ATTEMPTS TO LOCATE COMPLEMENTARITY-DETERMINING RESIDUES IN THE VARIABLE POSITIONS OF LIGHT AND HEAVY CHAINS *

Elvin A. Kabat and T. T. Wu

Departments of Microbiology, Neurology, & Human Genetics & Development, Columbia University, The Neurological Institute, Presbyterian Hospital New York, N.Y.
and Departments of Physics and Engineering Sciences Northwestern University Evanston, Ill.

Examination of the sequences of Bence-Jones proteins and myeloma immunoglobulin light chains from humans and from mice has led to the recognition of the variable and constant regions, and the accumulation of data on heavy chains indicates that they too contain a variable and a constant region.[1-5] The variable region comprises approximately the amino terminal half of the light chain and the amino terminal quarter of the heavy chain, and it is these regions which are generally believed to be responsible for antibody complementarity. The genetic control of the constant regions of both chains is readily explainable on classical genetic principles, but the genetics of the variable regions is far from clear and no generally agreed upon concept of the genetic determination of antibody complementarity has as yet been formulated. The recognition of subgroups in the variable regions of human (Reference 2, p. 133),[6, 7] and mouse [8] light chains and in human heavy chains,[9, 11] from sequence analyses, mainly of the first 20–25 amino terminal residues, has led to the designation of genes for these subgroups. It is generally accepted that the light chain and the heavy chain are each under the control of two genes, one for the variable and one for the constant region, and that a translocation results in the joining of these two genes.[12] These conclusions, however, do not account for antibody complementarity nor do they localize the combining site to any specific portions of the variable region.

When the subgroups of the variable regions were first recognized, it was noted [7, 4] that certain positions, notably those near 30 and 91–96, showed greater variability than could be accounted for by the subgroups. As further sequences accumulated, it became clearer that there were two regions of hypervariability, one following cysteine 23 and the other following cysteine 88 and comprising residues 24–34 and 89–96 respectively, and it was of special interest that these two regions were brought into close proximity by the disulfide bond I_{23}-II_{88}, and that insertions or deletions occurred in these regions.[13, 4]

A more detailed analysis indicated the presence of three hypervariable regions. Franěk (Reference 5, p. 311) tabulated the positions showing nonhomologous replacements and recognized a region from residues 50–55 in

* Aided by grants from the National Science Foundation, GB-8341 and GB-25686, a General Research Support Grant from the United States Public Health Service to Columbia University, and Biomedical Sciences Support grant FR 7028-03 from the National Institutes of Health to Northwestern University.

addition to the other two regions. A statistical analysis was made by Wu and Kabat [14, 5] of the complete and partial sequence data available on 77 Bence-Jones proteins and immunoglobulin light chains, considering human κ, human λ, and mouse κ chains of various subgroups as a single population which was aligned for maximum homology. Defining variability as

$$\frac{\text{Number of different amino acids at a given position}}{\text{Frequency of the most common amino acid at that position}},$$

three hypervariable regions were found involving residues 24–34, 50–56 and 89–97. It was proposed that at least the first and third of these regions, and possibly all three, together with similar regions in the heavy chains might be the complementarity determining regions and that amino acid side chains in these regions might make contact with the antigenic determinant, the remainder of the residues of the variable region being essentially structural and involved in three-dimensional folding.

The significance of the three hypervariable regions was greatly reinforced by the findings of Weigert and coworkers [15] who examined ten mouse λ chains; from composition analyses on peptides, six had apparently identical sequences in the variable regions and the remaining four showed variation only in one or another of three hypervariable regions. FIGURE 1 shows the sequences of Weigert and associates superimposed on the original plot of Wu and Kabat.

