Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 73–77, January 1985
Biochemistry

Complete amino acid sequence of human hemopexin, the heme-binding protein of serum

(protein structure/secondary structure/gene duplication/heme transport/galactosamine and glucosamine oligosaccharides)

NOBUHIRO TAKAHASHI, YOKO TAKAHASHI, AND FRANK W. PUTNAM*

Department of Biology, Indiana University, Bloomington, IN 47405

Contributed by Frank W. Putnam, September 4, 1984

ABSTRACT **We have determined the complete primary structure of human hemopexin, a plasma β-glycoprotein that specifically binds one heme with high affinity and transports it to hepatocytes for salvage of the iron. Human hemopexin (M_r ≈ 63,000) consists of a single polypeptide chain containing 439 amino acid residues with six intrachain disulfide bridges. The amino-terminal threonine residue is blocked by an O-linked galactosamine oligosaccharide, and the protein has five glucosamine oligosaccharides N-linked to the acceptor sequence Asn-X-Ser/Thr. The 18 tryptophan residues are arranged in four clusters, and 12 of the tryptophans are conserved in homologous positions. Computer-assisted analysis of the internal homology in amino acid sequence indicates that hemopexin consists of two similar halves, thus suggesting duplication of an ancestral gene. Limited tryptic digestion cleaves apohemopexin after arginine-216 into two half-molecules, whereas heme-saturated hemopexin is cleaved after lysine-101. The half-molecules are connected by a histidine-rich hinge-like region that contains two glucosamine oligosaccharides. A structural model for human hemopexin is proposed that is based on these properties and on computer-assisted predictions of the secondary structure and the hydrophilic/hydrophobic character. In this model α-helices and β-turns predominate, and the two halves are connected by an exposed connecting region in apohemopexin that becomes inaccessible to trypsin in heme-saturated hemopexin. Many segments of hemopexin are similar to sequences of other heme proteins, but no overall structural relationship of hemopexin to any other heme protein was identified.**

Hemopexin, the serum β-glycoprotein that binds one heme with high affinity, has been widely studied because of its physiological role in heme transport and catabolism and its decrease in hemolytic diseases (1–3). Although the protein was purified (4) and characterized by physical and chemical methods (5–7) soon after its discovery in 1958 (8), nothing except a partial amino-terminal sequence (9) was published about the structure of hemopexin until our recent reports on the *O*-glycosyl and *N*-glycosyl sites and the unusual clustering of tryptophan residues (10, 11). We now present the complete primary structure of human hemopexin, including the six sites of carbohydrate attachment and the location of the six intrachain disulfide bonds. We propose a structural model for hemopexin that is based on these data together with other results including: (*i*) studies of the sites of limited tryptic cleavage of apohemopexin and of heme-saturated hemopexin, both of which are called hemopexin in the literature (1); (*ii*) computer-assisted analysis of the internal homology of the amino acid sequence; (*iii*) predictions of the secondary structure and the hydrophilic/hydrophobic character (hydropathy profile) of the polypeptide chain.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Materials. Purified human hemopexin saturated with hemin was obtained from Behringwerke Laboratories (Marburg/Lahn, F.R.G.). Cohn fraction IV-4 of human plasma provided by Harold Gallick (Michigan Department of Public Health, Lansing, MI) was used for purification of apohemopexin by a method modified from that of Tsutsui and Mueller (12), in which we used 0.2 M sodium citrate buffer (pH 4.0) as the eluting buffer for the protein.

Methods. Limited tryptic digestion of apohemopexin was done with trypsin (enzyme-to-substrate weight ratio of 1:140) in 0.1 M ammonium bicarbonate at room temperature for 60 min. The reaction was stopped by addition of soybean trypsin inhibitor. The limited digest was separated by ion-exchange HPLC on a Spherogel-TSK IEX-540 DEAE column (Altex, Berkeley, CA) with a linear gradient from 0 to 0.15 M sodium acetate in 0.02 M Tris/acetic acid buffer, pH 8.0, during 75 min at a flow rate of 1.0 ml/min. The methods for sequence analysis of human hemopexin have been described by Takahashi *et al.* (10). To complete the final overlapping sequence, the carboxyl-terminal fragment obtained by limited digestion of apohemopexin was cleaved with dilute acid at 108°C for 5 hr (13). Apohemopexin was also digested extensively with trypsin after treatment with acid/acetone solution to establish the disulfide bridges in the protein.

