Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4761–4765, August 1984
Biochemistry

Structural model of human ceruloplasmin based on internal triplication, hydrophilic/hydrophobic character, and secondary structure of domains

(hydropathy profile of proteins/glucosamine oligosaccharides/gene duplication and evolution/copper oxidases/ferroxidase)

THOMAS L. ORTEL*, NOBUHIRO TAKAHASHI, AND FRANK W. PUTNAM†

Department of Biology, Indiana University, Bloomington, IN 47405

Contributed by Frank W. Putnam, April 30, 1984

ABSTRACT **A molecular model for the structure of human ceruloplasmin is proposed that is based on the determination of the complete amino acid sequence, studies of the products of limited proteolytic cleavage, calculations of the hydrophilic/hydrophobic character (hydropathy profile), and predictions of the local secondary structure. This multicopper oxidase ($M_r \approx 132,000$) consists of a single polypeptide chain (1046 amino acid residues) with four attached glucosamine oligosaccharides. Computer-assisted statistical analysis of the internal repetition in the amino acid sequence confirms that the entire polypeptide chain is divided into three contiguous homology units, each containing about 350 amino acid residues. Each homology unit is subdivided into three domains, designated A_1, A_2, and B, that differ in structure and probably in function. Calculations of the hydropathy profile and predictions of the secondary structure support a molecular model based on internal repetition of three homology units and help to identify characteristic features of the interdomain junctions. The alignment scores for internal duplication of pairings of the three homology units of ceruloplasmin exceed the scores yet reported for contiguous internal duplication of any other protein. This highly significant evidence for intragenic repetition suggests that the ceruloplasmin molecule evolved by tandem triplication of ancestral genes coding for a primordial copper oxidase.**

Ceruloplasmin [ferroxidase; iron(II):oxygen oxidoreductase, EC 1.16.3.1] is a sky-blue α_2-glycoprotein that contains at least six copper(II) ions of three types and binds about 95% of plasma copper (1). This protein is absent or deficient in Wilson disease, an autosomal trait in which deposition of copper occurs in many tissues. Multiple functions have been ascribed to ceruloplasmin, including (*i*) transport, mobilization, and homeostasis of copper; (*ii*) ferroxidase, amine oxidase, and possibly superoxide dismutase activity; and (*iii*) a role as an acute-phase reactant in the inflammatory response (1–3). Ceruloplasmin is the only known blue mammalian multicopper oxidase, and all of the proposed functions probably involve copper.

Although neither the three-dimensional structure of this large protein ($M_r \approx 132,000$) nor the coordination environments of the copper(II) ions have been established, we have determined the complete amino acid sequence of human ceruloplasmin (4). Our previous work has shown that ceruloplasmin is composed of a series of homologous domains that result from a remarkable degree of internal triplication in primary structure (4, 5). Here we summarize additional evidence for a structural model based on studies of the sites of limited proteolytic cleavage (6, 7), computer-assisted analysis of the internal homology in amino acid sequence, secondary structural characteristics, and the hydrophilic/hydrophobic profile of the molecule in terms of the hydropathic index, in which each amino acid has been assigned a value reflecting its relative hydrophilicity and hydrophobicity along the peptide chain. The model proposes that ceruloplasmin consists of a single polypeptide chain composed of three consecutive 350-residue homology units, each consisting of a series of domains. The high degree of internal homology indicates that ceruloplasmin arose early in vertebrate evolution by tandem triplication of a primordial gene that coded for a primitive copper oxidase.

MATERIALS AND METHODS

Materials. Two preparations of normal human ceruloplasmin were studied. One had undergone autolytic proteolysis and was a mixture of three main fragments (≈19, 50, and 67 kDa); each of these had been characterized (4, 8–10). The second preparation (4) was undegraded single-chain ceruloplasmin (≈132 kDa) and was subjected to limited digestion with trypsin to obtain the various fragments identified in Fig. 1 (6, 7). Details of the methods for enzymatic digestion and for purification, characterization, and sequence analysis of the latter fragments are given in the doctoral dissertation of Ortel (ref. 7; available from University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106).

