References and Notes

1. S. P. Ó Ríordáin and G. Daniel, *New Grange* (Praeger, New York, 1964).
2. G. Hawkins, *Stonehenge Decoded* (Doubleday, Garden City, N.Y., 1965).
3. A. Thom, *Megalithic Sites in Britain* (Oxford, London, 1967).
4. G. Daniel, *The Megalithic Builders of Western Europe* (Hutchenson, London, 1959), pp. 111–113.
5. Actually this tumulus represents what Thom calls a type D ring, one in which the two pivot stakes are placed at one-third the radius from a_1 rather than at the midpoint. As Fig. 3 shows, the pivots can be kept at the midpoint of the radius if a_2 is allowed to move inside the design.
6. L. B. Borst, *Science* **163**, 567 (1969).
7. The monument at Borrowston Rig requires anchor points that are placed farther apart on the circumference than the intersections made by the trisecting radii demand. In this design the circle of the smaller end arc passes through the center of the larger end arc, and the wider placement of the anchor stakes may have been made on this account. The site at Maen Mawr had anchor stakes that were closer together on the circumference. If the radial line from p_2 to either anchor is taken as a hypotenuse of a right triangle one side of which is half the anchor line, then the lengths of the sides of this triangle (in megalithic half-yards) are 14, 17, and 22, which is nearly Pythagorean ($14^2 + 17^2 = 485 \approx 484$).
8. M. Gardner, *Sci. Amer.* **221**, 239 (1969).
9. To mention one such consequence, with the megalithic method concentric rings, such as those found at Woodhenge, can be drawn without using any special mensuration technique. That is, a rope can be lengthened by an unspecified amount and a concentric design can be drawn at once. With a flexible compass, on the other hand, the large arc in a type II egg, for example, can be drawn without special measurement but the other arcs must be changed by x amount to remain equidistant from the perimeter of the original figure.
10. Thom presents a good argument for use of a standard length of 2.72 feet (one megalithic yard) in the construction of these structures, as well as of others in both the Old World and the New.

11. The writing of this article was supported in part by the Oklahoma State University Research Foundation.

Mechanism of Antibody Diversity: Germ Line Basis for Variability

Analysis of amino termini of 64 light chains indicates that much antibody variability is present in the germ line.

Leroy Hood and David W. Talmage

The nature of the genetic control of antibody variability is one of the most fascinating and approachable problems in mammalian genetics. Vertebrate organisms appear to be capable of synthesizing thousands of different antibody sequences, each presumably encoded by a different antibody gene. How then do these genes arise? The somatic theory of antibody diversity postulates that antibody genes arise by hypermutation from a few germ line genes during somatic differentiation. In contrast, the germ line theory postulates that vertebrates have a separate germ line gene for each antibody polypetide chain the creature is capable of elaborating.

We discuss here certain patterns that have emerged from amino acid sequence analysis of antibody polypeptide chains. These patterns indicate that much of the sequence diversity is present in the germ line. We also discuss why the germ line theory seems to be the simplest explanation for antibody diversity.

Dr. Hood is a senior investigator in the Immunology Branch of the National Cancer Institute, Bethesda, Maryland. Dr. Talmage is professor of microbiology and dean of the University of Colorado Medical School, Denver. Requests for reprints should be sent to Dr. Talmage.

General Immunoglobulin Structure

All five recognized classes of antibodies (immunoglobulins) in mammals contain two distinct polypeptide chains, called light and heavy chains (*1*). For example, there are two identical light and two identical heavy chains per molecule in the major serum immunoglobulins, the immunoglobulin G class; but the light and heavy chains differ chemically in different antibodies. Since normal antibodies produced against the simplest of antigenic determinants are generally heterogeneous, advantage has been taken of the homogeneous immunoglobulins produced in large quantities by plasmacytomas in humans and in the highly inbred BALB/c mouse (*2*). Light chains are frequently excreted in the urine of individuals with certain plasmacytomas; these light chains have been called Bence Jones proteins. This article deals only with light chains because relatively little information on comparative sequences is available for heavy chains.

Light chains from most mammalian species including man are of two types, lambda and kappa, which are readily distinguished by serological and chemical criteria (*1*, *3*). Light polypeptide chains have two parts: a common region (approximately residues 108 to 215), which is essentially invariant for a given light chain type and species and a variable region (approximately residues 1 to 107) which is different for each well-characterized protein (*4–6*). Presumably each variable region sequence directs the folding of a unique antibody light chain configuration.

Two patterns emerge when the amino acid sequences from many light chain variable regions are compared. (i) All kappa and lambda variable region sequences can be divided into subgroups on the basis of their similarity to one of eight prototype sequences (Fig. 1). (ii) Relatively minor deviation from the prototype sequences occurs among individual proteins of a given subgroup (Fig. 1) and is designated intrasubgroup variation. We first discuss the variable region subgroups and the reasons for concluding that each light chain subgroup must be encoded by at least one distinct germ line gene. We then discuss why we believe that the intrasubgroup variation may also be encoded by separate germ line genes.

Variable Region Subgroups of Kappa Chains

The nearly complete sequences of six variable regions of human kappa chains are known (Roy, Ag, Cum, Mil, Eu, and Ti), and partial sequences of more than 35 others have been determined. All variable regions of kappa chains can be assigned to one of three subgroups on the basis of their similarity to one of three sets of linked amino acid sequences (Figs. 1 and 2). These three sets of linked amino acids or three "prototype sequences" are derived by examining the proteins of

one subgroup and noting the major amino acid residue at each position (*3, 7*). For example, most V_{KI} proteins have Asp (*8*) at position 1, Ile at position 2, Gln at position 3, and Met at position 4 (Fig. 1). Thus the prototype V_{KI} sequence in this region would be Asp-Ile-Gln-Met. The three prototype sequences are given for the first 20 residues in Fig. 3, and it is evident that each prototype sequence differs from its counterparts by 5 to 6 residues in 20. At the amino terminus, most proteins within a subgroup deviate from their respective prototype sequence by approximately one residue in 20 (Fig. 1). Although the sequence data are more limited after position 20, these subgroup distinctions extend through the variable region at least to position 94 (beyond which subgroups cannot be distinguished in the present data). It is particularly important to note that in the carboxy-terminal portion of the V_K region, the 12 sequences available fall into the same kappa subgroups; that is, individual proteins do not switch subgroups (Fig. 2). The separation between subgroups is indicated by the fact that kappa chains within a subgroup are identical for 73 to 87 percent of their sequences, whereas comparisons between subgroups show a 51 to 68 percent identity (Table 1).

Two regions in light chains seem to be more variable than the remainder of the V_K region. Both "hypervariable" regions (residues 25 to 35 and 89 to 96) are just COOH-terminal to the cysteines (that is, Cys 23 and Cys 88) which form a disulfide bridge in the V region. Perhaps one or both of these sites is directly involved in the antigen-combining site. In spite of this variability, subgroup specific residues are present even in these regions (Fig. 2).

In addition to the linked amino acid residues (Fig. 3), yet another criterion may distinguish kappa subgroups. Homologous sequence gaps (*3*) may separate the proteins of one subgroup from those of a second. For example, the