Computer Analysis of Sequence Data. The sequence data base of the *Atlas of Protein Sequence and Structure* updated to June 1983 and the programs SEARCH, ALIGN, RELATE, and PRPLOT were provided by the National Biomedical Research Foundation (14). The programs SEARCH, RELATE, and ALIGN (15) were used either with the unitary matrix to detect identities or with the mutation data matrix, and all gave a score for statistical significance in standard deviations (SD) of the real score above a score of 100 random runs. We used PRPLOT to plot the hydrophilic/hydrophobic profile of the molecule in terms of the hydropathic index of Kyte and Doolittle (16), in which each amino acid has been assigned a value reflecting its relative hydrophilicity and hydrophobicity along the peptide chain. PRPLOT was also used to graph the secondary structure predicted by the procedure of Chou and Fasman (17) based on their values for the tendency of individual amino acids to appear in α-helix, β-sheet, and β-turn structures.

RESULTS AND DISCUSSION

Determination of the Complete Amino Acid Sequence. The amino acid sequence of human hemopexin (Fig. 1) was deduced from 11 CNBr fragments (which account for six methionine residues and include four additional fragments generated by acid cleavage of two Asp-Pro bonds in the protein) and also from sequence analysis of more than 100 tryptic,

*To whom reprint requests should be addressed.

FIG. 1. Summary of the complete amino acid sequence of human hemopexin. The sequence is shown along with all peptides necessary for the proof of sequence. The peptides obtained from different digestions are as follows: —, CNBr peptides; ---, tryptic peptides; —·—, chymotryptic peptides; ·····, *S. aureus* V8 peptides; -------, dilute acid peptides. An asterisk indicates *S. aureus* V8 peptides of the CNBr peptide. Small arrows under the amino acid sequence indicate the direct sequence determination of the fragments obtained by limited tryptic digestion of heme–hemopexin and apohemopexin. A solid vertical arrow marks the site of limited tryptic cleavage of heme–hemopexin and an open arrow, the site for apohemopexin. GalN indicates the attachment site for the galactosamine oligosaccharide and GlcN, the site for the glucosamine oligosaccharides.

chymotryptic, and *Staphylococcus aureus* V8 protease peptides of the alkylated or the citraconylated hemopexin. All purification of peptides was done by a combination of gel filtration and HPLC on a reversed-phase column (18, 19). Although we had earlier proposed a structural model based on sequence analysis of CNBr fragments and peptides purified from a series of enzymatic digests (10), some difficult overlapping sequence still had to be established in a CNBr fragment that covers the region from residue positions 233 to 352. The main difficulty was that the fragment from heme–hemopexin aggregated very easily and also tended to aggregate as a mixture with another CNBr fragment that covers

the region from residue 45 to residue 154. Furthermore, the latter fragment was resistant to proteolytic cleavage with enzymes such as trypsin and *S. aureus* V8 protease. However, apohemopexin was successfully cleaved in the middle of the molecule by limited tryptic digestion, and a stable fragment consisting of the carboxyl-terminal half was obtained and was used to complete the sequence. The final overlapping peptide, which covers the region from 287 to 334, was obtained by dilute acid cleavage of the carboxyl-terminal fragment (13). During *S. aureus* V8 digestion, one unusual split between Leu-87 and Ile-88 was observed. All overlaps were proven by a minimum of two unique residues. Furthermore, all peptides that were purified and subjected to sequence analysis have been placed in the complete sequence given in Fig. 1, where only peptides needed to establish the sequence are shown. We found no evidence for polymorphism in the amino acid sequence.

