*Proc. Nat. Acad. Sci. USA*
Vol. 68, No. 9, pp. 2019–2021, September 1971

# Localization of Two Additional Hypervariable Regions in Immunoglobulin Heavy Chains

**(myeloma proteins/antigen–antibody specificity/human/amino acid sequences)**

J. MICHAEL KEHOE AND J. DONALD CAPRA

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

*Communicated by Henry G. Kunkel, June 7, 1971*

**ABSTRACT** **Cyanogen bromide fragments were isolated from the heavy chains of three human IgG myeloma proteins of the $V_H$III subgroup, sequenced by an automated method, and localized to the variable region. Inspection of these sequences, together with corresponding stretches from both human and animal proteins (studied in other laboratories) led to the detection of two additional hypervariable regions characteristic of the $V_H$ segment of immunoglobulin heavy chains. These areas of hypervariability, involving heavy-chain residues 86–91 and 101–109, were separated by a region of relative constancy. The close relationship of these two hypervariable regions, and the previously described first heavy-chain hypervariable region (residues 31–37), to the first heavy-chain disulphide bridge implies that the three hypervariable areas might be in close steric approximation in native immunoglobulin molecules.**

**Examination of the sequences of the terminal portion of $V_H$ of all these proteins (the segment from residue 95 to the beginning of homology region $C_H1$) revealed that no subgroup-specific residues could be identified in this area. Thus, heavy-chain subgroup distinctions may not extend through the entire variable region.**

An important conceptual advance in our understanding of the nature of the antibody combining site has derived from the independent recognition of hypervariable areas within the variable region of immunoglobulin light chains by Milstein (1), Kabat (2), and Franěk (3). Three such regions have been found in light chains, which involve residues 24–34, 52–55, and 89–97 (4). Although the direct participation of such hypervariable regions in the combining site seems a reasonable *a priori* assumption, firm experimental support for this hypothesis has only recently been forthcoming. A previous report from this laboratory presented evidence for an association of particular amino-acid residues within light-chain hypervariable areas with particular antibody activities, namely, antigammaglobulin (5). Goetzl and Metzger (6) have localized an affinity label in close proximity to the first light-chain hypervariable region of a murine myeloma protein that possesses antidinitrophenyl activity.

In heavy chains, only a single hypervariable region (positions 31–37) has been described thus far (7), and its existence and location have recently been confirmed (8). Some indication of an association between specific amino-acid residues within this hypervariable area and particular antibody activities has also been obtained (5). Ray and Cebra (9) have recently located an affinity label in this region of guinea pig antidinitrophenyl preparations.

On the basis of these observations, we searched for additional heavy-chain hypervariable regions. The experimental approach depended upon cleavage of heavy chains from purified human myeloma proteins with cyanogen bromide, isolation of selected individual fragments, and determination of their amino-acid sequence with a Beckman model 890 protein sequencer.

A preliminary account of this work was presented before the American Association of Immunologists, April 15, 1971.

## MATERIALS AND METHODS

### Fragment preparation

Heavy chains were prepared from human IgG myeloma proteins *Jon*, *Tei*, and *Was* as described (10). Isolated chains were cleaved with cyanogen bromide (11), and the resulting individual fragments were purified by gel filtration. Fragments were completely reduced and alkylated in 8.0 M guanidine–0.1 M Tris·HCl, at pH 8.0. Intrachain half-cystines were labeled by a 5-min pulse of [$^{14}$C]iodoacetamide (New England Nuclear Corp.) before the addition of a 100-fold excess of cold iodoacetamide to facilitate the identification of the phenylthiohydantion (PTH) derivative of cystine.

### Sequence determination

Aminoacid sequences were determined as described (12), except that thin-layer chromatography was also used in the identification of certain PTH derivatives. For small peptides, the 890 Sequencer was used with a peptide program that used a volatile buffer (dimethylallylamine).

