Single-chain structure of human ceruloplasmin: The complete amino acid sequence of the whole molecule

Nobuhro Takahashi, Thomas L. Ortel*, and Frank W. Putnam

Department of Biology, Indiana University, Bloomington, IN 47405

Contributed by Frank W. Putnam, September 29, 1983

ABSTRACT We have determined the amino acid sequence of the amino-terminal 67,000-dalton (67-kDa) fragment of human ceruloplasmin and have established overlapping sequences between the 67-kDa and 50-kDa fragments and between the 50-kDa and 19-kDa fragments. The 67-kDa fragment contains 480 amino acid residues and three glucosamine oligosaccharides. These results together with our previous sequence data for the 50-kDa and 19-kDa fragments complete the amino acid sequence of human ceruloplasmin. The polypeptide chain has a total of 1,046 amino acid residues (M, 120,085) and has attachment sites for four glucosamine oligosaccharides; together these account for the total molecular mass of human ceruloplasmin (132 kDa). The sequence analysis of the peptides overlapping the fragments showed that one additional amino acid, arginine, is present between the 67-kDa and 50-kDa fragments, and another, lysine, is between the 50-kDa and 19-kDa fragments. Only two apparent sites of amino acid interchange have been identified in the polypeptide chain. Both involve a single-point interchange of glycine and lysine that would result in a difference in charge. The results of the complete sequence analysis verified that human ceruloplasmin is composed of a single polypeptide chain and that the subunit-like fragments are produced by proteolytic cleavage during purification (and possibly also in vivo).

Ceruloplasmin [Cp; ferroxidase; iron (II):oxygen oxidoreductase, EC 1.16.3.1] is a blue α₂-glycoprotein that binds 90–95% of blood plasma copper and has six (or seven) cupric ions per molecule. The various functions ascribed to this protein are not clearly understood, but at least four have been reported: ferroxidase activity, amine oxidase activity, copper transport and homeostasis, and superoxide dismutase activity (1–3). None of these functions is mutually exclusive, and it is not known whether the sites for different activities are located separately along the polypeptide chain or not. Ceruloplasmin is characteristically decreased in the hereditary disorder Wilson disease (1, 2), and there are unconfirmed reports that ceruloplasmin is abnormal or defective in this disease. Genetic polymorphism of ceruloplasmin is known at a low level in several populations. In black Americans, the frequency of the genetic variant CpA is 0.052, but in pooled plasma from the red Cross the incidence of CpA probably does not exceed 0.6% (1).

The conclusions on the number and size of subunit chains of ceruloplasmin reported in early studies (1, 4–7) were rather contradictory. The confusion originated from the extreme susceptibility of the ceruloplasmin molecule to limited proteolysis. The analysis was complicated further by the following factors: (i) difficulty in preventing autolytic degradation during purification, (ii) strong interaction among the degraded fragments, which could be dissociated only by use of strong denaturing agents, (iii) technical difficulties in sequence determination of such a large protein, and (iv) the extraordinary degree of internal duplication in the sequence of the polypeptide chain later recognized by us (8–11).

Although the single-chain structure was proposed in 1972 by Ryden (7), reports of multisubunit structures for ceruloplasmin continued to appear (1, 4, 5, 6) until Kingston et al. (12) demonstrated that human ceruloplasmin prepared by different procedures contained varying amounts of discrete, subunit-like fragments and showed that the form usually isolated contained three polypeptides with molecular masses of 67 kilodaltons (kDa), 50 kDa, and 19 kDa, respectively. Kingston et al. (12) also showed that the principal fragments could be aligned as a single chain with 67 kDa as the amino terminus, 50 kDa as the central segment, and 19 kDa as the carboxyl terminus. Our laboratory later reported the amino acid sequence of the 19-kDa fragment (8) and then the sequence of the 50-kDa fragment (9). Comparison of these revealed an internal duplication in sequence that reflected the existence of two homology regions and also indicated the probable location of the binding sites for type 1 blue copper ions (10). We also published part of the amino acid sequence of the 67-kDa fragment and showed that there was an internal triplication of the structure of ceruloplasmin (11). At the same time, we proposed a structural model of human ceruloplasmin that was based on our sequence analysis, on computerized analysis of the internal homology, and on the sites of autolytic and tryptic cleavage (11). The model showed the location of the probable disulfide bonds and the glucosamine oligosaccharide attachment sites and depicted an alternating structure of six domains of two different kinds (or possibly nine domains of three kinds). We have now completed the determination of the amino acid sequence of the 67-kDa fragment and have also established the sequences overlapping the 67-kDa and 50-kDa fragments and also the 50-kDa and 19-kDa fragments. The results verify that human ceruloplasmin consists of a single polypeptide chain with 1,046 amino acid residues and that the subunit-like fragments are produced by proteolytic cleavage.

