Proc. Natl. Acad. Sci. USA
Vol. 73, No. 6, pp. 2096–2100, June 1976
Immunology

Human and murine phosphorylcholine-binding immunoglobulins: Conserved subgroup and first hypervariable region of heavy chains

(Waldenström's IgM/NH_2-terminal sequence/germ line variable-region genes)

WALTER F. RIESEN*†, DIETMAR G. BRAUN*, AND JEAN-CLAUDE JATON*

* Basel Institute for Immunology, Postfach, 4005 Basel 5, Switzerland; and † Institute for Clinical and Experimental Cancer Research, University of Bern, Tiefenau-Hospital, 3004 Bern, Switzerland

Communicated by Elvin A. Kabat, March 18, 1976

ABSTRACT **The NH_2-terminal 36 residues of the heavy chain and the NH_2-terminal 40 residues of the light chain from a human Waldenström's IgM with binding activity for phosphorylcholine (phosphocholine) are compared with the published sequences of five mouse IgA myeloma proteins with the same activity. An extensive structural similarity, i.e., 3 amino acid interchanges within framework residues, and one in the hypervariable region, is noted between the heavy chains of both species. The light chains, however, show a considerable diversity and, in contrast to the heavy chain, no correlation between the primary structure of the first hypervariable region and the binding specificity is apparent.**

The finding of a very similar heavy chain variable region in two different species that are separated by about 75 million years in evolution favors the concept of stable transmission of variable region genes throughout evolution.

The mechanism by which diversity of immunoglobulins is generated remains a controversial issue; essentially two alternative hypotheses have been proposed. According to one theory, there is one germ line gene for each variable region (1–3); the alternative theory proposes a small number of germ line genes that is amplified by somatic mechanisms, e.g., point mutation (4–6), recombination (7), and excision of DNA followed by repair with error (8, 9). The possibility has also been raised that the hypervariable regions (10, 11) are encoded by a large number of small separate segments of DNA ("episomes") that are inserted into a limited number of germ line genes (10, 12).

Amino acid sequence analyses of antigen-binding monoclonal immunoglobulins from BALB/c mice (13–16) and from man (17) and of antibodies against streptococcal polysaccharides induced in part-inbred rabbits (18) have indicated the repeated occurrence of V-region genes, which can be interpreted as favoring the multiple germ line hypothesis.

The identification of a Waldenström's macroglobulin with specificity for phosphorylcholine‡ (PC) has recently been described (19). In the present report the NH_2-terminal amino acid sequences of the heavy and light chains of this IgM are reported and compared with the structural data obtained by Barstad *et al.* (13) on five murine anti-PC IgA myeloma proteins. The intriguing observation from this study is the extraordinary degree of homology of the human μ chain within its amino-terminal 36 residues with murine α chains from PC-binding proteins.

MATERIALS AND METHODS

The properties and purification of IgM FR, which has binding activity for PC, have been described elsewhere (19). The protein was partially reduced and alkylated; heavy and light chains were separated by gel filtration on a column of Sephadex G-200 (2 × 200 cm) in 5 M guanidine hydrochloride.

Preparation of Cyanogen Bromide (CNBr) Fragments. The partially reduced and alkylated heavy chain was treated at a 2:1 CNBr/protein ratio(wt/wt) in the presence of 70% formic acid for 16 hr at 4°. The isolation and characterization of the CNBr peptides will be described elsewhere.

Amino Acid Sequence Determination. Heavy and light chains (300–400 nmol) were subjected to 38 and 40 cycles, respectively, of automated Edman degradation in a Beckman model 890B sequenator (20). The conventional Beckman Quadrol program was used for the light chain, whereas the heavy chain was degraded in the presence of dimethylbenzylamine buffer. The COOH-terminal sequence of the CNBr peptide 1-34 of the heavy chain was established by digestion with carboxypeptidases A and B (21). The CNBr peptide spanning residues 35–85 of the heavy chain was degraded in the sequenator after reaction with 4-sulfophenyl isothiocyanate (22).

