EVIDENCE SUPPORTING SOMATIC ASSEMBLY OF THE DNA SEGMENTS (MINIGENES), CODING FOR THE FRAMEWORK, AND COMPLEMENTARITY-DETERMINING SEGMENTS OF IMMUNOGLOBULIN VARIABLE REGIONS*

By ELVIN A. KABAT,‡ TAI TE WU,§ and HOWARD BILOFSKY

From the National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205; Departments of Microbiology, Human Genetics, and Development, and Neurology, and the Cancer Center, College of Physicians and Surgeons, Columbia University, New York 10032; Departments of Biochemistry Molecular Biology, Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60201; and Bolt Beranek and Newman Inc., Cambridge, Massachusetts 02138

The variable regions[1] (V-regions)[2] of immunoglobulin light (V_L) and heavy (V_H) chains can be considered to be made up of four framework (FR) separated by three complementarity-determining (hypervariable [1-3, cf. 4]) regions or segments (CDR). This is supported by x-ray crystallographic studies (5-8). In the light chains, FR1, FR2, FR3, and FR4 comprise residues 1-23, 35-49, 57-88, and 98-107 and CDR1, CDR2, and CDR3 comprise residues 24-34, 50-56, and 89-97F, respectively. In the heavy chains, the FR are 1-30, 36-49, 66-94, and 103-113 and the CDR 31-35B, 50-65, and 95-102. Length differences are found in CDR1 and CDR3 of the light chain, and in all three CDR and FR3 of the heavy chain (1-3, cf. 4). Subgrouping of the light (9-12) and heavy chains (13, 14) has been based on the sequence differences and similarities largely found in FR1.

In a recent study we reported that the various FR segments could be grouped into sets that consisted of chains with identical sequence (15). Chains that were identical except for an uncertainty concerning the presence or absence of the amide group, Glx or Asx, were accepted. Sets contained from 1-18 members, and when each V-region was traced from one FR segment to the next, it was found that members from one set in FR1 could be associated with different sets in FR2, FR3, and FR4 (15). This apparently independent assortment of FR segments was observed with human $V_\kappa I$,

* Sponsored by the National Cancer Institute, National Institute of Allergy and Infectious Diseases, National Institute of Arthritis, Metabolism, and Digestive Diseases, the National Institute of General Medical Sciences, and the Division of Research Resources (contract NO1-RR-2147 and NO1-RR-8-2158), of the National Institutes of Health, Bethesda, Maryland.

‡ Recipient of grant BMS-76-81029 from the National Science Foundation and by a Cancer Center Support grant (CA13696) to Columbia University, N.Y. from the National Cancer Institute, Bethesda, Md.

§ Research Career Developmental Awardee (5-KO1-AI-70497), National Institutes of Health, Bethesda, Maryland. The work was aided in part by a grant from the National Institutes of Health, 5-RO1-GM21482-04.

[1] The data base of variable region sequences is maintained in the PROPHET computer system (34).

[2] *Abbreviations used in this paper*: CDR, complementarity-determining regions, FR, framework; V-region, variable regions; V_H, immunoglobulin heavy chain; V_L, immunoglobulin light chain.

J. Exp. Med. © The Rockefeller University Press · 0022-1007/79/06/1299/15 $1.00

Volume 149 June 1979 1299-1313

mouse V_κ, rabbit V_κ, and mouse and human V_HIII chains. The hypothesis was suggested that the individual FR segments were products of germ-line minigenes which, together with those for the CDR, were assembled somatically to form the genes for the V-region (15). A minigene is defined as a DNA segment coding for a portion of a complete V-region and which shows some evidence of segregation as a functional unit independent of the rest of the DNA coding for the V-region.

Attempts were made (16) to rearrange the individual FR sets in the order of increasing sequence differences from the set with the most members. This would be expected to eliminate the assortment, assuming that somatic mutation was determining the differences between the individual FR sets (16), but it only increased the complexity of the assortment patterns in tracing the chains from one FR to the next. Moreover, there were a substantial number of amino acid substitutions that would have required two base changes (16). These findings made it unlikely that somatic mutation was playing a fundamental role in the observed framework variability.

Tonegawa et al. (17) isolated a mouse λ-DNA clone, Ig 13 λ, from a 12-d-old mouse embryo DNA which contained the nucleotide sequences coding for the V-region from FR1 almost through CDR3, amino acid residues 1–96 (numbering as in reference 2), followed by an intervening sequence. More recently, Brack et al. (18) and Bernard et al. (19) described three other mouse $V_\lambda 2$ DNA clones, Ig 99λ and Ig 25λ from embryo DNA and Ig 303λ from adult myeloma H2020. The embryo clone, Ig 99λ, coded for amino acid residues 1–95 whereas Ig 25λ contained the nucleotide sequences coding for amino acids 96–107, termed the J segment, which includes all of FR4 and two residues of CDR3 followed by an intervening sequence of about 1.2 Kb after which the nucleotides coding for the C-region began with a repetition of Gly107. However, the clone from adult myeloma Ig 303λ had a contiguous nucleotide sequence coding for the entire V-region, residues 1–107, followed by the 1.2 Kb intervening sequence, and the C-region nucleotides as in Ig 25λ. Thus, the J segment had been joined to the rest of the V-region between the 12th d of embryonic life and the V_λ 2 adult myeloma, confirming the hypothesis of somatic assembly (15) for FR4 and that the J segment is a minigene. It should be emphasized that because only FR sets were used in demonstrating our original assortment (15), it would be independent of, and would be seen, whether or not any CDR residues assorted with FR4.

Recent findings by Seidman et al. (20, 21) with mouse V_κ chains indicate that even in an adult myeloma MOPC149, two clones, K2 and K3, contained the nucleotide sequences coding for amino acid residues 1–97 of the light chain preceded and followed by extensive flanking sequences. However, neither clone had the sequence corresponding to the secreted myeloma protein as determined on a cDNA copy from the mRNA, differences being seen in FR and CDR. The complete amino acid sequence of the protein was not available and whether their residues 96 and 97 correspond to the end of CDR3 is not clear. The possibility arises that the somatic joining of FR4 and the C-region to the third CDR is an essential step in synthesis of the intact myeloma light chain.

