Proc. Nat. Acad. Sci. USA
Vol. 71, No. 3, pp. 845–848, March 1974

Variable Region Sequences of Five Human Immunoglobulin Heavy Chains of the V_HIII Subgroup: Definitive Identification of Four Heavy Chain Hypervariable Regions

(myeloma proteins/amino acid sequences/antibody combining site)

J. DONALD CAPRA AND J. MICHAEL KEHOE

Department of Microbiology, Mount Sinai School of Medicine of the City University of New York, 10 East 102 Street, New York, N.Y. 10029

Communicated by Henry G. Kunkel, November 5, 1973

ABSTRACT **The variable regions of five human immunoglobulin heavy chains of the V_HIII subgroup have been totally sequenced. Three of the heavy chains belonged to the IgG class and two to the IgA class. Examination of these sequences, and comparison with additional published heavy chain sequences, showed that a total of *four* hypervariable regions is characteristic of human heavy chain variable regions.**

The relatively conserved character of large segments of the heavy chain variable region was very evident in these studies. The conserved segments, which are those sections located outside the hypervariable regions, comprise approximately 65% of the total heavy chain variable region. The following general structural pattern for antibody molecules emerges from this and related studies: an overall combining region superstructure is provided by the more conserved segments while the refinements of the active site specificity are a function of hypervariable regions.

The antibody combining site is now believed to reside exclusively in the variable regions of the heavy and light polypeptide chains of the immunoglobulin molecule. Evidence is accumulating from several laboratories which indicates that hypervariable regions within the variable region are directly involved in the antibody combining site as well as being responsible, at least in part, for the idiotypic determinants of myeloma proteins and specific antibodies (1–5).

The existence of three hypervariable regions in the variable region of human immunoglobulin heavy chains has been established by previous studies from this laboratory. Residues 31–37 were described as the first hypervariable region of the heavy chain (6), and, after fragmentation of IgG heavy chains with cyanogen bromide, two additional hypervariable regions were localized between residues 86–91 and 101–110 (7).

We have now completed the amino acid sequence from residues 41 to 84 of the three V_HIII proteins originally reported (6, 7) as well as the complete V region sequence of two IgA proteins with V_HIII variable regions. The data make apparent an additional area of sequence hypervariability between residues 51 and 68, thus supporting the observations of Cebra *et al.* made on pooled guinea-pig immunoglobulins (5). When these data on V_HIII proteins are included with that available for proteins of the V_HI and V_HII subgroups and analyzed by the method of Wu and Kabat (3), four distinct areas of sequence hypervariability are observed.

MATERIALS AND METHODS

Myeloma Proteins. Tei (IgG1 kappa, Gm az), Was (IgG1 kappa, Gm az), Jon (IgG3 lambda, Gm g), Zap (IgA1 kappa), and Tur (IgA1 kappa) were isolated from serum or plasma by zone electrophoresis on polyvinyl copolymer ("Pevicon") (8). After further purification by gel filtration chromatography, they were reduced with 0.1 M β-mercaptoethanol and alkylated with iodoacetamide. The heavy and light chains were separated by gel filtration in propionic acid (9, 10).

Fragment Preparation: Heavy chains were treated with cyanogen bromide (11) and the resulting individual fragments purified by gel filtration chromatography in 5 M guanidine·HCl. Three proteins (Tei, Zap, and Tur) yielded a large N-terminal fragment comprising residues 1–85. Proteins Was and Jon, which contain a methionine residue at position 34, gave fragments comprising residues 1–34 and 35–85. Since all human IgG myelomas have a methionine at position 252, Tei and Was yielded a large fragment comprising residues 86–252. In protein Jon, however, an additional methionine was present at position 111. Consequently two distinct fragments comprising residues 86–111 and 112–252 were obtained from this protein. IgA1 proteins contain a methionine at residue 426 (12) so proteins Zap and Tur both yielded a very large fragment composed of residues 86-426.

Sequencing Procedure. Positions 1–85: On both the intact heavy chain as well as on the 1–85 fragment, proteins Tei, Zap, and Tur were sequenced 60 steps on the automated sequencer (13, 14). Tryptic peptides were prepared and separated on Dowex 50×4 with a pyridine-formate buffer system. In proteins Tei and Zap two invariant peptides were aligned by homology alone (70–74 and 75–78) while in protein Tur,

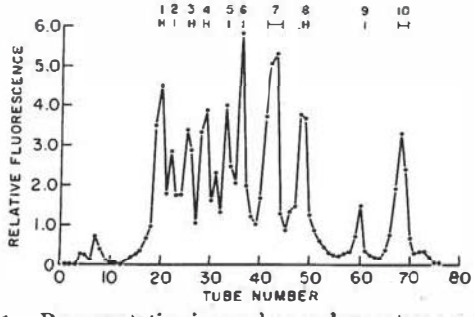

FIG. 1. Representative ion exchange chromatogram of tryptic hydrolysate of the amino terminal (1–85) cyanogen bromide fragment of α chain from protein Tur. Peptides were isolated from a Dowex 50X4 column and characterized and analyzed as described in the *text*.

