Chromosomal localization of mouse satellite DNA. Science 168:1356.
- PERNIS, B., G. CHIAPPINO, A. S. KELUS and P. G. H. GELL 1965 Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. J. Exp. Med. 122:853.
- PERNIS, B., L. FORNI and L. AMANTE 1971 Immunoglobulins as cell receptors. Ann. N.Y. Acad. Sci. 190:420.
- PETERSON, H. M. and J. R. LAUGHMAN 1963 Intrachromosomal exchange at the bar locus in Drosophila. Proc. Nat. Acad. Sci. U.S. 50:126.
- POLJAK, R. J., L. M. AMZEL, H. P. AVEY and L. N. BECKA 1972 Structure of Fab' new at 6 Å resolution. Nature N. B. 235:137.
- PORTER, R. R. 1971 The combining site of antibodies. Harvey Lect. 65:157.
- POTTER, M. 1967 The plasma cell tumors and myeloma proteins. Methods Cancer Res. 2:105.
- PRAHL, J. W. 1967 The C-terminal sequences of the heavy chains of human immunoglobulin G myeloma proteins of differing isotypes and allotypes. Biochem. J. 105:1019.

- PRAHL, J. W. and R. R. PORTER 1968 Allotype-related sequence variation of the heavy chain of rabbit immunoglobulin G. *Biochem. J.* 107:753.
- PRAHL, J. W., W. J. MANDY and C. W. TODD 1969 The molecular determinants of the A<sub>11</sub> and A<sub>12</sub> allotypic specificities in rabbit immunoglobulin. *Biochemistry* 8:4835.
- PRAHL, J. W., W. J. MANDY, G. S. DAVID, M. W. STEWARD and D. W. TODD 1970 Participation of allotypic markers in rabbit immunoglobulin classes. *Protides Biol. Fluids Proc. Colloq.* 17:125.
- PRESS, E. M., and N. M. HOGG 1969 Comparative study of two immunoglobulin G Fd-fragments. *Nature* 223:807.
- PUTNAM, F., A. SHIMIZU, C. PAUL, T. SHINODA and H. KÖHLER 1971 The amino acid sequence of human macroglobulins. *Ann. N.Y. Acad. Sci.* 190:83.
- RAFF, M. D. 1973 T and B lymphocytes and immune responses. *Nature* 242:19.
- RALPH, P. and A. RICH 1971 Immunoglobulin synthesis in a cell-free system. *Biochemistry* 10:4717.
- RAY, A. and J. J. CEBRA 1971 Localization of affinity-labeled residues in the primary structure of anti-dinitrophenyl antibody raised in strain 13 guinea pigs. *Biochemistry* 11:3647.
- SARMA, V. R., E. W. SILVERTON, D. R. DAVIES and W. D. TERRY 1971 The three dimensional structure at 6 Å resolution of a human  $\gamma$ G1 immunoglobulin molecule. *J. Biol. Chem.* 246:3753.
- SILVERSTEIN, A. M. and R. A. PRENDERGAST 1971 Lymphogenesis, immunogenesis and the generation of immunologic diversity. P. 69. In: *Development Aspects of Antibody Formation and Structure* (J. Sterzl and H. Riah, Eds.) 2nd ed., Vol. I. Academic Press, New York.
- SILVERSTEIN, A. M., J. W. UHR, K. L. KRANER and R. J. LUKES 1963 Fetal response to antigenic stimulus. *J. Exp. Med.* 117:799.
- SINGER, S. J., N. MARTIN and N. THORPE 1971 Affinity labeling of active sites of antibodies and myeloma proteins. *Ann. N.Y. Acad. Sci.* 190:342.
- SMITH, G., L. HOOD and W. FITCH 1971 Antibody diversity. *Ann. Rev. Biochem.* 40:969.
- SMITHIES, O. 1967 Antibody variability. *Science* 157:267.
- SMITHIES, O. 1970 Pathways through networks of branched DNA. *Science* 169:882.