Computer Analysis of Sequence Data and Structural Parameters. The sequence data base of the *Atlas of Protein Sequence and Structure* and the programs SEARCH, RELATE, ALIGN and PRPLOT were provided by the National Biomedical Research Foundation (11). The programs SEARCH, RELATE, and ALIGN were used (5, 10) 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 hydropathic index with the unified hydropathy scale of Kyte and Doolittle (12). When the same segment length was selected, the graph was identical to one earlier done for us with the computer program SOAP by Russell F. Doolittle (personal communication). PRPLOT was also used to graph the secondary structure predicted by the procedure of Chou and Fasman (13) 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

Molecular Model of Ceruloplasmin. Determination of the complete amino acid sequence of human ceruloplasmin (4) coupled with other experimental studies (6–9) permits predictions of secondary structural characteristics and proposal

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.

*Present address: Indiana University School of Medicine, Indianapolis, IN 46223.
†To whom reprint requests should be addressed.

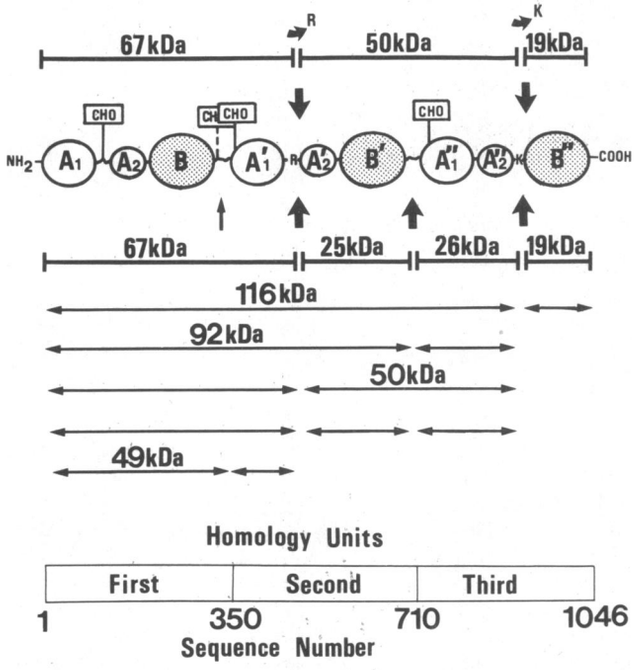

FIG. 1. Structural model of human ceruloplasmin based on sites of limited proteolytic cleavage, sequence analysis, internal homology, and parameters of secondary structure. The single polypeptide chain is cleaved autolytically into the three major fragments shown above the model (67 kDa, 50 kDa, and 19 kDa). In the intact molecule, these fragments are connected by the single amino acid residues arginine (R) and lysine (K). Limited cleavage with trypsin yields the same fragments except that the arginine and lysine are retained at the carboxyl-terminal portions of the 67-kDa and 50-kDa fragments, respectively. Other fragments released by trypsin are indicated; broad vertical arrows identify the major sites of cleavage. The locations of the four glucosamine oligosaccharides (CHO) are given. The carbohydrate that is missing in type II ceruloplasmin is indicated by the broken line; the presence of carbohydrate in type I ceruloplasmin confers increased resistance to cleavage at this site (narrow vertical arrow). The molecule consists of three homologous units of two (A, B) or three (A_1, A_2, and B) domains. The shading of the B domains indicates that they probably contain the type 1 copper(II) ligand sites.

of a molecular model showing the organization of the three tandem homology units and the domain structure (Fig. 1). The model shows the principal sites of limited cleavage of the undenatured single-chain molecule and the corresponding fragments obtained by digestion *in vitro* with trypsin. The amino acid sequence has been reported for the three major fragments derived by spontaneous proteolysis in plasma—i.e., 67 kDa (4), 50 kDa (9), and 19 kDa (8). Sequence analysis of peptides derived from intact ceruloplasmin (4, 7) proved that the protein consists of a single polypeptide chain in which the three fragments are aligned as shown with an arginine residue between the 67-kDa and 50-kDa fragments and a lysine residue between the 50-kDa and 19-kDa fragments. The single arginine and lysine residues are presumably removed during autolytic cleavage by a serum carboxypeptidase of type B that is specific for carboxyl-terminal basic residues. Many examples of such postproteolytic or postsynthetic processing are now known.