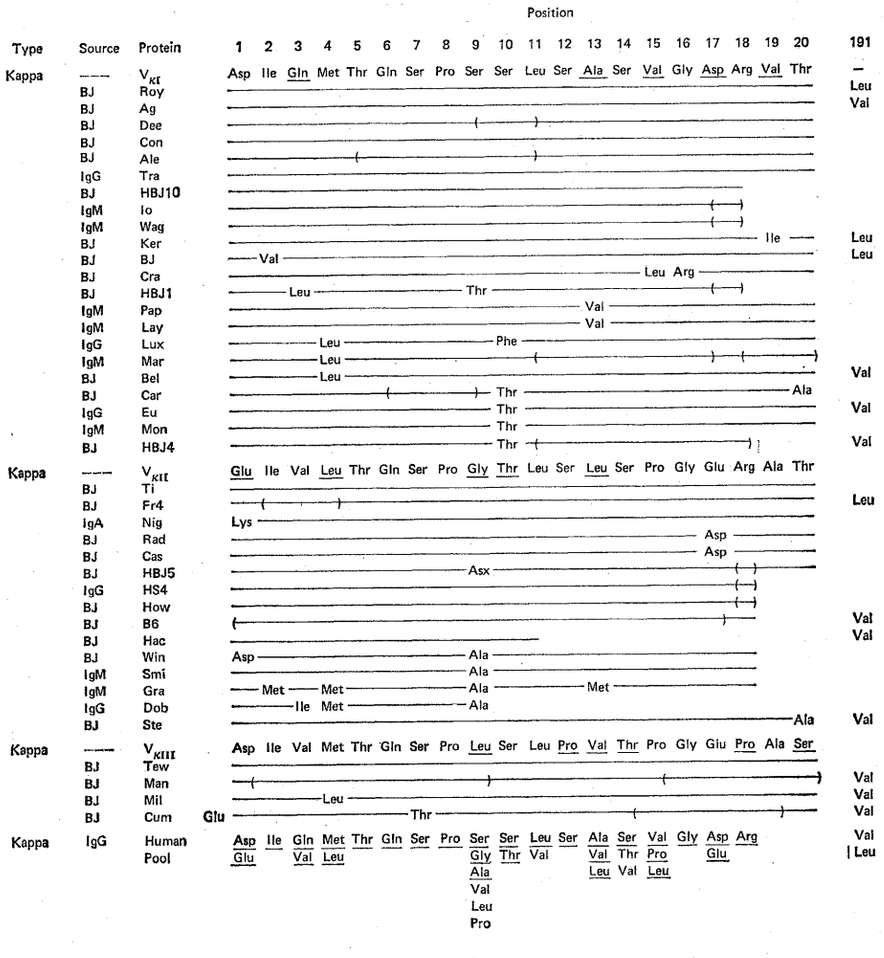

Fig. 1. Amino terminal sequences of human lambda and kappa chains divided into subgroups on the basis of associated amino acid residues at certain positions. The prototype sequence is given first with those residues unique to the subgroup underlined. In subsequent sequences, only those residues that differ from the prototype are included. The source of the light chain is given, with BJ indicating Bence Jones protein. Position 191 in kappa chains has a Leu-Val interchange (Inv marker). Position 190 in lambda chains has an Arg-Lys interchange (Oz group). Residues for positions 191 in kappa and 190 in lambda are in references listed below except for HBJ4, Hac, and Dob (*63*) and for the human pool (*13*). Most of this data is summarized in two papers (*17, 64*). Roy and Cum are taken from (*65*); Ag, Ha, Bo, and Sh (*6*); Ker and BJ (*66*); Cra, Pap, Lux, Mon, Con, Tra, Nig, Win, Gra, Cas, Smi, and human pool (*9*); Lay, Mar, Wag, Io, How, and Koh (*67*); Rad and Fr4 (*7*); Man, Bel, and B6 (*64*); Eu (*68*); Mil (*28*); Tew (*69*); Ti (*70*); HBJ1, HBJ4, HBJ10, HBJ5, and HS4 (*3*); HBJ2, HBJ7, HBJ8, HBJ11, and HBJ15 (*25*); HS92, HS78, HS94, HS68, HS77, HS70, and HS86 (*17*); Hac, Dob, and Pal (*71*); BJ98 (*72*); Hul (*73*); X (*74*); Kern and 111 (*75*); and Car, Dee, and Ale (*7*).

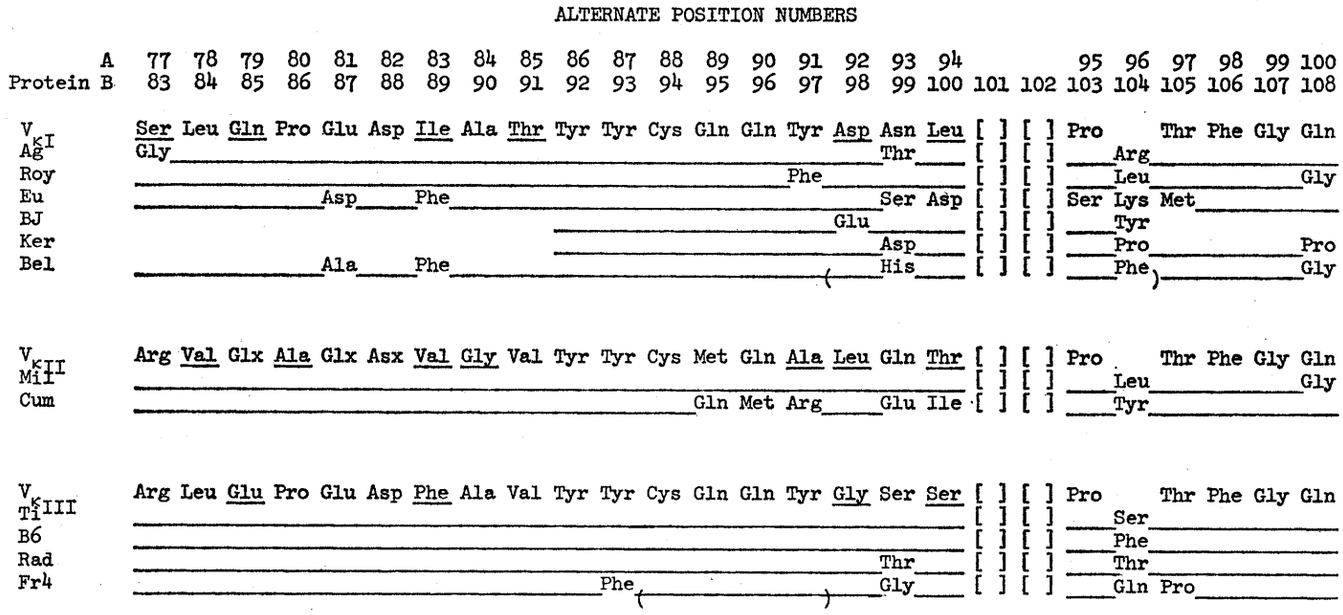

Fig. 2. Carboxy-terminal portion of variable region. The alternative numberings of positions are (A) that used in most previous references and (B) that used in the computer comparison of various chains mentioned in this paper (see Tables 1 to 3). Subgroups cannot be discerned beyond positions 94 to 100, the same point at which a deletion occurs in V_κ regions compared with $V\lambda$ regions. References are given in Fig. 1.

two sequenced $V_{\kappa III}$ proteins have an insertion of four to six amino acids at or near position 30, as compared with the $V_{\kappa I}$ proteins (the exact number of extra residues in the $V_{\kappa III}$ protein Mil is uncertain) (see Table 1). A single $V_{\kappa II}$ protein has an insertion of one amino acid in this region. Thus, precisely located sequence gaps and insertions may also distinguish the three variable region subgroups of human kappa chains.

What do subgroups mean in relation to gene structure? It is difficult to imagine a genetic mechanism that could generate, from a single gene, the sequences of $V_{\kappa I}$, $V_{\kappa II}$, and $V_{\kappa III}$ proteins with linked amino acid residues and gaps. Indeed, it seems that each V_κ subgroup is encoded by a separate germ line gene (or genes).

Are these distinct kappa subgroup genes alleles? Niall and Edman (*9*) found that all three kappa subgroups are present in a pool of light chains; for example, Glu at position 1 is characteristic of $V_{\kappa II}$ proteins, Gln at 3, of $V_{\kappa I}$ proteins, and Leu at 9, of $V_{\kappa III}$ proteins (see Fig. 1, Human Pool). Two experiments suggest that every individual can produce $V_{\kappa I}$ and $V_{\kappa II}$ proteins. Light chains from 13 normal individuals have the same major alternatives as the normal pool at each of the first four positions (*10*). With three alleles the probability of selecting seven consecutive identical heterozygotes from the normal population is considerably less than one in a thousand $[(2 \times 1/9)^7]$ if we assume an equal frequency. Furthermore, trypsin-digested peptides characteristic of $V_{\kappa I}$ and $V_{\kappa II}$ proteins have been isolated from the light chains of 17 normal individuals (*7*). The question of whether or not the minority group $V_{\kappa III}$ (11 percent of the pool of kappa myeloma chains according to Fig. 1) is found in all normal individuals has not yet been answered conclusively, but there is no reason to suspect that it is not. Thus, all individuals produce $V_{\kappa I}$ and $V_{\kappa II}$

Table 1. Comparison of individual human variable regions within and between groups. Proteins were translated to nucleotide sequence and aligned to show maximum homology; the percentages of nucleotide and amino acid homology were then determined. The numbering was made according to longest chain combination (116 amino acid residues); deletions were placed in all V_κ chains at position 113, in all V_κ chains and $V_{\lambda III}$ chains at positions 101 and 102, in all V_λ chains at position 9, and in $V_{\lambda III}$ and $V_{\lambda V}$ chains at position 1. Deletions around residue 30 were located as follows: $V_{\kappa I}$ 30 to 35; $V_{\kappa II}$ 32 to 36; $V_{\lambda I}$ (New) 31 to 34; $V_{\lambda I}$ (Ha) 31, 33 to 34; $V_{\lambda III}$ 31 to 36; $V_{\lambda IV}$ 31, 33 to 34, and $V_{\lambda V}$ 31 to 36. All chains were programmed for a computer to record comparisons as same nucleotide, different nucleotide, unknown, gap-gap, or gap-nucleotide. The percentage of nucleotide homology is the number of identical nucleotides times 100 divided by the sum of all known nucleotides. The single extra residue on Cum at the amino terminus was not counted. References are given in Fig. 1. Uncertainty regarding chain length of Mil and New and the location of a deletion in Ti gives only a small uncertainty in the percentage homology for these proteins.

