AN ANALYSIS OF THE SEQUENCES OF THE VARIABLE REGIONS OF BENCE JONES PROTEINS AND MYELOMA LIGHT CHAINS AND THEIR IMPLICATIONS FOR ANTIBODY COMPLEMENTARITY*

By TAI TE WU, Ph.D., and ELVIN A. KABAT, Ph.D

(From the Departments of Microbiology, Neurology, and Human Genetics and Development, College of Physicians and Surgeons, Columbia University, and the Neurological Institute, Presbyterian Hospital, New York 10032; the Biomathematics Division, Graduate School of Medical Sciences, Cornell University and the Sloan-Kettering Institute, New York 10021)

(Received for publication 26 March 1970)

The extraordinary versatility of the antibody-forming mechanism in producing an almost limitless number of specific receptor sites complementary for almost any molecular conformation of matter within a size range (1–3) represented by a hexa- or heptasaccharide as an upper and a mono- or disaccharide as a lower limit, is almost certainly related to the unique structural features of immunoglobulins and differentiates them from all other known proteins. These antibody-combining sites are formed as a consequence of the interaction of two polypeptide chains, a light and a heavy chain (2, 4, 5). The antibodies usually formed to various antigens often represent heterogeneous populations of immunoglobulin molecules of different classes, subclasses, and genetic variants and also show specificities toward different antigenic determinants (1, 2, 6, 7). In some instances, however, relatively homogeneous populations of antibodies with respect to many of these properties have been obtained. Among these have been human antibodies to dextran and levan (8, 9) and rabbit antibodies to the group-specific carbohydrate of streptococcus (10–12), antibodies to the Type III–specific capsular polysaccharide of pneumococcus (13, 14), rabbit antihapten (15), and specimens of antibodies and of Fab′ fragments which crystallized (Nisonoff et al., in references 16, 17), but sequence data on these are not yet available.

The large body of sequence data related to immunoglobulin structure comes from the analysis of urinary Bence Jones proteins and from the monoclonal immunoglobulins found in large amounts in the sera of patients with multiple myeloma and Waldenström macroglobulinemia (16, 18). While a substantial body of evidence was available relating these proteins to immunoglobulins, the recent demonstration that many myeloma globulins have specific ligand-binding properties like those of many antibodies provides increasing confidence that myeloma globulins represent homogeneous populations of antibody molecules (16, 18–27). The ability to produce in BALB/c

* Aided by grants from the National Science Foundation (GB-8341) and the National Cancer Institute (CA-08748), and a general Research Support Grant of the U. S. Public Health Service.

mice myelomas and macroglobulinemias (28) which produce myeloma globulins and Bence Jones proteins like those in the human, provides a source of data from which important evolutionary trends can be inferred.

Thus the extensive sequence data on Bence Jones proteins, which are considered to be light chains of myeloma globulins and Waldenström macroglobulins (29), and on various light and heavy chains, provide information clearly pertinent to the problem of the elucidation of the structure of antibody-combining sites.

The unique finding that distinguishes the immunoglobulins from all other proteins is that the N-terminal half of the light chains and the N-terminal quarter of the heavy chains vary in sequence in samples obtained from individual monoclonal immunoglobulins and that indeed no two such variable regions of any chain and no two myeloma immunoglobulins or Bence Jones proteins have thus far been found to be identical in sequence (30). The constant region, however, is essentially no different from other proteins in that the variation in the amino acids found at any position is ascribable to species and class variations or to genetic variants such as Inv factors. By comparison of sequence data on the variable and constant regions of Bence Jones proteins with amino acid composition of purified human antibodies, it could be shown that most of the compositional variation could only originate in the variable region (see Kabat in reference 18).

From sequence data, a variety of hypotheses have been advanced (7, 31–35) to explain the structural basis of antibody complementarity. All of these are selective theories, i.e. they consider that the information for complementarity is essentially built into the primary sequence of each chain and that a given antigen only triggers the biosynthesis of those antibody molecules having complementary receptor sites. There are two types of selective theories: germ line theories (36) and somatic mutation theories (37–39). At present no hypothesis is generally accepted. Excellent reviews (see above) are available.

The present communication is an extension of earlier efforts from this laboratory (18, p. 87, and 40–43) to locate more precisely those portions of the variable region which are directly responsible for antibody complementarity, that is which make direct contact with the antigenic determinant, and to explain the unique capacity of these proteins to have so many complementary regions.

As in the earlier studies, all human κ, human λ, and mouse κ Bence Jones protein and light chain sequences are aligned for maximum homology (44) and all variable regions are considered as a unit and compared with the constant regions. These earlier studies had called attention to the following:

(*a*) The variable regions had few if any species-specific positions while the constant regions of the human and mouse proteins had 36 species-specific amino acid substitutions per 107 residues (40, 45). A species-specific position is defined as one at which the amino acid residues in the mouse proteins differ from those in the human proteins.

(*b*) When the invariant residues of these two regions were compared, the latest tabulation (45) showed the variable regions to have 10 invariant and almost invariant glycines and no invariant alanines, leucines, valines, histidines, lysines, or serines while the constant regions had 3 each of invariant alanine, leucine, and valine, and 2 invariant histidines, 2 invariant lysines, and 5 invariant serines. It was suggested that the invariant glycines were important in contributing to the flexibility needed by the variable region in accommodating the numerous substitutions (41, 43) at the variable positions. It was also suggested that the invariant glycines near the end of the variable region at positions 99 and 101, plus the almost invariant glycine at position 100, provided a pivot upon which the complementarity-determining regions might move to make better contact with the antigenic determinant (43; 18, p. 87) just as the walls of the lysozyme site have been shown to adjust somewhat to accommodate its hexasaccharide substrate (46). The hydrophobic residues in the constant region were hypothesized to be involved in noncovalent bonding to the heavy chain.

(*c*) From an examination of sequences of the κI, κII, and κIII subgroups (Hood et al. in reference 16) (47, 48) of the human Bence Jones proteins in which many of the proteins in a subgroup had an identical sequence for the first 20–24 residues, it was postulated that there are two kinds of residues in the variable regions, those making direct contact with the antigenic determinant (complementarity determining) and those which are involved only in three-dimensional folding (42). The latter would be expected to have less stringent requirements, and more mutation noise would be permitted than with the complementarity-determining residues. This distinction led to the inspection of the sequences for short stretches showing very high variability and two of these were identified: the most variable beginning at residue 89 and ending at 97, the other running from residue 24 through 34. Each of these two unusually highly variable regions began after an invariant half-cystine and was followed by an invariant phenylalanine (residue 98) and an invariant tryptophane (residue 35) respectively. It is of interest that the two regions are brought close together by the S—S bond I_{23}–II_{88} (45). Milstein (47), Milstein and Pink (7), and Franěk (49) have also called attention to the highly variable positions in these regions and Franěk (49) has noted an additional highly variable region around residues 52–55. It was hypothesized (45) that these first two regions might represent the complementarity-determining regions and that complementarity might be acquired by the insertion of small linear sequences into the light and heavy chains by some episomal or other insertion mechanism. It is striking that the differences in chain length seen in the Bence Jones proteins arĕ confined to these two regions of the chain. The remaining portions of each chain would be essentially under the control of structural genes. The inserted sequences would be drawn from a large but finite set and either inserted under the influence of antigen, if antibody-forming cells are multipotent, or individual sequences might be distributed to immunoglobulin-forming cells during differentiation if the capacity of individual cells to synthesize antibody is restricted.

This working hypothesis offers several advantages:

(*a*) It is capable of providing the evolutionary stability and accounts for the universality of the antibody-forming mechanism throughout the vertebrates. Germ line theories (34–36) postulate one gene for each of the thousand or more variable regions (30). This would be expected to result in divergence during evolution since the loss by mutation of any one variable region would only minimally affect the capacity to form antibody and survival; thus individuals and populations lacking certain variable regions would arise.

(*b*) It offers a substantial simplification to the problem of producing a very large number of complementary sites. While it is known that in all proteins with specific receptors the site is formed by residues from widely separated portions of the chain, these sites are all formed by single chains. Thus, forming a three-dimensional site must involve residues from various regions. The antibody site being formed by a heavy and a light chain need not necessarily be so restricted.

Since much additional data on the light chains and a number of heavy chain sequences have been accumulated, the present communication represents a further attempt at analyzing the unique features of the variable regions of immunoglobulin chains. Among aspects considered are the role of glycine, invariant residues, and hydrophobicity patterns, and the highly variable portions, with a view to localizing the regions responsible for complementarity and evaluating various theories in terms of evolutionary origin and perpetuation of the antibody-forming mechanism.

Sequence Data Employed—Complete and partial sequence data have been published on 77 Bence Jones proteins and immunoglobulin light chains as well as on a number of heavy chains. Data were available on 24 human κI, 4 human κII, 17 human κIII, 10 human λI, 2 human λII, 6 human λIII, 5 human λIV, 2 human λV, 2 mouse κI, and 5 mouse κII proteins.[1]

The original light chain sequence data may be found in the following references.

HBJ 98: Baglioni, C. 1967. *Biochem. Biophys. Res. Commun.* **26**:82.

Eu: Cunningham, B. A., P. D. Gottlieb, W. H. Konigsberg, and G. M. Edelman. 1968. *Biochemistry.* **7**:1983.

Mil (human κII): Dreyer, W. J., W. R. Gray, and L. Hood. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:353.

Hac, Dob, Pal: Grant, A., and L. Hood. Unpublished work.

Roy, Cum: Hilschman, N., and L. C. Craig. 1965. *Proc. Nat Acad. Sci. U. S. A.* **53**:1403; Hilschmann, N. 1967. *Hoppe-Seyler's Z. Physiol. Chem.* **348**:1077; Hilschmann, N., H. U. Barnikol, M. Hess, B. Langer, H. Ponstingl, M. Steinmetz-Kayne, L. Suter, and S. Watanabe. 1968. *Fed. Eur. Biochem. Soc. Symp., 5th.* In press.

[1] The World Health Organization has recently changed the notation of subgroups so that human κII in this paper will become human κIII and human κIII will become human κII.

HS 78, HS 92, HS 94, HS 68, HS 70, HS 77, HS 86, HS 24: Hood, L., and D. Ein. 1968. *Nature (London)*. **220**:764.

HBJ 7, HBJ 11, HBJ 2, HBJ 8: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *J. Mol. Biol.* **22**:179.

MBJ 41, MBJ 70, MBJ 6: Hood, L., W. R. Gray, and W. J. Dreyer. 1966. *Proc. Nat'l Acad. Sci. U. S. A.* **55**:826.

HBJ 10, HBJ 1, HBJ 4, HBJ 6, HBJ 5, HS 4, HBJ 12, HS 6, HBJ 15: Hood, L., W. R. Gray, B. G. Sanders, and W. J. Dreyer. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:133.

Ste: Edman, P., and A. G. Cooper. 1968 *Fed. Eur. Biochem Soc. Letters.* **2**:33; Hood, L., and D. W. Talmage. 1969. *In* Developmental Aspects of Antibody Formation and Structure. Prague. In press.

Lay, Mar, Ioc, Wag, How, Koh: Kaplan, A. P. and H. Metzger. 1969. *Biochemistry.* **10**: 3944.

New, III, Mil (human λIV): Langer, B., M. Steinmetz-Kayne, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349**:945.

BJ, Ker: Milstein, C. 1966. *Biochem. J.* **101**:352.

Rad, Fr4: Milstein, C. 1967. *Nature (London)* **216**:330.

X: Milstein, C. 1968. *Biochem. J.* **110**:631.

Bel, Man, B6: Milstein, C. 1968. *Fed. Eur. Biochem. Soc. Symp. on γ-globulin*, Prague.

Day, MBJ46, Roy: Atlas of Protein Sequence and Structure, M. O. Dayhoff, Editor. 1969.

Mz: Milstein, C., B. Frangione, and J. R. L. Pink. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:31.

Ale, Car, Dee: Milstein, C., C. P. Milstein, and A. Feinstein. 1969. *Nature (London)* **221**:151.

Cra, Pap, Lux, Mon, Con, Tra, Nig, Win, Gra, Cas, Smi: Niall, H., and P. Edman. 1967. *Nature (London)* **216**:262.

MOPC 149, AdjPC 9, MOPC 157: Perham, R., E. Appella, and M. Potter. 1966. *Science (Washington)* **154**:391.

Kern: Ponstingl, H., M. Hess, and N. Hilschmann. 1968. *Hoppe-Seyler's Z. Physiol. Chem.* **349**:867.

