Variable Region Sequence of the Light Chain from a Waldenström's IgM with Specificity for Phosphorylcholine†

Walter F. Riesen* and Jean-Claude Jaton

ABSTRACT: The variable region sequence of the light chain from the human IgM FR with binding activity for phosphorylcholine has been determined. Automated Edman degradation was used for the whole chain and for a large cyanogen bromide fragment comprising the third hypervariable region and the entire constant part. The rest of the sequence was established by means of the "Dansyl-Edman" technique with tryptic peptides. The sequence of light chain FR can be assigned to the subgroup II of human light chains with which it shares 92% homology within the nonhypervariable (framework) residues. There is no apparent sequence homology between the variable region of the human light chain FR and the aminoterminal 41 residues of the light chains published so far from the mouse myeloma proteins TEPC 15, HOPC 8, S 107, and McPC 603 with phosphorylcholine binding activity. Recent data on the light chain of the phosphorylcholine binding mouse myeloma protein MOPC 167 (see Conclusion), however, indicate a considerable structural homology between the first hypervariable region of this murine protein and that of the human IgM FR, suggesting that both IgM FR and IgA MOPC 167 might have been selected by similar antigens.

Myeloma proteins produced by plasmocytomas in the BALB/c strain of mice have been shown to bind phosphorylcholine with relatively high incidence (Potter et al., 1973). Since by all available criteria these proteins behave like antibodies, they should provide excellent models for studies of the structural basis of antibody specificity. The analysis of the amino acid sequence of these mouse myeloma proteins is currently under way and partial data have been published (Barstad et al., 1974a; Rudikoff and Potter, 1974). The three-dimensional structure of one of these proteins, McPC 603, has been elucidated (Rudikoff et al., 1972; Padlan et al., 1973; Segal et al., 1974).

In man, antiphosphorylcholine activity has recently been detected in a Waldenström's IgM (Riesen et al., 1975). The phosphorylcholine-binding myeloma proteins in mouse and man offer the possibility to compare the variable (V)[1] domain of antibodies with the same activity in different species and are expected to shed light on the problem of the generation of antibody diversity.

We report here the amino acid sequence of the variable region (residues 1 to 117) of the light chain of the phosphorylcholine-binding IgM_{FR}.

Materials and Methods

Protein Characterization and Purification. The Waldenström's macroglobulin FR (IgM/κ) with activity for phosphorylcholine has been previously described (Riesen et al., 1975). The protein was purified by affinity chromatography on Sepharose–phosphorylcholine columns according to Chesebro and Metzger (1972).

Light Chain Preparation. IgM FR was reduced with 0.01 M dithiothreitol in 0.1 M Tris-HCl, pH 8.6, for 2 h at 37 °C, followed by alkylation with a 10% molar excess of iodoacetamide over dithiothreitol. The partially reduced and alkylated protein was dialyzed overnight against 5 M guanidine hydrochloride (Gdn·HCl). Heavy and light chains were separated on a column of Sephadex G-200 (2.5 × 200 cm) equilibrated with 5 M Gdn·HCl, pH 5.5. Complete reduction and alkylation with [2-^{14}C]iodoacetic acid was done by the method of O'Donnell et al. (1970).

Cyanogen Bromide (CNBr) Cleavage. Fully reduced and alkylated light chain was dissolved in 70% formic acid and CNBr was added at a 4:1 CNBr:protein (w/w) ratio. The reaction mixture was allowed to stand for 1 h at room temperature and 16 h at 4 °C. It was then diluted ten times with cold water and lyophilized.

Enzymic Digestions. Enzymic digestions with trypsin and α-chymotrypsin were carried out for 3 h at 37 °C by using a 1:100 ratio (w/w) of enzyme to protein solution of 1% NH_4HCO_3. Hydrolyses with carboxypeptidases A and B were conducted as recommended by Ambler (1967).

High-Voltage Paper Electrophoresis. Preparative high-voltage paper electrophoresis at pH 6.5 and 3.5 was done as described by Press et al. (1966). Peptides were detected by the ninhydrin–cadmium stain (Dreyer and Bynum, 1967) and those containing tyrosine, tryptophan, and arginine by specific stains (Smith, 1960). Radioactive peptides were revealed by radioautography using Kodak Royal Blue medical x-ray film. Elution of peptides from paper was done with 0.02 M NH_4OH.