## RESULTS AND DISCUSSION

Fig. 1 illustrates sequences from the carboxy-terminal third of the variable region of the three immunoglobulin heavy chains belonging to the $V_H$III subgroup (13); Fig. 1 includes other corresponding sequences available in the literature. The various proteins have been arranged according to the human subgroups $V_H$I, $V_H$II, and $V_H$III. Although the exact site of the $V_H$:$C_H1$ transition has not yet been definitively established, for the purposes of this report it has been placed at residue 122*. The methionine at position 85 is of particular interest since, in addition to occurring at a position within the mole-

* While the numbering system we have employed basically follows that given in ref. 14c, differences in deletion assignments beyond half-cystine 98 have resulted in our residue 122 corresponding to residue 118 of Cunningham *et al.*

```
VHI     85                 90                  95                  100                 105                 110                 115                 120
Eu      Leu Ser Ser Leu Arg Ser Glu Asp Thr Ala Phe Tyr Phe Cys Ala Gly Gly Tyr Gly Ile [                  ] Tyr Ser Pro Glu Glu Tyr Asn Gly Gly Leu Val Thr | Val Ser

VHII
Ou      Met Ile Asn Val Asn Pro Val Asp Thr Ala Thr Tyr Tyr Cys Ala Arg Val Val Asn Ser Val Met Ala Gly Tyr*Tyr Met Asp Val Trp Gly Lys Gly Thr Thr Val Thr | Val Ser
He      —— Thr ——Met Asp ——————————————————————Val His Arg His Pro Arg Thr Leu [      ] Ala Phe ——————————— Gln ———— Lys ——Ala | ————
Cor     ——[      ] Asp ———————————————————————————Ile Thr Val Ile Pro Ala Pro Ala Gly ——————————————— Arg ———— Pro ———— | ————
Daw     ——Asn Thr ——Gly——Gly ————————————————————Ser Cys Gly ——Gln [         ]——Phe ——Tyr ———— Gln ——Ile Leu ———— | ————
Rabbit  ————Thr Ser Pro Thr Gln ——————————Phe ———————                                          Asx Leu (——Gly Leu———) | ————

VHIII
Nie     Met Asn Ser Leu Arg Pro Glx Asx Thr Ala Val Tyr Tyr Cys Ala Arg Ile Arg Asp Thr Ala Met [      ] Phe Phe Ala His Trp Gly Gln Gly Thr Leu Val Thr | Val Ser
Jones†  ——Ile ——Val Thr ————Val ————Val Val ——————Val Val Val Ser Thr [          ] Ser Met Asp Val ———————————————— Pro ———— | ————
Was†    ———— Arg ——Glu Ala ———————————————————Phe —— Gln Pro Phe Val Gln [    ] ———Asp Val Phe ———
Tei†    ——Leu ————Glu ————————————————Thr ——Val Thr Pro Ala ——Ala Ser Leu Thr —— Ser ——
Mouse   —— Ser Lys Val —— Ser ————————Leu ———————Ser Pro Tyr Tyr ————
                                                                                                                                        VH <——|——> CH1
```

FIG. 1. Sequence stretches in variable region of heavy chains, showing hypervariable regions 86–91 and 101–109. Sequences were determined during the present study (marked by †), or in other laboratories (refs. 14, 15, 16).

cule critical for the preparation of certain cyanogen bromide fragments of the variable region, this methionine residue has also proved to be a nearly invariant amino acid in mammalian heavy chains.

Inspection of the sequences presented in Fig. 1 led to the detection of two areas of hypervariability, one comprising positions 86–91 and the second, positions 101–109. A previous comparison of two $V_H$II proteins (Daw and Cor) by Press and Hogg (14*b*) led these authors to suggest that residues corresponding to positions 101–106 of Fig. 1 might comprise a hypervariable region. Consideration of the additional sequences now available fully confirms this proposal, in addition to extending the limits of this region of hypervariability.

Within hypervariable area 86–91, only one gap for protein *Cor* is required to provide the most satisfactory sequence alignments. However, for area 101–109, gaps must be inserted in six of the nine sequences shown to maintain the most satisfactory homology in the more constant stretches immediately preceding and following this hypervariable region. Gaps have previously been associated with the first heavy-chain hypervariable region (7) and with light-chain hypervariable regions (4).

For the macroglobulin *Ou* (14*d*), a 3-residue insertion is required after 109 (see asterisk in Fig. 1) for optimal alignment with the other sequences. The significance of such an insertion at the end of this hypervariable area, if any, is not clear.