Amino Acid Composition. The amino acid composition calculated from the sequence analysis (Table 1) corresponds closely to that obtained by amino acid analysis of the protein (5–7). The content of tryptophan (18 of 439 residues) is unusually high. On the basis of the peptide content this corresponds to 7.46% by weight or 4.1 mol %, which may be compared to an average for proteins of 1.4 mol % (14). A search of the updated protein sequence data base (14) revealed that the tryptophan content of hemopexin is exceeded by that of many small biologically active peptides such as hormones and toxins. However, the only large polypeptide listed that exceeds hemopexin in tryptophan content is also from a heme protein—i.e., bovine cytochrome oxidase polypeptide VII (4.71 mol %); the next listed is polypeptide IV (4.08 mol %). The tryptophan content explains the unusually high absorbance at 280 nm of hemopexin, which is given by Muller-Eberhard and Liem (1) as $A_{1\,cm}^{1\%} = 19.7$ for the apoprotein and 21.8 for the equimolar heme–protein complex.

Polypeptide Chain Structure and Molecular Weight. Human hemopexin consists of a single polypeptide chain containing 439 amino acid residues. The molecular weight calculated from the amino acid sequence of the unmodified polypeptide chain is 49,295. This is considerably less than the molecular weight of 60,000 that we estimated by NaDodSO$_4$/polyacrylamide gel electrophoresis in the presence of mercaptoethanol, and it also is less than the molecular weight that has been reported, which ranges from 57,000 to 80,000 (4, 7). However, hemopexin contains one galactosamine oligosaccharide and five glucosamine oligosaccharides and binds heme, which has a molecular weight of 616. If a typical molecular weight of 600 is assigned for the galactosamine oligosaccharide and 2500 for each of the five glucosamine oligosaccharides, the molecular weight of human hemopexin will be approximately 63,000.

Table 1. Amino acid composition of human hemopexin based on the complete sequence determination

Amino acid	No. of residues	Amino acid	No. of residues
Aspartic acid	25	Valine	24
Asparagine	14	Methionine	6
Threonine	23	Isoleucine	9
Serine	30	Leucine	37
Glutamic acid	24	Tyrosine	16
Glutamine	13	Phenylalanine	19
Proline	35	Lysine	21
Glycine	43	Histidine	19
Alanine	28	Arginine	23
Half-cystine	12	Tryptophan	18

Molecular weight of unmodified polypeptide chain is 49,295; number of residues is 439. Thr-1 binds a galactosamine oligosaccharide. Asn-41, Asn-164, Asn-217, Asn-223, and Asn-430 bind glucosamine oligosaccharides.

Internal Homology of Hemopexin. The four clusters of tryptophan residues previously identified (10) suggested that hemopexin exhibits internal homology in amino acid sequence, which is characteristic of many plasma proteins (20). This was verified by a computer analysis of the amino acid sequence by use of the program ALIGN and the mutation data matrix (15). Fig. 2 shows that hemopexin exhibits internal duplication. In this alignment there are 56 identities out of 172 possible matches between residues, with 13 breaks. Thus, about 25% of the residues are paired. The alignment score is 6.58 SD, which is highly significant (15). Moreover, there are four sets of triple identities, and one sequence of five residues (Val-Asp-Ala-Ala-Phe) occurs three times. In addition, two-thirds of the tryptophan and cysteine residues are conserved (8 of the 12 cysteines and 12 of the 18 tryptophans). This is very significant because both are important for structure and function, and these amino acids are usually the least abundant in proteins. This statistical analysis of the amino acid sequence suggests that hemopexin evolved by gene duplication and that it is composed of two structural domains although it binds only one heme.