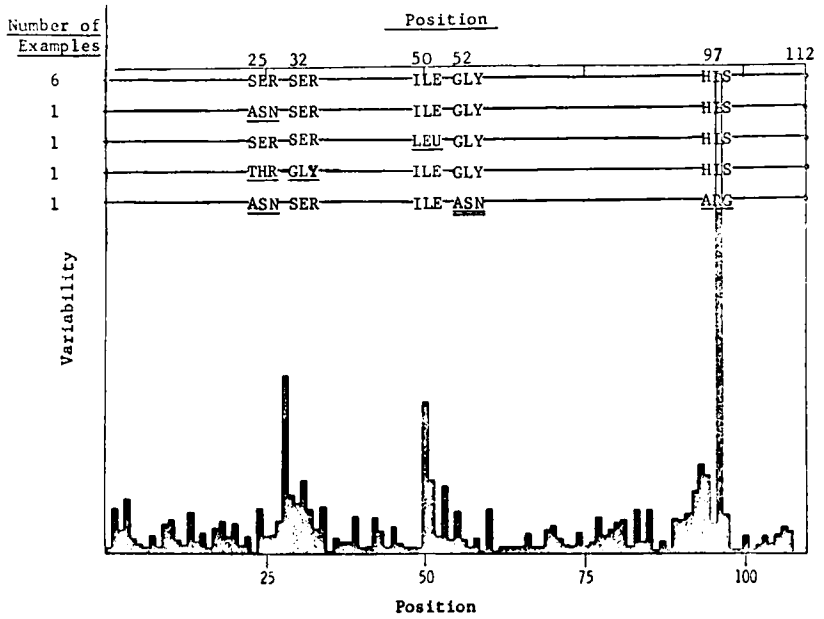

FIGURE 1. Mouse λ myeloma sequences of Weigert and coworkers [15] superimposed on the variability against position plot for human κ, human λ and mouse κ Bence-Jones proteins and light chains of Wu and Kabat.[14] The positions at which no differences among the mouse λ chains were found are indicated by a line. Sequences are given at positions at which differences occurred, and the number of nucleotide changes indicated by the underlining.[15]

It should be noted that their positions 50 and 52, if aligned for homology relative to our data, would be at 48 and 50. The mouse λ chains thus seem to be much more homogeneous in the variable region than mouse κ or human κ or λ chains, so that the changes appear to be restricted to the hypervariable regions.

It is also noteworthy that Capra and Kunkel [16] have found two cases of hypergammaglobulinemic purpura with antibodies of restricted specificity to have identical sequences for the first 40 N-terminal residues of their light chains that include the first hypervariable region; both chains were unusual in having Thr-Val at residues 13 and 14. It would be of great interest to establish whether the combining sites are identical and whether the sequence of the entire variable region of the light chains (and ultimately of the heavy chains) proves to be identical. Wang and colleagues [17, 18] have shown that in a patient with both a myeloma γG2κ and a macroglobulin γMκ protein, the κ light chains of both were identical in amino acid composition, bands in urea-starch gel, peptide maps, and in optical rotatory dispersion and circular dichroism. The light chains have not been sequenced, but the first 27 residues from the N-terminal region of the heavy chains and the idiotypic specificity of both proteins were identical.

We have continued to tabulate sequences of light chains and have now accumulated complete and partial sequences data on 121 human κ and λ and mouse κ and λ sequences. The tabulation only includes data for positions at which the sequence is reported as unequivocal. FIGURE 2 shows the variability vs. position plots for all 121 light chains as a single population (A), for 33 human κI chains (B), for 58 human κ chains (C), and for 58 human and 27 mouse κ chains (D). It is evident that the hypervariable regions may be seen in all plots. When two values are given for any position this is due to uncertainty in regard to Glx and Asx residues. It should be emphasized that apart from the N terminus, the number of complete sequences is still not as large as desirable and that the data for positions 40–85 are based on less than 20 proteins, while those at positions 1–23 are based on between 59 and 117 proteins.

It is of considerable interest that Singer and Thorpe found the invariant Tyr 86 of the light chain to be affinity-labeled in anti-DNP antibodies,[25] and Goetzl and Metzger [26] showed that position 34 in the above alignment (actually position 32) in the mouse λ myeloma protein with anti-DNP activity was labeled. Tyr 86 is very close to the hypervariable region, and position 34 is in the first hypervariable region. The identification of residues by affinity labeling in antibodies and myeloma proteins with different specificities will obviously be of importance in defining the relation of residues in the hypervariable regions to antibody complementarity.