All other proteolytic fragments illustrated in Fig. 1 have been characterized at least by partial sequence analysis, tryptic peptide profile, retention times with HPLC, and by other means (6, 7). And all fit into the complete sequence (4). Because all cleavage sites identified in fragments obtained by tryptic digestion appear to end with basic amino acids (4, 7), we conclude that serine proteases normally present in plasma are responsible for the apparent autolysis or spontaneous proteolysis. This can be simulated with plasmin, but others of the many known serine proteases of the coagulation and complement pathways may participate also in what is probably the initial step in the normal catabolism of ceruloplasmin.

Domain Structure of Ceruloplasmin. We propose that the single polypeptide chain of ceruloplasmin consists of three contiguous homology units, each of which contains a set of discrete homologous domains (Fig. 1). Each homology unit consists of a polypeptide chain of about 40 kDa with about 350 amino acid residues and one or more glucosamine oligosaccharides at nonhomologous sites. Cleavage between the individual homology units is not observed in the autolytically cleaved material, but trypsin rapidly cleaves at a site between the second and third homology units (5–7). In type I ceruloplasmin, which has carbohydrate attached at asparagine-339, tryptic cleavage between the first two units is hindered, but in type II ceruloplasmin, it occurs rapidly due to the absence of carbohydrate.

Each homology unit is further divided into either two (A and B) or three (A_1, A_2, and B) domains based on additional sites of cleavage within the second and third homology units (Fig. 1). Both during autolysis and limited tryptic digestion, the 19-kDa fragment is released most rapidly and is most stable (B″ in Fig. 1), but a corresponding fragment is not readily formed from the first and second homology units. Likewise, cleavage is rapid between the A'_1 and A'_2 domains of the second homology unit, but this site is resistant not only in the first homology unit (where it may be shielded by the carbohydrate) but also in the third homology unit. Thus, a further explanation of the differential sensitivity to proteolytic cleavage at homologous sites between proposed domains must be sought in differences in the primary and secondary structures at the interdomain boundaries.

Internal Triplication of Human Ceruloplasmin. The human ceruloplasmin molecule exhibits a remarkable internal triplication that is not matched in degree and extent by any other protein yet investigated. Previously, by computer-assisted analysis of the amino acid sequence of the 50-kDa and 19-kDa fragments (10) and of the partial sequence of the 67-kDa fragment in comparison to the latter two (5), we had established that there was a highly significant statistical correlation for a 3-fold internal homology in the ceruloplasmin molecule. Fig. 2 illustrates the almost unprecedented degree of internal homology when the comparison is carried out on the entire sequence of the 1046-residue polypeptide chain. In this alignment all three homology units exhibit about 30% identity in amino acid sequence, and each pair has about 40–45% identity. The alignment scores for statistical significance determined by the computer program ALIGN with the mutation data matrix are: first and second homology units (upper line and middle line), 36.2 SD; second and third homology units, 42.1 SD; and first and third homology units, 40.1 SD. These scores for duplications within a single polypeptide sequence are higher than any tabulated by Barker *et al.* (14) for all proteins listed in their 1978 data base with two exceptions. Human plasminogen has a score of 44.5 for five segments, each of about 80 residues in length, but these include only half the chain, whereas the entire ceruloplasmin chain has a contiguous internal triplication. The other case is the α_2 chain of human haptoglobin (SD = 37.1), where there are two identical repeats of 59 residues representing 83% of the chain, and these are attributed to a genetic crossover between two alleles of the α_1 chain. Furthermore, the internal repeat in ceruloplasmin shows a high degree of conservation of the least frequent amino acids cysteine, tryptophan, methionine, and histidine, which is regarded as an important criterion of relatedness (15). The odds for finding one or more pairs of identical hexapeptides in unrelated proteins in a data base with 10,000 residues are 1:1 (16). However, in a

