Proc. Nat. Acad. Sci. USA
Vol. 71, No. 4, pp. 1123–1127, April 1974

Amino-Acid Sequence of the Variable Region of the Heavy (Alpha) Chain of a Mouse Myeloma Protein with Anti-Hapten Activity

(CNBr cleavage/alpha immunoglobulin chains/anti-dinitrophenyl activity)

SHARRON H. FRANCIS*, R. GRAHAM Q. LESLIE†, LEROY HOOD‡, AND HERMAN N. EISEN§

Department of Microbiology, Washington University School of Medicine, St. Louis, Missouri; and ‡Division of Biology, California Institute of Technology, Pasadena, Calif.

Contributed by Herman N. Eisen, November 26, 1973

ABSTRACT **Cyanogen bromide cleavage of the heavy (alpha) chain of protein 315 (an immunoglobulin A mouse myeloma protein with anti-dinitrophenyl activity) yielded five fragments of which one (CN2), with 156 residues, contained the chain's entire variable region. Determination of the amino-acid sequence of CN2 showed that: (1) the variable region has appreciable homology (about 33% identities) with the variable region of the light chain from the same molecule; and (2) the constant-region sequence immediately following the probable transition from variable to constant domains is the same in the protein-315 α as in human γ1 and μ chains (-Val-Ser-Ser-). The sequence of the cyanogen bromide octapeptide (CN5) from the carboxy terminus of the protein-315 heavy chain closely resembles the corresponding segments of human α and μ chains.**

Multiple lines of evidence show that the ligand-binding specificities of antibody molecules are determined by about 107 residues at the amino terminus of light chains and about 115 at the amino terminus of heavy chains (called V_L and V_H regions, respectively) (1–3). The complete amino-acid sequences for V_L and V_H of the same molecule have been established for two myeloma proteins (4–6), but their ligand-binding specificities are not known.

The purpose of this paper is to report the amino-acid sequence of the variable region of the heavy chain (H^{315}) from protein 315. This sequence and that previously reported for the light chain from the same molecule (7) provide an essentially complete primary structure for the V_H and V_L regions of protein 315. The results are of particular interest because this protein, an IgA myeloma protein produced by mouse plasmacytoma MOPC-315, is known to have ligand-binding sites that resemble in many ways those of antibodies against Dnp and Tnp raised by conventional immunization procedures (8). For instance, protein 315 binds Dnp and Tnp ligands with high affinity, distinguishes sharply among ligands of similar structure (e.g., between 2,4- and 2,6-Dnp groups), and exhibits several of the characteristic spectral changes undergone by ligand and immunoglobulin in the immune complex. In addition, the specific-binding reactions of protein 315 are enthalpy-driven and show the same "strange" crossreactions between Dnp, 5-acetouracil, and menadione (2-methyl-1,4-naphthoquinone; vitamin K_3) that are also found with many conventionally raised antibodies against Dnp (8).

Protein 315 was purified from serum of BALB/c mice bearing the MOPC-315 tumor, and its fully reduced, aminoethylated chains were isolated as described (9). The fragments obtained from cyanogen bromide degradation of the heavy chain were subjected to sequence analysis in the Beckman sequenator and by the dansyl-Edman procedure (10, 11). Several of the fragments were degraded further with trypsin, chymotrypsin, and thermolysin into smaller peptides, which were purified by ion exchange chromatography on Dowex resins (12), gel filtration, paper chromatography, and high-voltage electrophoresis. The amino-acid sequences of these peptides were determined with the aid of dansyl-Edman degradation and by digestion with carboxypeptidases A and B. Analysis of heavy chains from intact molecules that had been labeled with a specific affinity-labeling reagent ([^{14}C]bromoacetyl-Dnp-L-lysine) aided in identification of the fragment with the variable region (13).