MATERIALS AND METHODS

Materials. Two preparations of human ceruloplasmin purified from normal pooled plasma were studied. One had undergone autolytic proteolysis and was a mixture of three main fragments with approximate molecular masses of 19 kDa, 50 kDa, and 67 kDa (12). This preparation was the source of the 19-kDa fragment and the 50-kDa fragment for which the complete sequences were reported by Kingston et al. (8, 12) and Dwulet and Putnam (9), respectively. We isolated the 67-kDa fragment and characterized it by methods they have described. The second preparation (Cp4) was undegraded single-chain ceruloplasmin with a molecular mass

Abbreviation: kDa, kilodalton(s).
*Present address: Indiana University School of Medicine, Indianapolis, IN 46223.
†To whom reprint requests should be addressed.
FIG. 1. Summary of the complete amino acid sequence of the 67-kDa fragment of human ceruloplasmin. Although >100 peptides were subjected to sequence analysis, only peptides needed to establish the sequence are shown. The peptides obtained from different digestions are as follows: —, cyanogen bromide peptides; —-, tryptic peptides; ——, chymotryptic peptides; | | | | | , * , S. aureus V8 peptides. One asterisk shows tryptic peptides of the citraconylated fragment, and two asterisks indicate subpeptides of tryptic peptides with S. aureus V8 protease.
Fig. 2. The complete amino acid sequence of the whole molecule of human ceruloplasmin. A double numbering system is used; the upper set represents the actual position beginning with the amino terminus. The three fragments that were subjected to sequence analysis independently are designated 67-kDa, 50-kDa, and 19-kDa above the amino terminus of the sequence. In the 50-kDa fragment, the lower set is numbered consecutively beginning with the amino terminus of the 50-kDa fragment (Ser-1). In the 19-kDa fragment, the lower set is numbered...
of 132 kDa and was prepared and characterized as described by Noyer et al. (13). These preparations, whether intact (single-chain) or spontaneously cleaved, consisted predominantly of type I ceruloplasmin with a small amount of type II. The two types were separated by gel chromatography (13); the only difference between them is reported to be in the carbohydrate content (14). All sequence analysis was done on a mixture containing predominantly type I.

**Methods.** The sequence analysis of the purified 67-kDa fragment was described previously by Takahashi et al. (11). To get overlapping peptides between the 67-kDa and 50-kDa fragments and between the 50-kDa and 19-kDa fragments, undegraded single-chain ceruloplasmin was reduced, carboxymethylated, and digested with *Staphylococcus aureus* V8 protease. The digest was first fractionated by gel filtration on Sephadex G-50 and separated by high-performance liquid chromatography with a linear gradient of n-propanol containing 0.1% trifluoroacetic acid. The final purification was performed by rechromatography with a linear gradient of n-propanol containing 0.1% heptfluorobutyric acid (15).

**RESULTS AND DISCUSSION**

The amino acid sequence of the amino-terminal 67-kDa fragment was deduced from >100 peptides obtained by cyanogen bromide cleavage and by trypic, chymotryptic, and *S. aureus* V8 protease digestions of the alkylated fragment or the citraconylated fragment (Fig. 1). Each assigned residue was established in a minimum of two peptides isolated from different digestes. All overlaps were also 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. One amino acid was missing in the sequence reported in our previous paper (11); however, we have confirmed the presence of residue Leu-325 between His-324 and Lys-326 in the sequence of the 67-kDa fragment. The amino-terminal sequence of the first seven amino acid residues of the 67-kDa fragment was identical with that of the undegraded single-chain ceruloplasmin, revealing that the 67-kDa fragment is the amino-terminal fragment of the native intact protein. The carboxy-terminal amino acid of the 67-kDa fragment was identified as serine by hydrazinolysis (16); no carboxy-terminal peptides obtained from the various enzymatic digests had any amino acid other than serine as the carboxy-terminal residue. The proof of the sequence analysis of the 67-kDa fragment is summarized in Fig. 1. The 67-kDa fragment is composed of 480 amino acid residues and has attachment sites for three GlcN oligosaccharides.