In the more constant region (92–100), between the two hypervariable segments, comparatively little sequence variability is seen. The functional importance of this region is not currently apparent, but its relative constancy, location between the two hypervariable areas, and presence in lower species as well as man (Fig. 1) could well signify its participation in some aspect of the combining site, as previously suggested by Bourgois and Fougereau (16). Within this region, position 95 is of special interest because it is the last residue from the amino terminus of $V_H$ for which human subgroup specific residues are evident (Phe for $V_H$I, Thr for $V_H$II, Val for $V_H$III).

Fig. 1 also shows sequence stretches in heavy chains corresponding to the human proteins from IgG of two lower species, rabbit (15) and mouse (16) (pooled IgG and a myeloma protein, respectively). Hypervariable area 86–91 is clearly evident in both rabbit and mouse chains. However, as already noted, residues 92–100 from both species show a marked similarity to the human heavy-chain sequences. The hypervariable region starting at position 101 is evident in the mouse IgG2a protein and can no doubt be inferred to be present in the rabbit, since Fruchter *et al.* (15) were unable to obtain a unique sequence in this area in pooled, heterogeneous rabbit

FIG. 2. Variability-factor values for the sequences shown in Fig. 1, determined according to the method of Wu and Kabat (4).

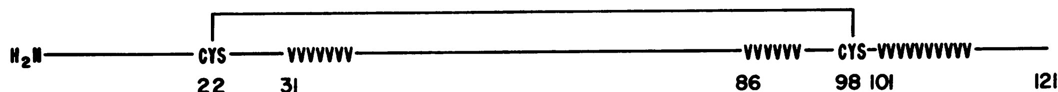


FIG. 3. Location of three hypervariable areas in the heavy chain, relative to half-cystines 22 and 98.

gammaglobulin, although they did report an unambiguous sequence on both sides of the region.

Wu and Kabat (4) have quantitated amino-acid sequence variability for immunoglobulin light chains by determining, for any position desired, a "variability factor" according to the number of different residues observed and the frequency of the most common residue. In their notation, increased variability leads to higher "variability factor" values. Fig. 2 shows variability factor values for positions in the carboxy-terminal third of $V_H$ of the proteins shown in Fig. 1; the two hypervariable regions are clearly identified.

A careful examination of all available $V_H$ sequences from position 96 to the beginning of homology region $C_H1$ indicates that *no* subgroup-specific residues can be identified in this area. Thus, from the available data, the last subgroup-specific residue of $V_H$ appears to be 95. The implications of such a pattern for proposed mechanisms of antibody diversity are obvious (e.g., potential for mutational drift and recombinational events in this region) and deserve further consideration. There is precedence from the analysis of light chain sequences for a loss of subgroup distinction near the variable region–constant region junction (17, 18).

Fig. 3 is a schematic, two-dimensional representation of the $V_H$ region of the immunoglobulin molecule, to illustrate the linear disposition of the three known heavy-chain hypervariable areas relative to the appropriate half-cystines. In Fig. 4*a*, the same representation is folded at the midpoint of the disulfide loop to show the planar juxtaposition of the first and second $V_H$ hypervariable regions that can be so obtained. Fig. 4*b* shows that the third hypervariable region can be brought into close relation to the first two by a reorientation of the polypeptide chain around half-cystine 98. While these are obvious oversimplifications of the tertiary structure, it nevertheless seems striking that such a close relation of either two, or all three, of these hypervariable areas can be obtained.

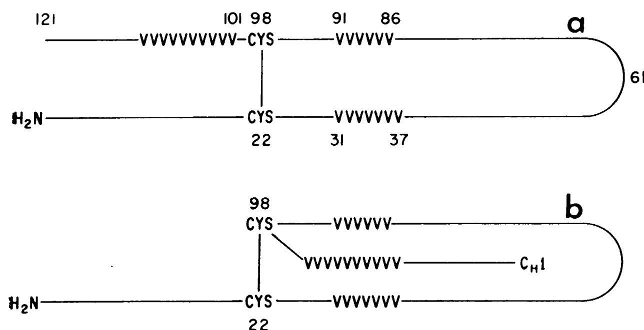


FIG. 4. (*a*) Approximation of the first two hypervariable areas in the heavy chain obtainable by folding the chain at the center of the first disulfide loop. (*b*) Approximation of all three hypervariable regions, close to the first disulfide bridge in the heavy chain.

