Proc. Nat. Acad. Sci. USA
Vol. 72, No. 12, pp. 5107–5110, December 1975
Immunology

Similarities among hypervariable segments of immunoglobulin chains

(immunoglobulin sequences/complementarity-determining segments)

TAI TE WU*, ELVIN A. KABAT†, AND HOWARD BILOFSKY‡

* Departments of Engineering Sciences and Biochemistry and Molecular Biology, Northwestern University, Evanston, Illinois 60201; † National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014; † Departments of Microbiology, Human Genetics and Development and Neurology, Columbia University, and the Neurological Institute, Presbyterian Hospital, New York, N.Y. 10032; and ‡ Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts 02138

Contributed by Elvin A. Kabat, October 7, 1975

ABSTRACT **A human λV (Mcg) and a human λII (Vil) myeloma protein have identical sequences in their first hypervariable segments although they differ at 21 positions throughout the variable region. If a different structural gene is responsible for each subgroup, the findings favor insertion of information for the hypervariable or complementarity-determining segments.**

The three hypervariable segments of the variable regions of light (1) and of the heavy (2, 3) chains of human and mouse immunoglobulins have been delineated from a statistical examination of sequences aligned for maximum homology and postulated to be the complementarity-determining residues (1, 2) of antibody combining sites. They comprise positions 24–34, 50–56, and 89–97 of the light chains with insertions of up to six residues 27A to 27F and up to four residues 97A to 97D (in the case of rabbit immunoglobulins) in the light chains, and positions 31–35, 50–65, and 95–102 in heavy chains, with possible insertions of two residues 35A and 35B, of three residues as 52A, 52B, 52C and as 82A, 82B, 82C (outside the hypervariable segments), and of up to eight residues 100A to 100H in the heavy chains. Affinity labeling studies of various antibodies and myeloma globulins have shown residues in all six hypervariable segments to be labeled (4). Moreover, x-ray diffraction studies of several groups have established the three-dimensional structure of the variable regions of a Bence Jones light chain dimer Mcg (5), of two Fab′ fragments, McPC 603 (6) and Newm (7), and of an Fv dimer REI (8). The Bence Jones dimer bound a number of low-molecular-weight substances. One of the Fab′ fragments was specific for phosphorylcholine and the other for a hydroxy derivative of vitamin K. No binding activity has been found for REI. In all three instances, the binding sites were formed by the hypervariable segments. However, the second hypervariable segment of the light chain in McPC 603 was shielded from the site by an insertion of six residues in the first hypervariable region, and that in Newm was removed from the site by a deletion of seven residues from position 55 to 61. The fourth region in the heavy chain showing hypervariability (2, 9), especially at positions 84 and 86, was not part of the combining site (10). It has been suggested that hypervariable segments are responsible for conferring idiotypic specificity (1, 10, 11). We shall refer to the three hypervariable regions of each chain making up the binding site as the complementarity-determining regions or segments.

Recently, we have begun to accumulate and store all available amino-acid sequences of immunoglobulin variable regions in the Prophet System (12). The stored data contain 95 completed first, 65 second, and 66 third complementarity-determining segments of 310 light chains, and 39 first, 25 second, and 21 third complementarity-determining segments of 133 heavy chains. It therefore became possible to look for similarities and differences in these segments in the various light and heavy chains.

Weigert and coworkers (13) previously noted that mouse lambda light chains were extremely restricted in sequence variability over the entire variable region, and thus many of their hypervariable segments were identical. Two mouse kappa light chains (14, 15) had identical sequences except for two differences in the first and one in the third complementarity-determining regions. Capra and Kunkel (16) found the antibodies in two cases of hypergammaglobulinemic purpura to have identical sequences from the NH_2-terminus through the first hypervariable segment. Similarly, three mouse kappa chains (17–19) of myeloma globulins binding phosphorylcholine were identical through residue 35. Their corresponding heavy chains as well as four other phosphorylcholine-binding mouse myeloma proteins are also identical through residue 36 including the first hypervariable segment, except for a Leu-Val replacement at position 4 in one of them. Two other human kappa chains were identical through residue 40 except for a Ser-Ala replacement at position 22 (9), and were identical to a third which has been sequenced up to residue 35 (20). Wang and his colleagues (21, 22) have shown in a patient with myeloma IgG2κ and a macroglobulin IgMκ protein, that the light chains of both were identical for the NH_2-terminal 37 residues, as were the variable regions of the heavy chains up to residue 34. The complementarity-determining segments showing such similarities were of the same subgroups, and thus can provide no insight as to whether the hypervariable segments are of separate genetic origin from the rest of the variable region (1).