Cyanogen bromide cleavage of the fully reduced, aminoethylated heavy chain yielded five fragments that were purified largely by gel filtration on Sephadex (G-50 and G-100) in 5 M guanidine·HCl (Fig. 1). This report focuses on the 2nd and 5th fragments, CN2 and CN5, which correspond, respectively, to a large amino-terminal (variable) segment and a small carboxy-terminal (constant) segment of the chain. The other fragments, all derived from within the constant region, will be described elsewhere: here we note only that CN1 has approximately 228 residues, CN3 is a mixture of peptides, including one with carbohydrate, and CN4 is an octapeptide.

CN2 contains the entire V_H region and the adjoining portion of the constant region, including the latter's first Cys peptide. It has 156 amino-acid residues and lacks carbohydrate. End-group analysis by the cyanate (R. A. Bradshaw, unpublished results) and methylisothiocyanate (10) procedures showed that the amino-terminal residue in both this fragment and in the intact heavy chain is aspartic acid. When the whole chain and CN2 were analyzed with the Beckman automatic sequenator, it was possible to identify the first nine residues of the chain and the first 49 of the frag-

Abbreviations: Ig, immunoglobulin; V_L and V_H, variable regions of light and heavy chains, respectively, of antibody molecules; H^{315}, variable region of the heavy chain of protein 315.

* Present address: National Heart and Lung Institute, National Institutes of Health, Bethesda, Md.

† Present address: Department of Chemical Pathology, Guys Hospital Medical School, London, England.

§ To whom requests for reprints should be sent. Present address: Center for Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

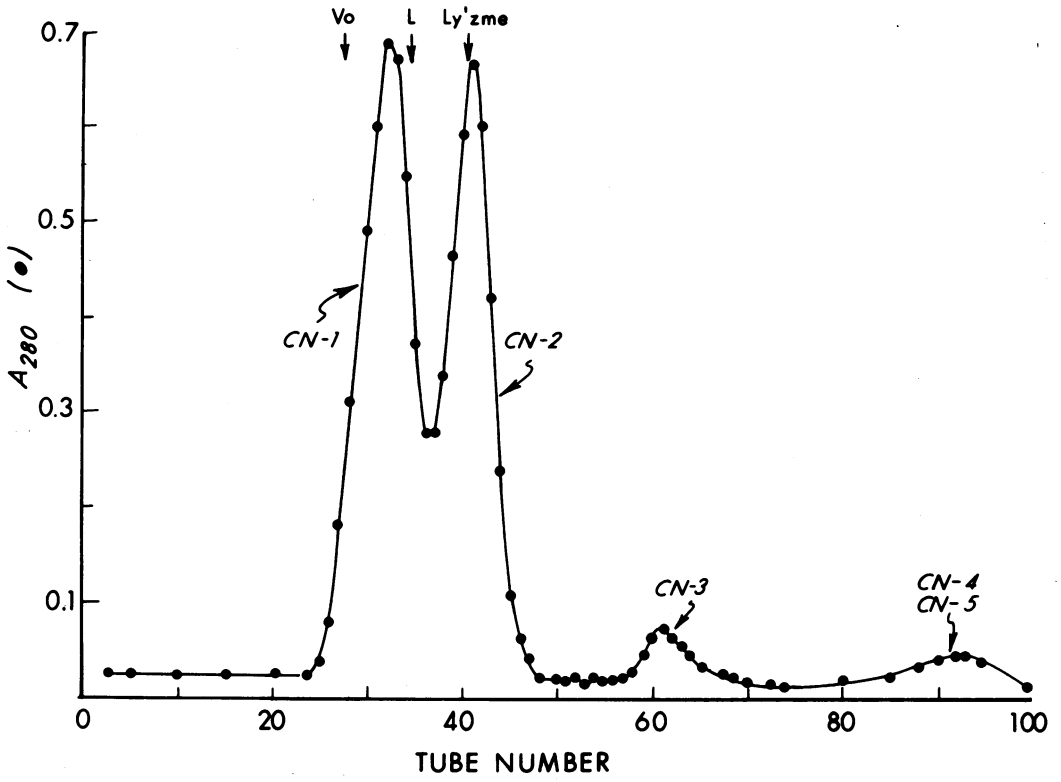

FIG. 1. Elution profile of the cyanogen bromide fragments of the heavy chain of protein 315. Heavy chain (7 mg) was cleaved with CNBr (CNBr/protein, 5/1, w/w) in 70% formic acid for 24 hr at 4°, and then placed on Sephadex G-100 (1.0 × 110 cm) in 5 M guanidine·HCl (Mann, Ultrapure) at a flow rate of about 5 ml/hr. 1.0-ml fractions were collected per tube. Volume markers are immunoglobulin light chain (*L*), molecular weight about 22,000, and lysozyme (*Ly'zme*), molecular weight 13,000; *Vo*, void volumes.