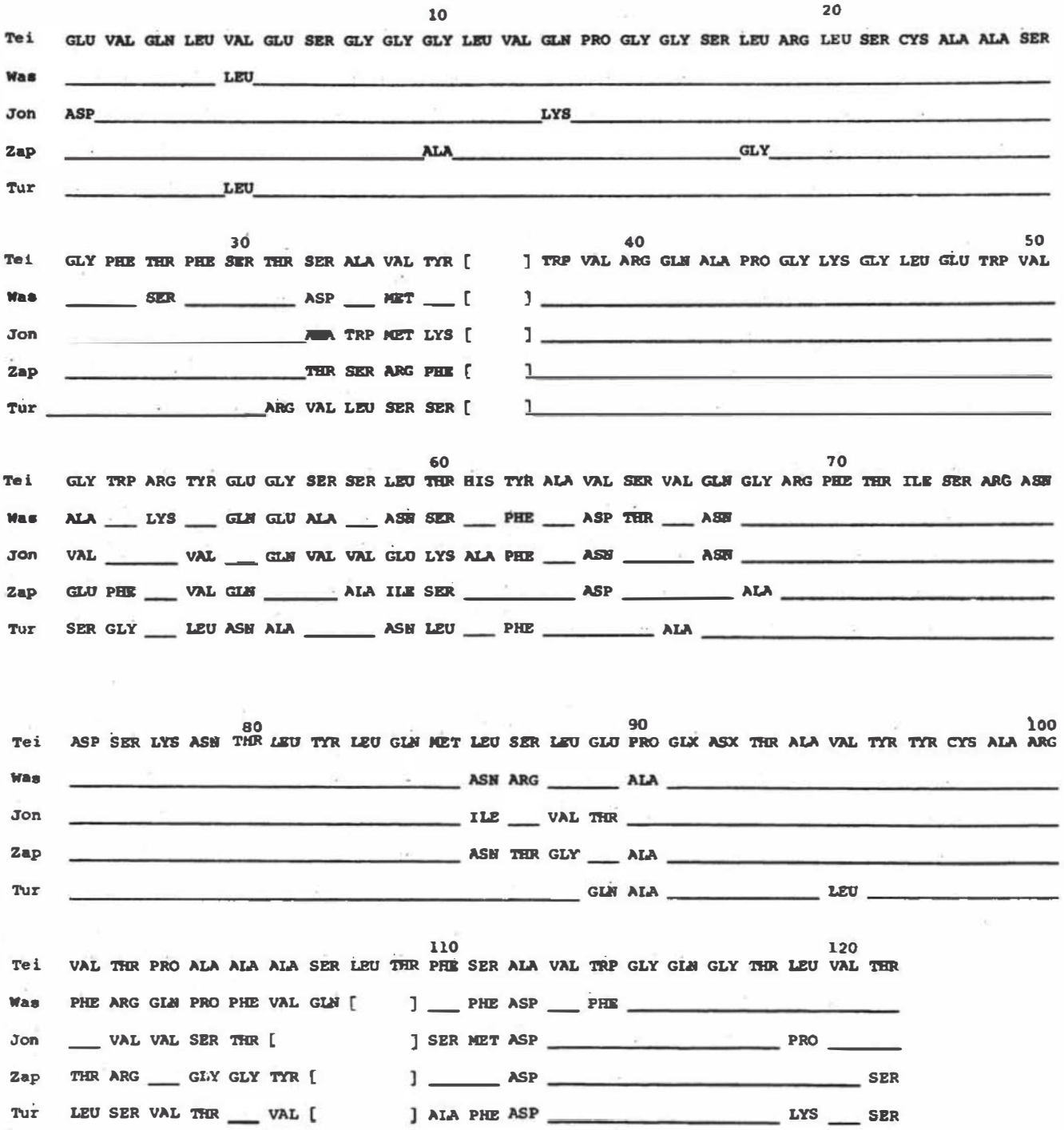

FIG. 2. The amino-acid sequence of the variable regions of five human immunoglobulin heavy chains.