- SMITHIES, O., D. GIBSON and M. LEVANON 1970 Linkage relationships in normal light chains. In: *Developmental Aspects of Antibody Formation and Structure* (J. Sterzl and I. Riha, Eds.) Academic Publishing House of the Czechoslovak Acad. of Sci. Prague, Vol. 1:339.
- SNAPPER, I. and A. KOHN 1971 *Myelomatosis*. University Park Press, Baltimore.
- SOLOMON, A. and C. L. MCLAUGHLIN 1971 Bence Jones protein and light chains of immunoglobulins. IV. Immunochemical differentiation among proteins within each of the three established kappa chain classes. *J. Immunol.* 106:120.
- STAVNEZER, J. and R.-C. C. HUANG 1971 Synthesis of a mouse Ig light chain in a rabbit reticulocyte cell-free system. *Nature N. B.* 230:172.
- STEINBERG, A. G. 1962 Progress in the study of the genetics of human gamma globulin types (the Gm and Inx groups). *Prog. Med. Genet.* 2:1.
- STERZL, J. and A. M. SILVERSTEIN 1967 Developmental aspects of immunity. *Adv. Immunol.* 6:337.
- STORB, U. 1972 Quantitation of immunoglobulin genes by nucleic acid hybridization with RNA from myeloma and spleen microsomes. *J. Immunol.* 108:755.
- SWAN, D., H. AVIV and P. LEDER 1972 Purification and properties of biologically active messenger RNA for a myeloma light chain. *Proc. Nat. Acad. Sci.* 69:1967.
- TERRY, W. D., L. E. HOOD and A. G. STEINBERG 1969 Genetics of Ig kappa chains: Chemical analysis of normal human light chains of differing InV types. *Proc. Nat. Acad. Sci. U.S.* 63:71.
- THOMAS, C. A., JR. 1970 The theory of the master gene. Pp. 973-998. In: *The Neurosciences: Second Study Program* (F. O. Schmitt, Ed.) Rockefeller Univ. Press, New York.
- WALKER, P., W. FLAMM and A. MCLAREN 1969 Highly repetitive DNA in rodents. P. 53. In: *Handbook of Molecular Cytology* (A. Lima-de-Faria, Ed.) North-Holland, Amsterdam.
- WANG, A.-C., H. H. FUDENBERG and J. R. L. PINK 1971 Heavy chain variable regions in normal and pathological immunoglobulins. *Proc. Nat. Acad. Sci. U.S.* 68:1143.
- WANG, A.-C., J. GERGELY and H. H. FUDENBERG 1973 Amino acid sequences at constant and variable regions of heavy chains of monotypic immunoglobulins G and M of a single patient. *Biochemistry* 12:528.

- WANG, A.-C., I. Y. F. WANG, N. MCCORMICK and H. H. FUDENBERG  
1969 The identity of light chains of monoclonal IgG  
and monoclonal IgM in one patient. *Immunochemistry*  
6:451.
- WANG, A.-C., S. K. WILSON, J. E. HOPPER, H. H. FUDENBERG and  
A. NISONOFF 1970 Evidence for control of synthesis of  
the variable regions of the heavy chains of immunoglobulin  
G and M by the same gene. *Proc. Nat. Acad. Sci. U.S.*  
66:337.
- WATERFIELD, M. D., J. E. MORRIS, L. HOOD and C. W. TODD 1973  
Rabbit immunoglobulin light chains: Correlations of  
variable region sequences with allotype markers. *J.*  
*Immunol.* 110:227.
- WEIGERT, M. G., I. M. CESARI, S. J. YONKOVICH and M. COHN  
1970 Variability in the lambda light chain sequences of  
mouse antibody. *Nature* 228:1045.
- WHITEHOUSE, H. L. K. 1967 Crossover model of antibody  
variability. *Nature* 215:371.
- WILKINSON, J. M. 1969a Variations in the N-terminal  
sequences of heavy chains of immunoglobulin G from rabbits  
of different allotype. *Biochem. J.* 112:173.
- WILKINSON, J. M. 1969b  $\alpha$ -Chains of immunoglobulin A from  
rabbits of different allotype: Composition and N-terminal  
sequence. *Nature* 223:616.
- WILKINSON, J. M. 1970 Genetic markers of rabbit immuno-  
globulins. *Biochem. J.* 117:3P.
- WU, T. T. and E. A. KABAT 1970 An analysis of the sequences  
of the variable regions of Bence Jones proteins and  
myeloma light chains. *J. Exp. Med.* 132:211.



Dr. Hood (right) with Dr. Schlessinger at the Symposium



Dr. A. E. Longley and his daughter Mrs. Mary Coe  
at the Symposium



Drs. Pittenger and Stadler



Drs. Roman and Rhoades