Sequence data on heavy chains have been accumulating, and we have examined these data for hypervariable regions. Complete sequences of seven heavy chain variable regions, including five γG1 (Eu, He, Daw, Cor, Nie), 1 γM (Ou), one mouse myeloma protein (MOPC 173), and partial sequences of 7 γG1, 2 γG2, 2 γG3, 1 γG4, 8 γM, 2 γA, 1 γE, four rabbit, two horse and one shark, were available. As for light chains, subgroups of the variable regions of the heavy chains have been recognized and termed V_{HI}, V_{HII}, V_{HIII}, and V_{HIV}.[9, 20, 10, 11] The degree of homology of these variable region sequences is truly extraordinary. FIGURE 3A shows a plot of variability against position considering all the heavy chain sequences available as a single population. In aligning for maximum homology, gaps of two residues are placed between

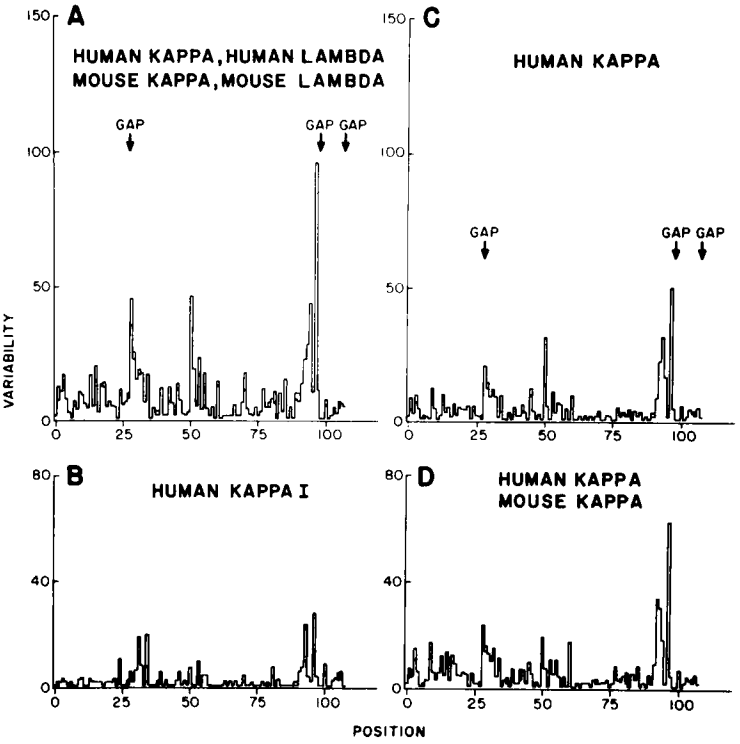

FIGURE 2. Variability at different amino acid positions for the variable region of light chains. GAP indicates positions at which insertions have been found. Data used: Wu and Kabat,[14] plus the following light chains: Pot, Die, Car, Tei, Joh (16); Dav, Fin (19); Ou (20); Hau (21); Til, Wil, Sal, Pom (22); Ful (23); G 173, F 31C, B J 149, B J 321, B J 63, A 603, A 870, A 384, A 467, F 47, H 37, B J 843, B J 674, A 8, A 15, B J 773, B J 265, GLP C 1, A 167 (24); S 104, X P 8, J 698, H 2061, J 558, HOPC 1, RPC 20, S 176, H 2020, S 178 (15).

residues 34 and 35 of V_{HI}, V_{HIII}, and mouse proteins, of one residue between positions 54 and 55, of three residues between positions 85 and 86 in some V_{HII} proteins, and of three to five residues between positions 100 and 101 in certain V_{HI}, V_{HII} and V_{HIII} proteins. Hypervariable regions are seen as for the light chains. The first hypervariable region comprises residues 31–35, and the last hypervariable region involves residues 95–102; in both instances, as with the light chains, gaps occur in both regions. There also are indications of two other regions which are somewhat more variable. One of these runs from residues 50–65 and corresponds approximately to the third light chain hypervariable region but is considerably larger; and the fourth, which was noted by Drs. Capra and Keogh, includes residues 81 and 83–85. A similar distribution is seen in FIGURE 3B, in which only the human heavy chain sequences are plotted; in this instance the hypervariable region extends from 50–54 rather than from 50–65, although somewhat more variability is seen for residues 61, 62, and 64, and residues 81, 83, 84, and 85 also show some-