Weigert et al. (22) have examined mouse $V_\kappa 21$ sequences from NZB mice, have included two residues, 96 and 97 of CDR3 in the J piece, and have shown these to assort together with FR4. They suggest, because residue 96 is the most hypervariable residue in the light chains (1, 2) that this assortment could contribute to the generation of diversity.

Examination by the Southern blot technic of EcoR1 fragments of embryonic mouse DNA, annealed to cloned constant and variable cDNA sequences of MOPC149 and MOPC41, which belong to different subgroups, showed that six to eight different sized EcoR1 fragments of DNA formed hybrids with the cloned cDNA. On this basis it was proposed (20, 21) that multiple, closely related V-genes exist for each subgroup and that the number of V_κ genes for the light chain in the mouse genome for the many subgroups may be ≃200 or more.

Although there are not enough clones (17–21) to define the extent of J segment unambiguously, it is clearly a minigene containing all of FR4. The one clone, sequenced Ig 25λ, appeared to contain the last two residues of CDR3. The other clones define J by difference and it always contains FR4 and possibly one or two residues of CDR3; much more data are obviously needed.

The V-region clones, for both V_κ and V_λ, contain contiguous nucleotide sequences coding for FR1 through most or all of CDR3. The question arises as to whether these are a direct reflection of the germ line. If so, it would be inferred that all of these genes for light chains, excluding J, already exist as such in the genome. This would be a stringent germ-line theory. On the other hand, the FR-assortment data suggest that the FR and CDR segments are separate in genome DNA and there must be some mechanism of joining FR and CDR segments.

We propose to examine the available amino acid sequences to evaluate data bearing on these divergent views.

Results

The amino acid sequence data are given in reference 2 plus additional published sequences (21–24), and other sequences which (25, 26) were made available by Doctors Martin Weigert, Lee Hood, Michael Potter, Stuart Rudikoff, E. Apella, and Rose Mage. The FR fragments of each sequence were grouped into sets of identical residues (15). The earlier study included only human V_κI, mouse V_κ, rabbit V_κ, and human and mouse V_HIII. The present study adds sets of human V_κII, V_κIII, and V_λII, mouse V_HIII, and rabbit V_H. Numerous sequences of myeloma proteins of the NZB strain have become available (22).

Fig. 1 shows the assortment data with the new NZB sequences (22) for mouse V_κ FR1 and FR2 segments together with the amino acid differences by which each set differs from the set with the largest number of members. Although the data are still limited because there are only two FR1 and two FR2 sets with multiple differences, sets containing single and multiple members of FR1 are associated with the same FR2 set and vice versa.

Table I summarizes data on sets of FR1 in relation to the CDR1. Assuming each choice of positions 1–96 of V-region sequence represents a germ-line gene, the minimum numbers of copies of identical FR1 sequences that are required to account for the differences in the V-region sequences through CDR1 are estimated. The human and rabbit data are not on inbred populations and the number of genes may be reduced by the existence of alleles if such are shown to exist. Because the basic pattern of assortment is seen within inbred mouse strains, definitive data on the numbers of copies are provided by these strains. The number of copies in the inbred and outbred species are comparable.

Thus, in Table I, the first human V_κI set contains 18 members with identical FR1;

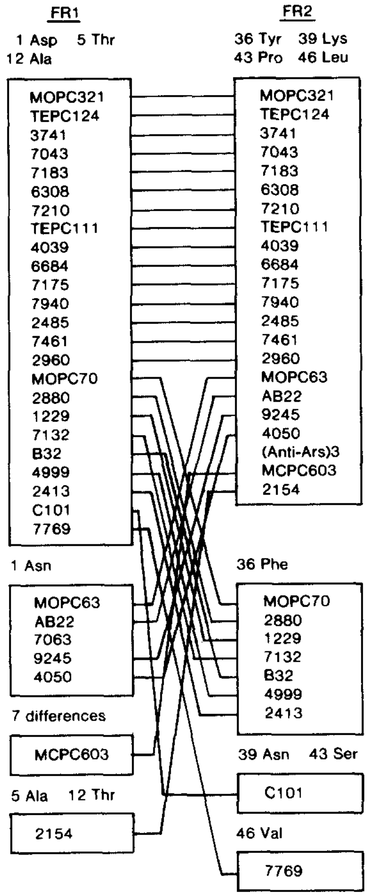

FIG. 1. Assortment of FR1 and FR2 in mouse V_κ light chains. Each set of identical sequences is enclosed in a box. The positions and the amino acid residues by which it differs from the other sets, are listed above the FR1 and FR2 set with the largest number of chains. Only the positions at which differences are found are given above the other sets.

12 of these have been completely sequenced through CDR1. All 12 differ from one another by one to five amino acids in CDR1. Thus, if each of these 12 CDR1 exist in the genome joined to an FR1, there must be 12 copies of the identical FR1 in the genome. Although the remaining six chains have only been sequenced through FR1, probably their CDR1 would also differ from the others, and the estimate of 12 copies is clearly a minimum. The other V_κI sets with multiple members also require multiple copies of FR1 in the genome to account for the different CDR. Only two chains, DAV and FIN (27), both of which show anti-human IgG activity, have an identical FR1 and CDR1 so that but one copy would be needed. Thus, of the 21 human V_κI chains sequenced through CDR1, 20 different genes were needed on a stringent germline theory. Because the remaining light chains were only sequenced through FR1, the number could even reach 27/28.