Regions compared	Homology (%)		Number of nucleotide-gap positions	Amino acid chain length difference
	Nucleotide	Amino acid		
Within subgroups				
$V_{\kappa I}$ (Roy) vs. $V_{\kappa I}$ (Ag)	92	87	0	0
$V_{\kappa I}$ (Roy) vs. $V_{\kappa I}$ (Eu)	80	73	0	0
$V_{\kappa I}$ (Ag) vs. $V_{\kappa I}$ (Eu)	81	74	0	0
$V_{\kappa III}$ (Mil) vs. $V_{\kappa III}$ (Cum)	85	82	0–6	0–2
$V_{\lambda I}$ (New) vs. $V_{\lambda I}$ (Ha)	84	75	0–3	0–1
$V_{\lambda III}$ (Kern) vs. $V_{\lambda III}$ (X)	79	73	0	0
Between subgroups				
$V_{\kappa I}$ (Roy) vs. $V_{\kappa II}$ (Ti)	74	61	3–9	1
$V_{\kappa I}$ (Eu) vs. $V_{\kappa II}$ (Ti)	76	68	3–9	1
$V_{\kappa I}$ (Roy) vs. $V_{\kappa III}$ (Cum)	64	51	18	6
$V_{\kappa III}$ (Cum) vs. $V_{\kappa II}$ (Ti)	72	64	15	5
$V_{\lambda I}$ (Ha) vs. $V_{\lambda III}$ (Kern)	68	55	18	6
$V_{\lambda I}$ (Ha) vs. $V_{\lambda IV}$ (Bo)	75	66	0	0
$V_{\lambda I}$ (Ha) vs. $V_{\lambda V}$ (Sh)	68	56	12	4
$V_{\lambda III}$ (Kern) vs. $V_{\lambda IV}$ (Bo)	65	58	18	6
$V_{\lambda III}$ (Kern) vs. $V_{\lambda V}$ (Sh)	72	64	6	2
$V_{\lambda IV}$ (Bo) vs. $V_{\lambda V}$ (Sh)	68	60	12	4

Subgroup	Position 1	3	4		9	10		12	13	14	15		17	18	19	20
$V_{\kappa I}$	Asp	-	Gln	Met	----	Ser	Ser	-	Ser	Ala	Ser	Val	-	Asp	Arg	Val	Thr
$V_{\kappa II}$	Glu	-	Val	Leu	----	Gly	Thr	-	Ser	Leu	Ser	Pro	-	Glu	Arg	Ala	Thr
$V_{\kappa III}$	Asp	-	Val	Met	----	Leu	Ser	-	Pro	Val	Thr	Pro	-	Glu	Pro	Ala	Ser

Fig. 3. Prototype sequences for three subgroups of human kappa variable regions. Hyphens represent residue identities in all prototype sequences.

proteins (by inference $V_{\kappa III}$ proteins); hence these alternatives cannot be alleles. Therefore it appears that at least three germ line genes must encode V_κ regions in each individual. A paradox arises, however, in that the common region of the kappa chains appears to be encoded by a single gene and not by three genes).

Common Region of Kappa Chains

The common regions of the kappa chains (residues 108 to 214) from light chains of different human myeloma proteins are identical except for a leucine-valine substitution at position 191 (*4, 11*). This subtle interchange can be detected in normal light chains by serological and chemical techniques. Using these techniques, one can do pedigree analyses which have shown that these two variants are inherited in Mendelian fashion, thus suggesting that the common region of human kappa chains is encoded by a single structural gene with two alleles (*12, 13*). The common region of the mouse kappa chains with a single amino acid sequence also appears to be encoded by a single gene (*5, 14*).

That the common region of the kappa chain is encoded by a single structural gene is also supported by evolutionary considerations. Genes for common regions of human and mouse light chains probably descended from a common ancestor (these common regions have the same number of residues and they are identical at 64 out of 107 positions, as indicated in Table 2). If multiple genes for the common region of the kappa chains existed in this ancestor and now exist in mouse and man, then it is difficult to explain how in the face of the normal mutation rate (40 percent of the residues differ in the chains of mouse and man) all copies within each species maintain the same amino acid sequence. Furthermore, if there are multiple genes for the common regions, it is difficult to see how the leucine and valine genetic markers evolved and why recombination has not randomized them. Thus the common region for kappa chains in man (and mice) apparently is encoded by a single gene (*15*).

Two Genes—One Polypeptide Chain

We have previously concluded that V_κ regions are encoded by at least three germ line genes. Since $V_{\kappa I}$, $V_{\kappa II}$, and $V_{\kappa III}$ sequences can be associated with a single variant for the common region (see kappa position 191 in Fig. 1), it seems that each kappa light chain is encoded by two genes (V_κ and C_κ), which are expressed as a single polypeptide chain. A similar picture seems to be true of human lambda chains.

Table 2. Comparison of regions in individual human and mouse kappa chains. M or m designates a mouse protein and h a human protein. M-41 gaps were placed at the same positions as $V_{\kappa I}$ proteins, M-70 gaps were placed at positions 33 to 34, 101 to 102, and 113. Otherwise see legend to Table 1.

Regions compared	Homology (%)		Number of nucleotide-gap positions	Amino acid chain length difference
	Nucleotide	Amino acid		
M-41 vs. M-70	64	57	12	4
M-41 vs. $hV_{\kappa I}$ (Ag)	77	64	0	0
M-41 vs. $hV_{\kappa II}$ (Ti)	71	58	3–9	1
M-41 vs. $hV_{\kappa III}$ (Cum)	59	49	18	6
M-70 vs. $hV_{\kappa I}$ (Ag)	71	60	12	4
M-70 vs. $hV_{\kappa II}$ (Ti)	72	62	9	3
M-70 vs. $hV_{\kappa III}$ (Cum)	71	59	6	2
mC_κ vs. hC_κ	70	60	0	0
mC_κ vs. hC_λ	57	44	6	2

Lambda Chains

The human lambda chain also has a common and a variable region (*6*). The common region is somewhat more complex in character than its kappa counterpart.

There are four or five subgroups in the variable region of the lambda chain, which again are defined by prototype sequences and gaps (Fig. 1 and Table 1) (*16, 17*). Some lambda subgroups appear less distinct than their kappa counterparts (compare $V\lambda_I$ and $V\lambda_{IV}$); nevertheless, statistical arguments support these subgroup distinctions (*17*). The existence of at least four lambda subgroups does require four germ line variable genes. Again it appears that every individual can synthesize most, if not all, of the lambda subgroups (*10*).

The common region of the lambda chain is more complex than its kappa counterpart in that it appears to be encoded by two separate germ line genes. The common region in the lambda chain is essentially invariant in light chains from myeloma proteins except for an amino acid interchange, arginine-lysine, at position 190 (*18*). Since every individual can synthesize both the arginine and lysine forms, two genes for the common region of the lambda chain genes are probably present in the germ line (*19*).

Each of the four or five subgroups of the lambda variable region can be found attached to either form of the lambda common region (see lambda position 190 in Fig. 1). Thus, the lambda chain must also be encoded by two distinct germ line genes (a V_λ and a C_λ) which are expressed as a single polypeptide chain. Otherwise each lambda chromosome would need eight copies of the lambda common region, an arginine and a lysine form for each subgroup. The evolution of such a system would be difficult if not impossible to explain.

The existence of light chain subgroups permits us to determine the minimum number of germ line genes encoding light chains— four for human kappa chains (three V_κ and one C_κ), six for human lambda chains (four V_λ and two C_λ).

Insertional Mechanism

For both types of human antibody light chains two separate light chain genes (variable and common region genes) or their products are joined

during differentiation of the immunocyte by a special insertional mechanism since the light chain is a single polypeptide chain. Although joining could occur at any stage of protein synthesis (DNA, RNA, or protein), four lines of evidence suggest that joining does not occur at the protein level. (i) If the light chain is synthesized as two separate units and joined during protein synthesis, it must have two growing points. Experiments designed to resolve the problem of the number of growing points suggest that there is a single light chain growing point (*20*). (ii) The size of the light and heavy chain polysomes is consistent with the synthesis of polypeptides having a molecular weight of 25,000 and 50,000 respectively and not with the synthesis of half molecules (*21*). (iii) The immunoglobulin products of myeloma tumors are stable through hundreds of transplantation generations, suggesting irreversible union of common and variable regions at the DNA level (*22*). (iv) Finally, one heavy chain disease protein appears to have a large segment of the V and C regions deleted (residues 19 to 217) (*23*). In this case the V region must have been joined to the C region at the DNA level before the deletion occurred, otherwise one must postulate that two independent deletions occurred in separate *V* and *C* genes. Furthermore, unless joining occurred at the DNA level one would have to postulate that new V region and C region recognition sites were generated so joining could occur either at the RNA or polypeptide levels (see next paragraph). Thus, it seems very likely that the joining of V and C regions occurs at the level of structural genes.