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

Ag, Ha, Bo, Sh: Putnam, F. W., K. Titani, M. Wikler, and T. Shinoda. 1967. *Cold Spring Harbor Symp. Quant. Biol.* **32**:9; Titani, K., T. Shinoda, and F. W. Putnam. 1969. *J. Biol. Chem.* **244**:3550.

TI: Suter, L., H. U. Barnikol, S. Watanabe, and N. Hilschmann. 1969. *Hoppe-Seyler's Z. Physiol. Chem.* **350**:275.

The accumulation of such large numbers of sequences makes it possible to use statistical criteria in defining the types of residues. Thus in earlier studies, an invariant residue was rigidly defined, e.g., a position at which all samples showed the same amino acid residue sometimes allowing a single exception. The definition of an invariant residue used in this paper is taken as a position at which 88-90% or more of the samples contain the same amino acid. This may allow potential functions to be recognized despite possible errors or uncertainties in sequence, or occasional substitutions compatible with function.

A summary of the sequence data is provided in Table I which lists the amino acids found at any position in any subgroup of human κ-, human λ-, and mouse κ-chains, the number of times each occurs, and the total number of sequences

TABLE I

Amino Acids Found at each Position in the Variable Region of the Various Subgroups of Human κ-, Human λ- and Mouse κ-Bence Jones Proteins