Table I gives similar data for mouse V_κ chains, the minimum number of copies required being 27 of the 49 chains sequenced through CDR1 or 27 of the 52 chains

TABLE I

Numbers of FR1 Copies Required to Provide for Each Unique CDR1 Sequence Assuming a Stringent Germ-Line Theory

Species and set	FR1	CDR1			Minimum number of FR1 copies required for CDR1	No. with FR1 completely sequenced		Minimum number of copies	
	No. of sequences in set	No. sequenced completely	No. identical	No. of amino acid differences found				No. of sequences	
Light Chains									
Human									
V_κI	18	12*	0	1-5	12	V_κI	29*	20/21	[20/29]
V_κI	7	5	0	2-3	5				
V_κI	2	2‡	0	2	2				
V_κI	2	2	2	0	1				
V_κII	7	2	0	1	2	V_κII	6	2/2	[2/6]
V_κIII	9	6	0	1-4	6	V_κIII	12	8/8	[8/12]
	3	2	0	7	2				
V_λII	2	2	0	5	2	V_λII	2	2/2	[2/2]
Mouse									
V_κ21	25	25	4, 4, 4, 4, 2	1-7	16	V_κ	52	27/49	[27/52]
V_κ21B	5	5	2	2	2				
V_κ4	7§	4‖	0	2-4	4				
V_κ22	4¶	4	4	0	1				
V_κ10	2**	2	2	0	1				
V_κ11	7‡‡	7	4 and 3	1	2				
V_κ20	2	2	2	0	1				
Rabbit									
V_κ	11	9	5 and 2§§	3	2	V_κ	44	14/22	[14/44]
	2	2‖‖	0	3	2				
	3	2¶¶	0	1	2				
	2	2***	0	3	2				
	2	2‡‡‡	0	5	2				
	4	3§§§	0	2, 5	3				
	2	2‖‖‖	2	0	1				
Heavy Chains									
Human									
V_HIII	3	3¶¶¶	0	3-5	3	V_HIII	25	6/6	[6/25]
	3	3****	0	3-5	3				
Mouse									
V_HIII	6	6‡‡‡‡	6	0	1 (3)	V_HIII	22	2/10	[2/22]
	4	4§§§§	4	0	1 (3)			(6/10)	(6/22)
V_HV	3	3‖‖‖‖	3	0	1				
Rabbit V_H									
	2	2¶¶¶¶	0	4	2	V_H	9	2/2	[2/19]

Values in brackets give minimum number of copies for all members of FR1 sets including those only sequenced through FR1.

Values in parentheses represent number of different germ-line genes required based on entire sequence if available.

[*] Two residues missing in CDR1 of Amyloid VIII-B, one difference found in CDR1 of those residues sequenced. Four residues missing in CDR1 of LOW, a cold agglutinin with anti-I activity; one difference found in those residues sequenced.

Table I—*continued*

‡ Three residues missing in BEL; two differences found in CDR1 of those residues sequenced.

§ Six anti-*β*1 → 6 galactan; one anti-*β*-D-GlcNAc.

‖ Two residues, 26 and 27 missing from S117.

¶ All anti-phosphocholine.

** All anti-*β*2 → 6 fructosans.

‡‡ All anti-*β*2 → 1 fructosans.

§§ K9-336 missing residue 32, 166 missing 31–34; includes anti-streptococcal group A variant carbohydrate, anti-*Micrococcus lysodeikticus*, anti-*p*-azophenylarsonate. K9-335 and K9-338 are identical throughout entire V-region.

‖‖ 3T74 anti-type III and 3322A anti-type VIII pneumococcal polysaccharide.

¶¶ BS-1, BS-5 anti-type III; 2348-3 anti-type VIII pneumococcal polysaccharide.

*** Residue 34 of K19 missing.

‡‡‡ Anti-streptococcal group A-variant carbohydrate.

§§§ Residue 31–34 of 722369 missing.

‖‖‖ Residues 30 and 34 missing in K31-147.

¶¶¶ Residue 34 in GR′ missing.

**** Tur cold agglutinin with anti-PR, POM cold agglutinin with anti-IgG1 activity.

‡‡‡‡ All anti-phosphocholine.

§§§§ All anti-*β*2 → 1 fructosans.

‖‖‖‖ Residues 31 and 35 missing in K2, 33 and 35 in C22.

¶¶¶¶ Both anti-pneumococcal type III polysaccharide.

in FR1 sets with multiple members. The footnotes to Table I show that some instances, with identical FR1 and CDR1, represent various myeloma proteins selected for antibody specificity. This would tend to reduce the number of copies more than if there had been random selection.

The rabbit V_κ data in Table I were also selected for antibody specificity, almost being antibodies to the pneumococcal type-specific and anti-streptococcal group-specific polysaccharides (2, 23). Nevertheless, of the 22 chains sequenced through CDR1, 14 copies of FR1 were required to account for the different CDR.

The same type of data are given for human, mouse, and rabbit V_H regions; sequence data are much more limited but multiple copies of FR1 are needed to join to the different CDR1.

Table II examines the FR2 set, residues 35–49, which was originally recognized (15) in one human, four mice, and eight rabbit light chains and thus, has been preserved over about 80 million yr or before these three species diverged in evolution. With the newer data (22, 24) the number of mouse light chains in this set has increased to 20; by tabulating differences in CDR1 and CDR2 associated with this single FR2 sequence, an estimate of 10 separate copies of FR2 are required for the 14 NZB mice and 5 separate copies for the 6 BALB/c mice based on the findings (17–21) that the DNA clones contain nucleotide sequences coding for the entire sequence from FR1 approximately through CDR3. To code for the 13 rabbit sequences, 12 FR2 copies would be needed despite the substantial selection for a few antibody specificities.

A second FR2 set was made up of seven mouse and four rabbit light chains. This FR2 set differs only at position 36 from the first set in having Phe instead of Tyr. Three copies of the six mouse chains and two copies of the four rabbit chains are required by the assumptions made above; Because the entire V-region of K9-335 and K9-338 were shown by Huser and Braun (25) to be identical, two copies would be needed for three sequences. For human light chains, with three $V_\kappa I$; with two of two

V_κII; and two of two V_λII sets each having an identical FR2, three, two and two copies, respectively, would be required.