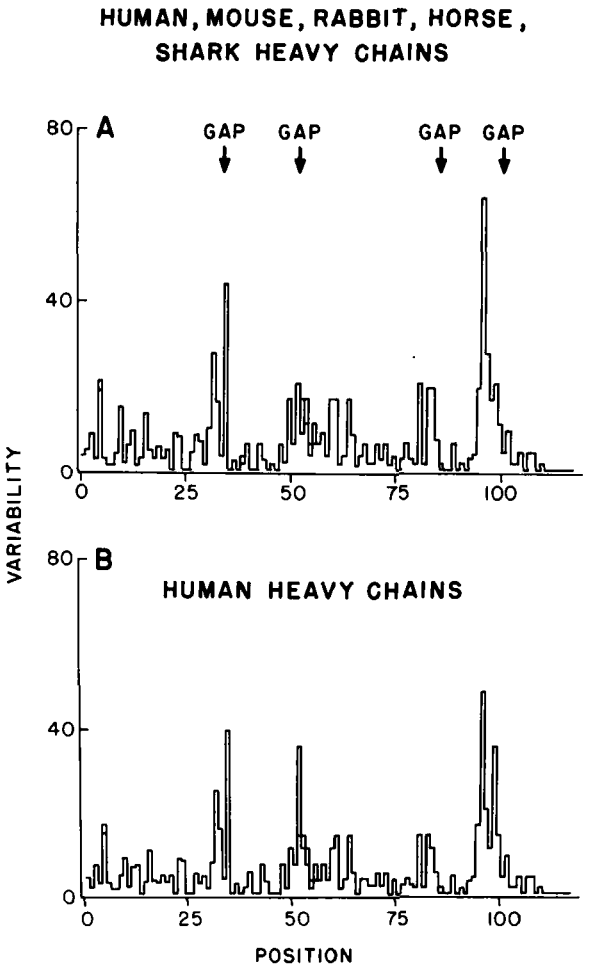

FIGURE 3. Variability at different amino acid positions for the variable regions of heavy chains. GAP indicates the positions at which insertions have been placed. Data used: Eu;[27] Ca;[28] Ste, horse γGab, horse γGT;[29] Dee;[30] He;[9] Daw, Cor;[31] Ou;[32] Car;[33] Sa;[34] Vin;[35] Til;[18] Sha;[36] Nie;[11] Tei, Was, Jon, Ben;[22] Fi, Vu;[37] Zuc;[38] Bus, Dau, Dos, Bal;[39] Di, Wo, Na, Hu, Re;[10] MOPC 173;[40] Rabbit;[41] Rabbit Aal, Aa2, Aa3;[42] Shark.[43]

what increased variability. Residue 5 in both plots shows substantial variability. The two major hypervariable regions are clearly evident. Unlike the light chain hypervariable regions, which begin after each Cys residue, the heavy chain hypervariability begins at residue 31 and 95 respectively, while the Cys residues are at positions 22 and 92. Thus while they too are brought into close proximity by the disulfide bond, there appears to be some displacement from the Cys residues.

The symmetry in the location of hypervariable regions in the light and heavy chains which are brought into relatively close proximity by the respective intrachain disulfide bonds can provide a three-dimensional site which could contain the complementarity determining residues, e.g., those which make contact with the antigenic determinant. As with the light chains, deletions or insertions occur in these two hypervariable regions. The role of the other hypervariable region is not clear. It is possible that it too is involved in site complementarity or, alternatively, it could contribute to the specificity of recognition of heavy and light chains. Its role will be more clearly established when some idea of its relation to the other hypervariable regions in three dimensions emerges. More sequence data on heavy chains are obviously needed. The hypervariability at positions 81 and 83–85, if substantial when more sequences are available, might also be related to site complementarity.

Species-Specific Residues

In 1967, when only a few human κ and two mouse κ Bence-Jones proteins had been sequenced, a comparison of the variable and constant regions showed that although mouse κ differed from human κ in the constant region at 43 positions, in the variable region there were many fewer positions, only about 10–14, at which the amino acids found in the mouse chains differed from those in the human chain.[44] This indicated a substantial deficit of species-specific residues in the variable region, and indicated substantial selective pressures in evolution to preserve the set of variable sequences. Subsequent analyses [13, 14] of all the human κ, human λ, and mouse κ chains as a single population reduced the number of species-specific residues to 2 out of the 107 in the variable region. It should be emphasized that this definition of species-specific residues is based essentially on structural considerations, i.e., do Bence-Jones proteins exist in either human or mouse with amino acids which differ at a given position in the variable region or, conversely, to form a functional chain is it necessary to restrict the amino acids present to such an extent that the human and mouse can have only the same limited set of amino acids at a given position?

