*Proc. Nat. Acad. Sci. USA*
Vol. 71, No. 3, pp. 974–978, March 1974

# "Alpha Chain Disease" Protein Def: Internal Deletion of a Human Immunoglobulin $A_1$ Heavy Chain

**(immunoglobulin structure/immunoglobulin synthesis/myeloma proteins)**

CARLOTA WOLFENSTEIN-TODEL*, EDITH MIHAESCO†, AND BLAS FRANGIONE*‡

* Department of Medicine, New York University School of Medicine, New York, N.Y. 10016; and † Laboratory of Immunochemistry, Research Institute on Blood Diseases, Hôpital Saint-Louis, 75475 Paris 10, France

*Communicated by Robert A. Good, October 29, 1973*

**ABSTRACT** **Protein Def is a human alpha chain disease protein related to $\alpha$1 immunoglobulin heavy chain. The molecular weight of the polypeptide portion of the monomeric molecule is 29,300, which is a little greater than half of a normal $\alpha$1 chain. The $NH_2$-terminal of the polypeptide is heterogenous and, after a short segment corresponding to the variable region, displays a gap which comprises the $C_H1$ constant domain. Normal synthesis resumes at a valine residue in the hinge region just before a segment which contains a partially duplicated fragment and the interheavy disulfide bonds. From there on, the molecule is apparently normal. Protein Def is therefore synthesized as an internally deleted $\alpha$1 heavy chain, followed by postsynthetic amino-terminal proteolysis. It is postulated that codon(s) specifying valine at the hinge region may be a recognition site for reinitiating synthesis after internal gaps equivalent to position 216 in gamma chain disease proteins.**

The discovery of human tumor cells which synthesize and secret, fragments of immunoglobulin heavy chains, "Heavy Chain Disease" proteins (HCD), offered an opportunity to determine if human cell mutants would provide the same valuable information as those obtained from mutants in bacteria. It is assumed that in these committed cells transcription and translation have occurred faithfully so that the aberrant polypeptide can be viewed as a primary gene product suitable for detailed chemical analysis.

HCD have been described for the three major Ig classes: $\gamma$, $\alpha$, and $\mu$ (1–3). Although $\alpha$HCD is probably the most frequent disorder among this group, most chemical studies have been carried out with $\gamma$HCD, and the results obtained from amino-acid sequence studies of several $\gamma$HCD proteins (4–7) have stimulated speculation concerning the origin of the incomplete chain and their relationship to the mechanisms controlling the synthesis of normal immunoglobulin H chain (see ref. 8). $\alpha$HCD appears as a condition primarily affecting the secretory IgA system (9). It is usually characterized by a lymphoplasmacytic proliferation involving the whole length of the small intestine, and its dominant clinical feature is a severe malabsorption syndrome (10). The clinicopathological features of patient Def have been previously published (11). One of the most intriguing observations is that most of the patients with $\alpha$HCD originated from and had been living in areas with a high rate of infestation by intestinal microorganisms, e.g., North Africa, the Middle East, and the Far East (12, 13). This geographic distribution suggested that environmental factors do play a role in the etiology and pathogenesis of $\alpha$HCD (13).

Antigenic analysis (12), molecular weight determination (14), and preliminary chemical studies (13) demonstrated that $\alpha$HCD molecules (including Def) were related to the Fc fragment of $IgA_1$. The N-terminal residues were shown to be heterogeneous (13), and attempts to obtain the $NH_2$-terminal sequence of several $\alpha$HCD proteins on an automated sequencer were unsuccessful. The question arose whether $\alpha$HCD proteins are normal $\alpha$ chains which undergo postsynthetic intracellular or extracellular degradation or whether they contain a primary defect followed and obscured by a secondary proteolysis (13). Biosynthetic studies of protein Def showed no evidence of post-ribosomal intracellular degradation of a normal size $\alpha$ chain (15). However, more confidence may be placed in the interpretation of the sequence data as an accurate reflection of immunoglobulin genes.

In this paper we present the amino-acid sequence of the hinge region of $\alpha$HCD protein Def. Its comparison with that of a normal $IgA_1$ defines the nature of the defect and indicates that the second hypothesis is correct.