**Single-Chain Structure of Ceruloplasmin.** The overlapping sequences between the 67-kDa and 50-kDa fragments and between the 50-kDa and 19-kDa fragments were both obtained from the *S. aureus* V8 protease digest of carboxymethylated single-chain ceruloplasmin. Although the digest was an extremely complex mixture of peptides produced from one of the largest known polypeptide chains (132 kDa), the purification of the peptides was greatly facilitated by prediction of the overlapping sequences expected from the results of sequence analysis of the spontaneously fragmented ceruloplasmin. The calculations of the molecular weights and hydrophobicities of the expected peptides made possible the prediction of the elution positions from the gel filtration column and from the reverse-phase column used (17). The peptides obtained covered the region from Gly-469 to Glu-494, which gave the overlapping sequence between the 67-kDa and the 50-kDa fragment, and the region from Leu-870 to Glu-896, which gave the overlapping sequence between the 50-kDa and 19-kDa fragments (Fig. 2). Additional amino acids, Arg-481 and Lys-510, were present in the 67-kDa and 50-kDa fragments and between the 50-kDa and 19-kDa fragments, respectively. Consequently, the three fragments produced spontaneously during purification (and possibly also in vivo) are aligned as a single chain with 67-kDa (480 residues) as the amino terminus, 50-kDa (405 residues) as the central segment, and 19-kDa (159 residues) as the carboxy terminus (Fig. 2). Although, the amino-terminal sequence of the 67-kDa fragment is consistent with that of the undegraded single-chain ceruloplasmin molecule, we have only weak evidence yet that the carboxy terminus of the 19-kDa fragment is identical with that of the undegraded single-chain molecule. Hydrazinolysis and digestion with carboxypeptidases A and B suggested that glycine is the carboxy-terminal residue of intact ceruloplasmin but the polypeptide chain is too large for accurate determination of the carboxy terminus by these methods.

The sequence of the 67-kDa fragment together with our previous results for the 50-kDa and 19-kDa fragments as well as the overlapping sequences between the three fragments give the complete single-chain sequence of human ceruloplasmin shown in Fig. 2. The single polypeptide chain of intact human ceruloplasmin contains 1,046 amino acid residues; the molecular mass calculated from the amino acid sequence of the polypeptide portion is 120,085 daltons. If the four GlcN glycans are taken into consideration, the molecular mass of human ceruloplasmin is about 132 kDa, which is comparable with the value of 130 kDa estimated by polyacrylamide gel electrophoresis in our previous report (12). The probable linkage of disulfide bridges is as follows: Cys-155 to Cys-181, Cys-257 to Cys-338, Cys-515 to Cys-541, Cys-618 to Cys-699, and Cys-855 to Cys-881, and there are probably free sulphydryl groups at Cys-221, Cys-319, Cys-680, and Cys-1021 (11).

**Internal Tripllication in the Single-Chain Molecule of Human Ceruloplasmin.** Computer-assisted analysis of the primary structure of human ceruloplasmin indicated that the entire molecule exhibits a 3-fold internal homology in amino acid sequence, with each homology unit containing about 350 residues (11). All three homology units exhibit about 30% identity in sequence, and each pair exhibits about 40% identity (11). This latter value is near the degree of identity between the α and β chains of human hemoglobin and the κ and λ light chains of human immunoglobulins. The extraordinary 3-fold internal homology supports the idea proposed by Dwulet and Putnam (10) that the ceruloplasmin molecule evolved by tandem triplication of ancestral genes.