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
0	74	Glu	1	0	1	0	0	0	0	0	0	0	0
		---	73	23	3	17	9	2	6	4	2	2	5
1	74	Lys	1	0	0	1	0	0	0	0	0	0	0
		PCA	17	0	0	0	8	2	6	0	0	0	1
		Asp	29	18	4	1	0	0	0	0	0	2	4
		Asx	5	5	0	0	0	0	0	0	0	0	0
		Glu	15	0	0	14	1	0	0	0	0	0	0
		Glx	1	0	0	1	0	0	0	0	0	0	0
		---	6	0	0	0	0	0	0	4	2	0	0
2	73	Ile	48	22	3	16	0	0	0	0	0	2	5
		Tyr	4	0	0	0	0	0	0	4	0	0	0
		Val	1	1	0	0	0	0	0	0	0	0	0
		Met	1	0	0	1	0	0	0	0	0	0	0
		Ser	19	0	0	0	9	2	6	0	2	0	0
3	72	Ile	1	0	0	1	0	0	0	0	0	0	0
		Pro	1	0	0	0	0	0	1	0	0	0	0
		Leu	1	1	0	0	0	0	0	0	0	0	0
		Val	33	0	3	16	8	0	0	1	0	0	5
		Ala	9	0	0	0	1	2	5	1	0	0	0
		Asp	1	0	0	0	0	0	0	1	0	0	0
		Glu	2	0	0	0	0	0	0	1	1	0	0
		Gln	21	19	0	0	0	0	0	0	0	2	0
		Glx	3	3	0	0	0	0	0	0	0	0	0
4	71	Leu	43	3	1	14	9	2	6	4	1	0	3
		Val	2	0	0	1	0	0	0	0	0	0	1
		Met	26	20	2	2	0	0	0	0	0	2	0
5	70	Ala	2	0	0	0	0	0	2	0	0	0	0
		Thr	67	23	3	17	9	2	3	4	1	2	3
		Ser	1	0	0	0	0	0	1	0	0	0	0
6	69	Gln	63	20	3	16	8	1	6	4	1	1	3
		Glx	6	3	0	1	1	1	0	0	0	0	0
7	63	Pro	20	0	0	0	9	2	6	3	0	0	0
		Thr	1	0	1	0	0	0	0	0	0	0	0
		Ser	41	22	2	14	0	0	0	0	0	1	2
		Asp	1	0	0	0	0	0	0	0	1	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa I	II	III	Human Lambda I	II	III	IV	V	Mouse Kappa I	II
8	64	Pro	58	22	3	15	9	2	0	3	1	1	2
		Ala	6	0	0	0	0	0	6	0	0	0	0
9	63	Leu	3	0	3	0	0	0	0	0	0	0	0
		Ala	7	0	0	4	0	0	0	0	1	0	2
		Thr	1	1	0	0	0	0	0	0	0	0	0
		Ser	41	21	0	0	9	2	6	2	0	1	0
		Gly	10	0	0	10	0	0	0	0	0	0	0
		Asx	1	0	0	1	0	0	0	0	0	0	0
10	63	Phe	1	1	0	0	0	0	0	0	0	0	0
		Thr	17	3	0	14	0	0	0	0	0	0	0
		Ser	25	18	4	0	0	0	0	0	0	1	2
		---	20	0	0	0	9	2	6	2	1	0	0
11	63	Leu	43	22	4	14	0	0	0	0	0	1	2
		Val	15	0	0	0	6	0	6	2	1	0	0
		Ala	5	0	0	0	3	2	0	0	0	0	0
12	61	Pro	4	0	4	0	0	0	0	0	0	0	0
		Ala	1	0	0	0	0	0	0	0	0	0	1
		Ser	56	21	0	13	9	2	6	2	1	1	1
13	61	Leu	12	0	0	12	0	0	0	0	0	0	0
		Val	11	2	4	0	0	0	0	2	1	0	2
		Met	1	0	0	1	0	0	0	0	0	0	0
		Ala	23	19	0	0	3	0	0	0	0	1	0
		Gly	14	0	0	0	6	2	6	0	0	0	0
14	61	Ala	6	0	0	0	5	0	0	0	1	0	0
		Thr	9	0	4	0	4	0	0	0	0	0	1
		Ser	46	21	0	13	0	2	6	2	0	1	1
15	61	Pro	36	0	4	13	8	2	6	2	0	0	1
		Leu	4	1	0	0	0	0	0	0	1	1	1
		Val	20	20	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	1	0	0	0	0	0	0
16	61	Arg	1	1	0	0	0	0	0	0	0	0	0
		Gly	60	20	4	13	9	2	6	2	1	1	2
17	61	Asp	23	21	0	1	0	0	0	0	0	0	1
		Glu	17	0	4	11	0	0	1	0	0	1	0
		Gln	19	0	0	0	8	2	5	2	1	0	1
		Glx	2	0	0	1	1	0	0	0	0	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studies	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
18	61	Pro	4	0	4	0	0	0	0	0	0	0	0
		Lys	1	0	0	0	1	0	0	0	0	0	0
		Ala	1	0	0	0	1	0	0	0	0	0	0
		Arg	40	21	0	13	4	0	0	0	0	1	1
		Thr	5	0	0	0	1	0	1	2	1	0	0
		Ser	9	0	0	0	1	2	5	0	0	0	1
		Gly	1	0	0	0	1	0	0	0	0	0	0
19	53	Ile	7	1	0	0	0	0	6	0	0	0	0
		Val	28	17	0	0	6	2	0	0	1	1	1
		Ala	18	0	4	10	1	0	0	2	0	0	1
20	53	Ile	1	0	0	0	1	0	0	0	0	0	0
		Val	1	0	0	0	0	0	0	1	0	0	0
		Ala	3	1	0	2	0	0	0	0	0	0	0
		Arg	1	0	0	0	0	0	0	0	1	0	0
		Thr	40	17	0	8	6	2	6	0	0	0	1
		Ser	7	0	4	0	0	0	0	1	0	1	1
21	43	Ile	30	14	4	0	5	2	0	3	1	0	1
		Leu	12	1	0	9	0	0	0	0	0	1	1
		Val	1	0	0	1	0	0	0	0	0	0	0
22	42	Ala	1	1	0	0	0	0	0	0	0	0	0
		Thr	19	14	0	0	0	0	0	3	1	1	0
		Ser	22	0	4	9	5	2	0	0	0	0	2
23	30	Cys	30	9	3	4	5	2	0	3	1	1	2
24	26	Arg	11	1	3	4	0	0	0	0	0	1	2
		Thr	2	0	0	0	0	2	0	0	0	0	0
		Ser	6	0	0	0	4	0	0	2	0	0	0
		Gly	1	0	0	0	0	0	0	1	0	0	0
		Gln	4	3	0	0	0	0	0	0	1	0	0
		Glx	2	2	0	0	0	0	0	0	0	0	0
25	25	Ala	13	6	0	4	0	0	0	0	0	1	2
		Ser	2	0	2	0	0	0	0	0	0	0	0
		Gly	10	0	0	0	4	2	0	3	1	0	0
26	24	Thr	2	0	0	0	0	2	0	0	0	0	0
		Ser	16	6	2	4	1	0	0	0	0	1	2
		Gly	2	0	0	0	2	0	0	0	0	0	0
		Asp	4	0	0	0	0	0	0	3	1	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
27	23	Lys	1	0	0	0	0	0	0	1	0	0	0
		Ser	5	0	0	0	3	1	0	0	1	0	0
		Asn	1	0	0	0	0	0	0	1	0	0	0
		Glu	1	0	0	0	0	0	0	1	0	0	0
		Gln	13	5	1	4	0	0	0	0	0	1	2
		Glx	2	1	1	0	0	0	0	0	0	0	0
a	22	Ser	5	0	1	3	0	0	0	0	0	0	1
		Asn	1	0	1	0	0	0	0	0	0	0	0
		---	16	6	0	1	3	1	0	2	1	1	1
b	22	Leu	2	0	2	0	0	0	0	0	0	0	0
		Val	1	0	0	0	0	0	0	0	0	0	1
		---	19	6	0	4	3	1	0	2	1	1	1
c	22	Leu	2	0	2	0	0	0	0	0	0	0	0
		---	20	6	0	4	3	1	0	2	1	1	2
d	22	Thr	1	0	0	0	1	0	0	0	0	0	0
		Ser	3	0	0	0	2	1	0	0	0	0	0
		Asp	1	0	1	0	0	0	0	0	0	0	0
		---	17	6	1	4	0	0	0	2	1	1	2
e	22	Asp	2	0	1	0	0	1	0	0	0	0	0
		Asn	3	0	0	0	3	0	0	0	0	0	0
		Asx	1	0	0	0	0	0	0	0	0	0	1
		---	16	6	1	4	0	0	0	2	1	1	1
f	22	Val	2	0	0	1	0	1	0	0	0	0	0
		Ser	2	0	1	0	0	0	0	0	0	0	1
		Gly	1	0	0	0	1	0	0	0	0	0	0
		Glx	1	0	1	0	0	0	0	0	0	0	0
		---	16	6	0	3	2	0	0	2	1	1	1
28	22	Ile	1	0	0	0	1	0	0	0	0	0	0
		Leu	4	0	0	1	0	0	0	2	1	0	0
		Val	2	0	0	2	0	0	0	0	0	0	0
		Met	1	0	0	0	1	0	0	0	0	0	0
		Thr	1	0	0	0	1	0	0	0	0	0	0
		Ser	3	1	1	1	0	0	0	0	0	0	0
		Gly	2	0	1	0	0	1	0	0	0	0	0
		His	1	0	0	0	0	0	0	0	0	0	1
		Asp	4	4	0	0	0	0	0	0	0	0	0
		Asx	3	1	0	0	0	0	0	0	0	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
29	21	Ile	8	6	0	0	0	0	0	0	0	1	1
		Arg	2	0	0	1	0	0	0	0	1	0	0
		Ser	3	0	0	3	0	0	0	0	0	0	0
		Gly	4	0	0	0	2	0	0	1	0	0	1
		Glu	1	0	0	0	0	0	0	1	0	0	0
		Asp	1	0	1	0	0	0	0	0	0	0	0
		Asx	2	0	1	0	0	1	0	0	0	0	0
30	21	Ile	2	1	0	0	0	0	0	0	0	0	1
		Lys	2	1	0	0	0	0	0	1	0	0	0
		Ser	2	1	0	0	0	0	0	0	0	0	1
		Gly	5	0	2	1	0	0	0	0	1	1	0
		Asp	1	0	0	0	0	0	0	1	0	0	0
		Asn	8	3	0	3	2	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
31	20	Tyr	1	0	0	0	0	0	0	0	1	0	0
		Phe	1	0	0	0	0	0	0	1	0	0	0
		Lys	2	1	0	0	0	1	0	0	0	0	0
		Thr	2	1	1	0	0	0	0	0	0	0	0
		Ser	7	1	0	2	2	0	0	0	0	1	1
		His	1	1	0	0	0	0	0	0	0	0	0
		Asn	4	1	0	2	0	0	0	0	0	0	1
		Asx	2	1	1	0	0	0	0	0	0	0	0
32	20	Trp	1	1	0	0	0	0	0	0	0	0	0
		Tyr	10	2	2	3	2	1	0	0	0	0	0
		Phe	5	3	0	1	0	0	0	0	0	0	1
		Leu	1	0	0	0	0	0	0	0	0	1	0
		Thr	1	0	0	0	0	0	0	1	0	0	0
		Asp	1	0	0	0	0	0	0	0	1	0	0
		Asn	1	0	0	0	0	0	0	0	0	0	1
33	18	Leu	11	4	2	4	0	0	0	0	0	0	1
		Val	4	0	0	0	2	1	0	1	0	0	0
		Met	1	0	0	0	0	0	0	0	0	0	1
		Ala	1	0	0	0	0	0	0	0	1	0	0
		Ser	1	0	0	0	0	0	0	0	0	1	0
34	18	Tyr	1	0	0	0	1	0	0	0	0	0	0
		Ala	6	1	0	4	0	0	0	0	1	0	0
		Ser	3	0	0	0	1	1	0	1	0	0	0
		Asp	1	0	1	0	0	0	0	0	0	0	0
		Asn	4	2	1	0	0	0	0	0	0	0	1
		Asx	3	1	0	0	0	0	0	0	0	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
35	17	Trp	17	4	2	4	2	1	0	1	1	1	1
36	17	Tyr	13	4	2	4	1	1	0	0	1	0	0
		Phe	2	0	0	0	0	0	0	1	0	0	1
		Leu	1	0	0	0	0	0	0	0	0	1	0
		His	1	0	0	0	1	0	0	0	0	0	0
37	16	Leu	2	0	2	0	0	0	0	0	0	0	0
		Gln	12	3	0	4	2	1	0	1	1	0	0
		Glx	2	0	0	0	0	0	0	0	0	1	1
38	16	His	1	0	0	0	1	0	0	0	0	0	0
		Gln	12	3	1	4	1	1	0	1	1	0	0
		Glx	3	0	1	0	0	0	0	0	0	1	1
39	16	Leu	2	0	0	0	2	0	0	0	0	0	0
		Lys	9	2	2	3	0	0	0	0	1	0	1
		Arg	2	0	0	1	0	0	0	1	0	0	0
		Gly	1	1	0	0	0	0	0	0	0	0	0
		His	1	0	0	0	0	1	0	0	0	0	0
		Asx	1	0	0	0	0	0	0	0	0	1	0
40	16	Pro	15	3	1	4	2	1	0	1	1	1	1
		Ala	1	0	1	0	0	0	0	0	0	0	0
41	16	Lys	1	1	0	0	0	0	0	0	0	0	0
		Gly	15	2	2	4	2	1	0	1	1	1	1
42	16	Lys	3	3	0	0	0	0	0	0	0	0	0
		Arg	1	0	0	0	0	1	0	0	0	0	0
		Thr	2	0	0	0	2	0	0	0	0	0	0
		Gln	7	0	1	4	0	0	0	1	1	0	0
		Glx	3	0	1	0	0	0	0	0	0	1	1
43	16	Pro	1	0	0	0	0	0	0	0	0	0	1
		Ala	11	3	0	4	2	1	0	0	1	0	0
		Thr	1	0	0	0	0	0	0	0	0	1	0
		Ser	2	0	1	0	0	0	0	1	0	0	0
		Gln	1	0	1	0	0	0	0	0	0	0	0
44	16	Ile	1	0	0	0	0	0	0	0	0	1	0
		Pro	15	3	2	4	2	1	0	1	1	0	1
45	16	Leu	2	0	0	0	0	0	0	1	1	0	0
		Lys	9	3	0	1	2	1	0	0	0	1	1
		Arg	3	0	0	3	0	0	0	0	0	0	0
		Ser	1	0	1	0	0	0	0	0	0	0	0
		Glx	1	0	1	0	0	0	0	0	0	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
46	14	Ile	1	1	0	0	0	0	0	0	0	0	0
		Leu	12	2	2	2	2	1	0	1	1	0	1
		Arg	1	0	0	0	0	0	0	0	0	1	0
47	14	Leu	11	3	2	2	2	0	0	0	0	1	1
		Val	3	0	0	0	0	1	0	1	1	0	0
48	14	Ile	12	2	2	1	2	1	0	1	1	1	1
		Met	2	1	0	1	0	0	0	0	0	0	0
49	14	Tyr	13	3	2	2	2	0	0	1	1	1	1
		Phe	1	0	0	0	0	1	0	0	0	0	0
50	14	Leu	1	0	1	0	0	0	0	0	0	0	0
		Val	1	0	0	1	0	0	0	0	0	0	0
		Lys	1	1	0	0	0	0	0	0	0	0	0
		Ala	2	0	0	0	0	0	0	0	0	1	1
		Arg	1	0	0	0	1	0	0	0	0	0	0
		Thr	1	0	1	0	0	0	0	0	0	0	0
		Ser	1	0	0	0	0	0	0	1	0	0	0
		Gly	3	0	0	1	0	1	0	0	1	0	0
		Asp	2	2	0	0	0	0	0	0	0	0	0
		Glu	1	0	0	0	1	0	0	0	0	0	0
51	14	Leu	1	0	1	0	0	0	0	0	0	0	0
		Val	2	0	0	1	0	1	0	0	0	0	0
		Ala	5	3	0	1	0	0	0	0	0	0	1
		Arg	1	0	0	0	0	0	0	0	1	0	0
		Thr	1	0	0	0	0	0	0	0	0	1	0
		Gly	1	0	1	0	0	0	0	0	0	0	0
		His	1	0	0	0	0	0	0	1	0	0	0
		Asp	2	0	0	0	2	0	0	0	0	0	0
52	14	Ser	10	3	2	2	0	1	0	0	0	1	1
		Asp	1	0	0	0	1	0	0	0	0	0	0
		Asn	2	0	0	0	1	0	0	0	1	0	0
		Glu	1	0	0	0	0	0	0	1	0	0	0
53	14	Tyr	1	0	1	0	0	0	0	0	0	0	0
		Lys	3	1	0	0	2	0	0	0	0	0	0
		Ser	4	1	0	2	0	0	0	0	0	1	0
		Asn	4	1	1	0	0	0	0	0	1	0	1
		Glu	1	0	0	0	0	1	0	0	0	0	0
		Glx	1	0	0	0	0	0	0	1	0	0	0
54	14	Leu	4	3	0	0	0	0	0	0	0	1	0
		Arg	9	0	2	2	2	1	0	1	1	0	0
		Gln	1	0	0	0	0	0	0	0	0	0	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
55	16	Pro	5	0	0	0	2	1	0	1	1	0	0
		Ala	6	0	2	4	0	0	0	0	0	0	0
		Gly	1	0	0	0	0	0	0	0	0	0	1
		Asx	1	0	0	0	0	0	0	0	0	1	0
		Glu	3	3	0	0	0	0	0	0	0	0	0
56	16	Ala	1	1	0	0	0	0	0	0	0	0	0
		Thr	5	1	0	4	0	0	0	0	0	0	0
		Ser	10	1	2	0	2	1	0	1	1	1	1
57	16	Thr	1	0	0	0	0	0	0	1	0	0	0
		Gly	15	3	2	4	2	1	0	0	1	1	1
58	16	Ile	6	0	0	3	1	0	0	1	1	0	0
		Val	9	3	2	0	1	1	0	0	0	1	1
		Thr	1	0	0	1	0	0	0	0	0	0	0
59	16	Pro	16	3	2	4	2	1	0	1	1	1	1
60	16	Val	1	0	0	1	0	0	0	0	0	0	0
		Lys	1	0	0	0	0	0	0	0	0	1	0
		Ala	1	0	0	0	0	0	0	0	0	0	1
		Ser	3	3	0	0	0	0	0	0	0	0	0
		Asp	8	0	1	3	2	1	0	0	1	0	0
		Asn	1	0	1	0	0	0	0	0	0	0	0
		Glu	1	0	0	0	0	0	0	1	0	0	0
61	16	Arg	16	3	2	4	2	1	0	1	1	1	1
62	16	Ile	1	0	0	0	1	0	0	0	0	0	0
		Phe	15	3	2	4	1	1	0	1	1	1	1
63	16	Ile	1	1	0	0	0	0	0	0	0	0	0
		Ser	15	2	2	4	2	1	0	1	1	1	1
64	16	Ala	1	0	0	0	1	0	0	0	0	0	0
		Gly	15	3	2	4	1	1	0	1	1	1	1
65	16	Thr	1	1	0	0	0	0	0	0	0	0	0
		Ser	15	2	2	4	2	1	0	1	1	1	1
66	16	Lys	3	0	0	0	2	1	0	0	0	0	0
		Arg	1	0	0	0	0	0	0	0	0	1	0
		Ser	2	0	0	0	0	0	0	1	1	0	0
		Gly	10	3	2	4	0	0	0	0	0	0	1
67	16	Phe	1	1	0	0	0	0	0	0	0	0	0
		Ser	15	2	2	4	2	1	0	1	1	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
68	16	Gly	15	3	2	4	2	0	0	1	1	1	1
		Asn	1	0	0	0	0	1	0	0	0	0	0
69	16	Ala	2	0	0	1	0	0	0	1	0	0	0
		Thr	11	3	2	3	2	0	0	0	0	0	1
		Sei	1	0	0	0	0	0	0	0	0	1	0
		His	1	0	0	0	0	0	0	0	1	0	0
		Asp	1	0	0	0	0	1	0	0	0	0	0
70	16	Thr	3	0	0	0	0	1	0	1	1	0	0
		Ser	2	0	0	0	2	0	0	0	0	0	0
		Asp	9	2	1	4	0	0	0	0	0	1	1
		Asx	1	0	1	0	0	0	0	0	0	0	0
		Glu	1	1	0	0	0	0	0	0	0	0	0
71	16	Tyr	1	0	0	0	0	0	0	0	0	1	0
		Phe	10	3	2	4	0	0	0	0	0	0	1
		Ala	5	0	0	0	2	1	0	1	1	0	0
72	16	Thr	11	3	2	4	1	0	0	1	0	0	0
		Ser	5	0	0	0	1	1	0	0	1	1	1
73	16	Phe	2	2	0	0	0	0	0	0	0	0	0
		Leu	14	1	2	4	2	1	0	1	1	1	1
74	16	Lys	1	0	1	0	0	0	0	0	0	0	0
		Ala	1	0	0	0	1	0	0	0	0	0	0
		Thr	12	3	1	4	0	1	0	1	1	1	0
		Gly	1	0	0	0	1	0	0	0	0	0	0
		Asn	1	0	0	0	0	0	0	0	0	0	1
75	16	Ile	15	3	2	4	2	0	0	1	1	1	1
		Val	1	0	0	0	0	1	0	0	0	0	0
76	16	Thr	2	0	0	0	1	0	0	0	1	0	0
		Ser	13	3	2	4	1	1	0	1	0	1	0
		His	1	0	0	0	0	0	0	0	0	0	1
77	17	Pro	1	0	0	0	0	0	0	0	0	0	1
		Arg	6	0	2	4	0	0	0	0	0	0	0
		Ser	4	3	0	0	0	0	0	0	0	1	0
		Gly	6	1	0	0	2	1	0	1	1	0	0
78	17	Leu	13	4	1	4	2	1	0	0	0	1	0
		Val	1	0	1	0	0	0	0	0	0	0	0
		Met	1	0	0	0	0	0	0	0	0	0	1
		Ala	2	0	0	0	0	0	0	1	1	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
79	17	Arg	3	0	0	0	2	1	0	0	0	0	0
		Glu	5	0	1	3	0	0	0	0	0	1	0
		Gln	6	4	0	0	0	0	0	1	1	0	0
		Glx	3	0	1	1	0	0	0	0	0	0	1
80	17	Pro	9	4	1	4	0	0	0	0	0	0	0
		Ala	3	0	1	0	0	1	0	0	1	0	0
		Thr	1	0	0	0	1	0	0	0	0	0	0
		Ser	3	0	0	0	1	0	0	1	0	1	0
		Glx	1	0	0	0	0	0	0	0	0	0	1
81	17	Val	1	0	0	0	0	0	0	1	0	0	0
		Ala	1	1	0	0	0	0	0	0	0	0	0
		Gly	1	0	0	0	1	0	0	0	0	0	0
		Asp	1	1	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	0	0	0	0	0	0	1
		Glu	10	2	1	4	1	0	0	0	1	1	0
		Glx	2	0	1	0	0	1	0	0	0	0	0
82	17	Asp	14	4	1	4	2	0	0	1	1	1	0
		Asx	3	0	1	0	0	1	0	0	0	0	1
83	18	Ile	2	2	0	0	0	0	0	0	0	0	0
		Phe	8	2	1	4	0	0	0	0	0	1	0
		Val	1	0	1	0	0	0	0	0	0	0	0
		Thr	1	0	0	0	0	0	0	0	0	0	1
		Glu	5	0	0	0	2	0	0	2	1	0	0
		Glx	1	0	0	0	0	1	0	0	0	0	0
84	18	Val	1	0	0	0	0	0	0	0	0	1	0
		Ala	16	4	1	4	2	1	0	2	1	0	1
		Gly	1	0	1	0	0	0	0	0	0	0	0
85	18	Val	6	0	2	4	0	0	0	0	0	0	0
		Met	1	0	0	0	0	0	0	0	0	0	1
		Thr	4	4	0	0	0	0	0	0	0	0	0
		His	1	0	0	0	1	0	0	0	0	0	0
		Asp	5	0	0	0	1	0	0	2	1	1	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
86	20	Tyr	20	6	2	4	2	1	0	2	1	1	1
87	20	Tyr	16	6	2	3	2	1	0	1	1	0	0
		Phe	3	0	0	1	0	0	0	1	0	0	1
		His	1	0	0	0	1	0	0	0	0	0	0
88	21	Cys	21	6	2	4	3	1	0	2	1	1	1