## MATERIALS AND METHODS

*Isolation of the Def Protein.* A three-step schedule was adopted for the purification of the protein (12): (*a*) Pevikon zone electrophoresis at pH 8.6; (*b*) gel filtration on Sephadex G-200 in 0.1M Tris·HCl buffer, pH 8.0, 1 mol/liter in NaCl and 1% in N butanol, after removal of low density lipoproteins by ultracentrifugation; and (*c*) the peak eluted between Ve/Vo 1.0 and 1.4 was again fractionated by Pevikon block electrophoresis. Segments of 2-cm width were eluted, concentrated, and tested for purity by gel diffusion techniques in order to avoid $\alpha$2M and IgM contaminants.

*Complete Reduction and Radioactive Alkylation.* Protein Def (500 mg) was dissolved at a protein concentration of 10 mg/ml in 6 M guanidine HCl·Tris, pH 8.2, and was reduced with 0.005 M dithiothreitol at room temperature under $N_2$. After 1 hr, reduction was terminated by the addition of [$^{14}$C]iodoacetic acid (0.01 M, specific activity 0.7 mCi/mmol).

Abbreviations: Nomenclature of immunoglobulins and their chains and fragments follows the recommendation of the World Health Organization, *Bull. WHO* **30,** 477 (1964); **33,** 721 (1965); **35,** 953 (1966); **38,** 151 (1968); HCD, heavy chain disease proteins; alpha chain disease proteins and myeloma proteins are designated by the first three letters of the patient's name.

‡ Present address: Imperial Cancer Research Fund, Lincoln's Inn Fields, London, WC2, England.

The reaction was allowed to proceed for 1 hr at room temperature, following which the protein was desalted by gel filtration on a column of Sephadex G-15 (2 × 35 cm) in 0.3 M acetic acid, and freeze dried.

*Enzyme Digestion and Separation of Peptides.* (*A*) The labeled protein was digested with L-(1-tosylamido-2-phenyl)-ethyl chloromethyl ketone trypsin (Worthington) in 0.2 M ammonium bicarbonate (pH 8.3) for 15 hr at 37°, enzyme substrate ratio 1:50 (w/w). The digest was centrifuged and the supernatant was freeze dried (10 ml) and submitted to chromatography on a column of Sephadex G-50 (3 × 130 cm) equilibrated in 1 M acetic acid at room temperature. Fractions (9 ml) were collected at a flow-rate of 27 ml/hr. The eluates were monitored by measuring the radioactivity on 50-μl aliquots. The insoluble material was further digested with pepsin (Worthington, twice crystallized), enzyme:substrate ratio 1:50 (w/w) in 5% formic acid for 15 hr at 37°.

(*B*) Protein Def (600 mg) was digested with L-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone-treated trypsin (Worthington) and pepsin (Worthington, twice crystallized), as previously described (16). The digest was fractionated by filtration on a Sephadex G-50 column (3.5 × 180 cm) with 1 M acetic acid as eluant. Fractions (6 ml) were collected at a flow-rate of 25 ml/hr. The eluate was monitored at 280 nm.

Partial reduction and radioactive alkylation were described previously (17). Purification of radioactive peptides, amino-acid analysis, and determination of amino-acid sequences were done as described previously (17). Subtilisin (Nutritional Biochem. Co.) digestion was done in 0.2 M ammonium bicarbonate (pH 8.3) for 15 hr at 37°, at an enzyme:substrate ratio of 1:20 (w/w).