**GlcN Oligosaccharide Attachment Sites of the Molecule.** All three glycopeptides purified from the 67-kDa fragment by Tetaert et al. (15) are placed in the sequence. Altogether there are four sites where sugar is attached, three in the 67-kDa fragment and one in the 50-kDa fragment (Fig. 2). However, comparison of the peptide maps of the two forms of human ceruloplasmin (type I and type II) suggested that the second oligosaccharide (Asn-339) is missing in type II, whereas type I (the predominant form) has all four oligosaccharides (18). Three other sites with the acceptor sequence Asn-X-Thr/Ser are not glycosylated (Asn-208, Asn-569, and Asn-907). The three unglycosylated sites are all in a β-turn structure and are close to or in a very hydrophobic region, in...
which they may be buried. Moreover, the three sites are in segments of sequence that have highest degree of homology in the whole molecule (11, 18). The carbohydrate of human ceruloplasmin is composed only of bi- and triantennary N-glycosidic glycans (19). However, it is not yet known what the structure of the oligosaccharide is at each attachment site.

**Polymorphism of Ceruloplasmin.** We have identified two apparent sites of amino acid interchange in the polypeptide sequence at positions 79 and 449, respectively. Both occur within the domains A1 and A2; proposed in our previous paper (11), and both involve a single-point interchange of glycine and lysine. The ratios of the yield of the two amino acids were about Gly/Lys = 20/80 in both sites. Both sites are in or adjacent to a β turn that lies between β-sheet structures, suggesting that both sites are exposed at the surface of the molecule. The two interchanges are associated with a potential charge difference in the protein. It is not clear if both interchanges in the 67-kDa fragment occur in the same molecule. However, the interchanges appear to be unrelated to the difference between type I and type II ceruloplasmin. As yet, we do not know if the glycine-lysine interchanges are correlated with known variants. Because phenotype CpB has a higher charge than CpA (based on electrophoretic mobilities), the presence of lysine should be associated with CpB, but the incidence of CpA is much too low to account for this. However, because the sequence of human ceruloplasmin is now completed, the identification of structural differences in genetic variants may be possible in the near future (17).

**Domain Structure and Proteolytic Degradation of the Single-Chain Ceruloplasmin.** The determination of the complete amino acid sequence of ceruloplasmin verifies that the protein is synthesized and secreted as a single-chain molecule; however it undergoes autolytic proteolytic cleavage during purification (and possibly also in vivo). This spontaneous cleavage produces three predominant fragments with molecular masses of 67 kDa, 50 kDa, and 19 kDa (12), and further degradation produces smaller fragments. These observations now turn out to be evidence that ceruloplasmin is composed of a series of discrete proteolysis-resistant domains (11). The intact native ceruloplasmin also can readily be cleaved in vitro by enzymes such as trypsin, plasmin, or kallikrein to produce domain-like fragments (18). However, the main differences between the spontaneous degradation during purification and the limited proteolysis in vitro are that (i) the limited proteolysis rapidly yields 25-kDa and 26-kDa fragments by an additional split in a fragment corresponding to the 50-kDa fragment (11, 18); (ii) the 67-kDa fragment also is cleaved slowly into 49-kDa and 18-kDa fragments by the limited cleavage (11, 18); and (iii) two basic amino acids, Arg-481 and Lys-887, are clipped out during the spontaneous degradation. We do not know what enzyme is responsible for the spontaneous cleavage of ceruloplasmin during purification (and possibly in vivo). However, after the spontaneous proteolytic cleavage produces the predominant 67-kDa, 50-kDa, and 19-kDa fragments by an unknown trypsin-like enzyme (many of which exist in plasma), the single arginine and lysine residues are excised by a carboxypeptidase B-like enzyme from the carboxyl-terminal positions of the 67-kDa and 50-kDa fragments, respectively. Although we cannot be sure of the physiological importance of the limited proteolysis, it is intriguing to speculate that the fragments may have different specific biological activities because of the multifunctional nature of ceruloplasmin. In that case, the process may represent some unknown regulatory mechanism for the functioning of ceruloplasmin, because the process is very similar to those for the activation and deactivation of C3a and C5a anaphylatoxins and Bradykinin.

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.