Excluding such cases of nearly identical variable regions, Capra and Kehoe (23) have sequenced two heavy chains from different patients which showed cross idiotypic specificity and were both IgM and anti-IgG, and noted that their second and third hypervariable segments were identical and the first hypervariable segments differed only by one residue, while there were seven amino-acid substitutions in the other parts of the variable region. Margolies *et al.* (24) noted that three rabbit kappa chains from antibodies for streptococcal group C carbohydrate-4135 (25), for pneumococcal type VIII carbohydrate-3115 and for *Micrococcus lysodeikticus*-120 (24, 26) have the same second hypervariable segment Arg-Ala-Ser-Thr-Leu-Ala-Ser, and postulated that it might not be complementarity-determining. However, three type III antipneumococcal antibodies, BS-1, K-25, and 3381

Abbreviation: PCA, pyrrolidone carboxylic acid.

Table 1. Sequences of the variable regions of the two human λII and two human λV immunoglobulin light chains

Position	1. Vil	2. Nei	3. Bo	4. Mcg	Position	1. Vil	2. Nei	3. Bo	4. Mcg	Position	1. Vil	2. Nei	3. Bo	4. Mcg
1	His	PCA	PCA	PCA	35	Trp	Trp	Trp	Trp	75	Ile	Ile	Val	Val
2	Ser	Ser	Ser	Ser	36	Phe	Tyr	Tyr	Tyr	76	Ser	Ser	Ser	Ser
3	Ala	Ala	Ala	Ala	37	Gln	Gln	Gln	Gln	77	Gly	Gly	Gly	Gly
4	Leu	Leu	Leu	Leu	38	Gln	Gln	Gln	Gln	78	Leu	Leu	Leu	Leu
5	Thr	Thr	Thr	Thr	39	His	Asn	His	His	79	Gln	Gln	Arg	Gln
6	Gln	Gln	Gln	Gln	40	Pro	Pro	Pro	Ala	80	Ala	Val	Ala	Ala
7	Pro	Pro	Pro	Pro	41	Gly	Gly	Gly	Gly	81	Glu	Glu	Glu	Glu
8	Ala	Ala	Pro	Pro	42	Thr	Lys	Arg	Lys	82	Asp	Asp	Asp	Asp
9	Ser	Ser	Ser	Ser	43	Ala	Ala	Ala	Ala	83	Glu	Glu	Glu	Glu
10	—	—	—	—	44	Pro	Pro	Pro	Pro	84	Ala	Ala	Ala	Ala
11	Val	Val	Ala	Ala	45	Lys	Lys	Lys	Lys	85	Asp	Asp	Asp	Asp
12	Ser	Ser	Ser	Ser	46	Leu	Leu	Leu	Val	86	Tyr	Tyr	Tyr	Tyr
13	Gly	Gly	Gly	Gly	47	Ile	Met	Val	Ile	87	Tyr	Tyr	Tyr	Tyr
14	Ser	Ser	Ser	Ser	48	Ile	Ile	Ile	Ile	88	Cys	Cys	Cys	Cys
15	Leu	Pro	Pro	Leu	49	Ser	Tyr	Phe	Tyr	89	Ser	Cys	Ser	Ser
16	Gly	Gly	Gly	Gly	50	Glu	Glu	Glu	Glu	90	Ser	Ser	Ser	Ser
17	Gln	Gln	Gln	Gln	51	Val	Gly	Val	Val	91	Tyr	Tyr	Tyr	Tyr
18	Ser	Ser	Ser	Ser	52	Arg	Asn	Ser	Asn	92	Thr	Ala	Val	Glu
19	Ile	Ile	Val	Val	53	Asn	Lys	Gly	Lys	93	Ser	Gly	Asp	Gly
20	Thr	Thr	Thr	Thr	54	Arg	Arg	Arg	Arg	94	Ser	Asx	Asn	Ser
21	Ile	Ile	Ile	Ile	55	Pro	Pro	Pro	Pro	95	Asn	Ser	Asn	Asp
22	Ser	Ser	Ser	Ser	56	Ser	Ser	Ser	Ser	96	Ser	Thr	Asn	Asn
23	Cys	Cys	Cys	Cys	57	Gly	Gly	Gly	Gly	97	Val	Arg	Phe	Phe
24	Thr	Thr	Thr	Thr	58	Val	Val	Val	Val	97A	Val	Val	Val	Val
25	Gly	Gly	Gly	Gly	59	Ser	Ser	Pro	Pro	97B	—	—	—	—
26	Thr	Thr	Thr	Thr	60	Asp	Asn	Asp	Asp	97C	—	—	—	—
27	Ser	Thr	Ser	Ser	61	Arg	Arg	Arg	Arg	97D	—	—	—	—
27A	—	—	—	—	62	Phe	Phe	Phe	Phe	98	Phe	Phe	Phe	Phe
27B	—	—	—	—	63	Ser	Ser	Ser	Ser	99	Gly	Gly	Gly	Gly
27C	—	—	—	—	64	Gly	Gly	Gly	Gly	100	Gly	Gly	Gly	Thr
27D	Ser	Ser	Ser	Ser	65	Ser	Ser	Ser	Ser	101	Gly	Gly	Gly	Gly
27E	Asp	Asp	Asp	Asp	66	Lys	Lys	Lys	Lys	102	Thr	Thr	Thr	Thr
27F	Val	Val	Val	Val	67	Ser	Ser	Ser	Ser	103	Lys	Arg	Lys	Lys
28	Gly	Gly	Gly	Gly	68	Ala	Gly	Asp	Gly	104	Leu	Val	Leu	Val
29	Gly	Ser	Asp	Gly	69	Asn	Lys	Asn	Asn	105	Thr	Thr	Thr	Thr
30	Tyr	Tyr	Asn	Tyr	70	Thr	Thr	Thr	Thr	106	Val	Val	Val	Val
31	Asn	Asn	Lys	Asn	71	Ala	Ala	Ala	Ala	106A	Leu	Leu	Leu	Leu
32	Tyr	Phe	Tyr	Tyr	72	Ser	Ser	Ser	Ser	107	Gly	Ser	Arg	Gly
33	Val	Val	Val	Val	73	Leu	Leu	Leu	Leu					
34	Ser	Ser	Ser	Ser	74	Thr	Thr	Thr	Thr					