Tables III and IV list the numbers of sets of FR1 and FR2, and the sequence differences from the set with the largest number of members. The data in Table III are for human V_κI sets, those in Table IV are for mouse V_κ sets but include human and rabbit sequences with identical FR2. The data show clearly that sets may differ one from another in from one to five amino acids. In both FR1 and FR2, differences involving two base changes from the set with most members were seen in three and four instances; one FR2 set had two base changes at two positions. The alternative forms of FR1 and FR2 occur in much lower frequency, as evidenced by the decreasing numbers of members of the other sets. The preserved FR2 set occurs in many more light chains than do the other alternatives.

The x-ray crystallographic data on MCPC603 (6, 28) show FR2 to be a loop in a region of the molecule away from the combining site with sufficient space to accommodate the various amino acid substitutions. Fig. 2 (28) shows the Fab fragment of MCPC603; its FR2 sequence of the light chain is that of set 1 present in one human, 20 mouse, and 13 rabbit chains (Tables II and IV).

Discussion

The data in Tables I and II clearly show that if mouse DNA contains linear sequences of nucleotides, each coding for a V-region from residues 1 to 95–97 as inferred from the data of Tonegawa et al. (17–19) and of Seidman et al. (20, 21) e.g., from FR1 to about the end of CDR3, then a large number of copies of nucleotides coding for identical FR1 and identical FR2 sets must be present in the genome. This number is far higher than was originally postulated, assuming a few identical FR, or subgroups, with a large number of different CDR.

This raises the question, as first put forward by Dreyer and Bennett (29) with respect to the existence of many V-regions if each were associated in the genome with an identical sequence for the C-region, as to what is preserving the multiple copies of the C-region intact over evolutionary time. This led to the inference, now clearly established by hybridization (30) and cloning (12, 17–21) that the V- and C-regions were encoded spearately, and that there is but one (or very few copies) of the C-region.

This dilemma applies even more forcefully to the present data. Not only would identical FR1 and FR2 sets have to exist in multiple copies, but there is no apparent reason for them to be preserved intact over evolutionary time because in each instance, a large number of alternative FR1 and FR2 sets are known. These may differ in from one to five amino acids. Moreover, in only 3 of the 22 positions of FR1, excluding Cys 23, and in only 2 of the 15 positions in FR2, have substitutions not been found. Thus, there is substantial capacity for variation without significantly affecting the capacity to form the proper three-dimensional structure. The FR2 set with one human, 20 mouse, and 13 rabbit V_κ sequences has been preserved intact over a period of about 80 million yr and thus, constitutes a primordial FR segment. Its association with different CDR1 and CDR2 in the inbred mouse leads to a minimum estimate of 10 copies in 14 NZB sequences and 5 copies in 6 BALB/c sequences. Other sets have thus far been seen only in a few copies and most only in a single copy. In the outbred

TABL

Variation in CDR1 and CDR2 Sequences of Human, Mouse, and Rabbit V_κ Chains Ha
Unique Sequence of CDR1 and/or CDR2

Residue number	Human V_κ	Mouse V_κ																						
	IV LEN	21B 4050	21B 9245	21B MOPC 63	21B AB 22	21C MOPC 321	21C TEPC 124	21C 3741	21C TEPC 111	21D 7043	21D 7183	21D 6308	21D 7210	21E 6684	21E 7175	21E 7940	21F 2485	21F 4039	21I 2960	21J 7461	■8 MCPC 603	● 3368	● BS-1	● BS-5
24	Lys	Arg	Arg	Arg	Arg	Arg	Arg	Arg	Arg	LYS	LYS	LYS	LYS	Arg	Arg	Arg	Arg	Arg	Arg	Arg	LYS	Gln	Gln	Gln
25	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	SER	Ala	Ala	Ala
26	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	ASN	Ser	Ser	Ser	Ser	Ser
27	Gln	Glu	Glu	Glu	Glu	Lys	GLN	Glu	Glu	GLN	GLN	GLN	GLN	Lys	Lys	Lys	Lys	Lys	Lys	Glu	GLN	GLU	Gln	Gln
C 27A	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	---	---	---
D B	Val	Val	Val	Val	Val	Val	Val	Val	Val	Val	Val	Val	LEU	Val	Val	Val	Val	Val	Val	Val	LEU	---	---	---
C	Leu	Asp	Asp	Asp	Asp	Asp	Asp	Asp	Asp	Asp	Asp	Asp	Asp	Ser	Ser	Ser	Ser	Ser	Ser	GLU	LEU	---	---	---
R D	Tyr	Ser	Ser	Ser	Ser	THR	TRP	Ser	Ser	Tyr	Tyr	Tyr	Tyr	Thr	Thr	ALA	Thr	Thr	Thr	Tyr	ASN	---	---	---
I E	Ser	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Ser	---	---	---
F	Ser	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Gly	---	---	---
28	Asn	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	ASP	ASP	ASP	ASP	SER	SER	PHE	SER	SER	ILE	PHE	ASN	Ser	Ser	Ser
29	Ser	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	Gly	GLN	Ile	Ile	Ile
30	Lys	Asn	Asn	Asn	Asn	Asn	Asx	Asn	Asn	Asp	Asp	Asp	Asp	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	THR	Lys	Gly	TYR	TYR
31	Asn	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	GLY	Ser	ASN	Asn	Ser	Ser
32	Tyr	Phe	Phe	Phe	Phe	Phe	Phe	Phe	Phe	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr	CYS	LEU	Phe	Glu	GLY	Asn
33	Leu	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	Met	LEU	Met	LEU	Leu	Leu	Leu
34	Ala	His	His	His	His	His	His	His	His	Asn	Asn	Asn	Asn	His	His	His	His	His	His	GLN	ALA	Ala	Ala	Ala
	Set I																							
35	Trp																							
36	Tyr																							
37	Gln																							
F 38	Gln																							
R 39	Lys																							
40	Pro																							
2 41	Gly																							
42	Gln																			All FR2 sequences identical to Set I				
43	Pro																							
44	Pro																							
45	Lys																							
46	Leu																							
47	Leu																							
48	Ile																							
49	Tyr																							
50	Trp	Leu	Leu	Leu	Leu	ARG	ARG	ARG	ARG	ALA	ALA	THR	ALA	Leu	Leu	Leu	Leu	Leu	Leu	Val	GLY	ARG	Lys	Lys
C 51	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
52	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser
D 53	Thr	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	Asn	SER	SER	Asn	Asn	THR	LYS	Thr	Thr
R 54	Arg	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	VAL	ARG	Leu	Leu	Leu
55	Glu	Glu	Glu	Glx	Glu	Glu	Glx	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	Glu	TYR	Glu	Glu	Ala	Ala	GLU
2 56	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser

No. of copies of FR2 needed/total sequences

1/1 10/14 NZB 5/6 BALB/c

... Deletion, chain continues.