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
89	22	Leu	1	0	0	0	0	0	0	0	0	1	0
		Met	1	0	1	0	0	0	0	0	0	0	0
		Ala	2	0	0	0	2	0	0	0	0	0	0
		Ser	1	0	0	0	0	1	0	0	0	0	0
		Asn	1	0	0	0	0	0	0	0	1	0	0
		Gln	15	6	2	4	1	0	0	2	0	0	0
		Glx	1	0	0	0	0	0	0	0	0	0	1
90	22	Met	1	0	1	0	0	0	0	0	0	0	0
		Ala	3	0	0	0	2	0	0	1	0	0	0
		Thr	2	0	0	0	1	0	0	1	0	0	0
		Ser	2	0	0	0	0	1	0	0	1	0	0
		Gln	13	6	2	4	0	0	0	0	0	1	0
		Glx	1	0	0	0	0	0	0	0	0	0	1
91	22	Trp	5	0	0	0	3	0	0	2	0	0	0
		Tyr	12	5	1	4	0	1	0	0	0	1	0
		Phe	1	1	0	0	0	0	0	0	0	0	0
		Ala	1	0	1	0	0	0	0	0	0	0	0
		Arg	2	0	1	0	0	0	0	0	1	0	0
		Ser	1	0	0	0	0	0	0	0	0	0	1
92	21	Leu	2	0	2	0	0	0	0	0	0	0	0
		Val	1	0	0	0	0	1	0	0	0	0	0
		Lys	1	0	0	0	0	0	0	0	0	0	1
		Ala	1	0	0	0	0	0	0	0	0	1	0
		Gly	3	0	0	3	0	0	0	0	0	0	0
		Asp	9	4	0	0	2	0	0	2	1	0	0
		Asn	1	1	0	0	0	0	0	0	0	0	0
		Glu	3	1	1	1	0	0	0	0	0	0	0
93	21	Tyr	1	0	0	0	1	0	0	0	0	0	0
		Thr	4	1	1	1	0	0	0	1	0	0	0
		Ser	7	1	0	2	1	0	0	1	1	1	0
		Gly	1	0	0	1	0	0	0	0	0	0	0
		His	1	1	0	0	0	0	0	0	0	0	0
		Asp	1	1	0	0	0	0	0	0	0	0	0
		Asn	2	2	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
		Glu	2	0	1	0	0	0	0	0	0	0	1
		Gln	1	0	1	0	0	0	0	0	0	0	0
94	21	Ile	2	0	1	0	0	0	0	1	0	0	0
		Leu	5	5	0	0	0	0	0	0	0	0	0
		Val	1	0	0	0	0	0	0	0	0	0	1
		Met	1	0	0	0	0	0	0	1	0	0	0
		Ala	1	0	1	0	0	0	0	0	0	0	0
		Arg	1	0	0	0	1	0	0	0	0	0	0
		Ser	8	0	1	4	1	0	0	0	1	1	0
		Asp	1	1	0	0	0	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0

TABLE I—*Continued*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
95	21	Pro	14	5	3	4	0	0	0	0	0	1	1
		Leu	2	0	0	0	2	0	0	0	0	0	0
		Thr	1	0	0	0	0	0	0	1	0	0	0
		Ser	2	1	0	0	0	0	0	1	0	0	0
		Gly	1	0	0	0	0	0	0	0	1	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
96	21	Trp	2	0	0	0	0	0	0	0	0	1	1
		Ile	1	0	1	0	0	0	0	0	0	0	0
		Tyr	2	1	1	0	0	0	0	0	0	0	0
		Phe	2	1	0	1	0	0	0	0	0	0	0
		Pro	1	1	0	0	0	0	0	0	0	0	0
		Leu	2	1	1	0	0	0	0	0	0	0	0
		Lys	2	1	0	0	0	0	0	0	1	0	0
		Arg	1	1	0	0	0	0	0	0	0	0	0
		Thr	1	0	0	1	0	0	0	0	0	0	0
		Ser	2	0	0	1	1	0	0	0	0	0	0
		Asn	1	0	0	0	1	0	0	0	0	0	0
		Asx	1	0	0	0	0	1	0	0	0	0	0
		Gln	1	0	0	1	0	0	0	0	0	0	0
		---	2	0	0	0	0	0	0	2	0	0	0
97	20	Phe	1	0	0	0	0	1	0	0	0	0	0
		Pro	1	0	0	1	0	0	0	0	0	0	0
		Met	1	1	0	0	0	0	0	0	0	0	0
		Ala	2	0	0	0	2	0	0	0	0	0	0
		Thr	12	4	3	3	0	0	0	0	0	1	1
		His	1	0	0	0	0	0	0	0	1	0	0
		---	2	0	0	0	0	0	0	2	0	0	0
a	19	Val	4	0	0	0	2	0	0	1	1	0	0
		Ala	1	0	0	0	0	0	0	1	0	0	0
		---	14	5	3	4	0	0	0	0	0	1	1
b	20	Ile	1	0	0	0	0	0	0	1	0	0	0
		Leu	1	0	0	0	0	0	0	0	1	0	0
		Val	4	0	0	0	2	1	0	1	0	0	0
		---	14	5	3	4	0	0	0	0	0	1	1
98	20	Phe	20	5	3	4	2	1	0	2	1	1	1
99	20	Gly	20	5	3	4	2	1	0	2	1	1	1
100	20	Pro	1	1	0	0	0	0	0	0	0	0	0
		Gly	11	2	1	0	2	1	0	2	1	1	1
		Gln	8	2	2	4	0	0	0	0	0	0	0
101	19	Gly	19	4	3	4	2	1	0	2	1	1	1

TABLE I—*Concluded*

Position	No. of Protein Sequences Studied	Amino Acids	Total	Human Kappa			Human Lambda					Mouse Kappa	
				I	II	III	I	II	III	IV	V	I	II
102	19	Thr	18	4	3	3	2	1	0	2	1	1	1
		Ser	1	0	0	1	0	0	0	0	0	0	0
103	19	Lys	14	4	1	3	1	1	0	1	1	1	1
		Arg	3	0	1	1	0	0	0	1	0	0	0
		Asn	1	0	1	0	0	0	0	0	0	0	0
		Gln	1	0	0	0	1	0	0	0	0	0	0
104	22	Leu	15	3	2	3	1	1	0	2	1	1	1
		Val	7	3	2	1	1	0	0	0	0	0	0
105	22	Thr	6	0	0	0	2	1	0	2	1	0	0
		Asp	4	3	0	1	0	0	0	0	0	0	0
		Glu	12	3	4	3	0	0	0	0	0	1	1
106	22	Ile	12	3	4	3	0	0	0	0	0	1	1
		Phe	1	1	0	0	0	0	0	0	0	0	0
		Leu	2	1	0	1	0	0	0	0	0	0	0
		Val	7	1	0	0	2	1	0	2	1	0	0
a	22	Leu	6	0	0	0	2	1	0	2	1	0	0
		---	16	6	4	4	0	0	0	0	0	1	1
107	22	Lys	15	6	3	4	0	0	0	0	0	1	1
		Arg	3	0	1	0	1	1	0	0	0	0	0
		Ser	2	0	0	0	0	0	0	2	0	0	0
		Gly	2	0	0	0	1	0	0	0	1	0	0

studied at the given position. Only data for which the sequence has been clearly assigned by the various authors have been included.

The Role of Glycine—It has been suggested that glycine plays a unique role in the structure of the variable region of immunoglobulin light chains (18, p. 87; 41–43, 45). Jukes (50) and Welscher (51) have generally agreed with this. A further careful analysis becomes essential for the understanding of the function of the glycines in the over-all structure and in relation to antibody-combining sites.

The basic property that differentiates glycine from all other amino acids structurally is the absence of a side chain. As a result, glycine can have many sterically allowable configurations. This has been verified experimentally in the case of lysozyme (46, 52) and tosyl-α-chymotrypsin (53). The two angles, Φ and ψ (54), which specify the conformation of the backbone of an amino acid have been calculated for each of the amino acids from the known tertiary structures of lysozyme (46, 52) and of tosyl-α-chymotrypsin (53). A typical plot of the permissible angles of the glycine as compared with the alanine residues is shown in Fig. 1. The allowable configurations of alanine are mostly

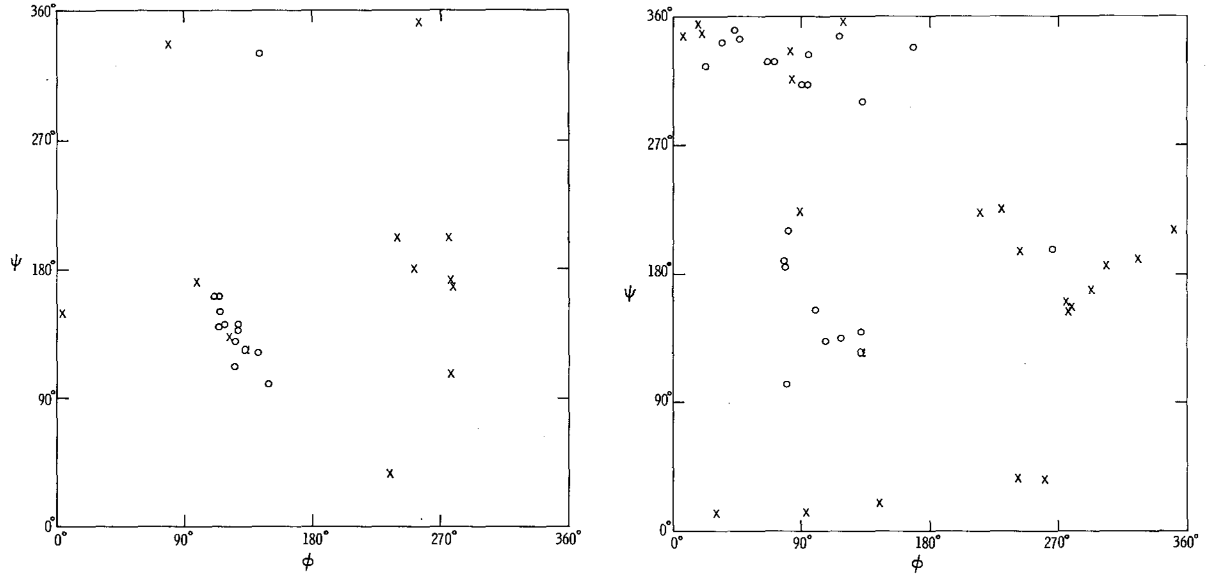

FIG. 1 *a*

FIG. 1 *b*

FIG. 1. Comparison of ϕ and ψ for glycine (X) and alanine (O) based on their occurrence in (*a*) lysozyme (data from references 46, 52), and in (*b*) tosyl-α-chymotrypsin (calculated from data given in reference 53). α indicates the values of ϕ and ψ for the α-helix.

clustered near the α-helical region of lysozyme (Fig. 1 *a* and reference 55), while those of glycine are widely distributed. Comparison with similar maps for other amino acids in lysozyme also shows them to be more restricted. The data for tosyl-α-chymotrypsin also show that glycine may have many more conformations (Fig. 1 *b*). This unique property of glycine thus may permit relative motion of the chains attached to the two ends of the molecule. With immuno-

TABLE II

Frequencies of Glycine Residues at Various Positions in the Variable Region of Light Chains

