*A COMPARISON OF INVARIANT RESIDUES IN THE VARIABLE AND CONSTANT REGIONS OF HUMAN K, HUMAN L, AND MOUSE K BENCE-JONES PROTEINS**

By Elvin A. Kabat

DEPARTMENTS OF MICROBIOLOGY AND NEUROLOGY, COLLEGE OF PHYSICIANS AND SURGEONS, COLUMBIA UNIVERSITY, AND THE NEUROLOGICAL INSTITUTE, PRESBYTERIAN HOSPITAL, NEW YORK

Communicated May 3, 1967

Studies of the tryptic peptides by fingerprinting[1-4] and determination of the sequences of human K, human L, and mouse K Bence-Jones proteins[5-15] have established that each of these three groups of proteins is made up of a variable and a constant region, comprising approximately the amino-terminal and carboxy-terminal halves of the molecules, respectively. In the variable region individual Bence-Jones proteins of each group may have different amino acid substitutions at a given position in the sequence, while in the constant region all proteins of each group have the same sequence except for position 191 in human K proteins, which is associated with Inv specificity.[8, 16]

With sequences arranged to give maximum homology, a comparison of these two regions in the various proteins showed that the variable region has very few species-specific residues, e.g., residues in which the human K and L proteins differed from the mouse K, while many such differences were present in the constant region;[17] with the publication of the complete sequence of a type L Bence-Jones protein,[15] these values are now at most 2 to 4 as compared with 36 species-specific residues, respectively. From this it was hypothesized that the variable region tended to be preserved during evolution perhaps because of its importance for antibody combining sites (cf. ref. 18), while the constant region, being less critical, had considerable freedom to mutate during evolution without seriously influencing the variable portion.

A further implication of this hypothesis would be that the invariant sequences of the two regions differed in composition and structure and that the variable regions of the three groups of Bence-Jones proteins might have a unique composition.

A compilation of the available sequence data on the human K, mouse K, and human L proteins[5-15] shows that 26 residues are invariant (e.g., the same amino acid occurs at the same position in the sequence) in the variable region (two uncertain because of amide groups) as compared with 33 in the constant region of all proteins studied. An identical amino acid is found at an additional 10 and 29 positions in the variable and constant portions, respectively, of all human K and mouse K and at 6 and 9 more positions in the respective regions of human L and mouse K proteins. The compositions are given in Table 1.

The amino acids making up the invariant residues in the human K and L and mouse K proteins differ to an extraordinary degree in the two halves of the chains. The variable region contains six glycines and no alanine, valine, leucine, or histidine, while the constant region has no glycine, three alanines, three valines, three leucines, and two histidines. There are, however, two invariant isoleucines and one invariant arginine in the variable but none in the constant region. On the other hand, the variable region has but two serines while the constant region has five.

The remaining amino acids in the two regions do not differ by more than one residue.

Among the residues identical in human K and mouse K Bence-Jones proteins, the largest difference is in serine and threonine, the variable region having two and one as compared with five and four, respectively, for the constant region. Since the latter has almost three times as many residues as the former, this may not be of significance. In addition, the variable portion contains no glutamic acid or asparagine, while three residues of each are found in the constant region. Of the residues identical in human L and mouse K, there is one glycine and three serines in each region.

TABLE 1

IDENTICAL RESIDUES IN THE VARIABLE AND CONSTANT REGIONS OF HUMAN K AND L AND MOUSE K BENCE-JONES PROTEINS

Variable Region		Constant Region	
Invariant in Human K, Human L, and Mouse K			
26 residues		33 residues	
6 Gly	2 Ser	0 Gly	5 Ser
0 Ala	2 Pro	3 Ala	3 Pro
0 Val	2 Tyr	3 Val	2 Tyr
0 Leu	2 Phe	3 Leu	2 Phe
0 His	2 Cys	2 His	3 Cys
2 Ile	1 Trp	0 Ile	1 Trp
1 Arg	1 Lys	0 Arg	2 Lys
1 Gln	1 Thr	0 Gln	1 Thr
0 Glu	1 Asp	2 Glu	1 Asp
2 Glx			
Identical in Human K and Mouse K			
10 residues		29 residues	
0 Gly	2 Ser	1 Gly	5 Ser
1 Ala	1 Pro	0 Ala	1 Pro
1 Val	0 Tyr	1 Val	1 Tyr
1 Leu	0 Phe	1 Leu	1 Phe
0 His	0 Cys	0 His	0 Cys
0 Ile	0 Trp	1 Ile	0 Trp
0 Arg	1 Lys	2 Arg	2 Lys
1 Gln	1 Thr	1 Gln	4 Thr
0 Glu	1 Asp	3 Glu?	2 Asp
0 Asn		3 Asn	
Identical in Human L and Mouse K			
6 residues		9 residues	
1 Gly	3 Ser	1 Gly	3 Ser
1 Leu	1 Lys	1 Val	1 Pro
		1 Glx	1 Tyr
			1 Thr