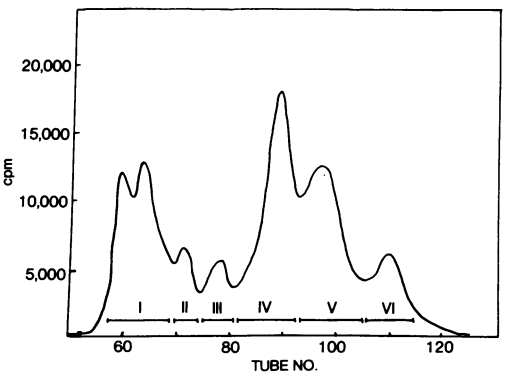


FIG. 1. Gel filtration of a tryptic digest of completely reduced and [$^{14}$C]carboxymethylated protein Def on Sephadex G-50. Digested protein was dissolved in 5 ml of 1 M acetic acid and centrifuged. The supernatant was added to the column (3 × 130 cm) equilibrated in the same solvent. Fractions of 5–6 ml were collected at a flow-rate of 25 ml/hr, and radioactivity was measured. The pooled fractions are indicated on the figure.

*Cystine Diagonal Maps.* These were carried out at pH 3.5 by the method of Brown and Hartley (18). Cysteic-acid peptides were isolated and purified by electrophoresis at pH 6.5 and 2.1 (17).

## RESULTS

Fig. 1 shows the separation of the soluble tryptic peptides obtained from the completely reduced and alkylated protein on a Sephadex G-50 column. The column fractions were combined as shown in the figure. The precipitate was further digested with pepsin, and the peptides were purified by high

TABLE 1. *Amino-acid composition of tryptic and tryptic-peptic S-carboxymethylcysteine peptides obtained after complete reduction and alkylation of protein Def*[a]

| Peak | I | II | III | IV | V | VI | | | Precipitate | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Peptide[b] | T 1[c] | T 2 | T 3 | T 4 | T 5 | T 6 | T 7 | T 8 | TP 9[c] | TP 10 | TP 11 | TP 12 | TP 13 |
| Cys[d] | 1.4 | 0.6 | 0.6 | 2.0 | 0.8 | 0.6 | 0.8 | 0.7 | 0.8 | 0.8 | 0.7 | 0.8 | 0.8 |
| Asp | | 2.1 | 1.0 | 2.0 | 1.0 | 2.0 | | | 0.9 | 0.9 | | | |
| Thr | 4.3 | 2 | 1.6 | | 1.1 | 1.8 | | | 0.9 | 0.9 | 3.0 | | 2 |
| Ser | 7.8 | 0.9 | 0.9 | 2.7 | 1.1 | | | | | | | 1.0 | |
| Glu | | 2.1 | 2 | 1.2 | | 1.1 | | | 1.0 | 1.0 | | | |
| Pro | 10.2 | 0.6 | 0.9 | 1.6 | | 2.0 | | | | | | | |
| Gly | 2.3 | 2.1 | 2 | 3.0 | 2.0 | | | | 1.1 | 1.1 | | | |
| Ala | 1.9 | 2.1 | 2 | 1.0 | | 2.2 | | 1.0 | 1.0 | 1.0 | | | |
| Val | 2.5 | | 1 | 2.0 | 0.7 | 0.8 | | | 0.8 | 0.8 | | | |
| Met | | | 0.8 | | 0.9 | 0.9 | | | | | | 1.0 | |
| Leu | | 5.7 | 1.8 | 1.7 | | 1.0 | | | | | | | 1.1 |
| Tyr | | | | 0.9 | | 0.9 | | | 0.8 | | | | |
| Phe | | | 1.7 | | 1.0 | | | | | | 0.7 | | |
| Lys | | | 2.0 | 1.0 | 1.0 | | 1.0 | | | | | | |
| His | 0.9 | | 1.3 | 0.9 | 1.0 | | | | | | | | |
| Arg | 1.1 | 0.6 | | | | | | 1.0 | | | | | |
| CHO[e] | + | + | | | | | | | | | | | |
| $NH_2$-terminal | Gly | Pro | Lys | Asp | Lys | Thr | Cys | Ala | Ala | Ala | Thr | Ser | ND[f] |

[a] Compositions are reported as mol of amino acid per mol of peptide.
[b] Hydrolysis for 20 hr. No corrections were made for destruction during hydrolysis.
[c] T, trypsin; TP, trypsin-pepsin.
[d] Determined as *S*-carboxymethylcysteine.
[e] +, carbohydrate present.
[f] ND, not done.