‾ Residue not identified.

Residues in capitals differ from most frequently occurring residue or residues within a species or subgroup.

● Anti-type III pneumococcal polysaccharide.

▲ Anti-streptococcal group A variant carbohydrate.

× Anti-*Micrococcus lysodeikticus.*

□ Anti-type II pneumococcal polysaccharide.

■ Anti-phosphocholine.

* FR2 of MIL must be identical to FR2 of NIM or of FR depending on whether it has GLN or GLU at position 42. In either,

E II

ving a Given FR2 Sequence: Estimation of Numbers of FR2 Copies Required for Each Assuming a Stringent Germ-Line Theory

Rabbit Vκ										Mouse Vκ						Rabbit Vκ			Human							
																			VκI			VκII			VλII	
3547	▲ K4820	▲ K9- 335-I	▲ K30- 267	× 120	● K-25	□ 4422	□ 311	□ 4363	□ 4192	21A 2880	21A 1229	21A 7132	21A B32	21A 70	21I 2413	K9-335 K9-338	K29- 213	K16- 167	ROY	AU	KA	MIL*	■ NIM or FR		BOH	BUR
Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Arg	Arg	Arg	Arg	Arg	Arg	Gln	Gln	Gln	Gln	Gln	Glu	Arg	Arg	Arg	Ala	ILE
Ala	Ala	Ala	Ala	SER	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ser	Ser	Ser	Gly	Gly
Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Thr	Thr
GLU	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Glu	Glu	Glu	Glu	GLN	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Gln	Ser	Ser
---	---	Ser	---	---	---	Ser	Ser	---	---	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	---	---	---	Asn	Ser	Ser	---	---
---	---	Val	---	---	---	Val	Val	---	---	Val	Val	Val	Val	Val	Val	Val	Val	Val	---	---	---	Leu	Leu	Leu	---	---
---	---	---	---	---	---	---	---	---	---	Asp	Asp	Asp	Asp	Asx	VAL	---	---	---	---	---	---	Leu	Leu	Val	---	---
---	---	---	---	---	---	---	---	---	---	Asn	Asn	Asn	Asn	Asx	Asn	---	---	---	---	---	---	Glx	Trp	Tyr	Ser	Ser
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Ser	Ser	Arg	Asp	ASN
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	Val	Val
ASP	Ser	TYR	ASN	Ser	ASX	TYR	TYR	ASN	Ser	Tyr	Tyr	Tyr	Tyr	SER	Tyr	Tyr	Tyr	Tyr	Asp	Asp	THR	ASX	Asp	Asx	Gly	Gly
Ile	Ile	SER	Ile	Ile	Ile	LYS	LYS	Ile	Ile	Gly	Gly	Gly	Gly	Gly	Gly	Ser	Ser	Ser	Ile	Ile	VAL	Gly	Gly	Gly	Gly	ASP
SER	Gly	Asn	Gly	Gly	TYR	Asn	Asn	TYR	Asn	Ile	Ile	Ile	Ile	Ile	VAL	Asn	Asn	Asn	Ser	Ser	LEU	---	TYR	ASX	Asn	TYR
ALA	Asn	Asn	Asn	Thr	Ser	Asn	Asn	Ser	Thr	Ser	Ser	Ser	Ser	Ser	Ser	Asn	Asn	Asn	Ile	ASP	SER	Asx	Lys	Thr	His	LYS
Asn	PHE	ARG	ARG	TYR	TYR	TRP	TRP	Asn	ALA	Phe	Phe	Phe	Phe	Phe	LEU	ARG	ARG	---	PHE	Tyr	Tyr	Tyr	Tyr	Tyr	PHE	Tyr
Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Met	Met	Met	Met	Met	Met	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Val	Val
Ala	SER	Ala	Ala	Ala	SER	Ala	Ala	Ala	GLY	Asn	Asn	Asn	Asn	Asn	HIS	Ala	Ala	SER	Asn	Asn	Asn	Asp	Asn	Asx	Ser	Ser
										Set 2																
										Trp									Trp			Trp			Trp	
										PHE									Tyr			Tyr			Tyr	
										Gln									Gln			LEU			Gln	Glx
										Gln									Gln			Glx	Gln	Gln	Gln	Glx
										Lys									Lys			Lys			HIS	
										Pro									Pro			Pro			Pro	
										Gly									Gly			Gly			Gly	
										Gln	All FR2 sequences identical to Set 2								LYS			Glx	Gln	Gln	LYS	
										Pro									ALA			SER			ALA	
										Pro									Pro			Pro			Pro	
										Lys									LYS			GLX	GLN	Glu	Lys	
										Leu									Leu			Leu			Leu	
										Leu									Leu			Leu			Ile	
										Ile									Ile			Ile			Ile	
										Tyr									Tyr			Tyr			Tyr	
ALA	Lys	ARG	ARG	ARG	Lys	Lys	Lys			Ala	Ala	Ala	Ala	Ala	Gly	Lys	Lys	Lys	Asp	Asp	ALA	Leu	Leu	Leu	Gly	GLU
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala			Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Ser	Val	Val
Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser			Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Asn	Ser
ASP	Thr	Thr	Thr	Thr	Thr	ASN	ASN			Asn	Asn	Asn	Asn	Asn	Asn	Thr	Thr	Thr	LYS	Asn	SER	Asn	Asn	Tyr	Lys	Ser
Leu	Leu	Leu	Leu	Leu	Leu	Leu	Leu			Gln	Gln	Gln	Gln	Gln	ARG	Leu	Leu	Leu	Leu	Leu	Leu	Arg	Arg	Arg	Arg	Arg
Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala			Gly	Gly	Gly	Gly	Gly	Gly	Ala	Ala	Ala	Glu	Glu	Glu	Ala	Ala	Asp	Pro	Pro
Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser			Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Ser	ALA	Ser	THR	Ser	Ser	Ser	Ser	Ser
									12/13				3/6				2/3			3/3			2/2			2/2

case CDR1 and CDR2 necessitate two copies.