The discovery of subgroups of κ and of λ chains have led other workers to consider that more residues in the variable region were species-specific. However, although the number of species-specific residues of the variable region may be questioned, there is little doubt that there are many fewer in the variable region than the 43 in the constant region.[44, 13, 14]

We have reexamined the question of species specificity, since many more sequences are now available. Unfortunately, however, there are still only two complete mouse κ sequences. TABLE 1 summarizes the results. Considering all human κ and λ and mouse κ proteins, there are but two positions, 50 and 96, in the variable region at which the mouse differs from the human. Position 96 is the most hypervariable position thus far reported, and it is doubtful whether it can be considered species-specific; the constant region has 36 species-specific positions computed on this basis. If one compares only human κ with mouse κ, there are 9 species-specific positions in the variable and 43 in the constant region. Comparing human κ_I with mouse κ_I, the species-specific residues are 24 and 43 respectively.

TABLE 1

SPECIES-SPECIFIC RESIDUES IN THE CONSTANT AND VARIABLE REGIONS OF HUMAN AND MOUSE BENCE-JONES PROTEINS AND IMMUNOGLOBULIN LIGHT CHAINS*

Proteins Compared	Species-Specific Residues	
	Variable Region	Constant Region
Human κ, human λ, mouse κ	2/107 (positions 50, 96)	36/107
Human κ, mouse κ	9/107 (positions 36, 43, 55, 60, 72, 80, 85, 92, 96)	43/107
Human κ_I, mouse κ_I	24/107 (positions 26, 30, 32, 33, 36, 39, 42, 43, 44, 46, 51, 55, 60, 66, 69, 71, 72, 80, 84, 85, 89, 92, 94, 96)	43/107

* Only one mouse κ_I and two mouse κ Bence-Jones proteins have been sequenced beyond residue 24.

The last two comparisons are of special interest since beyond residue 24, only two mouse κ and only one mouse κ_I proteins have been sequenced. It is of interest that no species-specific positions occur in the first 24 residues for which a number of mouse κ and κ_I proteins have been sequenced. Thus, the species-specific positions in TABLE 1 represent the maximum number, and this number can only be reduced as additional mouse sequences become available. However, even for the κ_I proteins, the lack of species-specific residues in the variable as compared with the constant region is highly significant. This becomes self-evident when one considers that the problem of alignment of variable regions to establish maximum homology is much simpler than with most other proteins of similar size and can be readily accomplished by inspection. Indeed, the similarity in sequences of the variable region of the heavy chains of different species suggests that very few residues of the heavy chain will show species specificity.

Role of Glycine

The unique pattern of invariant glycines in the variable regions of light chains of immunoglobulins and Bence-Jones proteins led to the suggestion that they might be responsible for conferring flexibility on the chains to permit three-dimensional folding to accommodate the numerous substitutions which were observed at the variable positions.[45, 13, 14] Consideration of the ϕ and ψ angles which glycine exhibited in other known proteins indicated that it had a much wider range than was seen for all other amino acids, so that it could indeed be responsible for conferring flexibility.[14] The unusual distribution of invariant glycines at positions 99 and 101 which—with the frequently occurring Gly at position 100 to give the sequences Gly-Gln-Gly or Gly-Gly-Gly near the end of the variable region and close to the two hypervariable regions joined by the disulfide bond I_{23}-II_{88}—led to the suggestion that these glycines might serve as a pivot to facilitate optimal contact of the complementarity

determining residues with the antigenic determinant [45] (Reference 2, p. 87). It was also noted that two glycines were present at an analogous position in all heavy chains [14] in a sequence Gly-Gln-Gly, Gly-Arg-Gly or Gly-Gly, and thus might also function as a pivot. It is well established that the walls of the lysozyme cleft move slightly to make optimal contact with its substrate.[46] The importance of good contact for binding of an antigenic determinant in the antibody combining site would tend to make such a pivot type of structure favored evolutionarily.