Position	Human Kappa			Human Lambda					Mouse Kappa		Total	%
	I	II	III	I	II	III	IV	V	I	II		
9			10/15								10/63	16
13				6/9	2/2	6/6					14/61	23
16	20/21	4/4	13/13	9/9	2/2	6/6	2/2	1/1	1/1	2/2	60/61	99
24							1/3				1/26	4
25				4/4	2/2		3/3	1/1			10/25	40
26				2/3							2/24	8
27f				1/3							1/22	5
28		1/2			1/1						2/22	9
29				2/2			1/2			1/2	4/21	19
30		2/2	1/4					1/1	1/1		5/21	24
39	1/3										1/16	6
41	2/3	2/2	4/4	2/2	1/1		1/1	1/1	1/1	1/1	15/16	94
50			1/2		1/1			1/1			3/14	21
51		1/2									1/14	7
55										1/1	1/16	6
57	3/3	2/2	4/4	2/2	1/1			1/1	1/1	1/1	15/16	94
64	3/3	2/2	4/4	1/2	1/1		1/1	1/1	1/1	1/1	15/16	94
66	3/3	2/2	4/4							1/1	10/16	62
68	3/3	2/2	4/4	2/2			1/1	1/1	1/1	1/1	15/16	94
74				1/2							1/16	6
77	1/4			2/2	1/1		1/1	1/1			6/17	35
81				1/2							1/17	6
84		1/2									1/18	6

TABLE II—*Concluded*

Position	Human Kappa			Human Lambda					Mouse Kappa		Total	%
	I	II	III	I	II	III	IV	V	I	II		
92			3/4								3/21	14
93			1/4								1/21	5
95								1/1			1/21	5
99	5/5	3/3	4/4	2/2	1/1		2/2	1/1	1/1	1/1	20/20	100
100	2/5	1/3		2/2	1/1		2/2	1/1	1/1	1/1	11/20	55
101	4/4	3/3	4/4	2/2	1/1		2/2	1/1	1/1	1/1	19/19	100
107				1/2				1/1			2/18	11

Fractions represent the number of instances in which glycine occurs to the total number of proteins studied at the given position for each subgroup. When values have not been given, glycine has not been reported at that position in the subgroup.

globulins, flexibility of the protein backbone can be one of the major factors that permit substitution of various amino acids at the variable positions; it also may allow movement of the site to make most favorable contact in combining with an antigenic determinant. Though a glycine residue confers maximum flexibility over all other amino acids, it might also arise from a random mutation in which the difference between glycine and other amino acids is not adverse for the over-all structure. In addition, some glycines might be complementarity determining. These latter two kinds of glycines must be distinguished from the first.

For the variable region of the light chains of human and mouse immunoglobulins and of Bence Jones proteins, alignment of amino acid sequences serves to identify the glycines which may be conferring flexibility as shown in Table II. All the glycines are listed. The frequencies, expressed as per cent, can roughly be divided into three categories:

A. 94–100%: Positions 16, 41 (or 39), 57, 64, 68, 99, and 101. Since glycine occurs at these positions in nearly all the proteins studied, it must have a fundamental structural significance and has been preserved in the evolution to man and mouse. These glycines are assumed to confer flexibility unique to antibodies.

It is of interest that glycine occurs at position 41 in 15 of 16 proteins studied (94%). The sequence at residues 39, 40, and 41 is Lys-Pro-Gly in 14 cases and Lys-Ala-Gly in one case. In the single exception, a human κI protein Ag, the sequence at 39, 40, and 41 is Gly-Pro-Lys.[1] Thus if the reported sequence is correct, the glycine at residue 39 might well serve the same function as the glycine at position 41. The two positions 41 and 39 may thus provide one invariant glycine (100%).

B. 35–62%: Positions 25, 66, 77, and 100. A careful examination indicates that glycines at these positions are at least group specific. Thus, within a group (κ or λ), they could serve the same function as the glycines of category *A*.

C. 4–24%: Positions 9, 13, 24, 26, 27f, 28, 29, 30, 50, 51, 55, 74, 81, 84, 92, 93, 95, and 107. These glycine residues are at variable positions. They therefore play a distinctly different role from those of the other two categories. They might either be related to antibody complementarity or, if not involved in the site itself, could have arisen from random mutation and be nevertheless compatible with three-dimensional folding.

Thus there are about 8 (human κ and mouse κ) to 10 (human λ) glycines in the variable region of the light chains of all proteins for which sequence data are available at these positions. They are as follows:

Human κ: Positions	16,		41 (or 39),	57,	64,	66,	68,		99,		and 101.
Human λ: Positions	16,	25,	41,	57,	64,		68,	77,	99,	100,	and 101.
Mouse κ: Positions	16,		41,	57,	64,		68,		99,	100,	and 101.

Positions 99, 100, and 101 have been postulated to function as a pivot permitting the combining regions of the light and heavy chains to make most favorable contact with the antigenic determinant (18, p. 87; and 41, 43). Examination of the heavy chain sequences reported to date (56, 57) shows Gly-Gln-Gly at positions 112, 113, and 114 in two instances (He and Daw), Gly-Arg-Gly in one (Cor) and Gly-Gly at positions 114 and 115 in Eu; but in the latter protein two gaps have been placed at positions 108 and 109 in aligning with He for maximum homology. Thus these glycines could also be functionally and positionally equivalent.

Invariant Residues—Earlier comparisons of the invariant residues of the variable and constant regions were based entirely on those positions at which only a single amino acid occurred. As more data accumulated this number diminished, until there are now only 11 such positions in the variable region: Gln 6, Cys 23, Trp 35, Pro 59, Arg 61, Asp (Asx) 82, Tyr 86, Cys 88, Phe 98, Gly 99, and Gly 101. However, if one accepts as essentially invariant those positions at which more than 88–90% of the proteins studied have the same amino acid at a given position, this number increases to 29. A comparison of these residues with those in the constant region is given in Table III. This procedure allows for possible errors as well as for the ability of some residues to substitute for others at a given position. The difference between the con-

[1] The sequence of Roy was originally reported as Gly-Pro-Lys but has been changed to Lys-Pro-Gly (see references to Roy in sequence data).

stant and variable regions is not much different than originally appeared (41) except that there is one invariant alanine and one invariant leucine in the variable region. However, the difference between the two regions is still quite clear, the variable region having in addition no invariant valine, lysine, and histidine, while the invariant residues in the constant region include three alanines, three leucines, three valines, two lysines, and two histidines.

TABLE III

Comparison of the Invariant Residues of the Variable with those of the Constant Region in Human κ-, Human λ-, and Mouse κ-Bence Jones Proteins

Amino Acid	Variable Region	Constant Region
Gly	7*	0
Ala	1	3
Leu	1	3
Val	0	3
Lys	0	2
His	0	2
Ile	1	0
Ser	3	5
Glu	0	2
Gln	2	0
Arg	1	0
Pro	3	4
Tyr	2	3
Cys	2	3
Phe	2	2
Trp	1	1
Thr	2	1
Asp	1	1
Total	29	35

* Not including positions 25, 66, 77, and 100.

Invariant residues are those in which 88–90% of the proteins analyzed contain the same amino acid residue at a given position.

Hydrophobicity Distribution of the Invariant Residues of the Variable Region—A parameter, $H\phi_{ave}$, based on the free energies of transfer of amino acid side chains from an organic to an aqueous environment has been introduced by Tanford (58) and applied by Bigelow (59) to the study of various proteins. $H\phi_{ave}$ is expressed in kilocalories per residue and varies from 3.00 for Trp to 0.45 for Thr and is very small, zero, or negative for Gly, Ser, His, Asp, Glu, Asn, and Gln; these have been taken as zero. These values have been used in an examination of the invariant residues of the variable region. Table IV summarizes the findings and tabulates HØ$_{\text{ave}}$ for the invariant residues and for the occasional other substituents found. In addition we have included data on two positions, 46 and 48, in which another substituent occurred with slightly lower frequency than that required by the definition of an invariant residue, and on three other positions 8, 12, and 37 in which the other substituents were confined to a single subgroup.

TABLE IV

Hydrophobicity, HØ$_{\text{ave}}$, of the Invariant Residues and almost Invariant Residues of the Variable Region

Position	Amino Acids		Hydrophobicity, HØ$_{\text{ave}}$	
	Invariant	Other	Invariant	Other
5	Thr 67*	Ala 2, Ser 1	0.45	0.75, 0.00
6	Gln 63		0.00	
16	Gly 60	Arg 1	0.00	0.75
23	Cys 30		1.00	
35	Trp 17		3.00	
38	Gln 12, Glx 3	His 1	0.00	0.00
40	Pro 15	Ala 1	2.60	0.75
41(39)	Gly 15	Lys 1	0.00	1.50
44	Pro 15	Ile 1	2.60	2.95
49	Tyr 13	Phe 1	2.85	2.65
57	Gly 15	Thr 1	0.00	0.45
59	Pro 16		2.60	
61	Arg 16		0.75	
62	Phe 15	Ile 1	2.65	2.95
63	Ser 15	Ile 1	0.00	2.95
64	Gly 15	Ala 1	0.00	0.75
65	Ser 15	Thr 1	0.00	0.45
67	Ser 15	Phe 1	0 00	2.65
68	Gly 15	Asn 1	0.00	0.00
73	Leu 14	Phe 2	2.40	2.65
75	Ile 15	Val 1	2.95	1.70
82	Asp 14, Asx 3		0.00	
84	Ala 16	Gly 1, Val 1	0.75	0.00, 1.70
86	Tyr 20		2.85	
88	Cys 21		1.00	
98	Phe 20		2.65	
99	Gly 20		0.00	
101	Gly 19		0.00	
102	Thr 18	Ser 1	0.45	0.00

TABLE IV—*Concluded*

Borderline for almost invariant.

Position	Amino Acids		Hydrophobicity, $H\emptyset_{ave}$	
	Invariant	Other	Invariant	Other
46	Leu 12	Ile 1, Arg 1	2.40	2.95, 0 75
48	Ile 12	Met 2	2.95	1.30

Invariant with a subgroup exception.

Position	Amino Acids		Hydrophobicity, $H\emptyset_{ave}$	
	Invariant	Other	Invariant	Other
8	Pro 58	Ala 6 (Human λ III)	2.60	0.75
12	Ser 56	Pro 4 (Human κ II)	0.00	2.60
		Ala 1 (Mouse κ II)		0.75
37	Glu 12, Glx 2	Pro 2 (Human κ II)	0.00	2.60

* Numbers next to the residue represent the number of samples in which the residue occurred.

It is of interest that 15 of the invariant residues have values of 0.00 or 0.45 and that 10 have values of 2.40–3.00, leaving only 4 residues of intermediate hydrophobicity: 2 half-cystines, 1 arginine, and 1 alanine. In most instances the other substituents reported as replacements at these positions generally had $H\emptyset_{ave}$ values not too different from the major substituent, but at positions 40, 41, 63, 67, 75, and 84 changes of over one unit were seen. (Position 41 may not be significant since, as discussed earlier, the exception had glycine in position 39.) The two borderline invariant residues both showed substantial $H\emptyset_{ave}$ differences and the three subgroup specific residues also varied substantially in $H\emptyset_{ave}$.

Of the 35 invariant residues in the constant region, 11 have values of 0.00 or 0.45, 13 have values of 2.40 to 3.00, and 11 have values of 0.75 to 1.70 (Table V). Thus the constant region has invariant residues which appear to be relatively uniformly distributed with respect to $H\emptyset_{ave}$ while in the variable region they are generally either very high or zero.

The average hydrophobicity for invariant residues of the variable region is about 1.09 while that of invariant residues of the constant region is 1.39.

Welcher (51) has computed $H\emptyset_{ave}$ for the entire light chain from sequence data on Bence Jones proteins and obtains values ranging from 0.970 to 1.04 kcal/residue, values in the same range as those for chains of other proteins. The variable regions range from 0.930 to 1.11 while the constant region values were

0.950 for mouse κ, 0.970 for human κ, and 1.02 for human λ. Thus the over-all hydrophobicities of the two regions are no different; Welcher also reported no difference between the two regions with respect to the pattern of nonpolar positions, while our data show a substantial difference in $H\emptyset_{ave}$ of the invariant residues of the two regions. Moreover, the average hydrophobicity of the entire constant region is about 0.98, significantly lower than 1.39 for the hydrophobicity of its invariant residues.