ment: both sequences were in complete agreement. Finally, when the affinity-labeled heavy chain (with [^{14}C]bromoacetyl-Dnp-L-lysine) was cleaved with cyanogen bromide and the fragments were separated as in Fig. 1, only CN2 contained the ^{14}C label; the label is known to attach to a lysine residue that, by homology with other V_H sequences, was previously localized in the V_H region (13).

CN2 was reacted with maleic anhydride and then digested with trypsin (14). The resulting arginine peptides, purified by ion exchange chromatography and gel filtration, accounted for the entire fragment. These peptides were subjected to sequence analysis and then aligned with the aid of other peptides derived from thermolytic and chymotryptic cleavages of CN2. The sequence from positions 1 to 156 is compared in Fig. 2 with the heavy chain of mouse protein 173 (an IgG myeloma protein produced by plasmacystoma MOPC-173) and with human heavy chains of various classes and V_H subgroups (4–6, 15, 16). The sequence from positions 38 to 67 in H^{315} corresponds to the affinity-labeled [^{14}C]bromoacetyl-Dnp-L-lysine peptide previously reported by Haimovich *et al.*; lysine 52 is the specifically labeled residue (13).

V_H of protein 315 resembles other V_H sequences: it contains the usual half-cystines and 20 of 22 residues found in nearly all V_H sequences so far determined (Fig. 2); the two differences are conservative replacements (Thr for Ser at position 25; Leu for Val at position 114). The switch from variable to constant regions has been established at position 115–116 for human γ1 chains (4, 5), and it may occur at the corresponding point in μ heavy chains (17). H^{315}, a mouse α chain, also has the Val-Ser-Ser sequence that appears to mark the beginning of the constant region. If the switch from variable to constant region really occurs at this point, it is striking that the constant regions of such diverse heavy chains as human γ1, human μ, and mouse α all begin with the same sequence. This sequence could be conserved because of constraints imposed by the putative joining mechanism that is thought to unite V and C genes (18–20).

The heavy chains of protein 315 and of protein 173 (15) are the only mouse heavy chains whose sequences have been extensively determined so far. Each has an unblocked amino-terminal (Glu or Asp), which is characteristic of human V_H sequences of the V_{HIII} subgroup (21–23), and it has been suggested that this subgroup may be the predominant one in some nonprimate species (24). However, the sequences of V_H173 and V_H315 differ markedly; the two are only 40% homologous and they differ by at least two sequence gaps (Fig. 2). V_H315 and V_H173 are thus likely to represent different subgroups in the mouse and, therefore, different germ-line genes. Moreover, both V_H315 and V_H173 differ significantly in sequence from the human V_{HIII} regions, and V_H315 has as much overall homology with a representative member of the human V_{HII} subgroup as with one from the V_{HIII} subgroup. Hence, not all unblocked heavy chains in the mouse correspond to the human V_{HIII} subgroup. It is possible that diversity of V_H sequences in the mouse may be as pronounced as the diversity in the mouse V_k family and in the human V_k, V_λ, V_H families (25–28) rather than being restricted as in mouse V_λ (29).