(24,27), have a Lys substitution for Arg at position 50 but are otherwise identical to the second hypervariable segment of 4135, 3315, and 120; there are many differences in the rest of the variable region.

From our collected data, it was observed that the first hypervariable segment of the two human lambda light chains Vil (28) and Mcg (29) which belong to λII (e.g., Nei) and λV (e.g., Bo) subgroups, respectively, are identical, though the rest of their variable regions differ by 21 amino-acid-residues (Table 1). Two human V_κI light chains, Ag (30) and Ni (31), have the same sequence of Asp-Ala-Ser-Asn-Leu-Glu-Thr for their second hypervariable segment, while their first and third complementarity-determining segments are different.

Fett and Deutsch (29) noted the similarity of the variable region of Mcg to the λII as well as the λV subgroup, 81% of the residues being identical to Vil (λII; λI in their nomenclature) and to Bo (λV; λIV in their nomenclature), but concluded from a comparison of the first 20 residues with various other λII and λV subgroup sequences that there was a closer resemblance to the latter. Since the two Bence Jones proteins Vil and Mcg came from different parts of the world and if they are indeed in different subgroups, the identity of the first hypervariable segments would be consistent with an insertional mechanism for incorporation of complementarity-determining segments (1, 9, 32).

Comparing the two λII sequences, Vil and Nei, and the two λV sequences, Bo and Mcg, with one another (Table 1), one finds 20 to 31 differences in residues between their variable regions (Table 2). Bo and Mcg show the lowest number of differences, 20, as compared with 21 to 31 for the others. If the same comparisons are made for the nonhypervariable regions, Bo and Mcg show only 11 differences while the others range from 14 to 17. This would tend to support the conclusion that Bo and Mcg are in the same subgroup and differ from Vil and Nei. Indeed, in the first 23 amino-acid resi-

Table 2. Triangular matrices for sequence differences in the variable regions of the four human λ myeloma proteins