Sequence Determination. The determination of sequences of small peptides by the "Dansyl-Edman" procedure was done as described (Gray, 1967). Amide residues were assigned whenever possible on the basis of electrophoretic mobility of peptides at pH 6.5 (Offord, 1966).

Automated sequence analyses of sections of the light chain and of large peptides were performed on a Beckman Model 890B sequencer, equipped with an undercut cup and N_2 flush. Protein or peptides were degraded by using the conventional Quadrol program (Edman and Begg, 1967). Alternatively, the volatile buffer *N,N'*-dimethylbenzylamine was employed for

† From the Basel Institute for Immunology, Postfach, 4005 Basel 5, Switzerland (J.-C.J.), and the Institute for Clinical and Experimental Cancer Research, University of Bern, Tiefenau Hospital, 3004 Bern, Switzerland (W.F.R.). *Received March 10, 1976.* Supported in part by the Swiss National Foundation for Scientific Research.

[1] Abbreviations used are: Gdn·HCl, guanidine hydrochloride; Pth, phenylthiohydantoin; V_L, variable region of light chain; dansyl, 1-dimethylaminonaphthalene-5-sulphonyl; Tris, tris(hydroxymethyl)aminomethane; PC, phosphorylcholine.

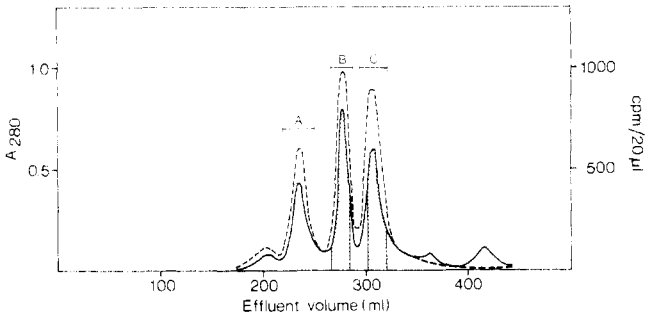

FIGURE 1: Elution profile of the CNBr cleaved reduced and 2-^{14}C-labeled alkylated light chain FR. The cleaved light chain (140 mg) was loaded on a Bio-Gel $A_{1.5m}$ column (2.5 × 200 cm) equilibrated and eluted with 5 M Gdn·HCl, pH 5.5. The peptide material under the hatched areas in peaks B and C was dialyzed against H_2O, lyophilized, and used for sequence analysis without any further purification.

peptides. The thiazolinone derivatives obtained after each degradation cycle were converted to the phenylthiohydantoin (Pth), by incubation in 1 N HCl at 80 °C for 10 min. The Pth-amino acids were extracted twice with ethyl acetate and dried under N_2. Pth-amino acids were identified by gas chromatography (Pisano and Bronzert, 1969) or by amino acid analyses after back hydrolysis to the free amino acids with 65% (w/v) HI for 20 h at 125 °C (Smithies et al., 1971). Half-cystine residues were identified as S-[^{14}C]carboxymethyl derivatives by liquid scintillation counting.

Results

Isolation and Characterization of Large Fragments from a CNBr Cleavage of [2-^{14}C] Carboxymethylated Light Chain FR. The amino acid composition of light chain FR indicated that it contains only 2 methionine residues/molecule. Advantage was taken of this fact by cleaving the light chain with CNBr. The components of the CNBr digest were separated by gel filtration into three major fractions: A, B, and C (Figure 1) which were all radioactive. Fractions A, B, and C were dialyzed against water to remove excess of salts and freeze-dried. They were recovered in a yield of 11, 46, and 43% respectively, on the basis of micromoles recovered per micromole of cleaved protein. Fraction A was shown by sequenator analysis to contain a polypeptide whose N-terminal sequence was identical with that of the native light chain. Its amino acid composition was very similar with that of the parent light chain FR and it therefore was assumed to represent uncleaved light chain. The peptide in fraction B (10 mg) was subjected to 23 cycles of automated Edman degradation (Figure 2). The amino acid composition of this peptide (not shown) and the sequence Lys-Arg-Thr-Val-Ala-Ala from positions 18–23, which is characteristic of the beginning of the C region in human κ chains (Dayhoff, 1972), identify this peptide as being the 95–214 residue fragment of light chain FR (see later).