The phenylthiohydantoin amino acid derivatives were identified by gas chromatography (23), liquid chromatography (24), in some cases by thin-layer chromatography (25), and by amino acid analysis after back hydrolysis to the free amino acids by HI (26). The repetitive yields varied between 93 and 95% for each sequenator run.

RESULTS

The amino acid sequence of the NH_2-terminal 38 residues of the FR μ chain is shown in Fig. 1. This sequence is compared with the prototype sequences of subgroups I, II, and III of human heavy chain variable regions (V_H) (27) and with the NH_2-terminal sequences of 5 α chains derived from PC-binding mouse myeloma proteins (HOPC 8, TEPC 15, S 107, MOPC 603, and MOPC 167) (13). This comparison clearly reveals that the V_H region of IgM FR belongs to the V_H III subgroup. V_HFR differs from the V_HIII prototype sequence by a single amino acid substitution in position 6, where an aspartic acid was found instead of a glutamic acid.

All the murine α chains shown in Fig. 1 also belong to the V_HIII subgroup; differences from the basic V_HIII sequence are noted in two positions: a Lys/Gln substitution in position 3 [most murine V_HIII sequences have Lys in this position (28)], and a Thr/Ala exchange in position 24. V_HFR differs in its framework residues from the PC-binding mouse myeloma proteins in three positions (Gln/Lys interchange in position 3, Asp/Glu substitution in position 6, and Ala/Thr exchange in position 24).

The amino acid sequence of V_HFR given here includes the first hypervariable region (residues 31–35). For protein MOPC 603 with activity against PC, it was shown that the first hypervariable region is directly involved in hapten binding (29).

Abbreviations: PC, phosphorylcholine; V, variable, and C, constant region of immunoglobulin chain; H, heavy, and L, light immunoglobulin chain.

‡ The Commission on Biochemical Nomenclature recommends the name *phosphocholine* for this compound.