Disulfide Bonds. Despite the conservation in cysteine residues and the internal duplication, only one pair of disulfide bonds is homologous (Fig. 2). In each half of the molecule the first and the last cysteine residues are linked by a homologous disulfide bond (Cys-27 and Cys-208; Cys-234 and

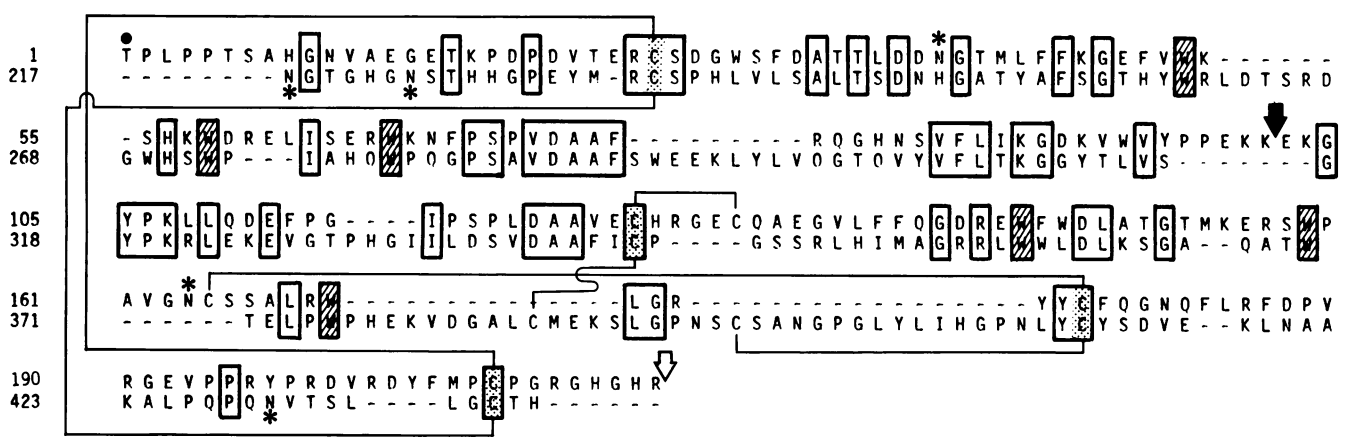

FIG. 2. Internal homology of human hemopexin. The standard one-letter code for amino acids is used. Residues that appear in equivalent positions in each segment of the protein are enclosed in solid boxes. Gaps have been inserted to maximize the homology. Conserved cysteines and tryptophans are shaded, and the disulfide bonds are shown. The attachment sites for galactosamine and for glucosamine oligosaccharides are indicated by ● and *, respectively.

Cys-437). However, the other disulfide bonds in the amino- and carboxyl-terminal domains are not homologous, even though some of the cysteines are paired as identities in Fig. 2. The linkages we found are: Cys-27 and Cys-208, Cys-126 and Cys-131, Cys-165 and Cys-177, Cys-234 and Cys-437, Cys-343 and Cys-385, and Cys-395 and Cys-412. Each half of the molecule has three intrachain disulfide bonds, but the two halves are not linked by a disulfide bond. Hence, limited proteolytic cleavage of the connecting region between the two halves or structural domains is possible in the unreduced protein.

Number and Location of Oligosaccharides. Fig. 1 shows the sites of attachment of the six oligosaccharides in the complete sequence. The galactosamine is O-linked to the amino-terminal residue, threonine. Glycosylation of the amino terminus is unusual and causes a virtual blocking of the end group. This causes difficulty in the Edman degradation of the intact protein (7, 9, 10). Fig. 2 shows that none of the glucosamine oligosaccharides are in homologous positions in the two halves of the molecule. However, although three glucosamine oligosaccharides are randomly distributed along the polypeptide chain (i.e., at Asn-41, Asn-164, and Asn-430), two are close together in the hinge region connecting the two domains (i.e., at Asn-217 and Asn-223). This requires that this section be exposed and hydrophilic.