The major N-terminal sequence of the peptide in fraction C was found to be the same as that of the native light chain. A minor sequence resulting from a partial split at Met-4 could be detected in about 20% yield. Fragment C is therefore the N-terminal 94-residue peptide.

The V region of L chain FR comprises CNBr fragment C (residues 1–94) and the N-terminal part of CNBr fragment B (residues 95 to about 113).

Sequence Determination of Fragment C (Residues 1–94). The sequence of the N-terminal 46 residues was determined by subjecting 300 nmol of intact light chain to automated sequence determination (Figure 2). The yields of the Pth derivatives at selected steps in the degradation of light chain FR are given in Figure 3. On the basis of 300 nmol of light chain FR, the absolute yield of Pth-Val at position 2 was 55% and a linear fall-off in yield can be seen throughout the sequencing run (repetitive yield: 94%). The rest of the primary structure of V_L FR was established by isolating tryptic peptides and sequencing them by the "Dansyl-Edman" procedure. Fragment C was digested with trypsin and the relevant peptides T_1–T_7 were isolated by high-voltage paper electrophoresis at pH 6.5 and/or 3.5. Their amino acid composition and mobility are given in Table I.

Peptides T_1 and T_4 are the only two radioactive peptides of the tryptic digest. Peptide T_1 contains homoserine and is therefore the C-terminal peptide of fragment C. The amino acid composition of peptide T_4 indicates 2.4 proline residues/molecule and a fractional methionine residue, not seen in any other tryptic peptides of the digest; this peptide contains about 21 amino acids and is assumed to represent the N-terminal section of light chain FR. A peptide was identified by electrophoresis at pH 3.5 which had a mobility slightly slower than neutral amino acids. This peptide, designated T_5, stained for Trp, Tyr, and Arg, and was contaminated by radioactive material (amino acid analysis not shown).

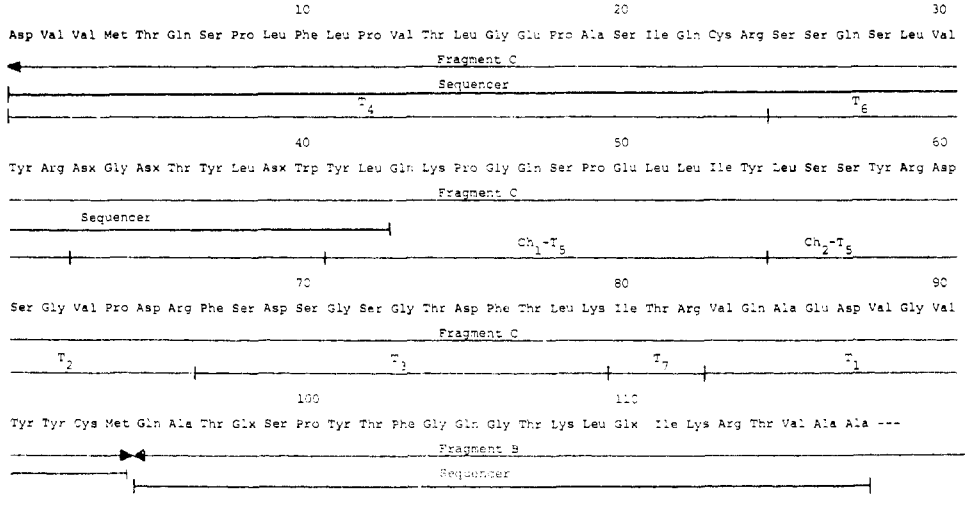

FIGURE 2: The primary structure of the V domain of light FR. The V_L region comprises CNBr fragment C (residues 1–94) and the N-terminal part of the CNBr fragment B (residues 95–117).