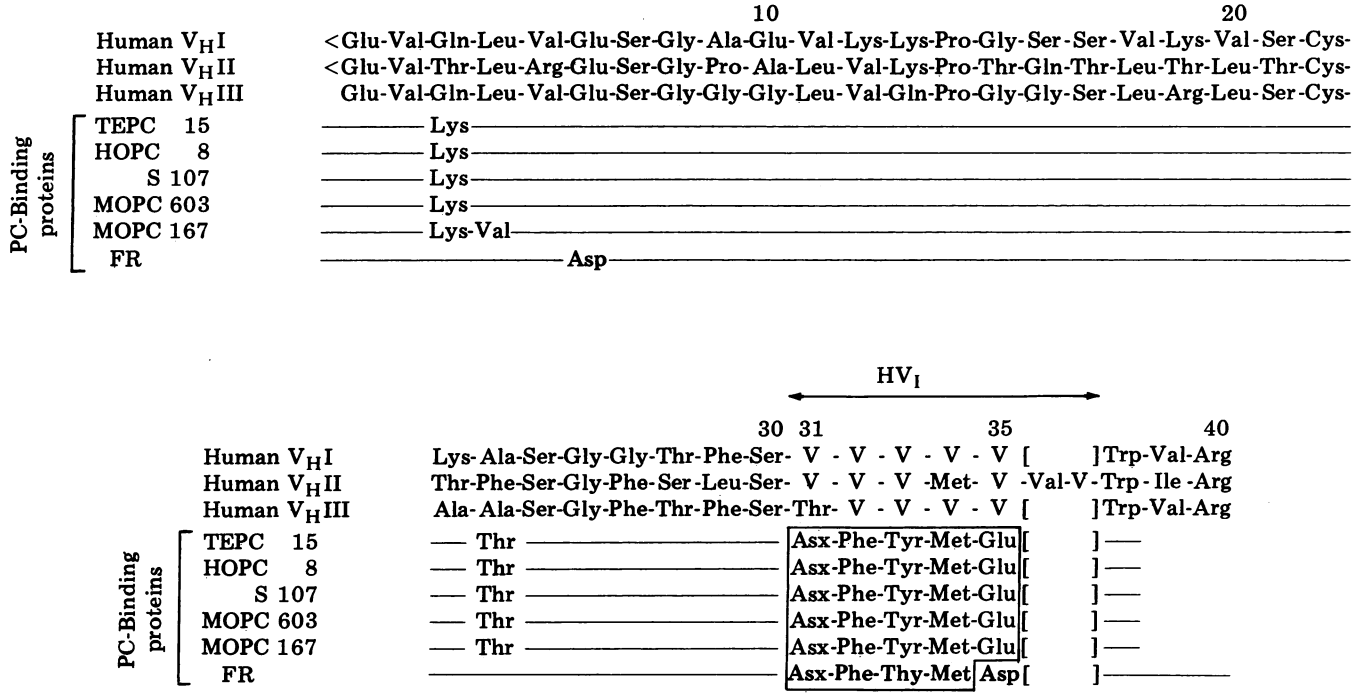

FIG. 1. NH_2-terminal sequences of the heavy chains of IgM FR, of five PC-binding IgA mouse myeloma proteins, and of the human V_H subgroups I, II, and III. Pyrrolidone carboxylic acid is symbolized by <Glu. A solid line represents identity with the V_HIII subgroup. Brackets, [ ], indicate deletions. V indicates variable residue positions. HV_I marks the extent of the first hypervariable region. The data for human V_H subgroups are from ref 27; for the murine myeloma proteins TEPC 15, HOPC 8, S 107, MOPC 603, and MOPC 167, from ref. 13. Identical residues within the hypervariable region have been framed.

Sequence analysis of V_HFR reveals identity within this region between murine and human proteins, with the exception of residue position 35, where an aspartic acid was found instead of a glutamic acid. This amino acid substitution would represent a single base exchange at the DNA level.

Comparisons among the NH_2-terminal sequence of V_HFR, the five PC-binding mouse myeloma proteins, and two Waldenström's IgM proteins (Pom and Lay) with anti-IgG activity (30) that belong to the V_HIII subgroup are given in Table 1. Within the framework section (positions 1–30) as much as 90% homology is found between V_HFR and the murine NH_2-terminal V_H regions. Homology (97% or 93%) is observed between the amino-terminal 30 residues of IgM FR and IgM Pom or IgM Lay, respectively. A high extent of homology (four identical residues out of five) is also noted between the first hypervariable regions (positions 31–35) of IgM FR and the mouse IgA proteins. The two human IgM proteins with different specificity, however, share only one residue out of five with IgM FR or with the murine proteins in this region.

Table 1. Homology between heavy chain NH_2-terminal regions
(expressed in % for positions 1–36 and 1–30 and as number of shared residues for positions 31–35)

	FR	H_8	T_{15}	S_{107}	M_{603}	M_{167}
Positions 1–36						
FR	—	89	89	89	89	86
H_8	89	—	100	100	100	97
Pom	86	83	83	83	83	81
Lay	83	81	81	81	81	78
Positions 1–30						
FR	—	90	90	90	90	87
H_8	90	—	100	100	100	97
Pom	97	93	93	93	93	90
Lay	93	90	90	90	90	87
V_HIII	97	93	93	93	93	90
Positions 31–35 (hypervariable region)						
FR	—	4	4	4	4	4
H_8	4	—	5	5	5	5
Pom	1	1	1	1	1	1
Lay	1	1	1	1	1	1

In contrast to this extraordinary degree of homology throughout the first hypervariable regions of human and murine heavy chains with the same activity, a considerable diversity was observed when the NH_2-terminal sequences of the light chains were compared. The sequences of the 40 NH_2-terminal residues (positions 1–35, according to the numbering scheme in ref. 31) of the light chain of IgM FR, (V_LFR), of the five mouse immunoglobulins with anti-PC activity and the prototype sequence of the human V_K subgroups I–IV (31) are given in Fig. 2. The sequence of the NH_2-terminal 29 residues places the light chain FR into the human V_KII subgroup, although differences at positions 2 (Val/Ile), 10 (Phe/Ser), and 15 (Leu/Pro) are present.