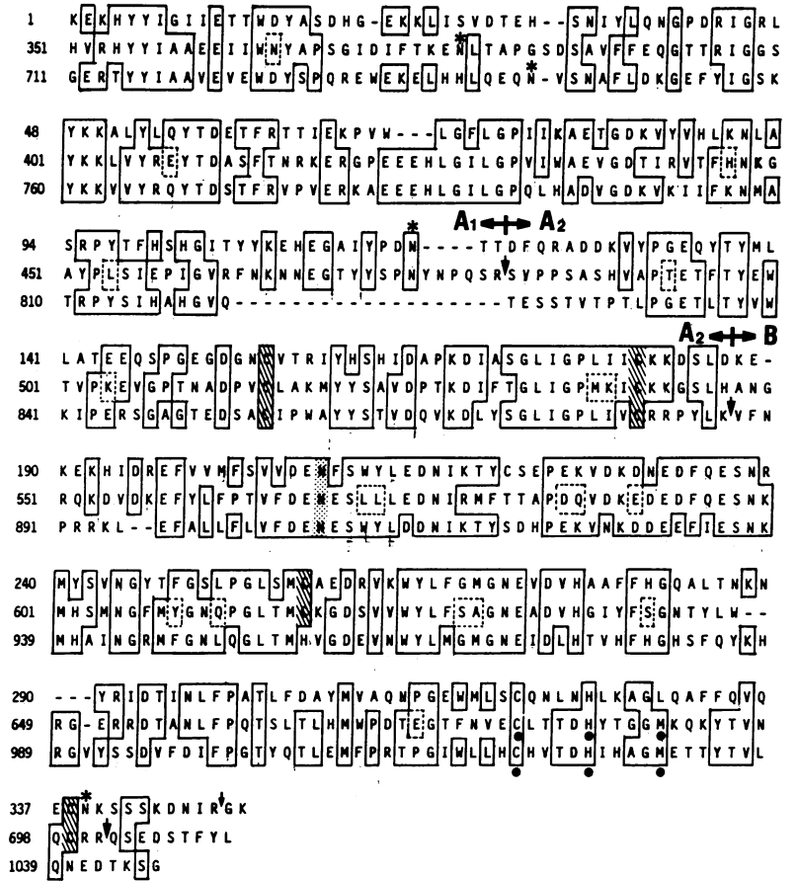

FIG. 2. Internal triplication in human ceruloplasmin as shown by homology in amino acid sequence. Except for a few gaps to maximize the homology, the complete amino acid sequence of Takahashi *et al.* (4) is given continuously in the one-letter code. The upper line in each block (residues 1–350) represents the amino-terminal homology unit containing domains A_1, A_2, and B of Fig. 1 for which the boundaries are shown. Similarly, the second line (residues 351–710) and the third line (residues 711–1046) represent the central and the carboxyl-terminal homology units, respectively, and the corresponding domains given in Fig. 1 (A_1', A_2', B′ and A_1'', A_2'', B″). The vertical arrows after residues 348, 481, 701, and 887 denote the four sites of tryptic cleavage shown in Fig. 1. Asterisks identify the locations of the four glucosamine carbohydrates (asparagine (N) residues 119, 339, 378, and 743). Three other sites with the potential acceptor sequence Asn-X-Ser that are homologous but are not glycosylated are shaded with dots (asparagine residues 208, 569, and 907). Residues boxed at a given position are identical except that nonidentical residues in the middle sequence are enclosed in dashed boxes. Homologous pairs of cysteine (C) residues that are probably joined by disulfide bonds are shaded. The solid circles in the lower right of the diagram denote the two sets of cysteine (C), histidine (H), and methionine (M) residues in the B′ and B″ domains, each of which is a putative binding site for a type 1 blue copper(II) ion. A, alanine; R, arginine; D, aspartic acid; Q, glutamine; E, glutamic acid; G, glycine; I, isoleucine; L, leucine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; V, valine; K, lysine.

total of 1046 residues, ceruloplasmin has one pair of 10 identical residues (second row of Fig. 2), one pair of 8 (fourth row), two pairs of 7 (fifth row), one pair of 6 (fifth row), and a number of pairs of identical pentapeptides, including one set of three (fourth row).