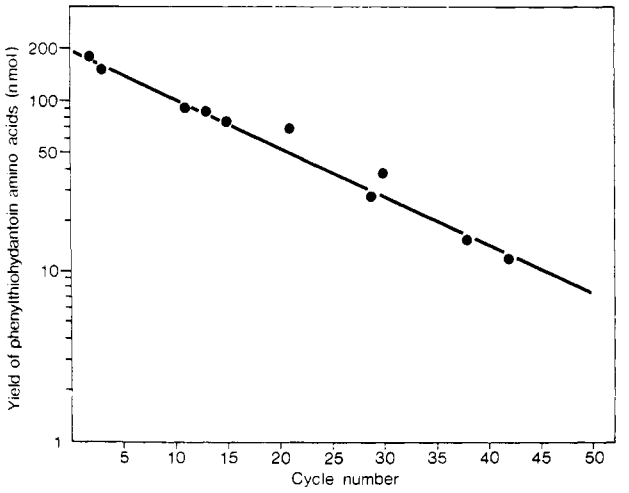

FIGURE 3: Quantitative yields of Pth-amino acids obtained at selected steps of the automated Edman degradation of light chain FR. Yields were computed by comparing peak heights of samples with those of relevant standard derivatives by gas chromatography.

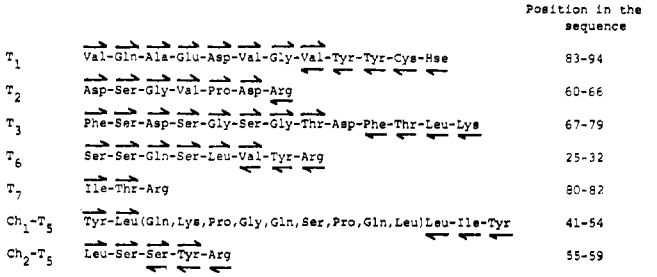

FIGURE 4: Amino acid sequences of some tryptic and chymotryptic peptides from CNBr fragment C of light chain FR. T and Ch are tryptic and chymotryptic peptides, respectively. (—) Amino acid residues identified by the "dansyl-Edman" procedure; (←) those identified by digestion with carboxypeptidases A and/or B. Alignment of peptides was done by homology with sequences of κ chains of subgroup II (Schneider and Hilschmann, 1975).

Peptide T_5 was eluted from paper in low yield and digested with chymotrypsin: two relevant peptides Ch_1-T_5 and Ch_2-T_5 were purified by high-voltage paper electrophoresis; their amino acid compositions are given in Table I. The amino acid sequences of the tryptic and chymotryptic peptides are reported in Figure 4.

For reasons of scarcity of material, no overlapping peptides could be prepared. The sequential order of the tryptic peptides T_2, T_3, T_4, and T_7 and of the chymotryptic peptides Ch_1-T_5 and Ch_2-T_5 relies therefore exclusively on homology with κ light chain sequences of subgroup II. These peptides comprise most of the nonhypervariable (framework) region; the extent of sequence homology of protein FR with V_κII basic sequence (Schneider and Hilschmann, 1975) is 92% (hypervariable regions excluded). Therefore, the ordering of these peptides within the sequence of light chain FR is most likely to be correct (Figure 2).

N-Terminal Sequence Determination of Fragment B (Residues 95–214). This was carried out by subjecting 10 mg of fragment B to automated Edman degradation. The N-terminal 23 residues were unequivocally identified (Figure 2). The absolute yield of Pth-Gln at step 1 was low (30%), an observation which is not unexpected by virtue of the possible cyclization of the glutamine residue into pyrrolidone carboxylic acid. The repetitive yield was 92% in two successive runs. This peptide comprises the third hypervariable section of the V region (residues 95 to about 113) and the entire C region.

Discussion

The present report describes the entire V region sequence of the light chain of the human IgM FR which displays binding activity for phosphorylcholine. A considerable number of human light chain sequences, notably of Bence Jones proteins, are known to date; however, with the exception of the sequence of protein New which binds vitamin K_1OH (Chen and Poljak, 1974), no entire V_L region of human immunoglobulins with defined activity has yet been reported.

TABLE I: Amino Acid Composition of Relevant Tryptic and Chymotryptic Peptides from CNBr Fragment C (Residues 1–94) of Light Chain FR.