The presence of the six invariant glycines at positions 16, 57, 64, 68, 99, and 101 in the variable regions of all human K and L and mouse K Bence-Jones proteins studied and the absence of the apolar residues, alanine, valine, and leucine, as well as of histidine might well be expected to give the variable regions of these molecules a degree of flexibility which would permit them to accommodate conformational changes caused by the amino acid substitutions at the variable positions. It is probably not coincidental that three of these glycines occur between the two cysteines 23 and 88 which must join to form the —S—S bond, that two others are not far from the beginning of the constant region, and that 30 of the 58 positions, known

to be variable in the three groups of proteins,[5-15] occur within ten residues of each of these two cysteines.

These data tend to confirm the suggestion[17] that the variable region may represent a unique type of structure which has been preserved during evolution, perhaps because of its importance for antibody combining sites. It does not necessarily imply that the variable region is completely immutable. The exact sequences involved in antibody combining sites are not known, and the remainder of the variable regions of both light and heavy chains might show considerable mutability without impairing that portion hypothesized to contribute to antibody complementarity. Indeed, such mutations might well have given rise not only to the differences found in the N-terminal amino acid of the light chains of rabbit antibody and γG immunoglobulin,[19] but also to the individual specificity of antibodies[20, 21] which does not involve the combining site but which enables the anti-A formed in one individual to be differentiated antigenically from the anti-A of other individuals.[21, 22] Studies on the variable portions of the light chains and of the Fd fragments of the heavy chains of antibodies or myeloma proteins of various species should provide additional information as to the extent of the underlying structural uniqueness of the variable regions.

A useful working hypothesis for the genetic basis of antibody complementarity would be that a gene arose early in vertebrate evolution, controlling the ability to synthesize variable regions, and then went through a few duplications and further evolution and duplication[23-25] to give the light chains and the various heavy chains as well as the different Gm factors.

The preponderance of alanine, leucine, valine, and histidine among the invariant residues of the constant region would tend to suggest that apolar side chains may be more important in its tertiary structure.

It is of interest that among the 35 invariant residues in all cytochrome *c* samples studied to date (cf. refs. 26 and 27) there are eight glycines as well as five lysines, three prolines, and three tyrosines. On the other hand, among the eight invariant residues common to all hemoglobins and myoglobin there is only one glycine (cf. ref. 26). Neither of these two kinds of protein has a disulfide bond bringing residues at distant points in the peptide chain close to each other as is found in both the variable and constant regions of Bence-Jones proteins.

If one compares the sequence of the 108 residues from the C-terminal end of the Fc fragment of the heavy chain[24] of rabbit γG immunoglobulin with the sequences of the constant regions of human K and L and mouse K Bence-Jones proteins, 15 of the 33 invariant residues are the same. These are as follows:

1 Val	3 Pro	1 Phe
2 Leu	2 Cys	1 Lys
1 His		1 Glu
2 Ser	1 Trp	

Thus, in addition to the cysteines, two of the three valines, two of the three leucines, two of the five serines, one of the two histidines characteristic of the constant region of human K and L and mouse K light chains, and all three prolines are retained in the C-terminal portion of the rabbit heavy chain.

Putnam *et al.*[28] list 20 residues instead of 15 as invariant in all four groups of proteins but this probably depends upon the amount of data then available. The conservation of invariant apolar residues in this portion of the Fc chain which is known not to be involved in antibody specificity contrasts sharply with the absence of many apolar amino acids among the invariant residues of the variable region as shown in Table 1.

Myeloma globulins which react specifically with antigen and possess many of the properties of antibodies have been described.[29,30] One of these, a γA protein from a patient with both myeloma and hyperlipidemia, reacted specifically with α and with β human lipoproteins. The other a γG myeloma protein, showed very high antistreptolysin activity which was located in the Fab fragment.[30] Characterization of these proteins with respect to heterogeneity of their specific combining sites and establishment of the sequences of their light and heavy chains should provide data on the structural basis of antibody complementarity and a further test of the hypothesis presented. The finding that protein A from cell walls of *Staphylococcus aureus* precipitates about 45 per cent of pooled human γG immunoglobulin by reacting with the Fc portion of the molecule[31] indicates that unique areas of complementarity of a nonantibody character also exist in this portion of the heavy chains. Indeed, the myeloma protein with antistreptolysin activity was also precipitable by staphylococcal protein A.[30]

Note added in proof: H. Metzger (*Federation Proc.*, **26**, 529 (1967)) has also described a Waldenström macroglobulin with antibody activity specific for the Fc fragment of human γG immunoglobulin.