Computer Analysis of Domain Structure. The program ALIGN was used with the mutation data matrix to evaluate the homology for all possible pairs of the nine domains shown in Fig. 1. The segments of sequence compared are given in Fig. 2 (e.g., positions 1–120 for A_1, 121–187 for A_2, and 188–348 for B, etc.). The alignment scores for any pairing of the three A domains ranged from 25.2 to 28.6 SD, where a score of 3.0 indicates a probable relatedness. Similarly, the scores for any pairing of the B domains ranged from 23.1 to 25.8 SD. In contrast, the pairings of A domains with B domains range from −0.55 to only 3.3 SD. These correlations indicate an internal triplication in which the three A segments are very homologous to each other and the three B segments are highly homologous to each other, but the A segments are not statistically related to the B segments. Similar calculations indicated that the three A_1 domains are very homologous in sequence (range of 13.3–20.2 SD) and likewise that the three A_2 domains are very homologous (range of 13.3–21.0 SD). On the other hand, the three A_1–A_2 interdomain segments are the least homologous regions of amino acid sequence and contain the longest gaps shown in Fig. 2. These computations support a molecular model of ceruloplasmin that consists of three contiguous homology units of about 350 residues, each of which is subdivided into A_1, A_2, and B domains that contain about 125, 65, and 160 residues, respectively.

Hydropathy Profile. The hydropathy profile of ceruloplasmin (Fig. 3) confirms a molecular model based on internal repetition of three homology units, each of which is subdivided into A_1, A_2, and B domains. As would be expected from the homology in sequence, the patterns for the three segments show considerable parallelism, particularly in the regions of highest and lowest hydropathic index. The profile also helps to identify some characteristic features of the interdomain junctions. At the boundary regions between the A_2 and B domains just after the disulfide bridge, the change in hydropathic index approaches a maximum positive value—i.e., high hydrophilicity between two very hydrophobic segments. This interdomain region corresponds to the short nonhomologous segments of sequence at the junctions of the A_2 and B domains in Fig. 2. This suggests that the two domains are in a compact conformation and are separated by short exposed hydrophilic sequences. Indeed, the highly charged basic sequence between A_2'' and B″ is the area most susceptible to proteolytic cleavage which results in excision of the 19-kDa fragment (the B″ domain). Likewise, all three B domains terminate in a very hydrophilic region that is highly susceptible to proteolytic cleavage between B′ and A_1'' but less so between B and A_1' because of the steric interference of the carbohydrate. As expected, the oligosaccharides (CHO in Fig. 3) are located in hydrophilic regions, whereas the three homologous unglycosylated potential acceptor sites for glucosamine oligosaccharide appear to be within a hydrophobic peak.

Secondary Structure. The secondary structure of a protein refers to the local spatial organization of the polypeptide chain backbone and thus includes: (*i*) well-defined conformations such as α-helices, and β-sheets, (*ii*) random regions, and (*iii*) β-turns, which are chain reversal regions consisting of tetrapeptides (13). In the absence of crystallographic data, we used the computer program PRPLOT (11) to predict the local secondary structure of the polypeptide chain by the empirical method of Chou and Fasman (13). Overall, this method predicts that human ceruloplasmin consists of about one-third β-sheet structures (33% of the residues), one-third