Amino Acid	T_1	T_2	T_3	T_4	T_6	T_7	Ch_1-T_5	Ch_2-T_5
Lys			0.7 (1)				0.9 (1)	
Arg		0.9 (1)		0.8 (1)	0.9 (1)	1.0 (1)		0.9 (1)
Cm Cys[a]	0.5 (1)			0.6 (1)				
Asp	1.1 (1)	2.0 (2)	2.3 (2)	1.2 (1)			0.2 (0)	
Thr			2.1 (2)	1.2 (2)		1.1 (1)	0.4 (0)	
Ser		0.9 (1)	2.9 (3)	1.6 (2)	3.0 (3)		0.8 (1)	1.7 (2)
Glu	1.8 (2)			3.2 (3)	1.3 (1)		2.7 (3)	
Pro		0.8 (1)		2.4 (3)			1.5 (2)	
Gly	1.2 (1)	0.9 (1)	2.3 (2)	1.2 (1)			1.4 (1)	
Ala	1.0 (1)			1.4 (1)			0.4 (0)	
Val[b]	2.4 (3)	1.1 (1)		2.2 (3)	0.8 (1)			
Met				0.3 (1)				
Ile				0.7 (1)		0.7 (1)	0.6 (1)	
Leu			0.9 (1)	3.0 (3)	0.9 (1)		3.4 (3)	1.0 (1)
Tyr	2.2 (2)				1.0 (1)		1.3 (2)	0.8 (1)
Phe			1.7 (2)	0.7 (1)				
Hse	0.5 (1)							
Total residues	10.7	6.6	12.9	20.5	7.9	2.8	13.6	4.4
Mobility at pH 6.5	−0.71	−0.35	−0.20	−0.15	+0.36	+0.63	+0.23	+0.41

[a] Not corrected for losses during hydrolyses. [b] Based on 20-h hydrolysis. Values are residues/molecule of peptide. Values in parentheses are integral values confirmed by sequence analysis. Chymotryptic peptides Ch_1-T_5 and Ch_2-T_5 are derived from the tryptic peptide T_5. Mobilities are expressed relative to Asp (= −1.0) or to Arg (= +1.0); T and Ch, are tryptic and chymotryptic peptides, respectively.

TABLE II: Degree (%) of Homology between V_LFR and Human Subgroups V_κI, V_κII, V_κIII, and V_κIV and Partial V_L Sequences of PC-Binding Mouse Myeloma Proteins.

		Roy(V_κI)	Tew(V_κII)	Ti(V_κIII)	Len(V_κIV)	TEPC 15	HOPC 8	S 107	MOPC 603	MOPC 167
Framework positions	FR	62	92	70	79	55[a]	55[a]	55[a]	55[a]	57[b]
Hypervariable positions	FR	10	43	19	29	11[c]	11[c]	11[c]	11[c]	

[a] Positions 1–29. [b] Positions 1–23. [c] First hypervariable region (positions 30, 31a–g, 32, according to the numbering scheme of Schneider and Hilschmann, 1975).

TABLE III: Degree (%) of Homology between Hypervariable and Framework Positions in Human Light Chains of Different Subgroups.[a]

	Hypervariable Positions									
	V_κI			V_κII				V_κIII		V_κIV
Framework Positions	Roy	Rei	Gal	Tew	Cum	Mil	FR	Ti	B_6	Len
V_κI Roy		52	43	10	9	19	10	43	33	18
Rei	91		57	19	18	29	20	43	43	36
Gal	85	83		24	14	24	14	43	33	23
V_κII Tew	61	62	63		56	67	43	24	24	32
Cum	60	60	60	94		56	56	23	23	36
Mil	62	59	62	92	88		52	29	29	41
FR	60	61	60	92	89	88		19	24	29
V_κIII Ti	70	67	71	72	68	70	66		71	32
B_6	65	66	72	71	69	67	73	92		36
V_κIV Len	73	73	74	77	75	73	76	76	75	

[a] The following positions were considered as hypervariable: Positions 30, 31a–g, and 32 (first hypervariable region), positions 50–56 (second hypervariable region), and positions 91–96 (third hypervariable region). The numbering scheme is according to Schneider and Hilschmann (1975).