Since kappa variable and common region genes only combine with one another and never with lambda variable or common genes (*24, 25*), the proposed insertional mechanism must have a means of recognizing and joining only appropriate gene pairs. The simplest model for this mechanism is the insertion of the lambda phage into the bacterial chromosome. The genome of this bacteriophage and the bacterial chromosome each must have a specific and complementary recognition site (*26*). As few as 12 nucleotides are sufficient for such a recognition site (*27*). These recognition sites in the immunoglobulin system must be in or adjacent to both common and variable genes. These sites could be expressed in the final polypeptide product at the NH_2-terminal portion of the common region, the COOH-terminal portion of the variable region, or under special circumstances in neither of these regions [see the model proposed in (*28*) and the discussion in (*29*)]. Perhaps our inability to recognize the COOH-terminal portion of the V_K region (residues 95 to 107) is a reflection in part of the insertional mechanism.

One other mechanism which may play a role in preventing the attachment of V_K regions to the C_λ region could be the physical separation in the mammalian genome of the C_λ and C_K genes (and presumably the corresponding V_λ and V_K genes). There is a preliminary indication that the genes for rabbit λ and κ light chains are indeed unlinked (*30*). In contrast the common genes of various classes and subclasses of heavy chains which may share V_H regions seem to be closely linked to one another (*31*).

POSITION

Protein	10	20	22	24	28	29	30	31	32	34	39	41	42	46	48	49	50
Ag	Ser	-------		Gln	--------		Asn-	His	Tyr	Asn	Gly-	Lys-	-----	Ile	--------		Asp
Roy	Ser	-------		Gln	----------		Ile		Phe	Asn	----------						Asp
Eu	Thr	-------		Arg	Ser-----		Asn-	Thr	Trp	Ala	----------				Met-	----	Lys
HBJ4	Thr	Ser-	Ala	Arg	Asx-	Val	-----	Asn	Trp	Ala	--------	Glx	---------		Ser		Lys
Number of Base Changes	1			1					2	2							2

Fig. 4. A comparison of the first 50 residues of four proteins formerly designated $V_{\kappa I}$ indicating further subdivision of the $V_{\kappa I}$ subgroup into two subgroups, $V_{\kappa I}$ and $V_{\kappa IV}$. Hyphens represent residues identical in two or more of the sequences. Subgroup-distinguishing residues are in boxes and are numbered. Number of base changes indicates the number of changes that would be required to convert the $V_{\kappa I}$ codon at a given position into a $V_{\kappa IV}$ codon. Protein HBJ 4 is taken from (*34*) and the others from the references given in Fig. 1.

Intrasubgroup Variability

Each light chain subgroup is probably encoded by one or more distinct germ line genes. We shall now examine the genetic basis for the variation which occurs within a subgroup. We have analyzed the intrasubgroup variability for 41 kappa and 23 lambda proteins over the 20 residues of the NH_2-terminal; thus more than 1200 residues were examined by comparing the amino acid sequence of each specific protein with its prototype subgroup sequence. There are 63 intrasubgroup amino acid interchanges which are indicated in Fig. 1 as breaks in the lines for individual proteins. Three generalizations emerge from these data; two of these generalizations suggest that much intrasubgroup variation is carried in the germ line.

1) The variation which occurs within the subgroups of antibody light chains is indistinguishable from that which occurs among evolutionarily related sets of proteins such as cytochromes, fibrinopeptides, and hemoglobulins. This is demonstrated (i) by the predominance of single base substitutions (60 of 63 intrasubgroup interchanges); (ii) by the random nature of transversional and transitional base changes (26 transversions and 15 transitions) (*32*); and (iii) by the striking tendency of G to mutate more frequently than is expected (13 G's mutated, whereas only 7 A's, 5 U's and 5 C's were substituted). Each of these properties is demonstrated by other evolutionarily related sets of proteins, such as the cytochromes (*33*), which are obtained from different species and are obviously encoded by distinct germ line genes.

2) The substitutions which occur at a single position within a subgroup are highly restricted, suggesting the presence of additional germ line genes. Seven positions in the kappa chains and eight in the lambda chains show two or more substitutions within a single subgroup (see $V_{\kappa I}$ positions 4, 10, and 13 in Fig. 1). Indeed, 47 of the 66 base substitutions are represented in these 15 positions. Furthermore, 29 of the 47 multiple substitutions are represented by identical replacements occurring at the same position within a single subgroup (see $V_{\kappa I}$ positions 4, 10, and 13 in Fig. 1).

Do those subsets of proteins with

identical replacements at a given position (for example, $V_{\kappa I}$ proteins Eu, HBJ4, Mon, and Car) have additional linked substitutions farther down the polypeptide chain? Both Eu and HBJ4 are compared for 50 residues with myeloma proteins Ag and Roy in Fig. 4. Proteins Eu and HBJ4 share five residues which distinguish them from the Roy and Ag proteins (*34*). Clearly Eu and HBJ4 belong to a fourth kappa subgroup in that they have linked amino acid residues which seem to distinguish them from other $V_{\kappa I}$ proteins. Preliminary data indicate that the $V_{\kappa II}$ proteins with Ala at position 9 will belong to a fifth kappa subgroup (*34*). Thus as additional sequence data are gathered, it seems likely that many of the positions at which multiple identical substitutions are present will indicate additional subgroups in the light chains. As additional subgroups are delineated from the major subgroups already defined, the intersubgroup differences will be less extensive. Ultimately the definition of a subgroup must come to rest on arbitrary statistical criteria related to the number of shared differences between two sets of proteins. In any case, it is clear that much of the light chain diversity indicated in Fig. 1 is encoded by multiple germ line genes. If each of the positions at which identical replacements occur do represent additional subgroups (for example, $V_{\kappa I}4$, 10, and 13; $V_{\kappa II}4$, 9, and 17; $V\lambda 11$, 13, and 14; and $V\lambda_{II}5$), human light chains would be encoded by a minimum of 18 germ line genes. This number will certainly increase as more data are gathered.

3) There is no evidence to suggest that the intrasubgroup variation has been generated by a somatic recombinational mechanism. No recombination has been detected between *V* genes of different subgroups (Figs. 1 and 2). Furthermore, in more than half of the intrasubgroup substitutions (26 out of the 47 substitutions in which the base change can be identified), it can be determined that the substituted base is not present at the corresponding position in any of the prototype genes of the same light chain type. Thus, simple recombination among three prototype V_{κ} genes cannot explain more than half of the intrasubgroup variability (*35*).

Genetic polymorphism probably does not account for most of the intrasubgroup variation. If much of the intrasubgroup variation is encoded in the germ line (rather than being the result of somatic differentiation), do these

	Position					
Protein	1	2	3	4	5	6
MOPC 149	Asp	Ile	Gln	Met	Thr	
MBJ 41	Asp	Ile	Gln	Met	Thr	Gln
MBJ 70	Asp	Ile	Val	Leu	Thr	Gln
MOPC 46	Asp	Ile	Val	Leu	Thr	Gln
Adj PC9	Glp	Ile	Val	Leu		
MOPC 157	Asp	Ile	Val			
MBJ 6	Asp	Ile	Val	Val	Thr	Gln

Fig. 5. Amino terminal sequences of BALB/c (mouse) kappa chains. MOPC 149, MOPC 46, Adj PC 9, and MOPC 157 were taken from (*37*); MBJ 41, 70, and 6 were taken from (*36*).

variant genes all occur in the same individual or do they represent genetic polymorphism in the human population? Genetic polymorphism certainly must be responsible for some of the sequence variation, but the variation at the amino termini of kappa chains from the inbred BALB/c mouse is similar to that seen (Fig. 5) in the human population (*36–38*). Thus if one accepts the genetic identity of individual BALB/c mice, polymorphism does not appear to be responsible for a majority of the V region substitutions.

Since the variation which occurs within the subgroups of antibody light chains is indistinguishable from that which occurs among evolutionarily related sets of proteins, random single base mutation followed by selection seems to be the obvious mechanism for generating antibody diversity. The question is whether antibody genes are produced by mutation and selection during somatic differentiation or during the evolution of higher creatures. The sequence data cannot yet give us a direct answer to this question, but the germline hypothesis appears to be compatible with the sequence data and somewhat simpler than its somatic counterparts.