TABLE V

Hydrophobicity, $H\emptyset_{ave}$, of the Invariant Residues of the Constant Region

Position	Amino Acid	Hydrophobicity $H\phi_{ave}$	Position	Amino Acid	Hydrophobicity $H\phi_{ave}$
111	Ala	0.75	149	Lys	1.50
112	Ala	0.75	151	Asp	0.00
113	Pro	2.60	168	Ser	0.00
115	Val	1.70	173	Tyr	2.85
118	Phe	2.65	176	Ser	0.00
119	Pro	2.60	177	Ser	0.00
120	Pro	2.60	179	Leu	2.40
121	Ser	0.00	181	Leu	2.40
123	Glu	0.00	189	His	0.00
125	Leu	2.40	192	Tyr	2.85
130	Ala	0.75	194	Cys	1.00
133	Val	1.70	197	Thr	0.45
134	Cys	1.00	198	His	0.00
139	Phe	2.65	203	Ser	0.00
140	Tyr	2.85	207	Lys	1.50
141	Pro	2.60	213	Glu	0.00
146	Val	1.70	214	Cys	1.00
148	Trp	3.00			

Variability—In considering the nature of the variable region, it is of importance to ascertain whether the variability is uniformly distributed or is confined to small segments of the variable regions. Thus, a quantity is defined for each amino acid position in the sequence

$$\text{Variability} = \frac{\text{Number of different amino acids at a given position}}{\text{Frequency of the most common amino acid at that position}} \quad [1]$$

in which the denominator is the number of times the most common amino acid occurs divided by the total number of proteins examined. Thus at position 7 (Table I) 63 proteins were studied, serine occurred 41 times and 4 different amino acids, Pro, Thr, Ser, and Asp, have been reported. The frequency of the most common is 41/63 = 0.65 and the variability is then 4/0.65 = 6.15. When there was uncertainty as to the number of amino acids, as in instances in which Glx or Asx has been reported, the extreme values of variability have been computed. For this equation an absolutely invariant residue would have a value of 1 while the theoretical upper limit for 20 amino acids randomly occurring would be 400.

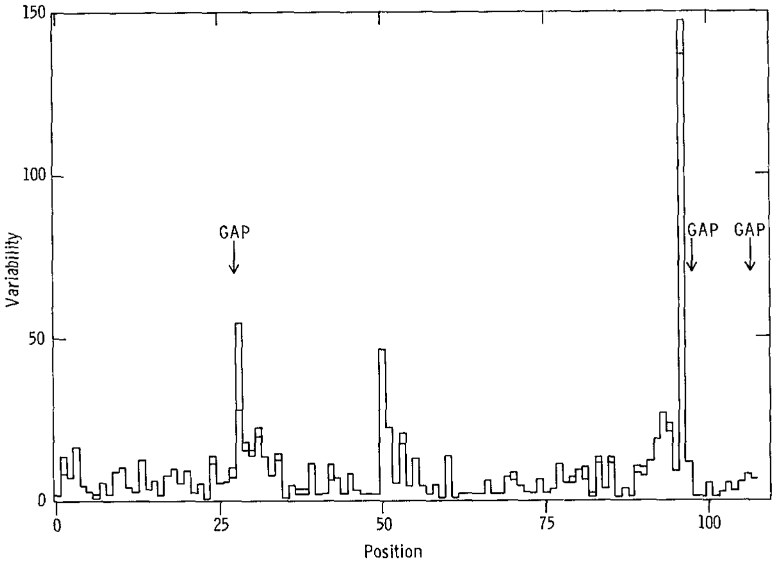

FIG. 2. Variability at different amino acid positions for the variable region of the light chains. GAP indicates positions at which insertions have been found.

Plotting variability against position for the 107 residues of the variable region (Fig. 2) shows three main peaks in the regions of residues 28, 50, and 96; two of these, 28 and 96, are the highly variable regions (7, 45, 47, 49) in which insertions occur, while position 50 has not been associated with an insertion. Franěk (49) has previously noted the high variability around position 50. The stretches of amino acid residues showing this high variability are 24–34, 50–56, and 89–97. The first and third regions begin after an invariant Cys and are followed by an invariant Trp and Phe respectively, at positions 35 and 98. The second region begins after an almost invariant position 49 (Tyr 13/14, Phe 1/14) and is followed by an invariant position 57 (Gly 15/16, Thr 1/16) (Table I).

The over-all sequence data were also examined to ascertain whether amino acid substitutions at each position were reflected in changes in hydrophobicity confined to certain stretches of the variable region. The findings which are not plotted showed that the same three regions associated with high variability were those with the greatest variation in $H\O_{ave}$.

Since a portion of the variability at many positions is group- or subgroup-specific and is therefore generally not complementarity determining, variability as defined in equation [1] was computed for the individual subgroups for which sufficient sequence data were available. A plot for κI showed high variability in the stretches 24–34 and 92–96 and at residues 53 and 56. κIII shows high variability at residue 96. Data are generally insufficient even for these two subgroups and the data for other subgroups do not permit such an analysis; the λI subgroup showed unusually high variability at position 18.

Classification of Variability at Individual Positions—The sequence data available at each position (Table I) were examined in an attempt to classify the position with respect to its role in the over-all structure. Invariant residues already considered are not included. The following categories were set up. (*a*) Invariant except for one subgroup. (*b*) κ- vs. λ-specific. (*c*) κ- vs. λ-specific with some subgroup variations. (*d*) Variation in κ- and λ-subgroups. (*e*) Variation in κ-subgroups with λ relatively constant. (*f*) Variation in λ-subgroups with κ relatively constant. (*g*) Unaccountable variability. (*h*) Insufficient data. (*i*) Possible species specificity. It has not been possible to set up absolute criteria for each of these classes and some difficulties were encountered in making these assignments, especially at positions for which the sequence data were relatively sparse. Occasional substitutions compatible with point mutations have often been neglected.

Examinations of each of these categories permit some interesting inferences to be made:

(*a*) There are five positions, 8, 12, 15, 37, and 54, which are essentially invariant except that in one subgroup another amino acid occurs. Thus at position 8, 58 proteins have Pro while 6, all of which belong to the λIII subgroup, contain Ala. At the other positions, additional amino acids sometimes occur in individual proteins. These positions are considered to be part of the basic skeleton of the variable region under the control of structural genes, the subgroup differences being the result of permissible mutations.

(*b*) Six positions, 7, 33, 71, 83, 105, and 106, show predominantly κ- vs. λ-specificity. Thus at position 7, 41 human κ- and mouse κ-proteins have Ser while 20 human λ-specimens have Pro. Two exceptions occur, a κII protein with Thr instead of Ser, and a λV with Asp instead of Pro.

(*c*) At five additional positions, 1, 2, 13, 25, and 27, evidence of κ- vs. λ-specificity persists but a substitution may occur in one or more subgroups. Thus at position 1, 16 human λ-proteins of subgroups λI, λII, λIII have PCA, while λIV and λV have no amino acid, and 31 human and mouse κI and κII proteins have Asp (5 additional have Asx) while 14 human κIII specimens have Glu

(1 additional with Glx). There are several exceptions: one human κIII with Lys and one with Asp and one mouse κII with PCA.

The 11 positions in categories *b* and *c* are of importance because they show that κ- and λ-specificity is not exclusively a property of the constant region but involves the variable region as well. These findings are consistent with the data indicating that the variable and constant regions of κ-chains always go together as do the variable and constant regions of λ-chains. They are also in accord with the immunochemical studies of Ruffilli and Baglioni (60) who showed that λ-specificity extended into the variable region, and with those of Ruffilli (61) that determinants in the variable region of κ cross-react with anti λ-sera. They also suggest a continuous evolutionary association for the variable and constant regions of the κ-chains as well as for the variable and constant regions of the λ-chains. The regions showing κ- and λ-specificity appear to be distributed at the beginning and end of the variable region.

(*d–f*) These three categories are instances of variation in composition ascribable to subgroup variation. At 9 positions, 3, 14, 18, 19, 22, 29, 42, 51, and 79, variation ascribable to subgroups is seen in both κ and λ chains. In an additional 10 positions, 4, 9, 10, 17, 20, 21, 55, 56, 77, and 85, subgroup variation is predominantly in κ-chains with λ relatively constant, and at 15 positions, 11, 26, 39, 47, 52, 66, 69, 72, 76, 78, 80, 89, 95, 97, and 107, the subgroup variation seems to involve λ-chains with κ relatively constant. The larger number of residues placed in category *f* is probably a consequence of the classification of λ-chains into five subgroups while κ-chains are only divided into three subgroups. Moreover the number of samples of λ-chains is fewer than for κ-chains so that other types of variation may be masked.

(*g*) Unaccountable variability. At 8 positions, 28, 30–32, 93, 94, 96, and 103, the variation within each subgroup appears to be greater than can be accounted for on any known basis, especially for those subgroups for which a sufficient number of samples have been examined. It is of interest that except for residue 103 these positions are clustered in the two regions of highest variability (45) and would be brought into close proximity by the disulfide bond I_{23}–II_{88}. It is postulated that these residues may be complementarity determining and actually be involved in making contact with the antigenic determinant. These residues are also very close to the position at which insertions occur (Table I).

(*h*) Insufficient data. At 15 positions, 24, 34, 36, 43, 45, 53, 58, 70, 74, 81, 87, 90–92, and 104, not enough sequences are available to assign them clearly to one of the other categories. Five of the residues, 24, 34, 90–92, occur in the two regions of highest variability close to most of those with unaccountable variability. The others are fairly well spread and only one, 53, occurs in the third highly variable region. Position 18, although placed in group *d* shows an extraordinary variability in the λI subgroup.

(*i*) Species-specific residues. At positions 50 and 60 the two mouse Bence

Jones proteins examined have residues which do not correspond to any reported for the human proteins. Thus these two positions are still classifiable as species-specific. Position 96, with its extraordinary variability, although placed in category *g* might also technically be called species specific since the two mouse samples both contain Trp which is absent in the human samples. At all three positions the human proteins show substantial variability in the number of substitutions which can occur—9 at position 50, 5 at position 60, and 11 at position 96. It thus seems very unlikely that the apparent species specificity of these positions will persist when more mouse sequences are available.

DISCUSSION

The ability to subject the large amount of sequence data on human and mouse Bence Jones proteins and light chains to a statistical analysis has supported the earlier conclusion about the lack of species specificity in the variable regions of human and mouse Bence Jones proteins (40). The data now available indicate at most 2 or 3 such residues in the variable region as compared with 36 in the constant region. Data in other species, although limited to the first few amino terminal residues, tend to support this. The rabbit has always been considered an important exception to the view that there was little if any species specificity in the variable region, since rabbit light chains had Ala and Ile as N-terminal residues. However, the demonstration by Hood et al. (62) that the N-terminal sequence of a homogeneous rabbit antibody to the C carbohydrate of the streptococcus was Ala-Asp-Val-Val-Met-Thr-Glu-Thr-Pro-Ala-Ser-Val indicates that the rabbit has merely added an N-terminal Ala to the N-terminal Asp, the other residues then being essentially similar to those in human light chains. Thus the rabbit is not an exception to the basic evolutionary unity of light chains.

The data now available (Table II) amply support the earlier suggestions for the role of the invariant glycines of the variable region both in conferring flexibility (Fig. 1) to permit substitutions at the variable positions, and for the glycines at positions 99 and 101, together with the frequently found glycine at residue 100, in functioning as a pivot to permit optimal fitting around the antigenic determinant (18, p. 87; and 41, 43). This postulate is now strongly supported by the finding of two glycines in an analogous region of the heavy chain. The heavy chain sequences thus far available also indicate that there may be other invariant glycines in the variable region (56, 57).

The data in Table IV show an unusual distribution of invariant residues in the variable region with respect to $H\emptyset_{ave}$ in that, with a few exceptions, these residues have either a high or a low or zero value, while $H\emptyset_{ave}$ values appear to be uniformly distributed throughout the invariant residues of the constant region (Table V). The significance and structural implications of this are not clear. This finding is not evident when one examines only the over-all hydrophobicity, which does not differ significantly for the two halves of the chain (51).

The variability at each position shows that variability is concentrated in three regions of the molecule (49), of which two have also been noted by other workers (7, 45, 47). Examination of the basis for the variability at each position shows that variation ascribable to κ and λ or to their subgroups does not readily explain all of the variation. Seven of the eight positions at which the variability cannot be otherwise accounted for occur in two of these regions. Of the 15 positions for which sufficient sequence data were lacking to permit clear assignment, 5 occur in the same 2 regions—24–34 and 89–97 and those are the 2 regions at which additional residues are found (Table I). Much more sequence data will be required to permit unequivocal elucidation of the role of each variable position.