Sites of Limited Proteolytic Cleavage. Like many proteins, hemopexin is subject to limited proteolytic cleavage that yields well-defined fragments. This suggests that the molecule consists of several structural domains connected by exposed susceptible peptide segments. Smith and Morgan (3) found that plasmin preferentially cleaved rabbit apohemopexin at a single unidentified site into two fragments, both of which had carbohydrate. The fragments were not linked by a disulfide bond; one had a molecular weight of ≈35,000, and the other had a molecular weight of ≈25,000. However, the rabbit heme–hemopexin complex resisted cleavage with plasmin. We found that limited digestion with trypsin cleaves human apohemopexin after Arg-216 into two unlinked half molecules, whereas heme-saturated human hemopexin is cleaved after Lys-101 (Fig. 1), and the two fragments can be separated only after reduction. Thus, the apohemopexins of both species are readily split by serine proteases into two fragments of similar size by cleavage at an exposed site in a connecting region, and conformational changes that result in a different cleavage pattern are induced by the binding of heme. However, the amino-terminal fragment of human apohemopexin is not stable, whereas the amino-terminal fragment of rabbit apohemopexin apparently is.

Secondary Structure. Little is known about the conformation of hemopexin or the heme-binding site (1). Therefore, we used the computer program PRPLOT (14) to predict and plot the local secondary structure of the polypeptide chain by the empirical method of Chou and Fasman (17), the parameters for which are based mainly on proteins that lack oligosaccharides. Overall, this procedure predicts that human hemopexin consists of about one-third α-helix (32% of the residues), with many β-turns (38% of the residues), and some β-sheet structures (17% of the residues). Either the remainder (13%) is in a random structure or a clear choice of structure cannot be made by the method applied. As expected, all six oligosaccharides are located in a β-turn.

The predicted secondary structure diagrammed in Fig. 3 shows a clear division of hemopexin into two structural domains that correspond to the amino-terminal half and the carboxyl-terminal half of the molecule. The two domains are joined by a hinge region that has a series of β-turns associated with two glucosamine oligosaccharides. Each domain consists essentially of a polypeptide segment of about 200 residues, the beginning and end of which is linked by one of the disulfide bonds that form the only homologous pair in the molecule (Cys-27 and Cys-208 in the first domain; Cys-234 and Cys-437 in the second). The secondary structure predicted for the two domains is very similar. The first domain has 34% α-helix, 40% β-turn, and 11% β-sheet; the second has 30% α-helix, 36% β-turn, and 22% β-sheet. The first domain is preceded by a proline-rich section that begins with the amino-terminal residue that is glycosylated with galactosamine; the second is preceded by the histidine-rich hinge that is glycosylated twice with glucosamine. The secondary structure predicted by the computer program PRPLOT thus supports the molecular model composed of two domains that was suggested by the internal duplication in amino acid sequence identified by the computer program ALIGN.

Hydropathy Profile. The hydropathy profile is in accord with a structural model composed of two domains linked by