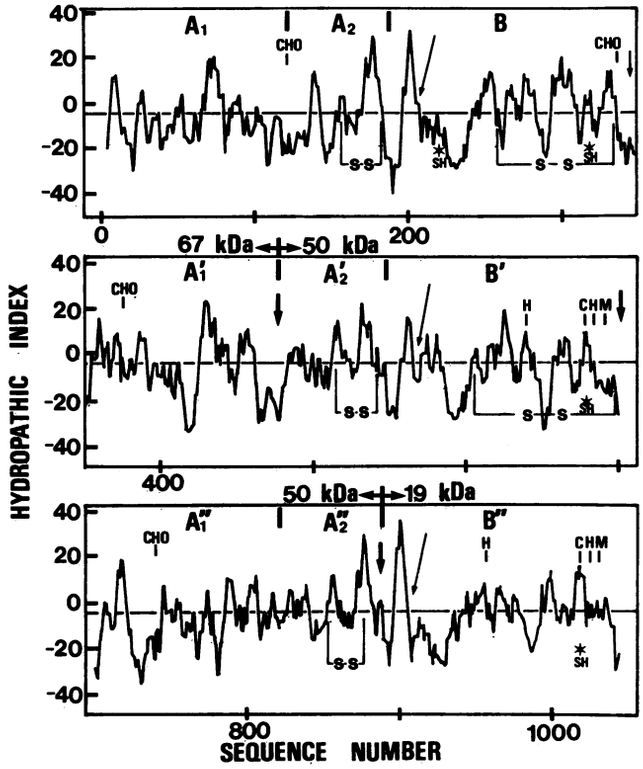

FIG. 3. Hydropathy profile for the entire human ceruloplasmin molecule calculated by the method of Kyte and Doolittle (12). The three segments represent the three homology units illustrated in the sequence comparison of Fig. 2. That is, the profile for positions 1–350 is given in the upper line of Fig. 2, and so forth. Vertical lines at the top of each profile block are placed at the interdomain junctions for the homologous A_1, A_2, and B domains. Vertical arrows correspond to the tryptic cleavage sites in Figs. 1 and 2, and the location of the four glucosamine oligosaccharides (CHO) is given. Diagonal arrows identify unglycosylated sites with potential acceptor sequences for glucosamine oligosaccharides. The location of the disulfide bridges is given, and the sites of free sulfhydryl (SH) groups are marked with an asterisk. The two homologous sets of cysteine, histidine, and methionine residues that are putative binding sites for type 1 blue copper(II) ions are designated C, H, and M. Negative values represent hydrophilic sections, and positive values represent hydrophobic sections.

β-turns (31%), and one-fifth α-helices (20%). Either the remainder (17%) is in a random structure or a clear choice of structure cannot be made by the method applied. These predictions, which emphasize the preponderance of β-sheet, β-turn, and random structure compared to a relatively low α-helix content, are in general accord with our earlier conclusions based on circular dichroism spectra (17) that suggested ceruloplasmin had approximately equal amounts of β and random structure and a small α-helix content (~10%).

Although the quantitative distribution of the several kinds of predicted secondary structures differs somewhat for each of the three homology units, these differences appear to be most significant in certain localized regions of the molecule. For brevity we illustrate this only for the A_2 domain and its adjoining boundaries with the A_1 and B domains (Fig. 4). This illustrates a significant point: although the change in hydropathy profile at the juncture of the three sets of A_2 and B domains is similar, the secondary structure differs for each interdomain segment as might be expected from the lack of homology in sequence for these regions shown in Fig. 2. The susceptibility to proteolytic cleavage at the juncture of A''_2 and B'' to yield the 50-kDa and 19-kDa fragments is explained by the accessibility of the basic highly charged connecting