* Aided by grants from the National Science Foundation (GB-3675) and the Office of Naval Research (Nonr 266(13)).

[1] Putnam, F. W., C. W. Easley, and J. W. Helling, *Biochim. Biophys. Acta*, **78**, 231 (1963).
[2] Putnam, F. W., and C. W. Easley, *J. Biol. Chem.*, **240**, 1626 (1965).
[3] Small, P. A., R. A. Reisfeld, and S. Dray, *J. Mol. Biol.*, **11**, 713 (1965).
[4] Cioli, D., and C. Baglioni, *J. Mol. Biol.*, **15**, 385 (1966); *J. Exptl. Med.*, **124**, 307 (1966).
[5] Hilschman, N., and L. C. Craig, these PROCEEDINGS, **53**, 1403 (1965).
[6] Titani, K., E. Whitely, Jr., L. Avogardo, and F. W. Putnam, *Science*, **149**, 1090 (1965).
[7] Milstein, C., *Nature*, **209**, 370 (1966); *J. Mol. Biol.*, **21**, 203 (1966).
[8] Titani, K., E. Whitely, Jr., and F. W. Putnam, *Science*, **152**, 1513 (1966).
[9] Hood, L. E., W. R. Gray, and W. J. Dreyer, these PROCEEDINGS, **55**, 826 (1966); *J. Mol. Biol.*, **22**, 179 (1966).
[10] Milstein, C., *Biochem. J.*, **101**, 338, 352 (1966); *Proc. Roy. Soc. Ser. B*, **166**, 138 (1966).
[11] Putnam, F. W., K. Titani, and E. Whitely, Jr., *Proc. Roy. Soc. Ser. B*, **166**, 124 (1966).
[12] Gray, W. R., *Proc. Roy. Soc. Ser. B*, **166**, 146 (1966).
[13] Gray, W. R., W. J. Dreyer, and L. Hood, *Science*, **155**, 465 (1967).
[14] Titani, K., M. Wikler, and F. W. Putnam, *Science*, **155**, 828 (1967).
[15] Wikler, M., K. Titani, T. Shinoda, and F. W. Putnam, *J. Biol. Chem.*, **242**, 1668 (1967).
[16] Baglioni, C., L. Alesco Zonta, D. Cioli, and A. Carbonara, *Science*, **152**, 1517 (1966).
[17] Kabat, E. A., these PROCEEDINGS, **57**, 1345 (1967).
[18] Porter, R. R., and R. C. Weir, *J. Cellular Comp. Physiol. Supp.* 1, **67**, 51 (1966).
[19] Doolittle, R. F., these PROCEEDINGS, **55**, 1195 (1966).
[20] Oudin, J., and M. Michel, *Compte rend.*, **257**, 805 (1963).
[21] Kunkel, H. G., M. Mannik, and R. C. Williams, *Science*, **140**, 1218 (1963).
[22] Kunkel, H. G., J. Killander, and M. Mannik, *Acta Med. Scand., Suppl.*, **445**, 63 (1966).
[23] Singer, S. J., and R. F. Doolittle, *Science*, **153**, 13 (1966).
[24] Hill, R. L., R. Delaney, R. E. Fellows, and H. E. Lebovitz, these PROCEEDINGS, **56**, 1762 (1966).

[25] Marchalonis, J., and G. M. Edelman, *Science*, **154**, 1567 (1966).
[26] Jukes, T. H., *Molecules and Evolution* (New York: Columbia Univ. Press, 1966).
[27] Margoliash, E., and A. Schejter, *Advan. Protein Chemistry*, **21**, 113 (1966).
[28] Putnam, F. W., K. Titani, M. Wikler, E. Whitely, Jr. and T. Shinoda, *Federation Proc.*, **26**, 419 (1967).
[29] Beaumont, J.-L., *Compte rend.*, **264**, 185 (1967).
[30] Kronwall, G., *Acta Pathol. Microbiol. Scand.*, in press (1967).
[31] Forsgren, A., and J. Sjöquist, *J. Immunol.*, **97**, 822 (1966).