Table III

FR1 Sets in Human $V_\kappa I$ Chains and Numbers of Chains in Each Set

No. in set	Residue number																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	Asp	Ile	Gln	Met	Thr	Gln	Ser	Pro	Ser	Ser	Leu	Ser	Ala	Ser	Val	Gly	Asp	Arg	Val	Thr	Ile	Thr	Cys
18																						Ile	
1																						Ser	
1																					Leu		
1																			Ile				
1													Val										
1										Thr													
7									Ala														
1																							
2				Leu																			
1													Val									Ala	
1										Thr										Ala			
1								Ala												Ile			
1										Pro								Ser					
1													Val				Glx						
1															Leu	Arg							
1													Thr	Val*									
2										Thr			Val										
1				Leu						Phe													
1														Thr							Leu	Leu*	
1				Leu					Thr												Phe		
1				Val									Val	Phe									
1		Val								Phe	Val												
1				Ile	Met						Val												
1												Val*	Ser	Pro				Leu					

Residues are identical to those in the top sequence except for substitutions listed.

* Two base changes.

rabbit, 12 copies would be required for the 13 sequences if none of these represent alleles. At the present, there is no basis for attributing the various copies to alleles even in outbred animals.

Huser and Braun (25) have found a single rabbit that makes antibody to the streptococcal group A variant carbohydrate with different FR1 and FR2. This rabbit has an identical CDR2; CDR1 differed only by a Gln Glu substitution at position 27. Moreover, different rabbits making antibody to the same antigen may have the same or different FR1 and FR2 sets. The existence in several species of identical FR2 sets which differ substantially in CDR1 and CDR2 suggests that the stability of FR2 is not maintained by selective pressure for certain antibody specificities. The X-ray crystallographic data which support a stable FR, upon which variation in CDR is responsible for specificity differences, also imply that the CDR do not determine selection or maintenance of a given FR2 set. All of these favor independent assortment of FR2 sets. Valbuena et al. (31) have shown by saturation hybridization analysis that there are no more than four to six germ-line genes for V-region sequences of $V_\kappa 21$. This number is substantially below that calculated in Table II as a minimum value for the $V_\kappa 21$ proteins having the preserved FR2 set. These data are consistent with the minigene concept.

Thus, the data in Tables I to IV require (*a*) many copies of genes for some FR1 or FR2 sets with no understanding of how they are maintained intact over evolutionary time, or (*b*) that in the genome, the V-region is made up of minigenes for each FR and by inference for each CDR segment, that these be present in only one or a few

TABLE IV

FR2 Sets in Mouse and Rabbit V_κ Chains and Number of Chains in Each Set Including Those Found in Human and Rabbit V_κ

Set	No. in set	Residue															Members of set
		35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	
1	1H, 20M 13R	TRP	TYR	GLN	GLN	LYS	PRO	GLY	GLN	PRO	PRO	LYS	LEU	LEU	ILE	TYR	See Table II
2	6M 4R		PHE														See Table II
3	1M									SER							MPC11 (V_κ19)
4	1M												VAL			PHE	7769 (V_κ21E)
5	1M															LYS*	2154 (V_κ21H)
6	1M					ASN				SER							CB101 (V_κ21E)
7	1M							GLU		SER							MOPC21 (V_κ15)
8	1M		LEU*			—	—	—	—	—	ILE		ARG				MOPC41 (V_κ9) (five residues missing)
9	1M			LEU		—	—	—	—	—	—	—	—				MOPC460 (V_κ1) (eight residues missing)
10	3M		PHE						LYS	ALA							UPC61, EPC109 (V_κ11); ABPC47 (V_κ)
11	1M							ASP	GLY*	THR	VAL*						MOPC173 (V_κ10)
12	1M		PHE	LEU		ARG				SER		GLN				SER	MOPC167 (V_κ24)
13	4M						SER		THR*	SER			PRO	TRP			XRPC44, XRPC24, TEPC109, J539 (V_κ4)
1	1R												VAL				3374‡
2	1R												GLY				4135§
3	1R						ALA						—	—	—	—	3T70‡ (four residues missing)
4	1R												ALA				2717‖
5	1R												GLY		LEU		XP-1‖
6	1R		PHE										ARG				3315¶
7	1R									ARG			VAL				4153 I§
8	1R		PHE										GLY*				AH80-5

* Two base changes. Amino acids listed in mouse sets 2–12 and rabbit sets 1–7 are those differing from set 1.

‖ Anti-*p*-azobenzoate

§ Anti-streptococcal group C carbohydrate.

¶ Anti-type VIII pneumococcal polysaccharide.

‡ Anti-type III pneumococcal polysaccharide.

Residues are identical to those in the top sequence except for substitutions listed.

copies in the germ line, and that they are assembled somatically as we have previously proposed (15) and as has since been shown unequivocally for the J segment (17–19) which includes all of FR4. Indeed, if many copies of genes coding for the preserved FR2 set were found already assembled, the explanation of why they are preserved intact, of their stabilization in evolution, and of the variation in the numbers of copies for each of the alternative FR2 sets will have to be at a level other than that of protein structure, probably in the genes themselves.