Total variable regions

Vil	Nei	Bo	Mcg
0	27	24	21 Vil
	0	31	26 Nei
		0	20 Bo
			0 Mcg

Nonhypervariable region

Vil	Nei	Bo	Mcg
0	14(20)	14(18)	14(15) Vil
	0	17(22)	16(19) Nei
		0	11(14) Bo
			0 Mcg

Complementarity-determining segments

Vil	Nei	Bo	Mcg
0	13(7)	10(6)	7(6) Vil
	0	14(9)	10(7) Nei
		0	9(6) Bo
			0 Mcg

First complementarity-determining segment

Vil	Nei	Bo	Mcg
0	3(3).	3(3)	0(3) Vil
	0	5(4)	3(3) Nei
		0	3(3) Bo
			0 Mcg

Second complementarity-determining segment

Vil	Nei	Bo	Mcg
0	3(2)	2(1)	2(1) Vil
	0	3(2)	1(2) Nei
		0	2(1) Bo
			0 Mcg

Third complementarity-determining segment

Vil	Nei	Bo	Mcg
0	7(2)	5(2)	5(2) Vil
	0	6(3)	6(2) Nei
		0	4(2) Bo
			0 Mcg

See *text* for explanation of values in parentheses.

dues, the usual basis of subgroup classification, Bo and Mcg differ only by a Leu-Pro substitution at position 15, while Vil and Nei differ by a His-PCA (pyrrolidone carboxylic acid) difference at position 1 and a Leu-Pro substitution at position 15. However, Vil and Nei differ from Bo and Mcg by Ala-Pro, Val-Ala, and Ile-Val substitutions at positions 8, 11, and 19.

When one examines the hypervariable segments (Table 2), one notes that Vil and Mcg show only seven differences while the other pairs range from 9 to 14. In the first complementarity-determining segment, Vil and Mcg are identical for all 14 positions, while other pairs show three and one shows five differences. In the second hypervariable segment, pairs involving Vil, Bo, and Mcg all show two differences while the Mcg–Nei pair shows but one difference, and the Nei–Vil and Nei–Bo pairs show three differences. In the third complementarity-determining segment, Bo–Mcg shows four differences, Vil–Mcg and Vil–Bo five differences and the other pairs six or seven differences.

In the first hypervariable segment, comparisons of Vil–Bo as well as Bo–Mcg show them to differ at positions 29, 30, and 31; the Vil–Nei and Nei–Mcg pairs differ at positions 27, 29, and 32, while Nei and Bo show five differences at positions 27, 29, 30, 31, and 32.

In the second hypervariable region Vil, Bo, and Mcg differ at positions 52 and 53. The Vil–Nei and Nei–Bo pairs each differ at positions 51, 52, and 53, while Nei and Mcg differ only at position 51.

In the third hypervariable region Bo and Mcg differ at position 92, 93, 94, and 95; Bo and Vil at 92, 93, 94, 96, and 97; Mcg and Vil at 92, 93, 95, 96, and 97; Bo and Nei at 89, 92, 93, 95, 96; Mcg and Nei at 89, 92, 94, 95, 96, and 97.

Thus the differences among proteins in each of the complementarity-determining regions appear to be independent of subgroup.

Values in parentheses (Table 2) show results to be expected if the 21 differences were distributed randomly among the 111 residues of the variable region. In the first hypervariable region one would expect almost three differences ($14/111 \times 21 = 2.6$) between Vil and Mcg, while none were found. All values in Table 2 for each of the three hypervariable segments show a greater variation than expected or are close to expected. The Vil–Mcg identity in the first hypervariable region is thus quite striking.

Since the Mcg dimer is known to bind a variety of low-molecular-weight substances, it would be of interest to examine Vil as well as an Mcg–Vil dimer for binding properties. One would also like to know whether they would exhibit cross idiotypic specificity. Available x-ray studies show that the first complementarity-determining segment of the light chain contributes substantially to the binding site (10).

A similar comparison between two human $V_\kappa I$ light chains, Ag and Ni, shows 23 differences, 12 of which are in the nonhypervariable regions. Ten of the remaining 11 are in the first hypervariable segment with an insertion of four residues at positions 27A, 27B, 27C, 27D and differences at 28, 29, 31, 32, and 34, while their second hypervariable segments are identical, and their third hypervariable segments differ only by one residue at position 96.

As additional sequences of other myeloma globulins and antibodies become available, one should be able to select and study differences and similarities in antibody and idiotypic specificities in terms of sequenes of complementarity-determining regions.