TABLE IV: Average of Extent of Homology (%) between Hypervariable and Framework Positions in Human Light Chains of Different Subgroups.

Framework Positions	Hypervariable Positions			
	V_κI	V_κII	V_κIII	V_κIV
V_κI	$\frac{51}{86}$	18	40	26
V_κII	61	$\frac{55}{93}$	24	35
V_κIII	70	70	$\frac{71}{92}$	34
V_κIV	73	75	76	

The amino acid sequence of light chain FR corresponds to the pattern of variation observed with other human κ chains (Schneider and Hilschmann, 1975), although a few substitutions not seen so far were found: Val-2, Phe-10, Glu-22, Asp-64, and Thr-76 (numbering scheme of Schneider and Hilschmann, 1975). The comparison of V_L FR with V_L of the subgroups I, II, III, and IV of human κ chains (Schneider and Hilschmann, 1975) clearly indicates that V_L FR belongs to the V_κII subgroup (Table II). Moreover, V_L FR resembles all of the known human V_κ subgroups more than the V_L regions of murine phosphorylcholine-binding immunoglobulins, at least over the N-terminal framework residues known so far (Barstad et al., 1974a). When this comparison is restricted to hypervariable positions (Wu and Kabat, 1970), light chain FR shows only 11% homology with light chains derived from PC-binding mouse immunoglobulins whereas 10–43% homology is found between light chain FR and light chains of subgroups I–IV from human immunoglobulins without known binding activity. The greatest homology (43%) is found with protein Tew belonging to the same subgroup (V_κII) as protein FR. This relatively high degree of homology is not unique to protein Tew, as an even higher extent of similarity is observed with two other proteins of subgroup II, i.e., Cum (56%) and Mil (52%). These data suggest a possible relationship between the sequence in the hypervariable regions and the framework sections of V_L domains. A detailed analysis of homologies between hypervariable regions and framework of human light chains of the same subgroup and of different subgroups is given in Tables III and IV. Within the same subgroup, hypervariable regions display a higher extent of homology than hypervariable regions from light chains of different subgroups. Evidence in favor of a direct association between subgroup and hypervariable or complementarity-determining region also comes from amino acid sequence data of immunoglobulins with different binding specificities (Barstad et al., 1974b). These authors showed that heavy chains from mouse myeloma proteins with the same binding activity belong to the same subgroup.

It is remarkable that human and mouse immunoglobulins with binding activity for phosphorylcholine share only 11% homology in a section of the light chain (first hypervariable region) which is part of the antigen binding site (Segal et al., 1974); the amino acid sequence of the entire V_L of the phosphorylcholine-binding murine immunoglobulins, however, has not yet been determined.

Crystallographic studies on the phosphorylcholine-binding

mouse myeloma protein McPC 603 have shown that the second hypervariable region of the light chain does not participate in the formation of the binding site because of the presence of the large loop including the first hypervariable region of this light chain (Padlan et al., 1973; Segal et al., 1974). It is relevant that the light chain of IgM_{FR} belongs to subgroup $V_\kappa II$ characterized by up to six insertions in the first hypervariable region (Schneider and Hilschmann, 1975) and that it is shorter only by one amino acid residue than the light chain of protein 603.

In addition, the three-dimensional structure of protein McPC 603 indicates that the phosphorylcholine hapten is bound asymmetrically in the cavity and seems to interact more with the heavy chain than with the light chain. Amino acid sequence analyses of the heavy chains of five phosphorylcholine-binding mouse immunoglobulins have shown that, with the exception of a single substitution in one chain, they share identical N-terminal sequences throughout the first hypervariable region (Barstad et al., 1974a). Recent findings on the N-terminal sequence of the heavy chain of IgM FR indicate that this chain differs from the mouse heavy chains by only four amino acids, three substitutions being located in the framework and only one—an aspartic acid instead of a glutamic acid—within the first hypervariable region (Riesen et al., 1976). Thus, in phosphorylcholine-binding immunoglobulins, a single aminoterminal heavy chain sequence may combine with several light chain sequences. A similar situation has recently been described by Capra et al. (1975) who found that at least four different light chains may be associated with a single heavy chain in mouse anti-*p*-azophenylarsonate antibodies bearing a cross-reactive idiotype.