The comparative analyses of the different light chains of PC-binding human and mouse proteins and of the Bence-Jones protein Cum (32), which belongs to the V_KII subgroup, are given in Table 2. Homology (80%) is observed between the framework residues (positions 1–29) of IgM FR and of protein Cum. The comparison of the V_LFR with the murine light chains however, reveals only about 50% homology. A similar extent

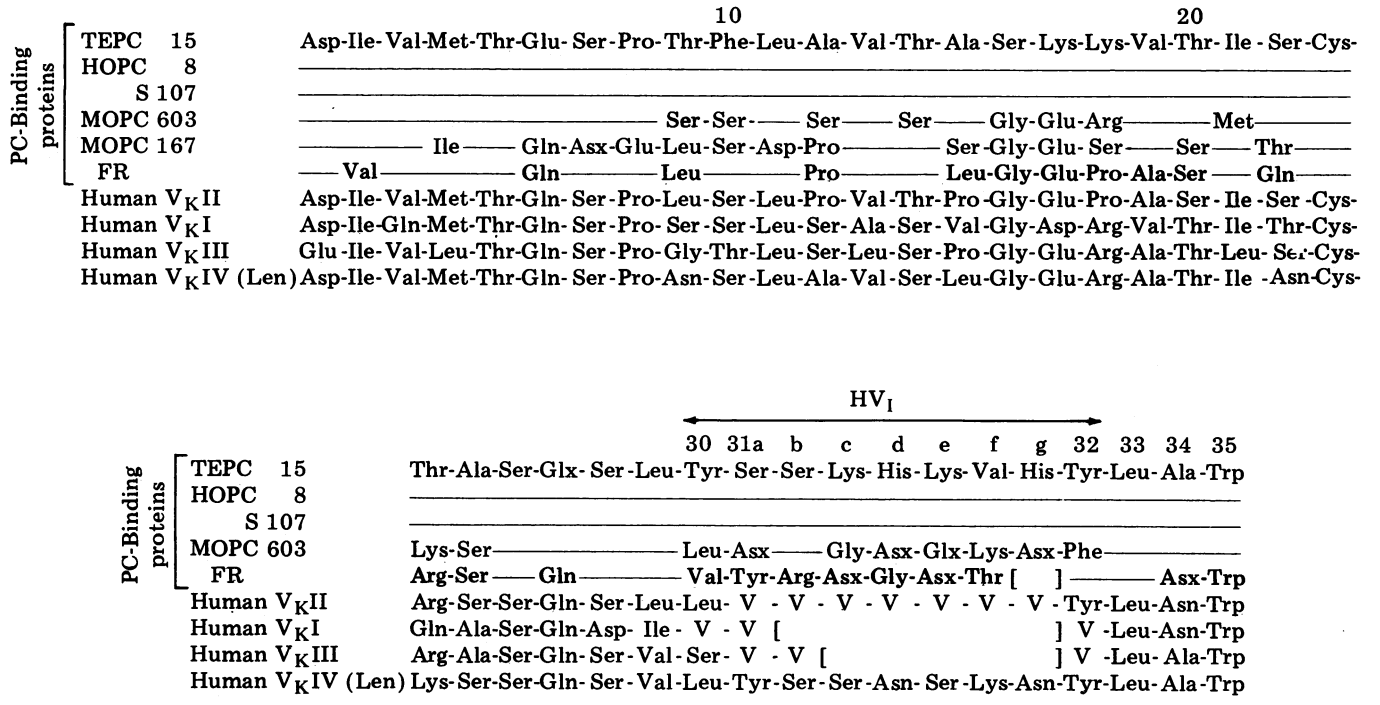

FIG. 2. NH_2-terminal sequences of the light chains of IgM FR, of five PC-binding IgA mouse myeloma proteins, and of the human V_K subgroups I, II, III, and IV. Brackets, [ ], indicate deletions. A solid line indicates identity with the first sequence listed. The data for human V_K subgroups are from ref. 31, for the murine proteins, from ref 13.

of homology is found between V_L from PC-binding mouse proteins of different idiotypes. Light chains from murine proteins with the same idiotype are identical in their framework residues and also within the first hypervariable region (positions 30, 31a–g, and 32). When these hypervariable positions are compared in proteins of different idiotypes (H_8 and M_{603}), only 11% homology is observed. A similar situation is found for V_LFR, which shows 0% or 11% homology with the mouse proteins. A degree of homology of 56% is found between the hypervariable regions of protein FR and Cum.