For the moment, if one accepts (*a*) the tendency of the positions of unaccountably high variability to be concentrated in the two short stretches of the chain at which insertions are found, (*b*) the finding that subgroup and group specificity occurs in the variable region and indeed probably over substantial portions of it, one might formulate the following working hypothesis extending the earlier concept (42, 45): The light chains of immunoglobulins except for the regions of unaccountable variability, 24–34 and 89–97, are governed by a number of structural genes, each chain being the product of two linked genes, one for the variable and one for the constant region (Hood et al. in reference 16). These structural genes are free to mutate and are limited only by the requirement that their product be capable of assuming the proper three-dimensional structure to permit an antibody site to be formed. By hypothesis the complementarity-determining residues are considered to be the result of the insertion into the DNA of the two short linear sequences, 24–34 and 89–97, which specifically determine what kind of antibody site will be formed. In the light chain the two insertions would be brought into close proximity by the disulfide bond I_{23}–II_{88}. A similar type of insertion would be made in the heavy chain, but thus far there is evidence for only one region of high variability (56, 57).

An insertion mechanism involving only short linear sequences provides a substantial simplification of the problem of providing a seemingly limitless number of complementary sites without the use of very large amounts of DNA. It would also be more likely to provide the necessary evolutionary stability and universality for the antibody-forming mechanism which cannot be adequately accounted for by the germ line hypothesis (36) of one gene for each variable region since such a system would diverge on an evolutionary time scale. The precision with which the insertion would have to be accomplished, (e.g. changes in length of one or two nucleotides would result in nonsense) in itself would tend to prevent evolutionary divergence since failures in the insertional mechanism would completely destroy the ability to form any antibody and individuals with such a defect would probably not survive. It eliminates difficulties in the translocation hypothesis which provides for the joining of one of many V genes with a C gene (33) and yet maintains the exclusive association of V_κ with C_κ and V_λ with C_λ genes. Indeed the finding that κ- and λ-specificity extends into the variable region supports the concept of one polypeptide chain resulting from the action of closely linked V and C genes. This also is consistent with the allotypic studies on γG, γM and γA (63–65). Moreover it has the additional merit of providing a fixed location for the antibody site while other theories (32, 34, 35, 39, 66) permit it to be formed by different portions of the variable region for various antibodies. Indeed Eisen (16) considers that three different sites, each of a different specificity, might be formed by a given V_L and V_H pair. This would make antibody sites completely different from all other sets of specific receptors known. The present hypothesis considers the site as involving a small fixed region of the molecule with specificity determined by the differences in side chains of complementarity-determining residues. This insertion hypothesis readily accounts for the variations in length occurring in these regions; no other hypothesis has explicitly considered this or adequately accounted for it.

On the basis of this hypothesis which ascribes only a role in three-dimensional folding to the first 23 N-terminal amino acid residues of the light chain, all recombinational theories of antibody formation (66, 67) become uninformative since they are based exclusively on sequence data in this region and thus are probably not dealing with a region involving antibody complementarity.

The present working hypothesis of linear regions determining antibody complementarity will stand or fall when adequate numbers of sequences for human and mouse light (and heavy) chains have been worked out.

The hypothesis makes certain predictions which also can be used to test its validity. One of these, since the insertion is hypothesized to determine complementarity, is that antibody molecules of a given specificity and with a uniform site can occur in any class or subclass of immunoglobulin by the insertion of the given short linear sequences. The finding (8) that human antidextran of α-(1 → 6) specificity may occur in γA, γM, γG2, and sometimes in γG1 immunoglobulins and may have κ- or λ-chains is consistent with but does not provide conclusive evidence for this hypothesis, since the human antidextran still represents mixtures with heterogeneous sizes of combining sites. More important, however, may be the findings of Pincus et al. (14) that rabbit type III and type VIII antibodies to the pneumococcal polysaccharide, which gave a straight line of slope not significantly different from 1.0 in a Sips plot for binding of an octasaccharide and could thus be considered quite homogeneous with respect to their antibody-combining sites, showed many light chains and several heavy chains on acrylamide gel electrophoresis. Thus a mixture of structurally different antibody molecules could have the same binding affinity and therefore probably have very similar or even identical combining sites. It would be especially important to determine whether such mixtures of antibodies also belonged to different classes and had antigenically different light and heavy chains.

Further studies on homogeneous antibodies of a given specificity but of different classes or subclasses would be predicted to show similar sequences in the insertional regions if their sites were identical. Conversely, antibodies of a given specificity but showing differences in the degree of cross-reactivity with related antigens would be expected to show smaller differences in their insertion regions than would antibodies of totally unrelated specificities. Such data may soon become available as sequences on myeloma proteins with antibody activity and on homogeneous antibodies are accumulated.

While such findings would provide strong support for the hypothesis, it should be borne in mind that the contour of an antibody site having a given specificity or binding affinity for a given determinant could conceivably be formed by several kinds of patterns of amino acid sequences. Under such circumstances, antibodies which were homogeneous in binding affinity but were mixtures of molecules with different sequences in the hypothesized insertional regions would be expected not to give unique sequences in these regions. Moreover, the possibility exists that binding could be influenced to some extent by different residues adjacent to a site but not in themselves complementarity determining.

The data on idiotypic specificity of myeloma globulins (68) and antibodies (69–73) are compatible with the insertion model and with the over-all concept of antibody structure proposed. Thus idiotypic determinants which are found in the variable regions would represent antigenic determinants formed by patterns of amino acid sequences involving some of the side chains of residues from the inserted regions—namely those forming the exterior portions of the site but also including some of the residues involved in three-dimensional folding and belonging to various subgroups, etc. This could give rise to a large number of determinants generally not related to specificity but influenced by or indeed partly created by the sequence of site-determining residues. In many instances in which immunodominant groups of the idiotypic determinants were from those of the inserted regions, one might expect the same idiotypic specificity to be manifested in several classes of immunoglobulins; this has been shown to be the case for γM and γG antibodies from the same rabbit (71). Thus the findings on idiotypic specificity provide further support for the uniqueness and universality of the antibody-forming mechanism.

Several models for insertion of information into DNA have been recognized. One of these is the episomal model (74, 75), and another involves two recombinations as in P1 phage transduction (76, 77). A self-perpetuating episome containing a large number of short nucleotide sequences the incorporation of any one of which into the structural genes of the light and heavy chains to provide the information for a given antibody complementarity provides a tempting mechanism. Such incorporation could be during embryogenesis, by which each cell receives the proper nucleotide sequences to program its structural genes for a given antibody specificity. Alternatively, if a cell is multipotent, it could have a number of sequences or the entire episome and the insertion of the proper sequence could be accomplished in some unknown manner after antigenic stimulation. While the evidence shows that one cell produces one kind of antibody at a time, and that myeloma cells produce populations of molecules which bind ligands in a homogeneous manner, such cells would already be programmed. The hypothesized programming could possibly result from the cell-cell interactions for which some evidence has been advanced. Systems of this type include transfer of information from macrophages to lymphoid cells (78) and two cell interactions such as thymus-bone marrow, etc. (16, p. 431, 79–81). There is no basis at present for any more than a passing reference to these as possibilities.

Placing the information for site complementarity in an independent self-duplicating mechanism would provide the evolutionary stability needed. It would lead to the universality which the antibody-forming system has clearly manifested, because the requirements for successful insertion would be relatively more stringent so that changes in the insertion material would have a higher frequency for completely destroying the capacity to form antibodies and such individuals would probably not survive.

There are difficulties in applying the episomal model to the antibody complementarity problem. The insertion in the antibody case would have to be by a recombination mechanism involving overlapping sequences on each side as in the P1 phage transduction, rather than as in the Campbell model. However, the degree of overlap on each side of the postulated insertions is very small, e.g., an invariant Cys at the beginning and invariant Trp 35 and Phe 98 at the end. Moreover episomes have not been found to date except in bacteria, although several possibly relevant systems in eukaryotes have been noted (74). It is also difficult to see how insertions to produce a given antibody specificity could be programmed for both the light and heavy chains since their contributions to binding are generally quite different.

The insertion model does not necessarily distinguish between germ line and somatic theories, for generating the diversity necessary to provide a large number of sites. Such diversity could be obtained by the existence of a large dictionary of insertions each of which determined a given specificity, e.g. by a germ line theory, or by the recombination of the nucleotides within the insertion material of a relatively small number of different sequences to produce a large number of recombinant sequences. This latter alternative might result in the formation of site-determining sequences which were not exactly the same for a given antibody in different species or in different individuals. Should sequence data on homogeneous antibodies to a given antigenic determinant formed in various species show that the same sequence is complementarity determining, the dictionary model would be favored over the recombinational version. Similarly, if homogeneous antibodies to a given determinant in the same individual but belonging to the various classes (γG, γM, γA) or subclasses (γG1, γG2, etc.) of immunoglobulins have the same complementarity determining residues, a dictionary model would also be favored.

The suggestions put forward are admittedly speculative as is the case at the moment with all other hypotheses. They do however present the problem of the antibody-combining site and of immunoglobulin structure in a different way, but in a way which is susceptible to verification or disproof by further data. Whether the model proposed ultimately stands or falls, the effort to assign each amino acid residue in the variable region a definite role by the statistical analysis employed should prove useful.

SUMMARY

In an attempt to account for antibody specificity and complementarity in terms of structure, human κ-, human λ-, and mouse κ-Bence Jones proteins and light chains are considered as a single population and the variable and constant regions are compared using the sequence data available. Statistical criteria are used in evaluating each position in the sequence as to whether it is essentially invariant or group-specific, subgroup-specific, species-specific, etc.

Examination of the invariant residues of the variable and constant regions confirms the existence of a large number of invariant glycines, no invariant valine, lysine, and histidine, and only one invariant leucine and alanine in the variable region, as compared with the absence of invariant glycines and presence of three each of invariant alanine, leucine, and valine and two each of invariant lysine and histidine in the constant region. The unique role of glycine in the variable region is emphasized. Hydrophobicity of the invariant residues of the two regions is also evaluated. A parameter termed variability is defined and plotted against the position for the 107 residues of the variable region. Three stretches of unusually high variability are noted at residues 24–34, 50–56, and 89–97; variations in length have been found in the first and third of these. It is hypothesized that positions 24–34 and 89–97 contain the complementarity-determining residues of the light chain—those which make contact with the antigenic determinant. The heavy chain also has been reported to have a similar region of very high variability which would also participate in forming the antibody-combining site. It is postulated that the information for site complementarity is contained in some extrachromosomal DNA such as an episome and is incorporated by insertion into the DNA of

the structural genes for the variable region of short linear sequences of nucleotides. The advantages and disadvantages of this hypothesis are discussed.

BIBLIOGRAPHY

1. Kabat, E. A. 1966. The nature of an antigenic determinant. *J. Immunol.* **97:**1.
2. Kabat, E. A. 1968. Structural Concepts in Immunology and Immunochemistry. Holt, Rinehart & Winston, Inc., New York.
3. Goodman, J. W. 1969. Immunochemical specificity: Recent conceptual advances. *Immunochemistry.* **6:**139.
4. Cohen, S., and R. R. Porter. 1964. Structure and biological activity of immunoglobulins. *Advan. Immunol.* **4:**287.
5. Kabat, E. A. 1966. Structure and heterogeneity of antibodies. *Acta haematol.* **36:**198.
6. Cohen, S., and C. Milstein. 1967. Structure and biological properties of immunoglobulins. *Advan. Immunol.* **7:**1.
7. Milstein, C., and J. R. L. Pink. 1970. Structure and evolution of immunoglobulins. *Prog. Biophys. Mol. Biol.* **21:**209.
8. Yount, W. J., M. M. Dorner, H. G. Kunkel, and E. A. Kabat. 1968. Studies on human antibodies. VI. Selective variations in subgroup composition and genetic markers. *J. Exp. Med.* **127:**633.
9. Dorner, M. M., W. J. Yount, and E. A. Kabat. 1969. Studies on human antibodies. VII. Acrylamide gel electrophoresis of purified human antibodies and myeloma proteins, their heavy and light chains. *J. Immunol.* **102:**273.
10. Fleischman, J. B., D. G. Braun, and R. M. Krause. 1968. Streptococcal group-specific antibodies: Occurrence of a restricted population following secondary immunization. *Proc. Nat. Acad. Sci. U. S. A.* **60:**134.
11. Braun, D. G., and R. M. Krause. 1968. The individual antigenic specificity of antibodies to streptococcal carbohydrates. *J. Exp. Med.* **128:**969.
12. Eichmann, K., H. Lackland, L. Hood, and R. M. Krause. 1970. Induction of rabbit antibody with molecular uniformity after immunization with group C streptococci. *J. Exp. Med.* **131:**207.
13. Pappenheimer, A. M., Jr., W. P. Reed, and R. Brown. 1968. Quantitative studies of the specificity of anti-pneumococcal polysaccharide antibodies, types III and VIII. III. Binding of a labeled oligosaccharide derived from S8 by anti-S8 antibodies. *J. Immunol.* **100:**1237.
14. Pincus, J. H., E. Haber, M. Katz, and A. M. Pappenheimer, Jr. 1968. Antibodies to pneumococcal polysaccharides: Relation between binding and electrophoretic heterogeneity. *Science (Washington)* **162:**667.
15. Brenneman, L., and S. J. Singer. 1968. The generation of antihapten antibodies with electrophoretically homogeneous L chains. *Proc. Nat. Acad. Sci. U. S. A.* **60:**258.
16. Frisch, L., editor. 1967. Antibodies. *Cold Spring Harbor Symp. Quant. Biol.* **32:**1–603.
17. Rossi, G., T. K. Choi, and A. Nisonoff. 1969. Crystals of fragment F_{ab}': Preparation from pepsin digests of human IgG myeloma proteins. *Nature (London)* **223:**837.