Theories of Antibody Diversity

Somatic theories of antibody formation suggest that a limited number of germline genes are responsible for antibody diversity (*29, 39–43*). This diversity is generated during differentiation of the immunocyte by an appropriate mutational mechanism (hypermutation, recombination, translational ambiguity). Although space does not permit a thorough discussion of the somatic theories individually, several general comments should be made. (i) It is generally agreed that somatic theories postulate that subgroup variability is encoded in the germ line, whereas an ad hoc mutational mechanism must generate the intrasubgroup variation. (ii) With a small number of genes this mechanism cannot be random in character because of the enormous amount of cell wastage that would ensue (*44*). (iii) The somatic mutational mechanism must be restricted to the variable gene (the common gene does not show similar variation). (iv) Although there is no direct evidence for a recombinational model, one cannot rule out the possibility that recombination occurs among a large number of genes—at least ten for each subgroup, which are themselves isolated from the genes of other subgroups. Hence the most attractive somatic theories, simple hypermutation and recombination, would need to be far more complex than was initially proposed (*39, 40, 42, 43*) [see (*45*) for a more thorough discussion of these points]. Thus we discuss below what appears to be a simpler solution to the problem of antibody diversity, the germ line theory.

Germ Line Theory

The germ line (multigene) hypothesis (*28, 46*) states that each organism has one variable gene in its germ line for each unique antibody polypeptide chain it can synthesize and that a separate gene encodes each type of common region. During somatic differentiation of the immunocyte, a single variable gene or its product is joined to the common gene or its product at some stage of protein synthesis, and the resulting differentiated immunocyte synthesizes one type of light chain and, presumably by a similar mechanism, one type of heavy chain. According to this theory, the variable genes arise by the normal process of chemical evolution, that is, gene duplication followed by mutation and selection. We first show that the sequence data are consistent with such a hypothesis and then discuss three questions concerned with the evolution of such a system.

Both the intra- and the intersubgroup variation is similar to that seen in evolutionarily related sets of proteins (see point 1 under intrasubgroup variation and remember that hemoglobin chains from different species can be divided by identical criteria into "subgroups"—such as the α, β, and δ "subgroups"). Furthermore, a typical phylogenetic tree

can be constructed from the 20 residues of the NH_2-terminal of 41 kappa and 23 lambda chains by the method of Fitch and Margoliash (*47*) (Fig. 6). While this tree is only an approximation of the real variable gene tree (only one-fifth of the variable region sequence is used), it does correlate perfectly with a similar light chain tree constructed from 13 complete variable sequences (*48*) and with data in Table 1. The kappa subgroups are seen as major branches on the evolutionary tree, whereas the intrasubgroup variations represent the terminal twigs of each subgroup branch.

Additional branching (subdivisions) of the major subgroups would be expected with a germ line model of antibody diversity as more sequence data accumulates. Such subdivisions are evident even now. For example, the $V_{\kappa I}$ branch designated 10 (enclosed by a dotted circle) in Fig. 6 is supported by a comparison of the first 50 residues of Eu, HBJ4, Ag, and Roy given in Fig. 4. As indicated earlier, five positions distinguish the Roy and Ag proteins from the HBJ4 and Eu proteins. This suggestion that Eu and HBJ4 belong to a fourth kappa subgroup ($V_{\kappa IV}$) is strengthened by the fact that three of the five substitutions are changes of two bases. Hence, as the intersubgroup subdivisions (branching) become more subtle, it will be impossible to distinguish them from the intrasubgroup variation. Indeed it seems reasonable to postulate that there is no real difference between the intra and the intersubgroup variation; rather these merely represent events that have occurred at different points of time in the evolution of multiple *V* genes. As will be discussed subsequently, gene duplication could generate many daughter genes which all show an early successful mutation (intersubgroup variation) whereas later mutations (intrasubgroup variation) may be present in just a single gene.

Complete sequence data on two mouse kappa chains provide an indication of the time scale of the evolution of light chains. Mouse protein MBJ-41 has striking similarities to human $V_{\kappa I}$ proteins over the first 60 residues, including the gap at residues 31 to 36. The remainder of protein MBJ-41 and most of protein MBJ-70 cannot be correlated with any of the kappa subgroups, an indication that some of the present subgroup distinctions had evolved before the human and mouse lines diverged. The remainder of the subgroup and all of the intrasubgroup variation are more recent events.

In human and mouse proteins, variable and common regions seem to be diverging at equal rates. For example, the mouse protein MBJ-41 differs from human $V_{\kappa I}$, $V_{\kappa II}$, and $V_{\kappa III}$ proteins by 36, 42, and 51 percent of its sequence respectively (Table 2). The mouse protein MBJ-70 differs from these same proteins by 40, 38, and 41 percent. The common regions of human and mouse proteins have diverged by 40 percent of their sequences. Thus mouse and

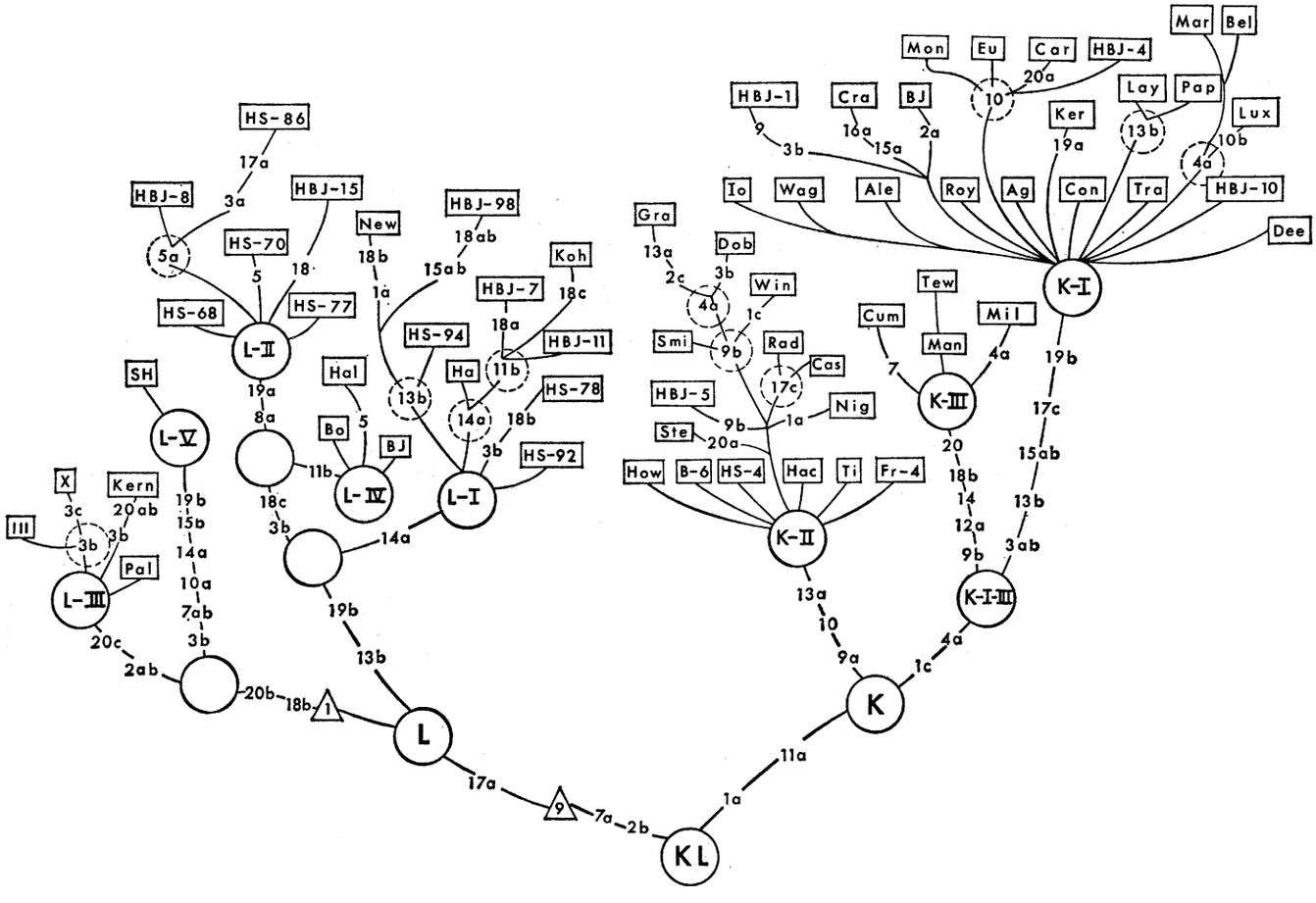

Fig. 6. A "phylogenetic" tree constructed from the amino terminal 20 residues of 41 kappa and 223 lambda proteins by the method of Fitch and Margoliash (*76*), which reduces the number of mutations to a minimum. The 64 proteins are indicated by closed rectangles. Deletions are indicated by triangles and mutations by numbers. The letters *a*, *b*, or *c* indicate which nucleotide is changed according to the genetic code. Closed circles indicate major classes and subclasses. Dotted circles indicate subdivisions of the subclasses, or highly improbable identical somatic mutations occuring in two different individuals.