TABLE 2 summarizes data on the occurrence of glycine in the heavy chains for those complete human sequences which are available as well as for partial sequences of human, mouse, rabbit, and shark immunoglobulins. Despite the few sequences available, the data fall into three groups with respect to the frequency of occurrence of glycines comparable to those found for the light chains:

A (80–100%): positions 8, 26, 42, 104, and 106. These may prove to be largely invariant glycines. Position 104 is not glycine in Eu but, as noted, there is a second glycine in position 107 that together with the glycine at 106 was hypothesized to function as a pivot. Their exact position is affected by the insertion of gaps in aligning for maximum homology.

TABLE 2

FREQUENCY OF OCCURRENCE OF GLYCINE RESIDUES AT VARIOUS POSITIONS IN THE VARIABLE REGIONS OF HEAVY CHAINS

Position	No. of Glycines / No. of Proteins Sequenced at that Position	Percent Glycine	Position	No. of Glycines / No. of Proteins Sequenced at that Position	Percent Glycine
6	1/22	5	49	2/7	29
8	21/22	96	50	1/7	14
9	14/22	64	55	1/7	14
10	10/22	45	60	1/7	14
13	1/22	5	65	2/7	29
15	16/21	76	85A	1/7	14
16	10/20	50	85C	1/8	13
17	1/19	5	94	2/10	20
26	18/18	100	95	1/8	13
27	2/17	12	96	1/8	13
29	1/15	7	97	2/8	25
31	1/15	7	100C	1/1	100
33	2/14	14	104	4/5	80 *
35	2/11	18	106	5/5	100
40	1/7	14	107	1/5	20 *
42	7/7	100			
44	4/7	57			

* The Gly at position 107 occurs in protein Eu, which lacks Gly at 104, thus all proteins have the sequence Gly-Gln-Gly, Gly-Arg-Gly or Gly Gly at positions 104, 105, 106, or 106, 107.

B (50–76%): positions 9, 10, 15, 16, 44. These glycines might also contribute to flexibility; although the number of sequences is limited, they appear to be specific for one or another of the variable region subgroups.

C (5–29%): positions 6, 17, 27, 31, 33, 35, 40, 49, 50, 55, 60, 65, 85A, 85C, 94, 95, 96, 97. These glycines might be involved in specificity of variable region subgroups or in site complementarity, or perhaps the total number of sequences available is too small to establish their role.

The total number of invariant glycines (A) in the variable region is somewhat lower for the heavy chains than for the light chains for which seven invariant glycines were found. The glycines in group B seem to be invariant in one or more of the heavy chain variable region subgroups as has been found for the light chains,[14] and thus may also contribute to flexibility. Moreover, there may be more of these than is apparent from the frequencies listed in TABLE 2, since so few complete sequences are available. Thus, of the glycines in group C above, at position 49, glycine is found in the single mouse and in the single V_{HI} chain sequenced. Similarly, at position 65, Gly is found in the V_{HI} and the V_{HIII} heavy chain sequenced. However, the four subgroups of the variable regions of heavy chains V_{HI}, V_{HII}, V_{HIII}, and V_{HIV} [9, 20, 10, 11] also seem to be unrelated to the class or subclass of the heavy chain, again favoring translocation as a mechanism for joining variable and constant regions.[12]

There is thus a remarkable degree of similarity between the variable regions of heavy and light chains with respect to hypervariable regions, the paucity of species-specific residues, and with respect to the distribution of glycines that probably indicates a unique evolutionary origin distinct from that of the constant regions. The genetic mechanism of this variability and antibody complementarity is still not clear.