In view of protein 315's ligand-binding activity, comparison of its V_H and V_L segments is of particular interest (Fig. 3). When these regions are aligned according to the method of Fitch (30), about 33% homology is evident, in contrast to the

somewhat lower frequency of identities in V_H and V_L sequences of protein Eu (23%) or of protein Ou (28%) (4, 17). It is only for three proteins (Eu, Ou, 315) that both V_H and V_L are known, and it is thus too soon to know the range of V_H–V_L sequence identities in chains from the same molecules or the significance of these homologies. Because all mouse λ chains, including protein 315, have remarkably similar V-region sequences (29), V_H315 shows essentially the same homology with the V regions of other mouse λ chains, e.g., protein 104 (31); but it is less similar to V segments of mouse k and of human k and λ chains (25–28% homology).

The homologous histidines at position 102 in the protein-315 chains may prove to be particularly interesting. Histidines are notably infrequent in variable regions (4), but in protein 315 they occur not only at homologous positions of V_H and V_L, but in the V-region sector that is subject to most variation in immunoglobulin chains (following Cys 96; see refs. 28 and 23). It is possible that these histidine residues interact with specifically bound ligands (2,4-Dnp and 2,4,6-Tnp groups).

One of the other cyanogen bromide fragments, CN5, was identified as the carboxy terminus of the heavy chain. It is the only fragment lacking homoserine: its carboxy-terminal residue is tyrosine, which is the same as that of intact H^{315}. The sequence of this fragment is shown in Fig. 4.

IgA has been detected so far only in mammals and in birds, while the phylogenetically antecedent amphibia have IgG-like and IgM-like proteins, and the lowest vertebrates (sharks) seem to have only IgM-like molecules (32, 33). It has therefore been suggested that by duplication the gene for μ gave rise to genes for γ and then, relatively recently, for α

```
                                          10                  20                  30
Mouse 315     (α,λ2)        D V Q L Q E S G P G L V K P S Q S L S L T C S V T G Y S I T
Mouse 173     (γ2a,k)       E V K L L E S G G P L V Q L G G S L K L S C A A S G F D F S
Human Nie     (γ1,k,VHIII)  Z V Q L V Q S G G G V V Q P G R S L R L S C A A S G F T F S
Human Ou      (μ,k,VHII)    Z V T L T E S G P A L V K P K Q P L T L T C T F S G F S L S
Human Eu      (γ1,k,VHI)    Z V Q L V Q S G A E V K K P G S S V K V S C K A S G G T F S

                                      40                  50
Mouse 315     (α,λ2)        S G Y F K N - W I R Q F P G N K L E W L G F I K Y D G S B -
Mouse 173     (γ2a,k)       R Y W M S - - W V R Q A P G K G L E W I G E I D P N S S T I
Human Nie     (γ1,k,VHIII)  R Y T I H - - W V R Q A P G K G L E W V A V M S Y B G B B K
Human Ou      (μ,k,VHII)    T S R M R V S W I R R P P G K A L E W L A R I - - B B B D K
Human Eu      (γ1,k,VHI)    R S A I I - - W V R Q A P G Q G L E W M G G I V P M F G P P

                  60                  70                  80
Mouse 315     (α,λ2)        -(Y.G)B P S L K N R V S I T R D T S E N Q F F L K L D S V T
Mouse 173     (γ2a,k)       N Y - T P S L K D K F I I S R N D A K N T L Y L Q M S K V R
Human Nie     (γ1,k,VHIII)  H Y - A D S V N G R F T I S R N D S K N T L Y L N M N S L R
Human Ou      (μ,k,VHII)    F Y W S T S L R T R L S I S K N D S K N Q V V L I M I N V N
Human Eu      (γ1,k,VHI)    N Y - A Q K F Q G R V T I T A D E S T N T A Y M E L S S L R

                      90                  100                 110
Mouse 315     (α,λ2)        (T.Z.B)T A T Y Y C A G D N D H L Y - - - - - - - F D Y W G Q
Mouse 173     (γ2a,k)       S E D T A L Y Y C A R S P Y Y A M
Human Nie     (γ1,k,VHIII)  P Z B T A V Y Y C A R I R D T A M F - - - - - - F A H W G Q
Human Ou      (μ,k,VHII)    P V D T A T Y Y C A R V V N S V M A G Y Y Y Y Y M D V W G K
Human Eu      (γ1,k,VHI)    S E D T A F Y F C A G G Y G I Y S - - - - - - - P E E Y N G

                                  120                 130
Mouse 315     (α,λ2)        G T T L T V S S E S A R N P T I Y P L(A,L,V,Z,K,S,B,Z,S,G)P
Human Nie     (γ1,k,VHIII)  G T L V T V S S A S T K
Human Ou      (μ,k,VHII)    G T T V T V S S G S A S A P T L F P L V S C E N S(D,P,S,S,T)
Human Eu      (γ1,k,VHI)    G - L V T V S S A S T K G P S V F P L A P S S K S T S G G T

                  140                 150
Mouse 315     (α,λ2)        V I I G C L I H B Y F P(S,G)- T M
Human Ou      (μ,k,VHII)    V A V G C L A Z D F L P D S I T F
Human Eu      (γ1,k,VHI)    A A L G C L V K D Y F P E P V T V
```