This work is sponsored by the National Cancer Institute, National Institute of Allergy and Infectious Diseases, National Institute of Arthritis, Metabolic and Digestive Diseases, National Institute of General Medical Sciences, Division of Research Resources (Contract no. N01-RR-4-2147), National Institutes of Health. T.T.W. is a Career Research Development Awardee (5-K01-AI-70497), National Institutes of Health. Aided in part by a grant from the National Institutes of Health, 1-R01-GM21482-01, to T.T.W. and by a grant from the National Science Foundation, BMS-72-02219-A02, to E.A.K.

1. Wu, T. T. & Kabat, E. A. (1970) *J. Exp. Med.* **132**, 211–250.
2. Kabat, E. A. & Wu, T. T. (1971) *Ann. N.Y. Acad. Sci.* **190**, 382–392.
3. Capra, J. D. & Kehoe, J. M. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 845–848.
4. Givol, D. (1974) *Essays Biochem.* **10**, 73–103.
5. Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973) *Biochemistry* **12**, 4620–4631.
6. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 4298–4302.
7. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1974) *Proc. Nat. Acad. Sci. USA* **71**, 3440–3444.
8. Epp, O., Colman, P. M., Fehlhammer, H., Bode, W., Schiffer,

M. & Huber, R. (1974) *Eur. J. Biochem.* **45,** 513–524.
9. Capra, J. D. & Kehoe, J. M. (1975) *Adv. Immunol.* **20,** 1–40.
10. Davies, D. R., Padlan, E. A. & Segal, D. M. (1975) *Annu. Rev. Biochem.* **44,** 639–667.
11. Capra, J. D., Tung, A. S. & Nisonoff, A. (1975) *J. Immunol.* **114,** 1548–1554.
12. Raub, W. F. (1974) *Fed. Proc.* **33,** 2390–2392.
13. Weigert, M. G., Cesari, I. M., Yonkovich, S. J. & Cohn, M. (1970) *Nature* **228,** 1045–1047.
14. McKean, D., Potter, M. & Hood, L. (1973) *Biochemistry* **12,** 749–759.
15. McKean, D., Potter, M. & Hood, L. (1973) *Biochemistry* **12,** 760–771.
16. Capra, J. D. & Kunkel, H. G. (1970) *Proc. Nat. Acad. Sci. USA* **67,** 87–92.
17. Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W. & Hood, L. (1974) *Science* **183,** 962–964.
18. Rudikoff, S. & Potter, M. (1974) *Biochemistry* **13,** 4033–4038.
19. Chaplin, J. L., Rudikoff, S., Potter, M. & Davie, J. M. (1975) *J. Exp. Med.* **141,** 608–621.
20. Gergely, J., Wang, A. C. & Fudenberg, H. H. (1973) *Vox Sang* **24,** 432–440.
21. Pink, J. R. L., Wang, A. C. & Fudenberg, H. H. (1971) *Annu. Rev. Med.* **22,** 145–157.
22. Wang, A. C., Gergely, J. & Fudenberg, H. H. (1973) *Biochemistry* **12,** 528–534.
23. Capra, J. D. & Kehoe, J. M. (1974) *Proc. Nat. Acad. Sci USA* **71,** 4032–4036.
24. Margolies, M., Cannon, L. E., Strosberg, A. D. & Haber, E. (1975) *Proc. Nat. Acad. Sci. USA* **72,** 2180–2184.
25. Chen, K. C. S., Kindt, T. J. & Krause, R. M. (1975) *J. Biol. Chem.* **250,** 3289–3296.
26. Haber, E., Margolies, M. N., Cannon, L. E. & Rosemblatt, M. S. (1975) *Miami Winter Symp.* **9,** 303–338.
27. Jaton, J. C. (1975) *Biochem. J.* **147,** 235–247.
28. Ponstingl, H. & Hilschmann, N. (1971) *Z. Physiol. Chem.* **352,** 859–877.
29. Fett, J. W. & Deutsch, H. F. (1974) *Biochemistry* **13,** 4102–4114.
30. Titani, K., Shinoda, T. & Putnam, F. W. (1969) *J. Biol. Chem.* **244,** 3550–3560.
31. Shinoda, T. (1973) *J. Biochem.* **73,** 433–446.
32. Capra, J. D. & Kindt, T. J. (1975) *Immunogenetics* **1,** 417–427.