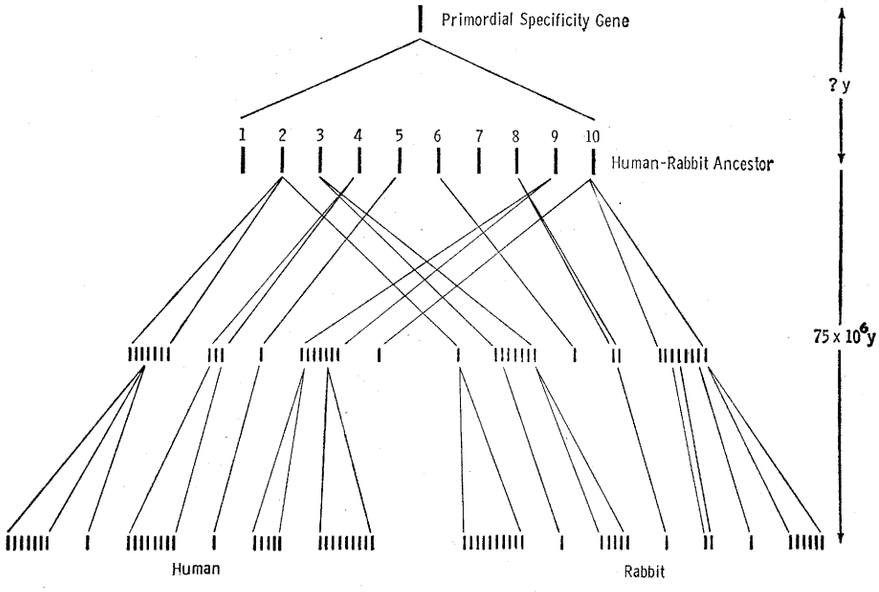

Fig. 7. Diagram of the evolution of a multigene system. This diagram does not imply that the common ancestor of human and rabbit had fewer germ line *V* genes than present-day man or rabbit. Rather, the intention is to show how a single gene in this ancestor could father a large number of contemporary genes and that different primordial genes could produce current rabbit and human *V* genes. This model can thereby explain species-specific residues and perhaps also provides an explanation for the apparent V region genetic markers (a1, a2, a3) in rabbit heavy chains (*61*).

human common and variable regions are equally distinct, and the corresponding genes seem to be diverging from one another at equal rates as would be expected in the evolution of independent but related genes. The same phenomenon is observed when a comparison of variable and common regions of human lambda and kappa chains is made (Table 3). Again both light chain regions are about equally different, and again common and variable genes seem to be diverging at equal rates. Similar comparisons of other lambda and kappa proteins show the same pattern.

Each of these observations is consistent with the hypothesis that separate germ line genes encode each distinct variable region and that the variable genes and the common genes are evolving independently but at similar rates. Furthermore, a germ line hypothesis is attractive because an ad hoc mutational mechanism is not required to explain intrasubgroup substitutions.

Table 3. Comparison of regions in individual human lambda and kappa chains. Common chains were numbered from 1 to 116 and aligned to maximize homology. An insertion of four amino acids was placed between positions 17 and 18, and deletions were placed in both C_κ and C_λ chains at positions 31 to 34, 71 to 75, 81, 90, and 96. In addition, deletions were placed in C_κ chains at position 116, and C_λ chains at 65 and 102 to 103. Otherwise see legend to Table 1.

Regions compared	Homology (%)		Number of nucleotide-gap positions	Amino acid chain length difference
	Nucleotide	Amino acid		
$V_{\lambda I}$ (Ha) vs. $V_{\kappa I}$ (Ag)	57	47	21	5
$V_{\lambda I}$ (Ha) vs. $V_{\kappa II}$ (Ti)	57	49	24	4
$V_{\lambda I}$ (Ha) vs. $V_{\kappa III}$ (Cum)	57	50	21	1
$V_{\lambda III}$ (Kern) vs. $V_{\kappa I}$ (Ag)	52	45	15	1
$V_{\lambda III}$ (Kern) vs. $V_{\kappa II}$ (Ti)	57	43	12	2
$V_{\lambda III}$ (Kern) vs. $V_{\kappa III}$ (Cum)	58	45	27	7
$V_{\lambda IV}$ (Bo) vs. $V_{\kappa I}$ (Ag)	55	49	21	5
$V_{\lambda IV}$ (Bo) vs. $V_{\kappa II}$ (Ti)	58	47	24	4
$V_{\lambda IV}$ (Bo) vs. $V_{\kappa III}$ (Cum)	57	46	21	1
$V_{\lambda V}$ (Sh) vs. $V_{\kappa I}$ (Ag)	55	44	21	1
$V_{\lambda V}$ (Sh) vs. $V_{\kappa II}$ (Ti)	58	45	18	0
$V_{\lambda V}$ (Sh) vs. $V_{\kappa III}$ (Cum)	58	44	33	5
C_λ vs. C_κ	55	39	6	2
C_λ vs. $V_{\kappa I}$ (Ag)	31	15	54	2
C_λ vs. $V_{\lambda I}$ (Ha)	26	12	51	7
C_κ vs. $V_{\kappa III}$ (Cum)	37	18	60	6

However, we must consider three of the specific questions regarding the evolution of a multigene system.

1) How many antibody genes might be required by a germ line hypothesis? There is reason to believe that the number of antibody genes required is considerably less than the number of antibody specificities that an individual can generate. First, each antibody molecule may have a broad range of specificities, some of which may seem unrelated. For example, MOPC 315, a homogeneous mouse myeloma protein with antibody activity, shows a high binding affinity (10^7) for molecules as diverse as those of dinitrophenol (hapten) and Vitamin K (*49*, also see *50*). Thus a single antibody molecule may have specificity for a large number of different antigenic determinants (sharing in part presumably a common tertiary structure) but exhibit different affinities for each. Second, 10,000 light chains and the same number of heavy chains with unrestricted combination of each light and heavy chain could generate 10^8 different antibody molecules ($10^4 \times 10^4$). There is some evidence which suggests that not all light and heavy chains combine equally well (*51*); nevertheless if only one combination in ten is effective, 20,000 antibody genes could generate 10^7 different antibody molecules.

Twenty thousand genes, each the size of a variable gene, do not comprise an unreasonable percentage of the human genome as is shown by the following calculation (*52*):

DNA content of human sperm cell (haploid)	$= 2.3 \times 10^{-11}$ g
Molecular weight of 1 base pair	$= 6.2 \times 10^2$
DNA content (per cell)	$= 3.7 \times 10^9$ base pairs
One variable region	= 107 amino acids
One variable gene	= 321 pairs
20,000 variable genes	$= 6.4 \times 10^6$ pairs
20,000 variable genes	= 0.2 percent of genetic material of human haploid cells

Thus each antibody chain type could have 1600 variable genes and use about 0.2 percent of the germ cell DNA (about 12 different common genes have been characterized). This is not an excessive demand, and it would be greatly reduced if all heavy chain classes share the same variable genes—as they may.

2) How might a multigene system evolve? Once a gene duplication has occurred, the genome can be expanded by nonhomologous pairing and recombination. An excellent example of this is the hemoglobin system where in some

animals there are two or three copies of the α gene (*53*), a myoglobin gene, and at least four different β-like chains (β, δ, γ, and ε). There are even additional hemoglobin genes in certain animals; for example, it appears that man has at least two and possibly four copies of the γ gene (*54*). Thus, man has at least seven hemoglobin genes, all of which are presumably derived from a single ancestoral gene. In a similar fashion the immune system may have evolved from a single gene, about 300 nucleotides in length (*55*), which duplicated to generate the primordial *V* and *C* genes. The *V* gene could then in time have been expanded by nonhomologous crossing over, and with the appropriate selective pressures a large pool of *V* genes could have been produced. This large pool of similar genes undergoes normal mutation and has an increased tendency for recombination. Thus the *V* gene pool is constantly in flux, generating new variants and losing old ones. Hence the *V* gene pool gives the organism a flexible and rapid means of generating new antibody genes in an ever-changing environment.

3) What selective forces can account for both the diversity and similarity of *V* genes in an evolving germ line set (*56*)? The individual light chain *V* genes are subject to two contrasting forces: toward diversity to increase the range of antibody activity, and toward similarity because of the structural demand for interaction with heavy chains. Each different *V* gene will be selected if it forms a useful antibody in combination with any one of the many different heavy chains. A large number of antibody molecules encoded by the light and heavy chain *V* genes would be expected to have useful functions of some sort in every individual.