It will obviously be of primary importance to reconcile the cloning and the sequence data for further understanding of the genetics of the generation of antibody diversity. By using mRNA or cDNA from one species having an FR2 present in several species or by using a synthetic DNA stretch from FR2 made to correspond to an actual nucleotide sequence determined by cloning (17–21, 2) one should be able to count the numbers of copies of FR2. If this were done in sperm DNA as well as in embryonic and adult myeloma DNA, some resolution of the dilemma might emerge. It should

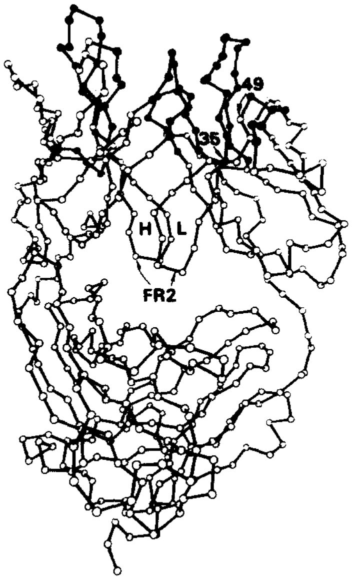

FIG. 2. Computer drawing of the α-carbon skeleton of the Fab fragment of MCPC603. The preserved FR2 loop of the L-chain residues 35-49 is indicated. Its location is such as to permit substitutions of side chains without necessarily changing the remaining structure. Multiple substitutions have been found at all positions except 35 and 38. Modified from reference 28.

also be possible to perform studies in families of inbred mice immunized with various antigens to determine whether assortment can be demonstrated in various generations.

The minigene hypothesis could account completely for the amino acid sequence data and would substantially reduce the amount of genetic material required for generating antibody diversity. For example, if each individual had 10 FR1, 10 FR2, and 10 FR3 sets, the hypothesis that they were already assembled in the germ line would necessitate 10^3 genes just for the three FR alone; indeed, 10 sets for each is a very low number (Tables III and IV). It would also relegate somatic mutation to a trivial role and to the extent that it occurred in the FR it would appear to increase the number of minigene sets. It should be noted also that hypotheses (17–21) based on sequencing clones also ascribe a minor role to somatic mutation as does a recent statistical examination of variability from the view point of population genetics (32). The latter analysis was considered to apply to the CDR as well as to the FR. It is of interest that Capra noted the low frequency of subgroup specific and of phylogenetically associated residues in the last 40 residues of the V_H region as compared to the first 40 residues, and suggested that these two segments might be under the control of separate genes (33).

Weigert et al. (22) have proposed that in the mouse V_κ 21 subgroups two residues of CDR3, in addition to FR4, are included in the J segment and diversity could be generated in this manner. They demonstrated independent assortment of this extended J segment. Because our original data (15) assorted only FR segments and did not include any CDR residues, the assortment would be independent of and would be seen whether or not any CDR residues were included together with FR4. As a result, the two sets of data (15, 22) are consistent. As mentioned earlier, the DNA clones that have been isolated (17–19), show that one and two residues in CDR3 may be included in the J segment of mouse V_λ clones. The two V_κ clones (20, 21) have been tabulated as terminating exactly at the end of CDR3. Thus, there is some uncertainty as to how many nucleotides coding for amino acids of CDR3 in addition to those for FR4 are present in various J segments. Resolution of this question will require sequencing of additional J containing clones. Whether the assumption of Weigert et al. (22) is justified will only become clear when such clones are available and when the mechanism of elimination of the intervening sequence between CDR3 and J is established.

Summary

Two sets of apparently conflicting data on the genes coding for the variable region are being accumulated. One suggests that the sets of nucleotides coding for the framework segments of immunoglobulin light and heavy (V_L and V_H) chains assort independently and are therefore germ-line minigenes which, together with sets of nucleotides coding for the complementarity-determining regions (CDR) or segments assemble to form complete variable (V)-region genes (15, 16, 33). The other, based on the findings with clones from 12-d-old embryo and adult mouse coding for V-regions, infer that the first three frameworks and the three complementarity-determining segments are already assembled as germ-line V-genes (17–21). It is now generally accepted that the J segment, which in the one instance sequenced (21) is made up of nucleotides coding for framework (FR)4 plus two residues of CDR3, is a minigene. An examination of sequences of human, mouse, and rabbit V-regions, assuming the latter hypothesis, indicates that individual framework sets would have to be present in many copies. The FR2 segment found in one human, 20 mice, and 13 rabbits would have to be present in at least 10/14 copies in the NZB, and 5/6 in the BALB/c mouse, and 12/13 in the rabbit. The X-ray crystallographic data show this region to be a loop, projecting out from the V-domain, capable of accommodating many substitutions and 12 and 8 alternative sequences for this FR2 segment have been found in mouse and rabbit V_κ chains with substitutions possible at 13 of the 15 positions. These alternative sequences occur much less frequently than the preserved FR2 segment. Thus, there is no basis in the protein structure to account for evolutionary stability of this FR2 segment if it occurs in so many copies in germ-line genes coding for residues 1–96, but its stability is easily explained if it were coded for by a separate germ-line minigene present as a single copy; the alternative forms could then have arisen by duplication and mutation of this minigene. Somatic assembly of the minigene segments for the three framework and three complementarity-determining segments during differentiation would account completely for our assortment data from which FR4 was inferred to be a minigene

The authors are indebted to Doctors Benvenuto Pernis, Rose G. Mage, David Davies, Michael Potter, Stuart Rudikoff, Sherie Morrison, and Christine Milcarek for helpful comments.

Received for publication 23 January 1979.

References

1. 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 and their implications for antibody complementarity. *J. Exp. Med.* **132:**211.