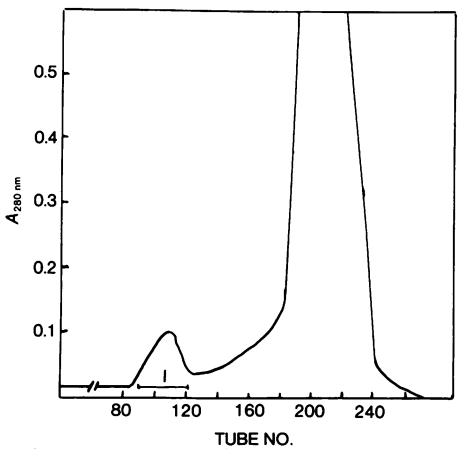


FIG. 2. Gel filtration of a peptic-tryptic digest of protein Def on Sephadex G-100 (3.5 × 180 cm) in 1 M acetic acid. Fractions of 5 ml were collected at a flow-rate of 25 ml/hr. The pooled fraction is indicated on the figure.

voltage paper electrophoresis. The amino-acid compositions, mobilities, and N-terminal amino acids of the major radioactive peptides isolated from each of the peaks and the core are shown in Table 1. A peptide (T1) containing the "hinge" region was localized in the first peak. Its partial sequence is: Gly-Thr-Ala-Gly-Ala-Val-Val-Ser - Val - Pro - Ser - Ser - ($Thr_3$, $Ser_5$, $Pro_{9-10}$, $Cys_2$, His, Arg) as determined by the dansyl-Edman procedure. Fig. 2 shows the elution pattern obtained by fractionation of a peptic-tryptic digest without previous reduction of protein Def on a column of Sephadex G-50. The "hinge" peptide was characterized in the first peak after partial reduction and alkylation with [$^{14}$C]iodoacetic acid, followed by paper electrophoresis at pH 3.5. The amino-acid compositions of the hinge peptide, and of the other radioactive peptide obtained from this fraction are shown in Table 2. The partial sequence of the hinge peptide is Val-Val-Ser-Val-Pro-Ser-Ser-Pro-Pro-Thr-Pro-Ser-Pro-Ser-Thr-Pro- Pro - Thr-Pro-Ser-Pro(Ser, Cys, Cys, His, Pro, Arg).

Another aliquot of the first peak was further digested with subtilisin. A diagonal map at pH 3.5 was performed (Fig. 3). The amino-acid composition, N-terminal residues, and electrophoretic mobility of high-yield peptides are shown in Table 3. The sequence of these peptides was established by the dansyl-Edman procedure as follows: TPS 2, Cys-Cys-His-Pro-Arg; TPS 1, Asp-Leu-Cys-Gly-Cys. Faint peptides were shown to derive from different splitting of peptides TPS 1 and TPS 2.

## DISCUSSION

The molecular weight of the monomeric polypeptide chain Def was reported to be 29,300 (14), and antigenic analysis (12), chemical typing, and diagonal maps (13) suggested that the entire Fc fragment, including the hinge peptide, was present. We may thus conclude that most of the isolated peptides obtained from protein Def and listed in Table 1 belong to the C-terminal of the H chain. In fact, preliminary sequence studies indicate that they are identical to homologous peptides obtained from an $IgA_1$ myeloma protein and located in the Fc fragment (19). In the absence of sequence data of the $NH_2$-terminal region and in view of the failure to show any idiotypic antigenic specificity (13), it was impossible to determine whether protein Def contained part of the Fd fragment. Furthermore, no data were available concerning the amino-acid sequence of the $NH_2$-terminal half of the $\alpha 1$ chain. However, the sequence of a tryptic peptide containing the hinge region of an $IgA_1$ myeloma protein is known (19), and comparison of this region with a similar but shorter tryptic peptide from Def indicates no homology in the first eight residues (Fig. 4). Starting at residue 9, the amino-acid sequence is identical with the exception of a substitution of Thr for Ser in position 12. The first cysteine residue of the hinge region is not present in protein Def, and the peptide containing the sequence Cys-Leu-Ala which was shown to be involved in a disulfide bridge with the first cysteine of the hinge of an intact $\alpha_1$ chain (20) is also absent. The only other tryptic-peptic peptide obtained together with the hinge from the first peak on Sephadex G-50 is the one containing the sequence Asp-Leu-Cys-Gly-Gys (TPS 1). By diagonal electrophoresis, after subtilisin digestion, this peptide was shown to be bound to the peptide containing the last two cysteines of the hinge region (Fig. 3) as it was in $IgA_1$ (20).