FIG. 2. Amino-acid sequence of the amino-terminal cyanogen bromide fragment (CN2) from the heavy chain of protein 315, compared with the corresponding sequences of another mouse and several human chains. Residues are numbered according to the sequence in protein Eu. Gaps are introduced to maximize homologies; residues identical in all these sequences are *underlined*. The *vertical* line between positions 115 and 116 marks the probable transition from V to C region. References are: mouse protein 173 (15); human proteins Nie (16), Ou (17, 35), Eu (4, 5). Abbreviations are from *Atlas of Amino Acid Sequence and Structure*, Dayhoff, M., ed. (National Biomedical Research Foundation, Silver Spring, Md.), Vol. 5 (1972).

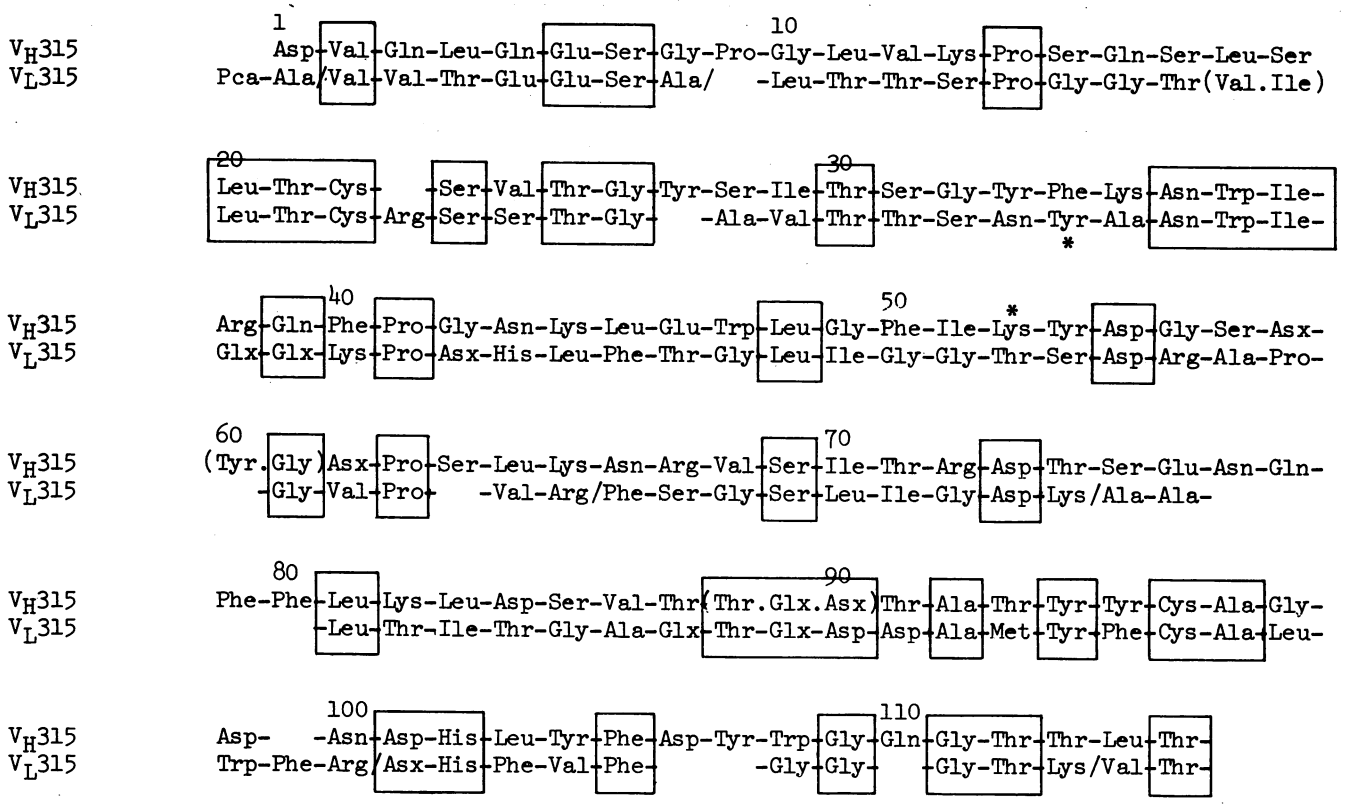

FIG. 3. Comparison of the variable regions of heavy and light chains (V_H and V_L) of protein 315. The V_H sequence is from Fig. 2, this paper; the V_L sequence is from ref. 7. The sequences are aligned according to Fitch (30). *Boxed positions* (33% of the total) are identical (28% are definite and 5% are probably identical). Numbering is according to protein Eu (see Fig. 1). (*) residues labeled specifically in the intact protein by [^{14}C]bromoacetyl Dnp affinity-labeling reagents (Tyr 34 of V_L and Lys 52 of V_H, see ref. 13).

Mouse α	Met-Ser-Glx-Gly-Asx-Gly-Ile-Cys-Tyr
Human α	Met-Ala-Glu-Val-Asp-Gly-Thr-Cys-Tyr
Human μ	Met-Ser-Asx-Thr-Ala-Gly-Thr-Cys-Tyr
Human γ_1	Gln-Lys-Ser-Leu-Ser-Leu-Ser-Pro-Gly