The close linear relationship of the three regions to the disulfide bridge (within eight, six, and two residues, respectively) adds to the probability of the proposed steric relationship. Kabat (19) has previously pointed out that *light-chain* hypervariable regions 24–34 and 89–97 are brought close together by light-chain disulfide bond $I_{23}$–$II_{88}$.

The variable region of immunoglobulin heavy chains thus appears to contain a minimum of three hypervariable regions. Evidence for an association of particular amino-acid residues within hypervariable regions of light and heavy chains with particular antibody activities (5), strongly implies that these regions are very intimately involved in the expression of antibody specificity. Such involvement could very well include antigen contact points, light chain–heavy chain recognition sites, or other, unappreciated aspects of immunoglobulin function.

We thank Dr. Elvin Kabat for a number of informative discussions, Drs. Sam Murphy and Gerald Penn for plasma from patient *Was*, and Dr. Jack Ohms for advice concerning automated sequence methodology. Mrs. Vicki Ecker and Miss Bonnie Gerber provided excellent technical assistance. Supported in part by National Science Foundation Grant no. GB-17046 and NIH Grant no. 09810.


1. Milstein, C., *Nature*, **216**, 330 (1967).
2. Kabat, E. A., *Proc. Nat. Acad. Sci. USA*, **59**, 613 (1968).
3. Franěk, F., *Symposium on Developmental Aspects of Antibody Formation and Structure*, Prague, 1969.
4. Wu, T. T., and E. A. Kabat, *J. Exp. Med.*, **132**, 211 (1970).
5. Capra, J. D., J. M. Kehoe, R. J. Winchester, and H. G. Kunkel, *Ann. N.Y. Acad. Sci.*, in press.
6. Goetzl, E. J., and H. Metzger, *Biochemistry*, **9**, 3862 (1970).
7. Capra, J. D., *Nature*, **230**, 62 (1971).
8. Wang, A. C., H. H. Fudenberg, and J. R. L. Pink, *Proc. Nat. Acad. Sci. USA*, **68**, 1143 (1971).
9. Ray, A., and J. J. Cebra, *Fed. Proc.*, **30**, 349 (1971).
10. Capra, J. D., and H. G. Kunkel, *Proc. Nat. Acad. Sci. USA*, **67**, 87 (1970).
11. Gross, E., and B. Witkop, *J. Biol. Chem.*, **237**, 1856 (1962).
12. Edman, P., and J. Sjoquist, *Acta Chem. Scand.*, **10**, 1507 (1956).
13. Kohler, H., A. Shimizu, C. Paul, V. Moore, and F. W. Putnam, *Nature*, **227**, 1318 (1970).
14. (*a*) Edelman, G. M., B. A. Cunningham, W. E. Gall, P. D. Gottlieb, U. Rutishauser, and M. J. Waxdal, *Proc. Nat. Acad. Sci. USA*, **63**, 78 (1969).
(*b*) Press, E. M., and N. M. Hogg, *Biochem. J.*, **117**, 641 (1970).
(*c*) Cunningham, B. A., M. N. Pflumm, U. Rutishauser, and G. M. Edelman, *Proc. Nat. Acad. Sci. USA*, **64**, 997 (1969).
(*d*) Wikler, M., H. Kohler, T. Shinoda, and F. W. Putnam, *Science*, **163**, 75 (1969).
(*e*) Ponstingl, H., J. Schwarz, W. Reichel, and N. Hilschmann, *Hoppe-Seyler's Z. Physiol. Chem.*, **351**, 1591 (1970).
15. Fruchter, R. G., S. A. Jackson, L. E. Mole, and R. R. Porter, *Biochem. J.*, **116**, 249 (1970).
16. Bourgois, A., and M. Fougereau, *FEBS Lett.*, **8**, 265 (1970).
17. Milstein, C., and J. R. L. Pink, *Progr. Biophys. Mol. Biol.*, **21**, 211 (1970).
18. Hood, L., and D. W. Talmage, *Science*, **168**, 325 (1970).
19. Kabat, E. A., *Ann. N.Y. Acad. Sci.*, **169**, 43 (1970).