These results allow us to make several deductions. Protein Def has an internal deletion that includes almost the whole

TABLE 3. *Amino-acid composition of cysteic-acid peptides derived from subtilisin digestion from peak I (Fig. 2) of a peptic-tryptic digest of protein Def*

| Peptide | TPS 1 | TPS 2 |
|---|---|---|
| Cys[a] | 2.05 | 2.00 |
| Asp | 1.0 | |
| Pro | | 1.15 |
| Gly | 1.16 | |
| Leu | 0.96 | |
| His | | 1.0 |
| Arg | | 1.16 |
| N-terminal | Asp | Cys |
| Mobilities[b] (pH 6.5) | 1.2 | 0.15 |

[a] Determined as cysteic acid.

[b] Mobilities are expressed as portions of the distance between ε-Dnp-Lysine and aspartic acid (24). TPS, Trypsin-pepsin-subtilisin.

TABLE 2. *Amino-acid composition of S-carboxymethylcysteine-containing peptides obtained from the first peak (Fig. 2) of a peptic-tryptic digest of protein Def*

| | Cys[a] | Asp | Thr | Ser | Pro | Gly | Val | Leu | Tyr | His | Arg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Peptide PT1 | 1.46 | | 3.08 | 6.68 | 10.51 | | 2.44 | | | 0.87 | 1.02 |
| Peptide PT2 | 1.74 | 1.13 | | 1.79 | | 1.08 | 0.93 | 1.06 | 0.88 | | |

[a] Determined as *S*-carboxymethylcysteine; PT, pepsin-trypsin.

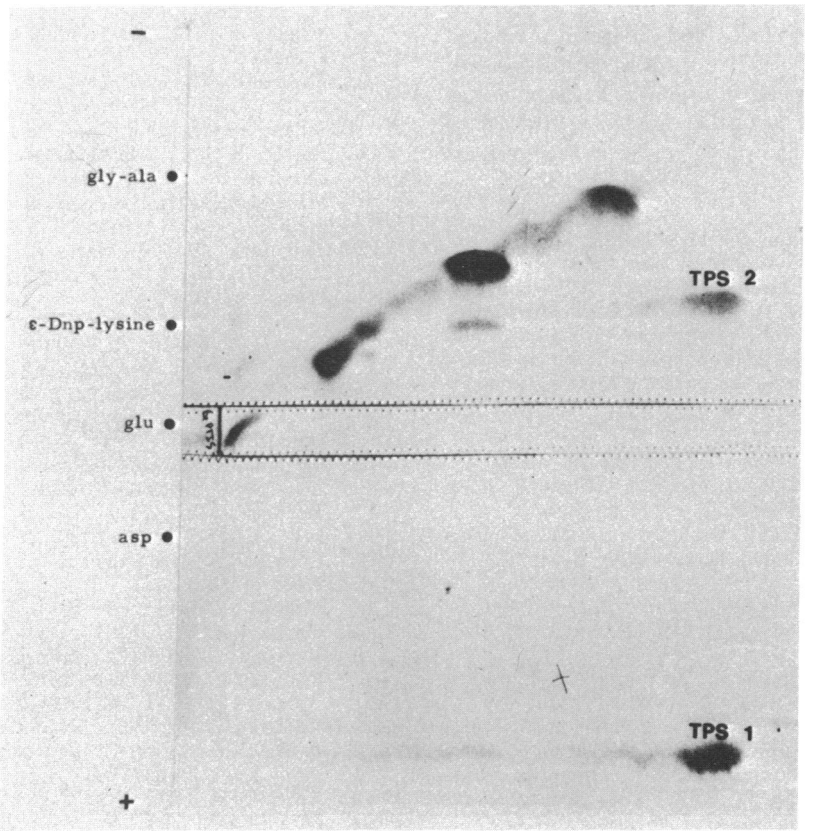


FIG. 3. Diagonal electrophoresis of a subtilisin digest of peak I (Fig. 2) obtained after peptic-tryptic digestion of protein Def. Electrophoresis in both dimensions was at pH 3.5. (TPS, trypsin-pepsin-subtilisin.)