References

1. KILLANDER, J. (Ed.). 1967. Nobel Symposium on Gamma Globulins. **3:** 17–643.
2. FRISCH, L. (Ed.). 1967. Cold Spring Harbor Symp. Quant. Biol. **32:** 1–603.
3. KABAT, E. A. 1968. Structural Concepts in Immunology and Immunochemistry. Chap. 9. Holt, Rinehart and Winston. New York.
4. MILSTEIN, C. & J. R. L. PINK. 1970. Prog. Biophys. Mol. Biol. **21:** 209.
5. ŠTERZL, J. & I. ŘÍHA. (Eds.) 1970. Developmental Aspects of Antibody Formation and Structure. Vols. I, II. Czechoslovak Academy of Sciences. Prague.
6. NIALL, H. & P. EDMAN. 1967. Nature (London) **216:** 262.
7. MILSTEIN, C. 1967. Nature (London) **216:** 330.
8. HOOD, L. E., M. POTTER & D. J. MCKEAN. 1970. Science **170:** 1207.
9. CUNNINGHAM, B. A., M. N. PFLUMM, U. RUTISHAUSER & G. M. EDELMAN. 1969. Proc. Nat. Acad. Sci. U.S.A. **64:** 997.
10. KÖHLER, H., A. SHIMIZU, C. PAUL, V. MOORE & F. W. PUTNAM. 1970. Nature (London) **227:** 1318.
11. PONSTINGL, H., J. SCHWARZ, W. REICHEL & N. HILSCHMANN. 1970. Z. Physiol. Chem. **351:** 1591.
12. EDELMAN, G. M. & J. A. GALLY. 1970. Nature (London) **227:** 341.
13. KABAT, E. A. 1970. Ann. N. Y. Acad. Sci. **169:** 43.

14. WU, T. T. & E. A. KABAT. 1970. J. Exp. Med. **132:** 211.
15. WEIGERT, M. G., I. M. CESARI, S. J. YONKOVICH & M. COHN. 1970. Nature **228:** 1045.
16. CAPRA, J. D. & H. G. KUNKEL. 1970. Proc. Nat. Acad. Sci. U.S.A. **67:** 87.
17. WANG, A. C., I. Y. F. WANG, J. N. MCCORMICK & H. H. FUDENBERG. 1969. Immunochemistry **6:** 451.
18. WANG, A. C., S. K. WILSON, J. E. HOPPER, H. H. FUDENBERG & A. NISONOFF. 1970. Proc. Nat. Acad. Sci. U.S.A. **66:** 337.
19. CAPRA, J. D. & J. OHMS. 1970. Fed. Proc. **29:** 703.
20. KÖHLER, H., A. SHIMIZU, C. PAUL & F. W. PUTNAM. 1970. Science **169:** 56.
21. WATANABE, S. & N. HILSCHMANN. 1970. Z. Physiol. Chem. **351:** 1291.
22. CAPRA, J. D. 1971. Nature, New Biology **230:** 61.
23. SOX, H. C. & L. E. HOOD. 1970. Proc. Nat. Acad. Sci. U.S.A. **66:** 975.
24. HOOD, L. E., M. POTTER & D. J. MCKERN. 1970. Science **170:** 1207.
25. SINGER, S. J. & N. O. THORPE. 1968. Proc. Nat. Acad. Sci. U.S.A. **60:** 1371.
26. GOETZL, E. J. & H. METZGER. 1970. Biochemistry **9:** 3872.
27. GALL, W. E. & G. M. EDELMAN. 1970. Biochemistry **9:** 3188.
28. PITCHER, S. E. & W. KONIGSBERG. 1970. J. Biol. Chem. **245:** 1267.
29. FOUGEREAU, M. 1970. Personal communication.
30. FRANGIONE, B. & C. MILSTEIN. 1967 Nature **216:** 939.
31. PRESS, E. M. & N. M. HOGG. 1970. Biochem. J. **117:** 641.
32. WIKLER, M., H. KÖHLER, T. SHINODA & F. W. PUTNAM. 1969. Science **163:** 75.
33. FRANGIONE, B. 1968. Ph.D. Thesis. University of Cambridge. England.
34. MILSTEIN, C. & B. FRANGIONE. Unpublished data.
35. PINK, G. R. L. & C. MILSTEIN. 1969. FEBS Symposium on γ-globulin. **15:** 177.
36. TERRY, W. D., K. OGAWA & S. KOCHWA. 1970. J. Immunol. **105:** 783.
37. MONTGOMERY, P. C., A. C. BELLO & J. H. ROCKEY. 1970. Biochim. Biophys. Acta **200:** 258.
38. FRANGIONE, B. & C. MILSTEIN. 1969. Nature **224:** 597.
39. BENNETT, C. 1968. Biochemistry **7:** 3340.
40. BOURGOIS, A. & M. FOUGEREAU. 1970. FEBS Lett. **8:** 265.
41. CEBRA, J. J., L. A. STEINER & R. R. PORTER. 1968. Biochem. J. **107:** 79; 1970. **116:** 249.
42. WILKINSON, J. M. 1969. Biochem. J. **112:** 173.
43. SURAN, A. A. & B. W. PAPERMASTER. 1967. Proc. Nat. Acad. Sci. U.S.A. **58:** 1619.
44. KABAT, E. A. 1967. Proc. Nat. Acad. Sci. U.S.A. **57:** 1345.
45. KABAT, E. A. 1967. Proc. Nat. Acad. Sci. U.S.A. **58:** 229.
46. PHILLIPS, D. C. 1967. Proc. Nat. Acad. Sci. U.S.A. **57:** 484.