2. Kabat, E. A., T. T. Wu, and H. Bilofsky. 1976. Variable regions of immunoglobulin chains; tabulations and analyses of amino acid sequences. *In* Medical Computer Systems. Bolt Beranek, and Newman, Cambridge, Mass.
3. Kabat, E. A., and T. T. Wu. 1971. Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains of immunoglobulins. *Ann. N. Y. Acad. Sci.* **190:**382.
4. Kabat, E. A. 1976. Structural Concepts in Immunology and Immunochemistry. Holt, Rinehart, & Winston, Inc., N. Y. 2nd edition. Chapter 9.
5. Poljak, R. J., L. M. Amzel, H. P. Avey, B. L. Chen, R. P. Phizackerley, and F. Saul. 1973. Three dimensional structure of the Fab′ fragment of a human immunoglobulin at 2.8 Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **70:**220.
6. Segal, D. M., E. A. Padlan, G. H. Cohen, S. Rudikoff, M. Potter, and D. R. Davies. 1974. The three dimensional structure of a phosphorylcholine-binding site. *Proc. Natl. Acad. Sci. U. S. A.* **71:**4298.
7. Epp, O., P. Colman, H. Fehlhammer, W. Bode, M. Schiffer, R. Huber, and W. Palm. 1974. Crystal and molecular structure of a dimer composed of the variable portions of the Bence Jones protein REI. *Eur. J. Biochem.* **45:**513.
8. Edmundson, A. B., K. R. Ely, R. L. Girling, E. E. Abola, M. Schiffer, F. A. Westholm, M. D. Fausch, and H. F. Deutsch. 1974. Binding of 2,4 dinitrophenyl compounds and other small molecules to a crystalline λ-type Bence Jones dimer. *Biochemistry.* **13:**3816.
9. Milstein, C. 1967. Linked groups of residues in immunoglobulin κ chains. *Nature* (*Lond.*). **216:**330.
10. Niall, H. D., and P. Edman. 1967. Two structurally distinct classes of kappa chains in human immunoglobulins. *Nature* (*Lond.*). **216:**262.
11. Hood, L., W. R. Gray, B. G. Sanders, and W. J. Dreyer. 1967. Light chain evolution. *Cold Spring Harbor Symp. Quant Biol.* **32:**133.
12. Potter, M. 1977. Antigen-binding myeloma proteins of mice. *Adv. Immunol.* **25:**141.
13. Barstad, P., V. Farnsworth, M. Weigert, M. Cohn, and L. Hood. 1974. Mouse immunoglobulin heavy chains are coded by multiple germ line variable region genes. *Proc. Natl. Acad. Sci. U. S. A.* **71:**4096.
14. Capra, J. D., and J. M. Kehoe. 1975. Hypervariable regions, idiotype, and the antibody-combining site. *Adv. Immunol.* **20:**1.
15. Kabat, E. A., T. T. Wu, and H. Bilofsky. 1978. Variable region genes for the immunoglobulin framework are assembled from small segments of DNA—A hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* **75:**2429.
16. Kabat, E. A. 1979. Implications of the assortment of framework segments for the assembly of immunoglobulin V_L and V_H regions and the generation of diversity. *In* Cells of Immunoglobulin Synthesis. B. Pernis and H. J. Vogel, editors. Academic Press, Inc., New York. In press.
17. Tonegawa, S., A. M. Maxam, R. Tizard, O. Bernard, and W. Gilbert. 1978. Sequences of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. U. S. A.* **75:**1485.
18. Brack, C., M. Hirama, R. Lenhard-Schuller, and S. Tonegawa. 1978. A complete immunoglobulin gene is created by somatic recombination. *Cell.* **15:**1.
19. Bernard, O., N. Hozumi, and S. Tonegawa. 1978. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell.* **15:**1133.
20. Seidman, G., A. Leder, M. H. Edgell, F. Polsky, S. M. Tilghman, D. C. Tiemeier, and P. Leder. 1978. Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis. *Proc. Natl. Acad. Sci. U. S. A.* **75:**3881.

21. Seidman, J. G., A. Leder, M. Nau, B. Norman, and P. Leder. 1978. Antibody diversity. The structure of cloned immunoglobulin genes suggests a mechanism for generating new sequences. *Science* (*Wash. D. C.*). **202:**11.

22. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. *Nature* (*Lond.*). **276:**785.

23. Brandt, D. Ch., and J.-C. Jaton. 1978. Identical V_L region sequences of two antibodies from two outbred rabbits exhibiting complete idiotypic cross-reactivity and probably the same antigen-binding site fine structure. *J. Immunol.* **121:**1194.

24. McKean, D. J., M. Bell, and M. Potter. 1978. Mechanisms of antibody diversity: multiple genes encode structurally related mouse κ variable regions. *Proc. Natl. Acad. Sci. U. S. A.* **75:** 3913.

25. Huser, H., and D. G. Braun. 1978. Rabbit variable kappa light chain regions: subgroups contain polypeptides encoded by multiple genes. *Hoppe Seyler's Z. Physiol. Chem.* **359:**1473.

26. Chersi, A., E. Appella, S. Carta, and R. Mage. 1979. The amino acid sequence of a variable region of rabbit b4 light chain from an anti-SIII antibody: comparison with light chains of the same subgroup from anti-A variant carbohydrate antibodies. *Immunochemistry.* In press.

27. Capra, J. D., and H. G. Kunkel. 1970. Amino acid sequence similarities in two human anti-gamma globulin antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **67:**87.

28. Davies, D. R., E. A. Padlan, and D. M. Segal. 1975. Immunoglobulin structures at high resolution. *In* Contemporary Topics in Molecular Immunology. Inman, F. P. and W. J. Mandy, editors, Plenum Publishing Corp., New York. **4:**127.

29. Dreyer, W. J., and J. C. Bennett. 1965. The molecular basis of antibody formation. A paradox. *Proc. Natl. Acad. Sci. U. S. A.* **54:**865.

30. Rabbitts, T. H. 1977. A molecular hybridization approach for the determination of the immunoglobulin V-gene pool size. *Immunological Rev.* **36:**29.

31. Valbuena, D., K. B. Marcu, M. Weigert, and R. B. Perry. 1979. The multiplicity of germ line genes specifying a group of related mouse kappa chains: implications for the generation of immunoglobulin diversity. *Nature* (*Lond.*). **276:**780.

32. Ohta, T. 1978. Sequence variability of immunoglobulins considered from the standpoint of population genetics. *Proc. Natl. Acad. Sci. U. S. A.* **75:**5108.

33. Capra, J. D. 1976. The implications of phylogenetically associated residues and idiotypes on theories of antibody diversity. *In* The Generation of Antibody Diversity: A New Look. A. J. Cunningham, editor, Academic Press, Inc., New York. 65.

34. Raub, W. F. 1974. The PROPHET system and resource sharing. *Fed. Proc.* **33:**2390.