FIG. 4. Comparison of the carboxy-terminal nonapeptides of the heavy chains of protein 315 (α), human α, human μ, and human γ_1 (5, 34–36).

(34). This proposal is in accord with the striking similarity between the carboxy-terminal sequences of human α, mouse α, and human μ chains (Fig. 4); five of nine residues are identical. At the beginning of the constant region, however, the protein-315 α chain is hardly more homologous with human μ than with human γ chains (Fig. 1) and, overall, α and γ are alike in that both are shorter by about one constant-region domain (about 110 residues) than μ chains. Hence contemporary α and γ might have arisen from a common precursor (derived from μ), with conservation of the carboxy-terminal sequence in α for a specific biologic function. Additional sequence data on the constant region of H^{315} may help clarify the evolution of immunoglobulin chains.

We thank Ralph A. Bradshaw for valuable suggestions, Walter M. Fitch for assistance in aligning V_H and V_L of protein 315, Walter Gray for preparation of protein 315 and separation of its chains, and Richard Pinkston for amino-acid analyses. This work was supported in part by Research Grant (AI-03231, AI-10781) and Training Grant (AI-00257) from the National Institute of Allergy and Infectious Disease, National Institutes of Health, a research grant from The National Science Foundation (GB-27605) and by postdoctoral fellowship support for S.H.F. from the U.S. Public Health Service (no. 10-F02-CA50956-01) and the American Cancer Society (no. PF-761).