Apart from the more specific selective forces imposed on individual *V* genes, there is selection for general functions that are common to all V regions (for example, combination with heavy chains). Since selection for general functions places identical constraints on all variable regions, it is easy to understand how invariant residues can be present in the V region (see kappa residues 5 to 8 in Fig. 1). The fact that structural requirements for these general functions can change in different species suggests how species specific residues might be selected in a multigene system; for example, rabbits have predominantly Ala at the NH_2-terminus of kappa chains whereas man has Asp. Again, suppose a mutation occurred in the common gene of the heavy chain of rabbit protein which caused a more effective light-heavy chain fit with light chains having Ala at the NH_2-terminus. In time, gene duplication would continue to generate new Ala genes, and natural selection would preserve these new copies. This process, which is illustrated in Fig. 7, suggests that significant shifts in gene frequency could have occurred during the 75 million years since the divergence of man and rabbit. Although we know very little about the nature of the selective forces operating in the immune system, the evolution of a multigene system does not seem an impossible task.

The germ line theory permits one to make predictions that can be tested in at least three ways independently.

1) Theoretically one should be able to hybridize messenger RNA from one myeloma tumor with the DNA from a germ cell of the same species and thus differentiate between a large and small number of similar genes. Admittedly, serious technical problems exist both in the isolation of mammalian messenger RNA and mammalian DNA-RNA hybridization. Nevertheless, this experiment should be possible in the near future.

2) One can search for identical $V_{\kappa I}$ chains in the highly inbred BALB/c mouse (presumably having no genetic polymorphism). At the 95 percent confidence level about 35 randomly selected proteins must be analyzed if there is to be a repeat in a pool of 200 different proteins, about 55 in a pool of 500, and about 80 in a pool of 1000 (*57*). If two identical proteins are found in the first 80 examined, this would render very unlikely any somatic theory (which should generate huge numbers of light chain genes) (*58*). On the other hand, if no identity is found the question of a germ line mechanism as opposed to a somatic mechanism would still be unresolved (remember that 10^5 variable genes require less than 1 percent of the germ line DNA).

3) If an authentic genetic marker, similar to the allelic Leu-Val interchange in kappa chains at position 191 is found in one of the variable region subgroups, this would suggest that the entire subgroup is encoded by a single gene (*59–61*). For example, let us suppose an individual should be found in whom half of the $V_{\kappa I}$ proteins begin with His instead of Asp. This would indicate a single mutation of the Asp codon of one of the two parental $V_{\kappa I}$ genes and should distribute itself in a classic Mendelian fashion among the progeny of this individual.

Obviously additional sequence data may reveal new patterns which could place additional constraints on one or more of the various theories. It will be interesting to see whether or not more extensive data on the inbred BALB/c mouse system will yield a picture similar to that of human light chains.

Thus a germ line theory is attractive in that: (i) it requires no ad hoc mutational mechanism to explain intrasubgroup variation (chemical evolution generates this diversity); and that (ii) the light chain sequence variations are similar to those seen in other sets of evolutionarily related proteins; and that (iii) evolution in a multigene system provides, through frequent gene duplication, an extremely rapid and flexible response to the environment.

Summary

Immunoglobulin light chain sequences have given us our first glimpse into the complexity of the genetic mechanism responsible for antibody diversity (*62*). Although various somatic models cannot be excluded (hypermutation and recombination), the germ line theory is attractive in that it is consistent with the sequence data, which seems to indicate an ever increasing number of germ line genes for all theories. Furthermore, it seems unnecessary to postulate a new ad hoc mutational mechanism for the generation of antibody diversity when a well-documented mechanism, chemical evolution, seems applicable. In any case, experiments are suggested which may allow us to distinguish unequivocally among the various theories of antibody diversity.

References and Notes

1. F. W. Putnam, *Science* **163**, 633 (1969); G. M. Edelman and W. E. Gall, *Ann. Rev. Biochem.* **38**, 415 (1969); *Symposium on Developmental Aspects of Antibody Formation and Structure*, J. Sterzl and H. Rhia, Eds. (Academic Press, New York, in press); S. Cohen and C. Milstein, *Advan. Immunol.* **7**, 1 (1967); E. Lennox and M. Cohn, *Annu. Rev. Biochem.* **36**, 365 (1967).
2. M. Potter, *Meth. Cancer Res.* **2**, 105 (1967).
3. L. Hood, W. R. Gray, B. G. Sanders, W. J. Dreyer, *Cold Spring Harbor Symp. Quant. Biol.* **32**, 133 (1967).
4. N. Hilschmann and L. C. Craig, *Proc. Nat. Acad. Sci. U.S.* **53**, 1403 (1965).
5. W. R. Gray, W. J. Dreyer, L. Hood, *Science* **155**, 465 (1967).
6. F. W. Putnam, K. Titani, M. Wikler, T. Shinoda, *Cold Spring Harbor Symp. Quant. Biol.* **32**, 9 (1967).
7. C. Milstein, *Nature* **216**, 330 (1967); C. Milstein, C. P. Milstein, A. Feinstein, *ibid.* **221**, 151 (1969). Only a subset of the $V_{\kappa I}$ proteins with a particular sequence around