the isolation of chymotryptic peptides established the sequence unambiguously. In all cases, tryptic peptides were sequenced in the automated sequencer, often using 4-sulfophenylisothiocyanate (Pierce Chemical) on the lysine peptides (15). In proteins Was and Jon, which contained cyanogen bromide fragments 1–34 and 35–85, the first 60 residues were established by automated sequencing of the intact heavy chain. Thus, in these two proteins, sequencing cyanogen bromide fragment 1–34 was superfluous since its composition agreed with the previously determined sequence. Fragment 35–85 of proteins Was and Jon was sequenced 35 and 40 residues respectively; this, together with the C-terminal tryptic peptides mentioned above gave the complete sequence for this section. *Residues 86–121:* In proteins Tei, Was, Zap, and Tur the sequence was established by a continuous automated run of 45 steps from residue 86 into the C_H1 domain. In both Zap and Tur, tryptic digestion and isolation of the resulting peptides confirmed a few questionable positions. In protein Jon, residues 86–111 were obtained disulfide linked to residues 1–34 after cyanogen bromide

digestion. This sequence was obtained by difference since residues 1–34 were known from the initial study of the intact heavy chain. Jon fragment 112–253 was subjected to a long sequencer run which definitely established the sequence of residues 112–121 as well as providing sequence data into the C_H1 domain.

Ion Exchange Chromatography. An example of a Dowex 50×4 chromatogram is shown in Fig. 1 for a tryptic digest of the Tur 1–85 fragment; 6.5-ml fractions were collected and 0.1 ml of each fraction analyzed by the fluorescamine procedure initially described by Udenfriend *et al.* (16). Ninhydrin analysis was also performed after alkaline digestion of 0.5-ml aliquots. In most analyses, only the fluorescamine procedure was employed since it was much more sensitive. As shown in Fig. 1, 10 fractions were pooled. Each was subjected to amino-acid analysis and several useful peptides were isolated and sequenced. T-1 (Asn Thr Leu Tyr Leu Gln Hsr) (79–85), T-3 (Asn Asp Ser Lys) (75–78), T-7 (Gly Leu Gly Trp Val Ser Gly Arg) (46–53), and T-10 (Phe Thr Ile Ser Arg) (70–74).

RESULTS AND DISCUSSION

The amino-acid sequences of the variable regions of the five human myeloma proteins is displayed in Fig. 2. The variability-factor values determined by the method of Wu and Kabat (3) for these as well as all the other human V region sequences available is shown in Fig. 3. These calculations were based on 25 sequences from residues 1 to 34, 11 sequences from residues 35 to 85, and 14 sequences from residues 86 to 122. Previous to this study there were only six published complete V region sequences, all but one (Nie) of the V_HI (Eu) or V_HII (Daw, Cor, He, Ou) subgroup (for references see legend to Fig. 3). With five additional V_HIII sequences the variability within and between subgroups can now be compared more meaningfully. In addition, with the availability of 11 complete sequences and several fragments, the Wu–Kabat plot becomes more statistically significant.

A discussion of the sequences can be conveniently divided into those sections of the V region which are relatively constant (1–30, 38–50, 69–83, 92–100, and 111–121), and the hypervariable regions (31–37, 51–68, 84–91, and 101–110).

About 65% of the variable region of the heavy chain shows limited variation. In fact, there are 17 positions (14%) which have been absolutely invariant in all human heavy chains regardless of their V region subgroup assignment. Certain positions are subgroup specific since at these positions all, or nearly all, of the members of one subgroup have a particular amino acid, while members of the other subgroup contain a different amino acid. Utilizing the four available V_HII proteins, positions 3, 9,16, 17, 19, 21, 23, 28, 29, 39, 42, 46, 50, 80, 81 and 82 appear to be subgroup specific. As noted previously, no subgroup specific residues are identifiable in the C terminal portion of the V region (7). There are thus 33 positions (27%) in the V region which are either invariant or subgroup specific. A comparison with the published sequences of myeloma proteins (17, 18), pooled immunoglobulins (5, 19, 20), and specifically purified antibodies (5, 21–23) from lower species, indicates that the particular amino acids found at these positions are characteristic of a wide variety of mammals and have been faithfully conserved during evolution. Such residues may have extremely important attributes for variable region function such as, for example, the provision of a distinct backbone structure which is crucial to antibody function.