Table 2. Degree (%) of homology between light chain NH_2-terminal regions

Positions 1–35

	FR	H_8	T_{15}	S_{107}	M_{603}
FR	—	46	46	46	46
H_8	46	—	100	100	56
M_{603}	46	56	56	56	—
BJ Cum (V_KII)	76	48	48	48	62

Positions 1–29

	FR	H_8	T_{15}	S_{107}	M_{603}	M_{167}
FR	—	55	55	55	55	57*
H_8	55	—	100	100	66	39*
M_{603}	55	66	66	66	—	43*
M_{167}	43*	39*	39*	39*	43*	—
BJ Cum (V_KII)	80	53	53	53	60	62*

Positions 30,31a,b,c,d,e,f,g, and 32 (hypervariable region)

	FR	H_8	T_{15}	S_{107}	M_{603}
FR	—	11	11	11	0
H_8	11	—	100	100	11
M_{603}	11	11	11	11	—
BJ Cum (V_KII)	56	22	22	22	56

Asx was taken identical to Asp or Asn, Glx identical to Glu or Gln. H_8, M_{603}, and M_{167} proteins do not share idiotypic specificities. H_8, T_{15}, and S_{107} have identical idiotypes.
* Only residues 1–23.

DISCUSSION

In the present report, human and murine immunoglobulins with binding activity for PC have been found to share extraordinary structural similarities within their NH_2-terminal V_H regions. For instance, comparative analysis of residue positions 1–36 shows 86–89% homology (Table 1); this comparison includes both framework and hypervariable region residues. Within framework residues protein FR displays 90% homology to the mouse heavy chains and 97% to the human V_HIII prototype sequence; this clearly indicates that both V_HFR and the V_H region of PC-binding mouse myeloma proteins belong to the same subgroup. This finding is consistent with the suggestion that subgroups evolved before speciation of mammals (27, 33).

An apparent correlation between particular antigen binding specificities and heavy chain subgroups has also been observed by Barstad *et al.* (28) in mouse myeloma proteins with different binding activities. According to recent computer analysis on various segments of 310 light and 133 heavy chains, however, the complementarity-determining (hypervariable) regions appear to be independent of subgroup (34). The comparison of the first hypervariable regions of heavy chains derived from PC-binding immunoglobulins of mouse and man reveals as many as four identical residues out of five, whereas only one identical residue out of five is observed between IgM FR and two Waldenström's macroglobulins of the same V_H subgroup but with a different specificity (anti-IgG). This suggests a close relationship between primary structure and specificity for the first hypervariable region of the heavy chain in PC-binding immunoglobulins. The only substitution between mouse and human heavy chains in this region is the Asp/Glu exchange in

position 35; the presence of an aspartic acid is expected to confer a slightly higher degree of acidity on the combining site. Glu-35 and Tyr-33 have been shown for the heavy chain of the mouse myeloma protein MOPC 603 to make contact with the PC hapten (29). This implies that the requirement for the contact amino acid in position 35 is merely its acidity, provided that it is located within an appropriate environment.

The corresponding evaluation of the NH_2-terminal light chain sequences indicates a much greater diversity (Table 2). It is remarkable that the first hypervariable regions of the light chains Cum (32) and the mouse immunoglobulins binding PC are more homologous than the corresponding V_L region of protein FR and those of the murine proteins.