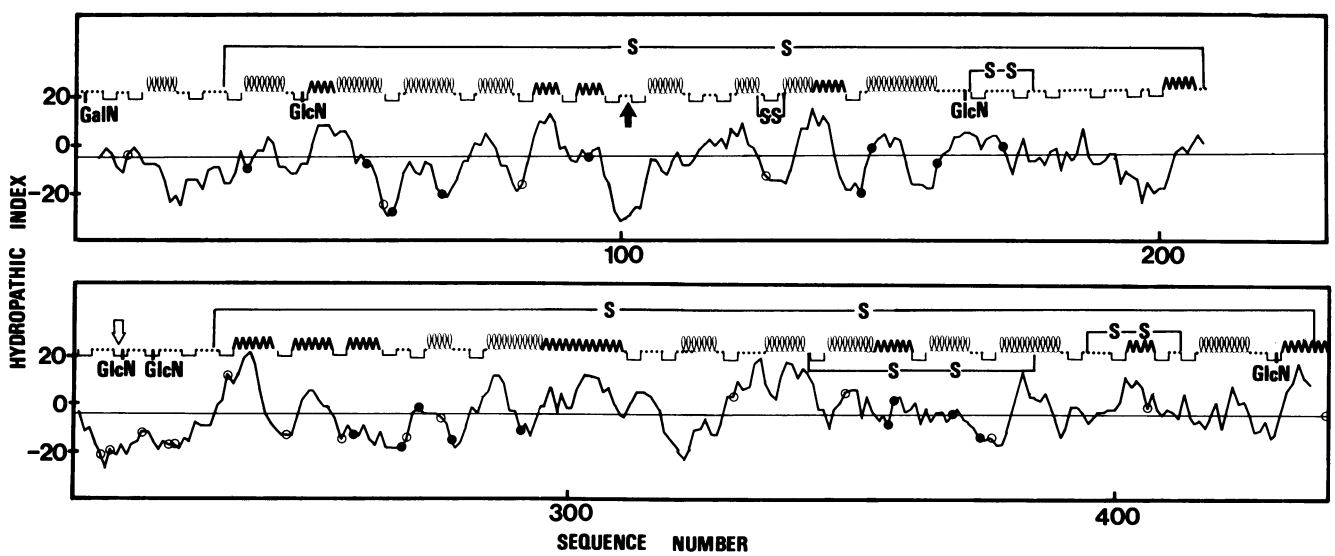

FIG. 3. Hydropathy profile (16) and secondary structure predictions (17) for human hemopexin calculated by use of the PRPLOT program (15). In the secondary structure prediction the residues are represented by dots and are shown in α-helical (wwww), β-sheet (ʌʌʌʌ), and β-turn (⌴) conformational states. The locations are given for the oligosaccharides (galactosamine and glucosamine) and for the disulfide bonds. A solid vertical arrow marks the site of limited tryptic cleavage of heme–hemopexin and an open arrow, the site for apohemopexin. In the hydropathy profile open circles identify the 19 histidine residues, and solid circles are used for the 18 tryptophan residues.

a hydrophilic exposed region. However, the hydropathy profiles of the two domains do not resemble each other strongly, as they do in the case of the series of homologous domains in ceruloplasmin, a protein that exhibits internal triplication with a much higher degree of internal homology (≈35 SD) (21). As expected, the carbohydrates of hemopexin (with one exception) are associated with hydrophilic regions. The exception (the second glucosamine) occurs just before a short disulfide bridge and a series of β-turns that would probably expose the carbohydrate.

Because it has been suggested that certain tryptophan residues are essential for the interaction of heme with hemopexin (1), all tryptophan residues are identified in the structural model of Fig. 3. Almost half of these are in the α-helix. Several associated with a β-turn are in very hydrophilic areas. Although most of the tryptophan residues are in regions that have a hydropathic index close to zero, it should be recalled that the hydropathic parameter for tryptophan has some ambiguity (16).

The hydropathy profile and the secondary structure predictions provide an explanation for the susceptibility of the two sites of limited tryptic cleavage of hemopexin. Specific cleavage of apohemopexin occurs after Arg-216 in the histidine- and carbohydrate-rich hinge region that is exposed because of a series of β-turns and is in the longest hydrophilic section of the molecule. In heme-saturated hemopexin specific cleavage occurs after Lys-101, which is in a β-turn and is in the most hydrophilic section of the molecule. The fact that different peptide bonds are scissile in apohemopexin and heme–hemopexin must be attributed to a conformational change induced by the binding of the heme.

Domain Structure. The preceding results and discussion support the structural model for hemopexin given in Fig. 3. The model illustrates a division of the molecule into two domains connected by an exposed hydrophilic hinge region that contains a single scissile bond which is protected when heme is bound. The evidence for this model is based on four sets of data: (*i*) the primary structure, including the complete amino acid sequence, the carbohydrate sites, and the disulfide bonding pattern; (*ii*) identification of a single site of preferential cleavage by trypsin that differs in apohemopexin and heme–hemopexin; (*iii*) a statistically significant alignment of the amino acid sequence as two homologous segments of similar size (each about 200 residues); and (*iv*) the secondary structure and hydropathy profiles calculated by use of computer programs. Because hemopexin consists of two domains but binds only one heme, the question arises whether the functions of heme binding by hemopexin and of hemopexin binding by the hepatocyte receptor reside in separate domains or are governed by both domains.