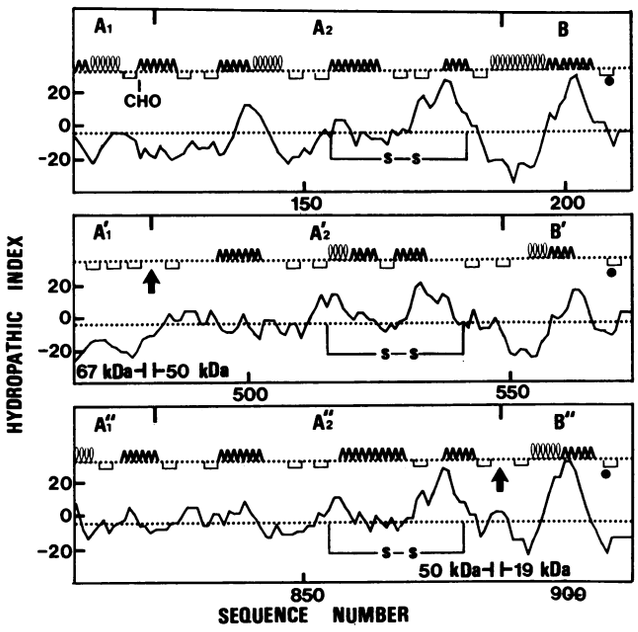

FIG. 4. Hydropathy profile (12) and secondary structure predictions (13) for the A_2 domains and the interdomain boundaries for human ceruloplasmin calculated with the PRPLOT program (11). The hydropathy profile is depicted as in Fig. 2, but is shown on an expanded scale for the segments illustrated. In the secondary structure prediction, the residues are represented by dots and are shown in α-helical (wwww), β-sheet (WWW), and β-turn ('⊔') conformational states. Other symbols have the same significance as in Fig. 2 except that the solid circles (at far right) indicate the three glucosamine acceptor sites that are not glycosylated.

segment that has a random structure and two β-turns. In contrast, the α-helix connecting the A_2 and B domains is resistant to proteolytic cleavage, but unexpectedly so is the segment connecting the A'_2 and B' domains despite its apparently open structure. Hence, an additional factor must be involved at this site.

In the case of the juncture of the three sets of A_1 and A_2 domains, the predicted secondary structural pattern (Fig. 4) is more consistent with the observed susceptibility to proteolytic cleavage. The carbohydrate at the juncture of A_1 and A_2 and the highly organized secondary structure confer resistance to proteolytic cleavage. In contrast, the random structure with four β-turns in the interdomain region between A'_1 and A'_2 facilitates proteolytic cleavage to yield the 67-kDa and 50-kDa fragments. However, the sequence connecting the A''_1 and A''_2 domains is in a β-sheet and resists tryptic cleavage because it doesn't have a single susceptible peptide bond; the only nearby basic amino acid is arginine-811, which is followed by proline (see Fig. 2).

Copper-Binding Sites. Although the ceruloplasmin molecule consists of three homologous units, the distribution of copper(II) ions among these units is apparently nonuniform. The exact stoichiometry is unclear, but the six (or seven) intrinsically bound copper(II) ions can be spectroscopically differentiated into several types usually given as: two type 1, one type 2, and three (or four) type 3 (2, 3, 18). The type 1, or blue copper(II) ion, has the strong absorption at about 610 nm and the hyperfine EPR signal characteristic of all blue proteins, including azurin and plastocyanin. The latter contain a single blue copper(II) ion, and crystallographic studies have identified and ligands as a clustered group of cysteine, methionine, the histidine residues and a second histidine. Because of sequence homology to azurin and plastocyanin, we have proposed (5, 8, 10) that a type 1 blue copper(II) ion in ceruloplasmin is bound to the clustered cysteine, methio-

nine, and histidine residues of the 19-kDa fragment (designated C, H, and M in the B″ domain in Fig. 3). The second type 1 blue copper(II) ion could be bound to the homologous set of cysteine, methionine, and histidine residues in the B′ domain. However, the two blue copper(II) ions of ceruloplasmin can be differentiated spectroscopically into type 1a "fast," which does not require methionine as a ligand, and type 1b "slow," which does (18). Hence, the B domain of the first homology unit also is a potential site, although the homologous methionine is not present. In addition to binding a type 1 blue copper(II) ion, the B″ domain probably binds several nonblue copper(II) ions because of its high histidine content and its sequence homology to superoxide dismutase and cytochrome oxidase in a histidine-rich area (8). Thus, active sites with functions differing in degree or kind may be located in different domains of this multicopper multifunctional oxidase.