It thus appears for the light chains of the studied proteins that, in contrast to their heavy chains, there is no obvious relationship between primary structure and specificity within the first hypervariable region; this might suggest that the heavy chain plays a more important role in determining this specificity, a concept already proposed by Barstad *et al.* (13). Crystallographic studies on the PC-binding mouse IgA myeloma protein MOPC 603 indeed have indicated that the PC hapten is bound asymmetrically in the cavity and interacts more with the heavy chain than with the light chain (29). Similarly, structural studies on antibodies to *p*-azophenylarsonate induced in mice (35) indicated that antibody populations with shared idiotypic specificities and an identical sequence of the NH_2-terminal one-third of the heavy chains contain at least four different light chains. This, however, is not a general situation. It has been shown, for example, that mouse myeloma proteins with anti-$\alpha(1 \rightarrow 3)$-dextran activity were extremely restricted in sequence variability over the entire V_L region (14); identical NH_2-terminal light chain sequences through the first hypervariable region have also been found in two human IgG anti-γ-globulins (17) and in rabbit antibodies against streptococcal polysaccharides (18). There are, in addition, examples of mouse myeloma proteins that bind $\beta(1 \rightarrow 6)$D-galactan in which both heavy and light chains are similar (36).

Apart from the extraordinary extent of sequence homology of the NH_2-terminal V_HFR and the murine V_H regions, evidence suggesting a considerable extent of structural similarity throughout the entire V domain of the human and murine PC-binding immunoglobulins comes from xenogeneic chain recombination experiments. Recombinants prepared from the polypeptide chains of IgM FR and the mouse IgA myeloma protein TEPC 15 recovered as much as 41% of the binding activity of homologous recombinants (W. F. Riesen, *J. Immunol.*, in press) although these two proteins do not share idiotypy (H. Cosenza and W. F. Riesen, unpublished data). Carson and Weigert (37) demonstrated that as few as three substitutions in the light chain V regions of mouse IgA myeloma proteins with anti-$\alpha(1 \rightarrow 3)$dextran activity caused a loss of about 50% of the binding activity of homologous recombinants; this indicates that even a small number of amino acid substitutions may alter the specificity of the antigen-binding site.

The finding of similar V regions associated with different C regions is in accordance with previous data based on sequence analysis (38, 39) and on idiotype studies (40) and is consistent with the accepted view that at least two genes are coding for one polypeptide chain (41).

The NH_2-terminal 36 residues of the human μ chain FR document the extraordinary fact that within an evolutionary span of about 75 million years (42) a sequence was preserved that differs from the mouse heavy chain sequences by only four amino acid substitutions. Three of these are located in framework positions, one within the hypervariable region. With regard to this finding, theories suggesting somatic mutation restricted to the complementarity-determining regions (5) will have to propose a mechanism by which such similarity is either maintained throughout evolution or newly created. The probability for parallel somatic mutations to occur in a random sample of one species and in five random samples of another species over such a long evolutionary time seems to be low.

Two of the three substitutions in the framework of the heavy chain FR are in line with the pattern of human V_HIII; the additional exchange in position 6 has not been observed in any other V_HIII region. It is possible that the aspartic acid in this position may have evolved in human V_HIII sequences specific for PC. If this were true, other examples of human antibodies specific for PC with heavy chains belonging to the V_HIII subgroup should display this substitution. In view of the sequence identity within the NH_2-terminal 36 residues in the heavy chains from mouse proteins with PC-binding activity, a possible linkage of residue Asp-6 with Asp-35 would be of great value in supporting one or the other theories that deal with the origin of antibody diversity.

The authors thank Dr. G. Frank (Eidgenössische Technische Hochschule, Zürich) for help in analyzing by liquid chromatography and W. Hochstrasser, W. Hunkeler, and R. Knecht for excellent technical assistance. This work was supported in part by the Swiss National Foundation for Scientific Research.