the region of Cys at 23 could be detected in this study.
8. The abbreviations for amino acid residues are: Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Glp, pyrrolidonecarboxylic acid; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine. The subgroup designations for immunoglobulin V (variable) regions are those proposed by the Conference on Nomenclature for Animal Immunoglobulins, Prague, June 1969. Abbreviation for bases are A, adenine; G, guanine; C, cytosine; and U, uracil.
9. H. Niall and P. Edman, *Nature* **216**, 262 (1967).
10. L. Hood, J. A. Grant, H. C. Sox, in *Symposium on Developmental Aspects of Antibody Formation and Structure*, J. Sterzl and H. Rhia, Eds. (Academic Press, New York, in press); J. A. Grant and L. Hood, unpublished results.
11. C. Baglioni, L. A. Zonta, D. Cioli, A. Carbonara, *Science* **152**, 1517 (1967); C. Milstein, *Biochem. J.* **101**, 338 (1966).
12. C. Ropartz and J. Lenoir, *Nature* **189**, 586 (1961); A. G. Steinberg, J. A. Wilson, S. Lanset, *Vox Sang.* **7**, 151 (1962).
13. W. D. Terry, L. Hood, A. G. Steinberg, *Proc. Nat. Acad. Sci. U.S.* **63**, 71 (1969).
14. M. Potter, W. J. Dreyer, E. L. Kuff, K. R. McIntire, *J. Mol. Biol.* **8**, 814 (1964).
15. If one evokes a master-slave model of chromosome structure where many germ line copies (slaves) of a single structural gene are corrected at each DNA division against a single copy (master) [see H. C. Callan, *J. Cell Sci.* **2**, 1 (1967)], then multiple germ line genes could evolve as a single gene. This might explain the identity of multiple common genes, but obviously it does not apply to the *V* genes.
16. Von B. Langer, M. Steinmetz-Kayne, N. Hilschmann, *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 945 (1968).
17. L. Hood and D. Ein, *Nature* **220**, 764 (1968).
18. D. Ein and J. Fahey, *Science* **156**, 947 (1967);
19. E. Appella and D. Ein, *Proc. Nat. Acad. Sci. U.S.* **57**, 1449 (1967); L. Hood and D. Ein, *Science* **162**, 679 (1968).
20. D. Ein, *Proc. Nat. Acad. Sci. U.S.* **60**, 982 (1968).
21. E. S. Lennox, P. M. Knopf, A. J. Munro, R. M. E. Parkhouse, *Cold Spring Harbor Symp. Quant. Biol.* **32**, 249 (1967); P. M. Knopf, personal communication.
22. M. D. Scharff, A. L. Shapiro, B. Ginsberg, *Cold Spring Harbor Symp. Quant. Biol.* **32**, 235 (1967).
23. M. Potter and R. Lieberman, *ibid.*, p. 187.
24. B. Frangione and C. Milstein, *Nature* **224**, 597 (1969). These generalizations also seem to hold for a second heavy chain disease protein (W. D. Terry, personal communication).
25. Although amino acid sequence homology suggests that lambda and kappa variable genes share a common ancestor (Fig. 1) (*25, 6*), it is equally obvious that lambda and kappa chains are now encoded by distinct sets of genes. Certain features are shared by all lambda (but not kappa) chains; for example, all lambda variable regions have three sets of gaps or insertions when compared with their kappa counterparts (residue 9 in Fig. 1). In addition, unique residue or residue alternatives exist at a number of positions (for example, lambda chains have Glp or a deletion at position 1, whereas kappa chains have Asp or Glu).
26. L. Hood, W. R. Gray, W. J. Dreyer, *J. Mol. Biol.* **22**, 179 (1966).
27. A. M. Campbell, *Adv. Genet.* **11**, 101 (1962).
28. C. A. Thomas, Jr., *Progr. Nucl. Acid Res. Mol. Biol.* **5**, 315 (1966).
29. W. J. Dreyer, W. R. Gray, L. Hood, *Cold Spring Harbor Symp. Quant. Biol.* **32**, 353 (1967).
30. M. Cohn, in *Rutgers Symposium on Nucleic Acids in Immunology*, O. J. Plescia and W. Braun, Eds. (Springer-Verlag, New York, 1968), p. 671.
31. R. Mage, G. O. Young, J. Rejnek, R. A. Reisfeld, E. Appella, *Protides Biol. Fluids Proc. Colloq. Bruges*, in press.
32. L. Herzenberg, H. McDevitt, L. Herzenberg, *Annu. Rev. Genet.* **2**, 209 (1968); J. B. Natvig, H. G. Kunkel, S. P. Litwin, *Cold Spring Harbor Symp. Quant. Biol.* **32**, 173 (1967).
33. One would expect a 2:1 ratio of transversions to transitions if the tendency of a mutating base to change to a purine or a pyrimidine is random. For example, G can mutate to T or C (transversions) or A (transitions), hence twice as many transversions as transitions would be expected.
34. T. J. Jukes, *Molecules and Evolution* (Columbia University Press, New York, 1966); C. Nolan and E. Margoliash, *Annu. Rev. Biochem.* **37**, 727 (1968); W. M. Fitch, *J. Mol. Biol.* **26**, 499 (1967).
35. H. C. Sox and L. Hood, in preparation.
36. A 50 percent success here is not meaningful because of the small number of alternatives.
37. L. Hood, W. R. Gray, W. J. Dreyer, *Proc. Nat. Acad. Sci. U.S.* **55**, 826 (1966).
38. R. Perham, E. Appella, M. Potter, *Science* **154**, 391 (1966).
39. F. Melchers, *Biochemistry* **8**, 938 (1969).
40. J. Lederberg, *Science* **129**, 1649 (1959).
41. O. Smithies, *ibid.* **157**, 267 (1967).
42. M. Potter, E. Appella, S. Geisser, *J. Mol. Biol.* **14**, 361 (1965); J. Campbell, *J. Theor. Biol.* **16**, 321 (1967); H. L. K. Whitehouse, *Nature* **215**, 371 (1967).
43. G. M. Edelman and J. Gally, *Proc. Nat. Acad. Sci. U.S.* **57**, 353 (1967).
44. S. Brenner and C. Milstein, *Nature* **211**, 242 (1966).
45. With a random mutational mechanism (single base changes) operating throughout the entire variable gene, about six different amino acids on the average could be generated at each codon. With six variants at each of 107 positions, one could generate 6^{107} ($\simeq 10^{83}$) variable region sequences (or cells if one sequence is allowed per cell). If three residue alternatives at each position are compatible with function (the sequence variation within a subgroup appears much more highly restricted) (*1*) then 3^{107} ($\simeq 10^{51}$) sequences (cells) are useful to the organism. Hence the ratio of useless to functional cells (sequences) is $6^{107}/3^{107}$ or more than 10^{30} useless cells would be produced for each functional one. Since the average man only generates 5×10^{18} cells during his entire life, a random mutational mechanism poses an impossible cell wastage problem.
46. L. Hood and D. Talmage, in *Developmental Aspects of Antibody Formation and Structure*, T. Sterzl and H. Rhia, Eds. (Academic Press, New York, in press).
47. W. J. Dreyer and J. C. Bennett, *Proc. Nat. Acad. Sci. U.S.* **54**, 864 (1965).
48. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).
49. M. Dayhoff, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Md., 1969), p. 29.
50. H. Eisen, *Fed Proc.* **29**, 78 (1970).
51. D. Schubert, A. Jobe, M. Cohn, *Nature* **220**, 882 (1968).
52. H. Grey and M. Mannik, *J. Exp. Med.* **122**, 619 (1965).
53. Calculation by W. R. Gray. The DNA value is taken from *The Chemical Basis of Heredity*, W. R. McElroy and B. Glass, Eds. (Johns Hopkins Univ. Press, Baltimore, 1957).
54. K. Hilse and R. A. Popp, *Proc. Nat. Acad. Sci. U.S.* **61**, 930 (1968).
55. W. A. Schroeder, T. H. Huisman, J. R. Shelton, J. B. Shelton, E. F. Kleihaner, A. M. Dozy, B. Robberson, *ibid.* **60**, 537 (1968).
56. R. L. Hill, R. Delaney, R. E. Fellows, Jr., H. E. Lebovitz, *ibid.* **56**, 1762 (1966); S. J. Singer and R. F. Doolittle, in *Structural and Molecular Biology*, A. Rich and N. Davidson, Eds. (Freeman, San Francisco, 1968).
57. There are, of course, examples of multigene systems with many copies of similar genes. About 10 percent of the DNA in mouse cells is "satellite DNA" which appears to consist of about a million similar copies of a short nucleotide sequence [see R. J. Britten and D. E. Kohne, *Science* **161**, 529 (1968)]. A second example is that of the ribosomal genes in *Xenopus laevis*. There appear to be about 450 copies of both the 28*S* and the 18*S* ribosomal cistrons which are linked in tandem along the chromosome with alternating 28*S* and 18*S* genes [D. Brown and C. S. Weber, *J. Mol. Biol.* **34**, 681 (1968)]. Furthermore DNA-RNA hybridization experiments suggest that individual members of each ribosomal set are very similar in sequence to one another [M. Birnstiel, M. Grunstein, J. Speirs, W. Hennig, *Nature* **223**, 1265 (1969)].
58. S. Cohen and C. Milstein, *Advan. Immunol.* **7**, 1 (1967).
59. O. Smithies (personal communication) has pointed out that if two identical sequences are also identical to a prototype subgroup sequence, one could argue that the corresponding subgroup gene may have just failed to undergo the hypermutation process. In this special case, two identical light chains would not, therefore, argue against the somatic theories.
60. It is beyond the scope of this paper to discuss the implications of the limited sequence information available on human and rabbit heavy chains. The data on the human heavy chain are consistent with the light chain analysis presented here (*60*). Interpretation of the data on the rabbit heavy chain is difficult in that heterogeneous mixtures of proteins have been used for sequence analysis. Wilkinson (*61*) has reported multiple sequence differences in the NH_2-terminal region of rabbit heavy chains which correlate with differing allotypes. These multiple amino acid substitutions do not constitute a simple genetic marker similar to the single residue interchange in the common region of human kappa chains (Inv marker). Consequently the meaning of Wilkinson's data is subject to a number of different interpretations at the genetic level. If, however, these differences do constitute a genetic marker in the variable region, it would render a germ line model of heavy chain diversity unlikely.
61. See G. M. Edelman and E. Press, in *Developmental Aspects of Antibody Formation and Structure*, J. Sterzl and H. Rhia, Eds. (Academic Press, New York, in press).
62. J. M. Wilkinson, *Biochem. J.* **112**, 173 (1969).
63. Sequence analysis of immunoglobulin light chains has given us a glimpse of differentiation in a complex mammalian system and has certainly raised a question as to whether other differentiating systems have followed a similar course.
64. L. Hood, unpublished results.
65. C. Milstein, *Fed. Eur. Biochem. Soc. Symp.* **15**, in press.
66. N. Hilschmann, *Hoppe-Seyler's Z. Physiol. Chem.* **348**, 1077 (1967).
67. C. Milstein, *Biochem. J.* **101**, 352 (1966).
68. A. Kaplan and H. Metzgar, *Biochemistry* **8**, 3944 (1969).
69. B. A. Cunningham, P. D. Gottlieb, W. H. Konigsberg, G. M. Edelman, *ibid.* **7**, 1983 (1968); G. M. Edelman, B. A. Cunningham, W. E. Gall, P. D. Gottlieb, U. Rutishauser, M. J. Waxdal, *Proc. Nat. Acad. Sci. U.S.* **63**, 78 (1968).
70. F. W. Putnam, *Science* **163**, 633 (1969).
71. N. Hilschmann, *Fed. Eur. Biochem. Soc. Symp.* **15**, in press.
72. A. Grant and L. Hood, unpublished results.
73. C. Baglioni, *Biochem. Biophys. Res. Commun.* **26**, 82 (1967).
74. A. B. Edmundson, F. A. Sheber, K. R. Ely, N. B. Simonds, N. K. Hutson, J. L. Rossiter, *Arch. Biochem. Biophys.* **127**, 725 (1968).
75. C. Milstein, *Biochem. J.* **110**, 631 (1968).
76. H. Ponstingl, M. Hess, N. Hilschmann, *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 867 (1968).
77. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).

Supported in part by NIH grants 5 RO1 AI 03047-11 and FR 00404-02. We thank S. Levine for assistance in preparing the amino acid sequence data for computer programming.