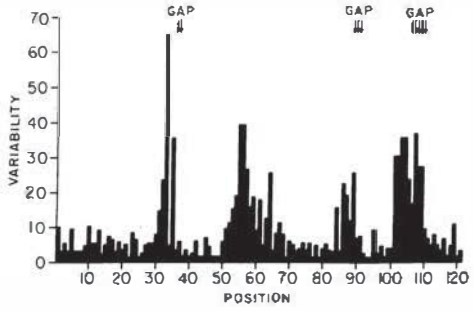

FIG. 3. Variability-factor values for the sequences shown in Fig. 1 as well as several other published sequences (36) determined according to the method of Wu and Kabat (3).

As can be seen on inspection of Figs. 2 and 3, about a third of the heavy chain variable region can be considered "hypervariable." These regions deserve special consideration because of their specific implications for the formation of the antibody combining site, the nature of idiotypic determinants, and various theoretical conceptions of the origin of antibody diversity.

In light chains, affinity labels have been localized near or within hypervariable regions (23–25), thus providing direct support for the general concept that hypervariable regions participate directly in the antibody-combining site. For the heavy chain, recent work has also been consistent with this idea. For example, Ray and Cebra localized affinity labels to the first (31–37) and the fourth (101–110) heavy chain hypervariable regions (26), Haimovich *et al.* (27) localized an affinity label to residue 54 of the mouse myeloma protein 315 (which has anti-dinitrophenol activity), and Press and coworkers have localized affinity labels at or near the fourth hypervariable region in rabbit antibodies (28). Therefore, although the primary structure and affinity labeling studies of these proteins was being carried out independently, and even in different laboratories in many instances, there is a general implication from the experimental observations that the same regions of the molecule which show the highest degree of sequence variation are near or part of those particular regions of the heavy chain where affinity labels have been localized.

A second piece of evidence linking the antibody combining site to the hypervariable regions has come from comparisons of sequences obtained from pooled immunoglobulin heavy chains with those of specifically purified antibody heavy chains. Sequence analyses of rabbit (29), guinea pig (5), and other mammalian heavy chain pools (19), indicate that a definitive sequence cannot be obtained within those regions which have been identified as hypervariable on the basis of studies with myeloma proteins. However, when specifically purified antibodies are studied, a single major sequence can be determined, as has been shown most definitively by Cebra and his coworkers (5).

Additional support for the functional significance of hypervariable regions has been provided by current notions concerning the tertiary structure of the immunoglobulin molecule. Crystallographic analysis of human immunoglobulins has now advanced to the point where it has been possible to assign the residues which may line a "pocket" within the

immunoglobulin molecule which presumably represents the combining site itself (30, 31). In each instance, the major residues which line the pocket are associable with hypervariable regions. In addition, the conformational models generated by the nearest neighbor calculations of Kabat and Wu (32) place hypervariable regions in close association with the putative combining site.

There is also growing evidence that at least some of the hypervariable regions are involved in the idiotypic determinants of myeloma proteins and antibodies. Cross idiotypic specificity among the cold agglutinins (33) and the anti-gamma globulins (34) is believed to be related to the combining site. In at least two distinct anti-gamma globulin molecules, the hypervariable regions show striking sequence similarities (7, 35).

The genetic origin of hypervariable regions remains unclear. The variability within heavy chain hypervariable regions seems more marked than that of light chain hypervariable regions. Of the 11 proteins which have now had their V regions completely sequenced, if one considers the 43 hypervariable positions of the heavy chain, there are no two proteins which have more than 10 residues in common. It seems likely that hundreds, or even thousands, of proteins would have to be sequenced in order to find two which are identical if no preselection bias (such as selection by idiotypic antisera or for combining specificity) is involved. This implies either that there are a very large number of germ line genes or that somatic processes are necessary to explain the diversity in the heavy chain hypervariable regions. Regardless of their origin, the hypervariable regions clearly play a crucial role in the antigen binding function of immunoglobulin molecules.

We thank Dr. Henry Kunkel for the subclass and genetic typing of the myeloma proteins. Bonnie Gerber, Ellen Bogner and Donna Atherton rendered invaluable technical assistance. This work was aided by grants from the National Science Foundation (GB 17046) and the U.S. Public Health Service (AI 09810) and a Grant-in-Aid from the New York Heart Association. J.D.C. is the recipient of National Institutes of Health Career Development Award 6-K4-GM-35, and J.M.K. is an Established Investigator of the American Heart Association.