Fd segment. The peptides corresponding to the heavy-light region (13) and the Fd constant loop are missing (19). Furthermore, since the sequence of the first eight residues does not have any homology with the region immediately before the hinge, they must come from the variable region. It can be argued that the difference observed in this region can be due to subclass specificity and that indeed Def belongs to an unknown $\alpha$-chain subclass. However, there is no indication in favor of this hypothesis, and Def appears to belong to the $\alpha 1$ subclass. Furthermore, there is a constant feature present in all $\alpha$ chains studied so far: the presence of the first cysteine residue of the hinge region (see above) which is absent in protein Def. How much of the variable region is present in protein Def is not known, since attempts to use the automated sequencer have been unsuccessful due to heterogeneity at the $NH_2$-terminal, low yields recovery, and probably a high content in carbohydrate (13). The absence of a variable intrachain loop and the molecular weight data suggest that it contains a very short variable region. However, one should emphasize that a limited proteolysis of the N-terminal end of protein Def presumably takes place and that incomplete proteins seem to be quite susceptible to degradation as a secondary phenomenon after ribosomal release (21). The replacement of Thr by Ser in position 12 of the hinge peptide can be explained by a single nucleotide base change; the possibility that this amino-acid substitution is an allotypic marker should be further investigated. It is interesting to point out that this amino-acid replacement is located at the beginning of a short segment of the hinge region which is duplicated (see Fig. 4) and was shown to be a common feature of $\alpha 1$ chains (16). Thus, the duplicated fragment as well as the CHO moiety has been preserved in protein Def. Val at position 9 of the hinge peptide, where the identity with a complete $\alpha 1$ chain starts, could be the equivalent of Glu at position 216 of $\gamma$ chains (22, 23), the site where normal synthesis resumes in several $\gamma$HCD proteins with internal deletions (6). Position 216 is the begin-

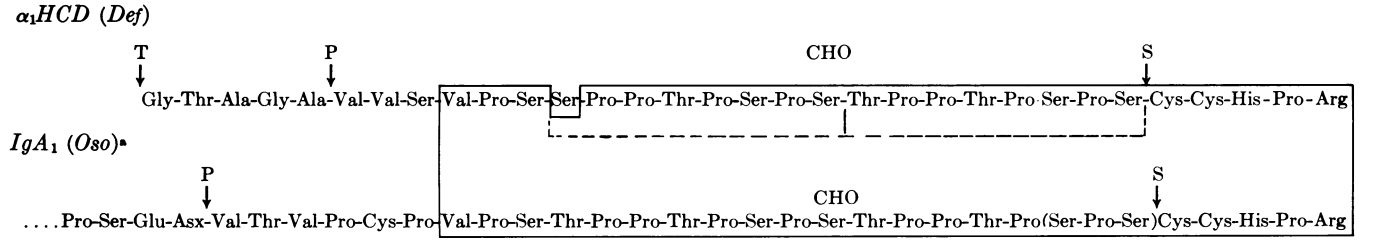


FIG. 4. Amino-acid sequence of the "hinge" region of an $\alpha_1$HCD protein (Def) and of $IgA_1$ (Oso). Identical residues are in the *box*. *Dashed line* indicates duplicated fragments. Note substitution of Thr for Ser in protein Def. *Arrows* indicate the sites of pepsin, P; trypsin, T; and subtilisin, S, digestion. Pepsin and trypsin split the hinge region of protein Def and Oso at different sites. (CHO) carbohydrate; its position is not known. [a] Ref. 19.

ning of the hinge region in $\gamma$ chains, and this region's unusual features have raised several possible theories with respect to its origin (8, 16). The biological and evolutionary significance of the hinge region should be further analyzed in order to understand the genetic mechanisms involved in the synthesis of antibody molecules.