These data could suggest that the hypervariable regions of heavy and light chains recognize different parts of an epitope.

Conclusion

Since the submission of this manuscript, the sequence of the light chain first hypervariable region of the phosphorylcholine-binding mouse myeloma protein MOPC 167 has become available (Potter, M., Padlan, E., and Rudikoff, S. (1976), *J. Immunol.* (in press)). The comparison of light chain FR and light chain MOPC 167 through the first hypervariable region reveals important similarities: (1) both hypervariable regions have the same length; they are one amino acid residue shorter than all other light chain first hypervariable regions of phosphorylcholine-binding immunoglobulins known so far, (2) six out of nine positions in the first hypervariable region are identical in both chains. These are both primary structural homologies and secondary structural similarities (same length; probably the loop has a similar configuration). While the light chain first hypervariable region is not directly implicated in phosphorylcholine binding, it is part of the combining site (Segal et al., 1974) and may be involved in antigen-carrier interaction: hence both IgM FR and IgA MOPC 167 might have been selected by similar antigens.

Acknowledgments

The authors thank R. Knecht for excellent technical assistance, Drs. D. G. Braun and J. R. L. Pink for reviewing the manuscript, and Drs. M. Potter, S. Rudikoff, and E. Padlan for making their sequence data available to us prior to publication.

References

Ambler, R. P. (1967), *Methods Enzymol. 11,* 135.

Barstad, P., Farnsworth, V., Weigert, M., Cohn, M., and Hood, L. (1974b), *Proc. Natl. Acad. Sci. U.S.A. 71,* 4096.

Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W., and Hood, L. (1974a), *Science 183,* 962.

Capra, J. D., Tung, A. S., and Nisonoff, A. (1975), *J. Immunol. 115,* 414.

Chen, B., and Poljak, R. J. (1974), *Biochemistry 13,* 1295.

Chesebro, B., and Metzger, A. (1972), *Biochemistry 11,* 766.

Dayhoff, M. O. (1972), Atlas of Protein Sequence and Structure of the National Biomedical Research Foundation.

Dreyer, W. J., and Bynum, E. (1967), *Methods Enzymol. 11,* 32.

Edman, P., and Begg, G. (1967), *Eur. J. Biochem. 1,* 80.

Gray, W. R. (1967), *Methods Enzymol. 11,* 469.

O'Donnell, I. J., Frangione, B., and Porter, R. R. (1970), *Biochem. J. 116,* 261.

Offord, R. E. (1966), *Nature (London) 211,* 591.

Padlan, E. A., Segal, D. M., Spande, T. F., Davies, D. R., Rudikoff, S., and Potter, M. (1973), *Nature (London), New Biol. 245,* 165.

Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244,* 5597.

Potter, M., Mushinski, E. B., Rudikoff, S., and Appella, E. (1973), Specific Receptors of Antibodies, Antigens and Cells, 3rd International Convocation of Immunology, Buffalo, N.Y., 1972, pp 270–284.

Press, E. M., Piggott, P. J., and Porter, R. R. (1966), *Biochem. J. 99,* 356.

Riesen, W., Rudikoff, S., Oriol, R., and Potter, M. (1975), *Biochemistry 14,* 1052.

Riesen, W. F., Braun, D. G., and Jaton, J.-C. (1976), *Proc. Natl. Acad. Sci. U.S.A. 73,* 2096.

Rudikoff, S., and Potter, M. (1974), *Biochemistry 13,* 4033.

Rudikoff, S., Potter, M., Segal, D. M., Padlan, E. A., and Davies, D. R. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,*

Schneider, M., and Hilschmann, N. (1975), *Hoppe-Seyler's Z. Physiol. Chem. 356,* 507.

Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M., and Davies, D. R. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 4289.

Smith, I. (1960), in Chromatographic and Electrophoretic Techniques, Vol. 1, Smith, I., Ed., New York, N.Y., Interscience, pp 96–97.

Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman J. C., and Ballantyne, D. L. (1971), *Biochemistry 10,* 4912.

Wu, T. T., and Kabat, E. A. (1970), *J. Exp. Med. 132,* 211.