1. Hood, L. & Talmage, D. W. (1970) *Science* **168**, 325–334.
2. Smith, G. P., Hood, L. & Fitch, W. M. (1971) *Annu. Rev. Biochem.* **40**, 969–1012.
3. Hilschmann, N., Ponstingl, H., Watanabe, S., Barnikol, H. U., Baczko, K., Braun, M. & Leibold, W. (1972) *FEBS Symp.* **26**, 31–48.
4. Lederberg, J. (1959) *Science* **129**, 1649–1653.
5. Cohn, M. (1971) *Ann. N.Y. Acad. Sci.* **190**, 529–584.
6. Jerne, N. K. (1971) *Eur. J. Immunol.* **1**, 1–9.
7. Gally, J. A. & Edelman, G. M. (1972) *Annu. Rev. Genet.* **6**, 1–46.
8. Brenner, S. & Milstein, C. (1966) *Nature* **211**, 242–243.
9. Baltimore, D. (1974) *Nature* **248**, 409–411.
10. Wu, T. T. & Kabat, E. A. (1970), *J. Exp. Med.* **132**, 211–250.
11. Capra, D. & Kehoe, M. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 845–848.
12. Capra, J. D. & Kindt, T. J. (1975) *Immunogenetics* **1**, 417–427.
13. Barstad, P., Rudikoff, S., Potter, M., Cohn, M., Konigsberg, W. & Hood, L. (1974) *Science* **183**, 962–964.
14. Weigert, M. G., Cesari, I. M., Yonkovich, S. J. & Cohn, M. (1970) *Nature* **228**, 1045–1047.
15. McKean, D., Potter, M. & Hood, L. (1973) *Biochemistry* **12**, 749–759.
16. McKean, D., Potter, M. & Hood, L. (1973) *Biochemistry* **12**, 760–771.
17. Capra, J. D. & Kunkel, H. G. (1970) *Proc. Natl. Acad. Sci. USA* **67**, 87–92.
18. Braun, D. G., Huser, H. & Jaton, J.-C. (1975) *Nature* **258**, 363–365.
19. Riesen, W., Rudikoff, S., Oriol, R. & Potter, M. (1975) *Biochemistry* **14**, 1052–1057.
20. Edman, P. & Begg, C. (1967) *Eur. J. Biochem.* **1**, 80–91.
21. Ambler, R. P. (1967) in *Methods in Enzymology*, ed. Hirs, C. W. (Academic Press, New York), Vol. 11, pp. 135–166.
22. Braunitzer, G., Schrank, B. & Ruhfus, A. (1970) *Hoppe-Seyler's Z. Physiol. Chem.* **351**, 1589–1590.
23. Pisano, J. J. & Bronzert, T. J. (1969) *J. Biol. Chem.* **244**, 5597–5607.
24. Frank, G. & Strubert; W. (1973) *Chromatographia* **6**, 522–524.
25. Boigné, J. M., Boigné, N. & Rosa, J. (1970) *J. Chromatogr.* **47**, 238–246.
26. Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G. & Ballantyne, D. L. (1971) *Biochemistry* **10**, 4912–4921.

27. Capra, J. D. & Kehoe, J. M. (1975) *Adv. Immunol.* **20,** 1–40.
28. Barstad, P., Farnsworth, V., Weigert, M., Cohn, M. & Hood, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4096–4100.
29. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4298–4302.
30. Capra, J. D. & Kehoe, J. M. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4032–4036.
31. Schneider, M. & Hilschmann, N. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356,** 507–557.
32. Hilschmann, N. (1967) *Hoppe-Seyler's Z. Physiol. Chem.* **348,** 1718–1722.
33. Bourgois, A. & Fougereau, M. (1970) *FEBS Lett.* **8,** 265–269.
34. Wu, T. T., Kabat, E. A. & Bilofsky, H. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 5107–5110.
35. Capra, J. D., Tung, A. S. & Nisonoff, A. (1975) *J. Immunol.* **115,** 414–418.
36. Rudikoff, S., Mushinski, E. B., Potter, M., Glaudemans, C. P. J. & Jolley, M. E. (1973) *J. Exp. Med.* **138,** 1095–1106.
37. Carson, D. & Weigert, M. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 235–239.
38. Köhler, H., Shimizu, A., Paul, C., Moore, V. & Putnam, F. W. (1970) *Nature* **227,** 1318–1320.
39. Wang, A. C., Pink, J. L. R., Fudenberg, H. H. & Ohms, J. (1970) *Proc. Natl. Acad. Sci. USA* **66,** 657–663.
40. Oudin, J. & Michel, M. (1969) *J. Exp. Med.* **130,** 619–642.
41. Hood, L. (1972) *Fed. Proc.* **31,** 177–187.
42. Dayhoff, M. O., ed. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C.), Vol. 5.