1. Edelman, G. M. & Gall, W. E. (1969) *Annu. Rev. Biochem.* **38**, 415–466.
2. Porter, R. R. (1973) *Science* **180**, 713–716.
3. Pink, J. R. L. & Milstein, C. (1970) *Progr. Biophys. Mol. Biol.* **21**, 209–263.
4. Edelman, G. M. (1970) *Biochemistry* **9**, 3197–3205.
5. 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.
6. Kohler, H., Shimizu, A., Paul, C. & Putnam, F. W. (1970) *Science* **169**, 56–59.
7. Schulenburg, E. P., Simms, E. S., Lynch, R. G., Bradshaw, R. A. & Eisen, H. N. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2623–2626.
8. Eisen, H. N., Michaelides, M. C., Underdown, B. J., Schulenburg, E. P. & Simms, E. S. (1970) *Fed. Proc.* **29**, 78–84.
9. Underdown, B. J., Simms, E. S. & Eisen, H. N. (1971) *Biochemistry* **10**, 4359–4368.
10. Waterfield, M. & Haber, E. (1970) *Biochemistry* **9**, 832–839.
11. Gray, W. R. (1967) in *Methods in Enzymology*, eds. Colowick, S. P. & Kaplan, N. O. (Academic Press, New York), Vol. XI, pp. 139–151.
12. Bradshaw, R. A., Babin, D. R., Nomoto, M., Srinivasin, N. G., Ericsson, L. H., Walsh, K. A. & Neurath, H. (1969) *Biochemistry* **8**, 3859–3871.
13. Haimovich, J., Eisen, H. N., Hurwitz, E. & Givol, D. (1972) *Biochemistry* **11**, 2389–2398.
14. Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1967) *Biochem. J.* **103**, 78P–79P.
15. Bourgeois, A., Fougereau, M. & DePreval, C. (1972) *Eur. J. Biochem.* **24**, 446–455.

16. Ponstingl, H., Schwarz, J., Reichel, W. & Hilschmann, N. (1970) *Hoppe-Seylers Z. Physiol. Chem.* **351**, 1591–1594.
17. Putnam, F. W., Florent, G., Paul, C., Shinoda, T. & Shimizu, A. (1973) *Science* **182**, 287–291.
18. Fudenberg, H. H., Wang, A. C., Pink, J. R. L & Levin, A. S. (1971) *Ann. N.Y. Acad. Sci.* **190**, 501–506.
19. Nisonoff, A., Fudenberg, H. H., Wilson, S. K., Hopper, J. E. & Wang, A. C. (1972) *Fed. Proc.* **31**, 206–209.
20. Hood, L. (1972) *Fed. Proc.* **31**, 177–187.
21. Kohler, H., Shimizu, A., Paul, C., Moore, V. & Putnam, F. W. (1970) *Nature* **227**, 1318–1320.
22. Pink, J. R. L., Buttery, S. H., DeVries, G. M. & Milstein, C. (1970) *Biochem. J.* **117**, 33–47.
23. Kehoe, J. M. & Capra, J. D. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 2019–2021.
24. Kehoe, J. M. & Capra, J. D. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 2052–2055.
25. Hood, L., McKean, D., Farnsworth, V. & Potter, M. (1973) *Biochemistry* **12**, 741–749.
26. Hood, L. & Talmage, D. W. (1970) *Science* **168**, 325–334.
27. Dayhoff, M. (1972) *Atlas of Protein Sequence and Structure* (Silver Spring, Maryland, National Biomedical Research Foundation).
28. Kabat, E. A. & Wu, T. T. (1971) *Ann. N.Y. Acad. Sci.* **190**, 382–393.
29. Cesari, I. M. & Weigert, M. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2112–2116.
30. Fitch, W. M. (1966) *J. Mol. Biol.* **16**, 9–16.
31. Appella, E. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 590–594.
32. Marchalonis, J. & Edelman, G. (1965) *J. Exp. Med.* **122**, 601–618.
33. Clem, L. W. & Small, P. A., Jr. (1967) *J. Exp. Med.* **125**, 893–920.
34. Grey, H. M. (1969) in *Advances in Immunology* (Academic Press, New York), Vol. 10, pp. 51–104.
35. Shimizu, A., Putnam, F. W., Paul, C., Clamp, J. R. & Johnson, I. (1971) *Nature New Biol.* **231**, 73–76.
36. Press, E. M., Piggot, D. J. & Porter, R. R. (1966) *Biochem. J.* **99**, 356–366.