18. Killander, J., editor. 1967. Gamma Globulins. *Nobel Symp.* **3:**17–643.

19. Beaumont, J. L. 1967. Une spécificité commune aux α- et β-lipoprotéines du sérum révélée par un autoanticorps de myélome-L'antigène Pg. *C. R. Acad. Sci Ser. D.* **264:**185.

20. Eisen, H. N., E. S. Simms, and M. Potter. 1968. Mouse myeloma proteins with antihapten antibody activity. The protein produced by plasma cell tumor MOPC-315. *Biochemistry.* **7:**4196.

21. Metzger, H., and M. Potter. 1968. Affinity site labeling of a mouse myeloma protein which binds dinitrophenyl ligands. *Science* (*Washington*). **162:**1398.

22. Potter, M., and M. A. Leon. 1968. Three IgA myeloma immunoglobulins from BALB/c mouse:Precipitation with pneumococcal C polysaccharide. *Science* (*Washington*). **162:**369.

23. Schubert, D., A. Jobe, and M. Cohn. 1968. Mouse myeloma producing precipitating antibody to nucleic acid bases and/or nitrophenyl derivatives. *Nature* (*London*). **220:**882.

24. Ashman, R. F., and H. Metzger. 1969. A Waldenström macroglobulin which binds nitrophenyl ligands. *J. Biol. Chem.* **244:**3405.

25. Cohn. M., G. Notani, and S. A. Rice. 1969. Characterization of the antibody to the C-carbohydrate produced by a transplantable mouse plasmacytoma. *Immunochemistry.* **6:**111.

26. Capra, J. D., D. Wertheimer, W. J. Yount, and H. G. Kunkel. 1969. Monoclonal γG anti-γ globulins in hypergammaglobulinemic purpura. *Clin. Res.* **17:**351.

27. American Association of Immunologists Symposium on Experimental Approaches to Homogeneous Antibody Populations. 1970. *Fed. Proc.* **29:**55–91.

28. Potter, M. 1967. The plasma cell tumors and myeloma proteins of mice. *Methods Cancer Res.* **2:**106.

29. Edelman, G. M., and J. A. Gally. 1962. The nature of Bence Jones proteins. Chemical similarities to polypeptide chains of myeloma globulins and normal γ-globulins. *J. Exp. Med.* **116:**207.

30. Quattrocchi, R., D. Cioli, and C. Baglioni. 1969. A study of immunoglobulin structure. III. An estimate of the variability of human light chains. *J. Exp Med.* **130:**401.

31. Lennox, E. S., and M. Cohn. 1967. Immunoglobulins. *Annu. Rev. Biochem.* **36:**365.

32. Cohn, M. 1968. The molecular biology of expectation. *In* Nucleic Acids in Immunology. O. J. Plescia, and W. Braun, editors. Springer-Verlag, New York. 671.

33. Edelman, G. M., and W. E. Gall. 1969. The antibody problem. *Annu. Rev. Biochem.* **38:**415.

34. Hood, L., and D. W. Talmage. 1969. On the mechanism of antibody diversity: Evidence for the germ line basis of antibody variability. *In* Symposium on Developmental Aspects of Antibody Formation and Structure, Prague. In press.

35. Hood, L., and D. W. Talmage. 1970. Mechanism and antibody diversity: Germ line basis for variability. *Science* (*Washington*) **168:**325.

36. Dreyer, W. J., and J. C. Bennett. 1965. The molecular basis of antibody formation: A paradox. *Proc. Nat. Acad. Sci. U. S. A.* **54:**864.

37. Lederberg, J. 1959. Genes and antibodies. *Science* (*Washington*). **129:**1649.

38. Burnet, M. 1966. A possible genetic basis for specific pattern in antibody. *Nature (London)*. **210:**1308.

39. Brenner, S., and C. Milstein. 1966. Origin of antibody variation. *Nature (London)*. **211:**242.

40. Kabat, E. A. 1967. The paucity of species-specific amino acid residues in the variable regions of human and mouse Bence Jones proteins and its evolutionary and genetic implications. *Proc. Nat. Acad. Sci. U. S. A.* **57:**1345.

41. Kabat, E. A. 1967. A comparison of invariant residues in the variable and constant regions of human K, human L and mouse K Bence Jones proteins. *Proc. Nat. Acad. Sci. U. S. A.* **58:**229.

42. Kabat, E. A. 1968. Unique features of the variable regions of Bence Jones proteins and their possible relation to antibody complementarity. *Proc. Nat. Acad. Sci. U. S. A*.**59:**613.

43. Kabat, E. A. 1969. Antibody complementarity and light chain structure. *In* Symposium on the Evolution of Immune Response. Oct. 20–22, 1969. U.S. Argonne National Laboratories, Lemont, Ill.

44. Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. *Science (Washington)*. **155:**279.

45. Kabat, E. A. 1970. Heterogeneity and structure of antibody-combining sites. *Ann. N. Y. Acad. Sci.* **169:**43.

46. Phillips, D. C. 1967. The hen egg-white lysozyme molecule. *Proc. Nat. Acad. Sci. U. S. A.* **57:**484.

47. Milstein, C. 1967. Linked groups of residues in immunoglobulin K chains. Nature *(London)*. **216:**330.

48. Niall, H., and P. Edman. 1967. Two structurally distinct classes of kappa-chains in human immunoglobulins. *Nature (London)*. **216:**262.

49. Franěk, F. 1969. The character of variable sequences in immunoglobulins and its evolutionary origin. *In* Symposium on Developmental Aspects of Antibody Formation and Structure, Prague. In press.

50. Jukes, T. H. 1969. Evolutionary pattern of specificity regions in light chains of immunoglobulins. *Biochem. Genet.* **3:**109.

51. Welscher, H. D. 1969. Correlations between amino acid sequence and conformation of immunoglobulin light chains. I. Hydrophobicity and fraction charge. *Int. J. Protein Res.* **1:**235. II. Sequence comparison and the pattern of nonpolar residues. *Int. J. Protein Res.* **1:**267.

52. Blake, C. C. F., A. G. Mair, A. C. T. North, D. C. Phillips, and V. R. Sarma. 1967. On the conformation of the hen egg-white lysozyme molecule. *Proc. Roy. Soc. Ser. B. Biol. Sci.* **167:**365.

53. Birktoft, J. J., B. W. Matthews, and D. M. Blow. 1969. Atomic coordinates of tosyl-α-chymotrypsin. *Biochem. Biophys. Res. Commun.* **36:**131.

54. Edsall, J. T., P. L. Flory, J. C. Kendrew, A. M. Liquori, G. Némethy, G. N. Ramachandran, and H. A. Scheraga. 1966. A proposal of standard conventions and nomenclature for the description of polypeptide conformations. *Biopolymers* **4:**121.

55. Ramachandran, G. N., and V. Sasisekharan. 1968. Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23:**283.

56. Press, E. M., and N. M. Hogg. 1969. Comparative study of two immunoglobulin G Fd-fragments. *Nature* (*London*). **223:**807.

57. Cunningham, B. A., M. N. Pflumm, U. Rutishauser, and G. M. Edelman. 1969. Subgroups of amino acid sequences in the variable regions of immunoglobulin heavy chains. *Proc. Nat. Acad. Sci. U. S. A.* **64:**997.

58. Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Amer. Chem. Soc.* **84:**4240.

59. Bigelow, C. C. 1967. On the average hydrophobicity of proteins and the relation between it and protein structure. *J. Theor. Biol.* **16:**187.

60. Ruffilli, A., and C. Baglioni. 1967. Subgroups of L type Bence Jones proteins. *J. Immunol.* **98:**874.

61. Ruffilli, A. 1968. The antigenic determinants of the variable half of a Bence Jones protein of type L. *J. Immunol.* **100:**201.

62. Hood, L., H. Lackland, K. Eichmann, J. Ohms, and R. M. Krause. 1970. Amino acid sequences from rabbit antibody light chains and gene evolution. In preparation.

63. Todd, C. W. 1963. Allotypy in rabbit 19S protein. *Biochem. Biophys. Res. Commun.* **11:**170.

64. Todd, C. W. 1966. Discussion. *J. Cell. Physiol.* **67**(Suppl. *1*):95.

65. Koshland, M. E., J. J. Davis, and N. J. Fujita. 1969. Evidence for multiple gene control of a single polypeptide chain: The heavy chain of rabbit immunoglobulin. *Proc. Nat. Acad. Sci. U. S. A.* **63:**1274.

66. Smithies, O. 1963. Gamma-globulin variability: A genetic hypothesis. *Nature* (*London*). **199:**1231.

67. Edelman, G. M., and J. A. Gally. 1967. Somatic recombination of duplicated genes: An hypothesis on the origin of antibody diversity. *Proc. Nat. Acad. Sci. U. S. A.* **57:**353.

68. Slater, R. J., S. M. Ward, and H. G. Kunkel. 1955. Immunological relationships among the myeloma proteins. *J. Exp. Med.* **101:**85.

69. Kunkel, H. G., M. Mannik, andR. C. Williams. 1963. Individual antigenic specificities of isolated antibodies. *Science* (*Washington*) **140:**1218.

70. Oudin, J., and M. Michel. 1963. Une nouvelle forme d'allotypie des globulines γ de sérum de lapin apparemment liéé à la jonction et à la specificité anticorps. *C. R. Acad. Sci.* **257:**805.

71. Oudin, J., and M. Michel. 1969. Idiotypy of rabbit antibodies. I. Comparison of idiotype of antibodies against *Salmonella typhi* with that of antibodies against other bacteria in the same rabbit, or of antibodies against *Salmonella typhi* in various rabbits. *J. Exp. Med.* **130:**595. II. Comparison of idiotypy of various kinds of antibodies formed in the same rabbit against *Salmonella typhi. J. Exp. Med.* **130:**619.

72. Daugharty, H., J. E. Hooper, A. B. MacDonald, and A. Nisonoff. 1969. Quantitative investigations of idiotypic antibodies. I. Analysis of precipitating antibody populations. *J. Exp. Med.* **130:**1047.

73. Hurez, D., G. Meshaka, C. Mihaesco, and M. Seligmann. 1968. The inhibition by normal γG-globulins of antibodies specific for individual γG myeloma proteins. *J. Immunol.* **100:**69.

74. Campbell, A. M. 1962. Episomes. *Advan. Genet.* **11:**101.

75. Campbell, A. M. 1969. Episomes. Harper & Row, Publishers, New York.

76. Lennox, E. S 1955. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology.* **1:**190.

77. Ozeki, H., and H. Ikeda. 1968. Transduction mechanisms. *Annu. Rev. Genet.* **2:**245.

78. Adler, F. L., M. Fishman, and S. Dray. 1966. Antibody formation initiated *in vitro.* III. Antibody formation and allotypic specificity directed by ribonucleic acid from peritoneal exudate cells. *J. Immunol.* **97:**554.

79. Miller, J. F. A. P., and G. F. Mitchell. 1968. Cell to cell interaction in the immune response. I. Hemolysin-forming cells in neonatally thymectomized mice reconstructed with thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128:**801.

80. Mitchell, G. F., and J. F. A. P. Miller. 1968. Cell to cell interaction in the immune response. II. The source of hemolysin-forming cells in irradiated mice given bone marrow and thymus or thoracic duct lymphocytes. *J. Exp. Med.* **128:**821.

81. Claman, H. N., E. A. Chaperon, and R. F. Triplett. 1966. Immunocompetence of transferred thymus-marrow cell combinations. *J. Immunol.* **97:**828.