Discussion

DR. HOOD: Have you done a similar analysis on cytochromes because I know they have between six and eight hypervariable positions and I don't remember the placement and so forth, but with the three dimensional model for example it would be very interesting to look at those hypervariable positions and look and see what kind of role they assume and I wonder how these results then would compare with your light and heavy chain result?

DR. KABAT: I haven't done exactly that. Immunoglobulin hypervariable regions really pose exactly the reverse problem relative to cytochrome c and other proteins. Cytochrome c has a well defined site and when you find a lot

of variability, you are actually seeing the permissible noise which allows the three dimensional structure of the cytochrome site to be retained.

Here we have exactly the opposite, we have many genes from many kinds of chains and yet the antibody site is almost certainly in terms of our knowledge of structural chemistry and in terms of the results with myeloma proteins, in exactly the same place three dimensionally, but the amino acids determining its complementarity are varying so that with the cytochromes or any other enzyme one sees essentially the noise outside the site, with the site remaining relatively constant, whereas the hypervariability is making evident the multiplicity of the sites in the immunoglobulin.

DR. COHN: I would like Dr. Hood and Dr. Kabat to tell us what is the difference that you have concerning the definition of the species-specific residue.

DR. KABAT: I have defined mine.

DR. HOOD: I was interested in Dr. Kabat's new slide because I think it can be used very nicely to point out what I mean by species specificity. I think what's important to compare here is individual gene products, one with another and I think in Dr. Kabat's initial analysis when he tried to compare species specificity all kappa chains of mouse were lumped together.

DR. KABAT: There wasn't much to lump we only had two.

DR. HOOD: Well okay. But the point is, if one then starts breaking the mouse kappa chains up into their respective genes sub-groups 1, 2, and 3, the number of species specific residues went up from what? Nine to twenty-four?

DR. KABAT: Well we didn't have any sequences beyond 24 except for one kappa 1.

DR. HOOD: Right, but I think the point here is if one takes the kappa 1 sequences in human and mouse and says it's possible that there are a great many more genes here and starts correspondingly asking to compare the individual gene products in kappa 1 in mouse with the individual gene products in kappa 1 in human, then I think you are going to start expanding out the number of species specific residues. What I'm saying is, that there are in kappa 1 proteins, now there is good evidence that there are at least four sub-subgroups and I think most people would agree probably there are four additional germ line genes. And what one has to do is to compare these sequences at a much finer level and I think it's going to go from nine to 24 to I don't know what.

DR. KABAT: It's never going to be 43.

DR. HOOD: Well no that's not true at all; I mean if we take the extreme and we say that each different sequence in fact represents a different gene, it could very easily be 43.

DR. KABAT: No. I think that what you're really saying is, that if you can make an almost infinite number of sub-groups then between the mouse and the human, it's going to be very hard to demonstrate species specificity. My definition is a definition based on residue for residue and it tells what is permissible, and if you find for example that at position 22 both serine and threonine occur both in the human and in the mouse and that these are the only two substitutions there is obviously no species specificity by this definition. Now occasionally these two seem to be different but if the rabbit turns out in a comparable position to have one of the two there will be no species specificity either so what I am talking about are the restrictions on three dimensional folding of this molecule as expressed in different species.

DR. HOOD: Yes and I guess the only point I'd make in addition to that, what's important to consider is not all of these things lumped together but the linked relationships, that is, it's perfectly conceivable in certain cases that glycine 21 is permitted if you have a valine at position 57, and that's why I'm saying it's important not to pool all the data together but rather to consider what I think are individual gene products but at the very least are individual polypeptide chain products and I think that's where the difference lies.