We thank Dr. M. Seligmann for his constant support and encouragement and Mrs. B. Wood for secretarial assistance. This work was supported by grants from the USPHS National Institutes of Health (AM-01431, AM-05064, and AM-02594), Centre National de la Recherche Scientifique (ERA 239), and Institut National de la Sante et de la Recherche Medicale (Grant 73.4.033.2). B.F. is a Senior Investigator of the Arthritis Foundation.

1. Franklin, E. C. (1964) *J. Exp. Med.* **120,** 691–709.
2. Seligmann, M., Danon, F., Hurez, D., Mihaesco, E. & Preud'Homme, J. L. (1968) *Science* **162,** 1396–1397.
3. Forte, F. A., Prelli, F., Yount, W. J., Jerry, L. M., Kochwa, S., Franklin, E. C. & Kunkel, H. G. (1970) *Blood* **36,** 137–144.
4. Frangione, B. & Milstein, C. (1969) *Nature* **224,** 597–599.
5. Franklin, E. C. & Frangione, B. (1971) *Proc. Nat. Acad. Sci. USA* **68,** 187–191.
6. Cooper, S. M., Franklin, E. C. & Frangione, B. (1972) *Science* **176,** 187–189.
7. Frangione, B., Lee, L., Haber, E. and Bloch, K. J. (1973) *Proc. Nat. Acad. Sci. USA* **70,** 1073–1077.
8. Frangione, B. & Franklin, E. C. (1973) *Seminars in Hematology* **10,** 53–64; Smithies O. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38,** in press.
9. Seligmann, M. & Rambaud, J. C. (1969) *Israel J. Med. Sci.* **5,** 151–157.
10. Rambaud, J. C., Bognel, C., Prost, A., Bernier, J. J., Le Quintrec, Y., Lambling, A., Danon, F., Hurez, D. & Seligmann, M. (1968) *Digestion* **1,** 321–336.
11. Rambaud, J. C., Matuchansky, C., Bognel, J. C., Bognel, Caroline, Bernier, J. J., Scotto, J., Perol, C., Ferrier, J. P., Mihaesco, Edith, Hurez, D. & Seligmann, M. (1970) *Ann. Med. Interne* **121,** 135–148.
12. Seligmann, M., Mihaesco, E., Hurez, D., Mihaesco, C., Preud'Homme, J. L. & Rambaud, J. C. (1969) *J. Clin. Invest.* **48,** 2374–2389.
13. Seligmann, M., Mihaesco, E. & Frangione, B. (1971) *Ann. N.Y. Acad. Sci.* **190,** 487–499.
14. Dorrington, K. J., Mihaesco, E. & Seligmann, M. (1970) *Biochim. Biophys. Acta* **221,** 647–649.
15. Buxbaum, J. & Preud'Homme, J. L. (1972) *J. Immunol.* **109,** 1131–1137.
16. Frangione, B. & Wolfenstein-Todel, C. (1972) *Proc. Nat. Acad. Sci. USA* **69,** 3673–3676.
17. Wolfenstein, C., Frangione, B., Mihaesco, E. & Franklin, E. C. (1971) *Biochemistry* **10,** 4140–4145.
18. Brown, J. R. & Hartley, B. S. (1966) *Biochem. J.* **101,** 214–228.
19. Mendez, E., Frangione, B. & Franklin, E. C. (1973) *Biochemistry* **12,** 5186–5194.
20. Wolfenstein-Todel, C., Prelli, F., Frangione, B. & Franklin, E. C. (1973) *Biochemistry* **12,** 5195–5197.
21. Natori, S. & Garen, A. (1970) *J. Mol. Biol.* **49,** 577–588.
22. Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. B., Rutishauser, H. & Waxdal, M. J. (1969) *Proc. Nat. Acad. Sci. USA* **63,** 78–85.
23. Press, E. M. & Hogg, N. M. (1969) *Nature* **223,** 807–810.
24. Offord, R. E. (1966) *Nature* **211,** 591–593.