1. Milstein, C. (1967) *Nature* 216, 330-332.
2. Franek, F. (1969) *Symposium on Developmental Aspects of Antibody Formation and Structure, Prague*.
3. Wu, T. T. & Kabat, E. A. (1970) *J. Exp. Med.* 132, 221–250.
4. Capra, J. D., Kehoe, J. M., Winchester, R. & Kunkel, H. G. (1971) *Ann. N.Y. Acad. Sci.* 190, 371–381.
5. Cebra, J. J., Ray, A., Benjamin, D. & Birshtein, B. (1971) *Progr. Immunol.* (First International Congress of Immunology), 269–284.
6. Capra, J. D. (1971) *Nature New Biol.* 230, 61–63.
7. Kehoe, J. M. & Capra, J. D. (1971) *Proc. Nat. Acad. Sci. USA* 68, 2019-2021.
8. Kunkel, H. G. (1954) *Methods Biochem. Anal.* 1, 141–155.
9. Fleischman, J. G., Porter, R. R. & Press, E. M. (1963) *Biochem. J.* 88, 220-228.
10. Capra, J. D. & Kunkel, H. G. (1970) *J. Clin. Invest.* 49, 610-621.
11. Gross, E. & Witkop, B. (1962) *J. Biol. Chem.* 237, 1856–1863.
12. Chuang, C. Y., Capra, J. D. & Kehoe, J. M. (1973) *Nature* 244, 158–160.
13. Edman, P. & Begg, F. (1967) *Eur. J. Biochem.* 1, 80–91.
14. Capra, J. D. & Kunkel, H. G. (1970) *Proc. Nat. Acad. Sci. USA* 67, 87–92.
15. Inman, J. K., Hannon, J. E. & Appella, E. (1972) *Biochem. Biophys. Res. Commun.* 46, 2075–2081.
16. Udenfriend, S., Stein, S., Bohlen, P., Dairman, W., Leimgruber, W. & Wiegele, M. (1972) *Science*, 178, 881–882.
17. Kehoe, J. M. & Capra, J. D. (1972) *Proc. Nat. Acad. Sci. USA* 69, 2052–2055.
18. Bourgois, A., Fougereau, M. & de Preval, C. (1972) *Eur. J. Biochem.* 24, 446–455.
19. Capra, J. D., Wasserman, R. W. & Kehoe, J. M. (1973) *J. Exp. Med.* 138, 410–427.
20. Mole, L. E., Jackson, S. A., Porter, R. R. & Wilkinson, J. M. (1971) *Biochem. J.* 124, 301–318.
21. Fleischman, J. B. (1973) *Immunochemistry* 10, 401-407.
22. Strosberg, A. D., Jaton, J. C., Capra, J. D. & Haber, E. (1972) *Fed. Proc.* 31, 771.
23. Goetzl, E. J. & Metzger, H. (1970) *Biochemistry* 9, 1267–1278.
24. Franek, F. (1971) *Eur. J. Biochem.* 19, 176–183.
25. Chesebro, B. & Metzger, H. (1972) *Biochemistry* 11, 766-771.
26. Ray, A. & Cebra, J. J. (1972) *Biochemistry* 11, 3647–3657.
27. Haimovich, J., Eisen, H. N., Hurwitz, E. & Givol, D. (1972) *Biochemistry* 11, 2389-2397.
28. Press, E. M., Fleet, G. W. J. & Fisher, C. E. (1971) in *Progress in Immunology*, ed. Amos, B. (Academy Press, New York), p. 233.
29. Cebra, J. J., Givol, D. & Porter, R. R. (1968) *Biochem. J.* 107, 69-70.
30. Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B., personal communication.
31. Poljak, R. J. (1973) *Abstracts of Ninth International Congress of Biochemistry-Stockholm*, p. 31.
32. Kabat, E. A. & Wu, T. T. (1972) *Proc. Nat. Acad. Sci. USA* 69, 960–964.
33. Williams, R. C., Kunkel, H. G. & Capra, J. D. (1968) *Science* 161, 379–381.
34. Kunkel, H. G., Agnello, V., Joslin, F. G., Winchester, R. J. & Capra, J. D. (1973) *J. Exp. Med.* 137, 331–342.
35. Capra, J. D. & Kehoe, J. M., unpublished observations.
36. (*a*) Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U., & Waxdal, M. J. (1969) *Proc. Nat. Acad. Sci. USA* 63, 78–85; (*b*) Press, E. M. & Hogg, N. M. (1970) *Biochem. J.* 117, 641–660; (*c*) Cunningham, B. A., Pflumin, M. N., Rutishauser, U. & Edelman, G. M. (1969) *Proc. Nat. Acad. Sci. USA* 64, 997–1003; (*d*) Wikler, M., Kohler, H., Shinoda, T. & Putnam, F. W. (1969) *Science* 163, 75–78; (*e*) Ponstingl, H., Schwarz, J., Reichel, W. & Hilschmann, N. (1970) *Hoppe-Seyler's Z. Physiol. Chem.* 351, 1591.