Sequence Homology to Other Proteins. The computer program SEARCH was used with the unitary data matrix to compare 15 successive 30-residue segments of hemopexin with the entire updated protein sequence data base (14). In the computer printout a heme-containing protein was listed first three times, was second twice, and was within the first 10 listed 12 times for the 15 segments of hemopexin compared. With two exceptions each successive 30-residue segment of hemopexin scored eight or more identities with one or more of a series of heme- or heme-pigment-containing proteins such as cytochromes, globins, and phycocyanins. Many different eukaryotic and prokaryotic cytochromes scored high. The most striking finding was that the covalent heme-binding site of the *c*-type cytochromes scored high for five segments and gave the highest score of all proteins for the hemopexin sequence for residues 392–422. Yet, despite the structural similarity of segments of hemopexin to segments of many heme proteins, there was no statistically significant evidence for a continuous homologous alignment of hemopexin with any other heme protein. However, further study of well-defined fragments of hemopexin such as those reported here should facilitate identification of the heme-binding domain and also the site for uptake of hemopexin by the hepatocyte receptor.

We thank P. H. Davidson, S. A. Dorwin, J. Madison, and K. Huss for excellent technical assistance and Dr. H. Gallick for human plasma fraction IV-4. This work was supported by Grant AM 19221 from the National Institutes of Health.

1. Muller-Eberhard, U. & Liem, H. H. (1974) in *Structure and Function of Plasma Proteins*, ed. Allison, C. A. (Plenum, London), Vol. 1, pp. 35–53.
2. Smith, A. & Morgan, W. T. (1981) *J. Biol. Chem.* **256,** 10902–10909.
3. Smith, A. & Morgan, W. T. (1984) *Protides Biol. Fluids Proc. Colloq.* **31,** 219–224.
4. Schultze, H. E., Heide, K. & Haupt, H. (1961) *Naturwissenschaften* **48,** 696–697.
5. Heimburger, N., Heide, K., Haupt, H. & Schultze, H. E. (1964) *Clin. Chim. Acta* **10,** 293–307.
6. Hrkal, Z. & Muller-Eberhard, U. (1971) *Biochemistry* **10,** 1746–1750.
7. Hayem-Levy, A. & Havez, R. (1973) *Clin. Chim. Acta* **47,** 113–122.
8. Neale, F. C., Aber, G. M. & Northam, B. E. (1958) *J. Clin. Pathol.* **11,** 206–219.
9. Morgan, W. T., Capra, J. D., Kehoe, J. M. & Muller-Eberhard, U. (1974) *FEBS Lett.* **48,** 109–113.
10. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2021–2025.
11. Takahashi, N., Takahashi, Y. & Putnam, F. W. (1984) *Protides Biol. Fluids Proc. Colloq.* **31,** 215–218.
12. Tsutsui, K. & Mueller, G. C. (1982) *Anal. Biochem.* **121,** 244–250.
13. Inglis, A. S. (1983) *Methods Enzymol.* **91,** 324–332.
14. Barker, W. C., Hunt, L. T., Orcutt, B. C., George, D. G., Yeh, L. S., Chen, H. R., Blomquist, M. C., Johnson, G. C. & Dayhoff, M. O. (1983) *Protein Sequence Database: Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Version 7.
15. Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 359–362.
16. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
17. Chou, P. Y. & Fasman, G. O. (1978) *Annu. Rev. Biochem.* **47,** 251–276.
18. Tetaert, D., Takahashi, N. & Putnam, F. W. (1982) *Anal. Biochem.* **123,** 430–437.
19. Takahashi, N., Takahashi, Y., Ortel, T. L., Lozier, J. N., Ishioka, N. & Putnam, F. W. (1984) *J. Chromatogr.* **317,** 11–25.
20. Putnam, F. W., ed. (1984) *The Plasma Proteins: Structure, Function, and Genetic Control* (Academic, Orlando, FL), 2nd Ed., Vol. 4.
21. Ortel, T. L., Takahashi, N. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4761–4765.