Evolution of Ceruloplasmin. The preceding results have shown that ceruloplasmin consists of three contiguous repeating homology units that are so closely related in amino acid sequence that they appear to have been derived by successive duplication of a primordial gene coding for a copper-binding protein of about 350 amino acid residues in length. Because the alignment scores for the pairwise comparisons of the homology units are so similar (36–42 SD), the order of the gene duplications cannot be decided. Two consecutive elongations probably occurred close together on an evolutionary time scale. Without comparative sequence data and molecular weights for ceruloplasmin of other species, the time of the first event is difficult to estimate. However, Doolittle (19) suggests it probably occurred at about the same time as for transferrin, some 400 million years ago. Although our statistical comparison by the computer program SEARCH did not identify any significant homology of the sequences of human transferrin and ceruloplasmin, there are several points of relationship besides the fact that both are metal-binding plasma glycoproteins. Their genes are linked (20) and, thus, are both probably on chromosome 3 (21) to which transferrin and its receptor have been mapped. The ferroxidase activity of ceruloplasmin on ferrous iron facilitates uptake and transport of ferric iron by transferrin (1–3). Furthermore, human transferrin consists of two homologous halves, each of about 340 residues, that are about 42% identical in sequence (22).

Our determination of the complete amino acid sequence of human ceruloplasmin should facilitate the cloning and nucleotide sequence determination of the entire ceruloplasmin gene. This would provide a test of our hypothesis that ceruloplasmin evolved by triplication of an ancestral gene and may show a relationship between the domains proposed in our structural model and the exons in the genomic structure.

We thank P. H. Davidson, S. A. Dorwin, J. Madison, Y. Takahashi, and K. Huss for their excellent technical assistance. This work was supported by Grant AM 19221 from the National Institutes of Health.

1. Poulik, M. D. & Weiss, M. L. (1975) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, New York), 2nd Ed., Vol. 2, pp. 51–81.
2. Zgirslti, A. M., Chidambarom, M. & Frieden, E. (1982) in *Inflammatory Diseases and Copper*, ed. Sorensen, J. R. J. (Humana, Clifton, NJ), pp. 171–181.
3. Frieden, E. (1980) *Ciba Found. Symp.* **79**, 93–124.
4. Takahashi, N., Ortel, T. L. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 390–394.
5. Takahashi, N., Bauman, R. A., Ortel, T. L., Dwulet, F. E., Wang, C.-C. & Putnam, F. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 115–119.
6. Ortel, T. L., Takahashi, N. & Putnam, F. W. (1983) *J. Chromatogr.* **266**, 257–263.
7. Ortel, T. L. (1983) Dissertation (Indiana Univ., Bloomington).
8. Kingston, I. B., Kingston, B. L. & Putnam, F. W. (1980) *J. Biol. Chem.* **255**, 2886–2896.
9. Dwulet, F. E. & Putnam, F. W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 790–794.
10. Dwulet, F. E. & Putnam, F. W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2805–2809.
11. 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) in *Atlas of Protein Sequence and Structure, Protein Sequence Database*, Version 7 (Natl. Biomed. Res. Found., Washington, DC).
12. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
13. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47**, 251–276.
14. 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.
15. Doolittle, R. F. (1981) *Science* **214**, 149–159.
16. Kabsch, W. & Sander, S. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1075–1078.
17. Noyer, M. & Putnam, F. W. (1981) *Biochemistry* **20**, 3536–3542.
18. Hervé, M., Garnier, A., Tosi, L. & Steinbuch, M. (1981) *Eur. J. Biochem.* **116**, 177–183.
19. Doolittle, R. F. (1984) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, New York), 2nd Ed., Vol. 4, in press.
20. Westkamp, L. R. (1983) *Ann. Hum. Genet.* **47**, 293–297.
21. Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., van Bragt, P. H., Baldwin, W. D. & Bowman, B. H. (1984) *ICSU Short Rep.* **1**, 268–269.
22. MacGillivray, R. T. A., Mendez, E., Shewale, J. G., Sinha, S. K., Lineback-Zims, J. & Brew, K. (1983) *J. Biol. Chem.* **258**, 3543–3553.