Proc. Natl. Acad. Sci. USA
Vol. 78, No. 5, pp. 2805–2809, May 1981
Biochemistry

Internal duplication and evolution of human ceruloplasmin

(protein structure/internal homology/gene evolution/copper oxidases/ferroxidase)

FRANCIS E. DWULET* AND FRANK W. PUTNAM†

Department of Biology, Indiana University, Bloomington, Indiana 47405

Contributed by Frank W. Putnam, January 29, 1981

ABSTRACT With the completion of the primary structure of the 50,000- and 19,000-dalton fragments of human ceruloplasmin [ferroxidase; iron(II):oxygen oxidoreductase, EC 1.16.3.1], over half of the covalent structure of the single polypeptide chain of this protein is known. Visual and computer analysis of the sequence of the 564 amino acid residues in the two fragments gives clear evidence of statistically significant internal homology suggestive of evolutionary replication of two smaller units. Two homology regions, each composed of 224 residues, were defined by an intrasequence alignment that required only three gaps in each 224-residue segment. The two homology regions exhibited 43% identity in sequence, and 13% of the remaining positions had similar residues. The sequence of a 160-residue segment in ceruloplasmin exhibits significant homology to the active (copper-binding) sites of blue electron-transfer proteins such as azurins and plastocyanins and multicopper oxidases such as cytochrome oxidase and superoxide dismutase. It is proposed that a primitive ceruloplasmin gene was formed by the fusion of two genes coding, respectively, for proteins about 160 and 190 amino acid residues in length and that this precursor gene coding for about 350 amino acids was later triplicated to form the gene for the present-day ceruloplasmin molecule of about 1050 amino acids.

Analytical and statistical comparison of amino acid sequences has been a powerful method for detection and estimation of structural similarities among and within proteins, for evaluation of structure–function relationships in a family of proteins, and for study of genetic change in the course of evolutionary development of a class of proteins (1–3). The comparison may be done by visual alignment of two sequences for best fit as judged both by identity and similarity of paired amino acids at corresponding positions; gaps in the sequences are sometimes inserted to maximize the fit. Statistical methods for assessing relatedness of proteins are based on computer programs that accumulate pair scores for two amino acids, using a mutation data matrix based on physical and structural parameters and codon differences for all possible pairs of amino acids (4) or from actual accepted point mutations accumulated from related sequences (5). By such means the degree of intersequence or intrasequence homology may be assessed; here "homologous" is used to mean matching (identical or similar) in structure, position, physical characteristics, and codons.

With the recent completion of the primary structures of the 50,000-dalton (50-kDal) (6) and 19,000-dalton (19-kDal) (7–9) fragments of human ceruloplasmin [ferroxidase; iron(II):oxygen oxidoreductase, EC 1.16.3.1], 564 residues of the amino acid sequence of the single polypeptide chain of this protein are known. These fragments are from the carboxyl terminus of the molecule and account for more than half of the primary structure of this blue copper oxidase. During our study we identified a remarkable degree of internal homology in ceruloplasmin, which guided our sequence determination, and we noted significant homology to other copper-containing proteins (10). This report presents evidence for these observations and offers a model for the evolutionary development of ceruloplasmin and other multicopper oxidases.

METHODS

During the course of sequence analysis of the 50-kDal fragment it became apparent that many peptides were homologous to segments of sequence in the 19-kDal fragment (10). In fact, the placement of many peptides in the final sequence of the 50-kDal fragment was predicted from this homology. On completion of the sequence a search for intrasequence homology within the 50-kDal fragment and for intersequence homology between the two fragments was made by manual alignment of short segments against the entire structure until the optimal fit was found. This led to the alignment in Fig. 1, in which the carboxy-terminal 65 amino acids of the 50-kDal fragment followed by the 159 in the 19-kDal fragment are aligned with positions 1–224 of the 50-kDal fragment, leaving an unmatched stretch of 116 amino acids.

In order to test the validity of the alignment in Fig. 1 and to evaluate its statistical significance, the consecutive sequences of the 50-kDal and 19-kDal fragments (564 residues) were subjected to a search for intrasequence homology by using the program RELATE of the *Atlas of Protein Sequence and Structure* (5). This program compares all possible segments of a given length (in this case, 25 residues) both in the real sequence and in 100 random runs, and, using a mutation data matrix for all possible pairs of amino acids, it accumulates segment scores. The mean of a predetermined number of highest segment scores is expressed in standard deviation (SD) units. Also, by use of the program SEARCH, selected 25-residue segments of the 50-kDal and 19-kDal sequences were compared with all known protein sequences to seek hints of evolutionary similarity. The amino-terminal sequence of the single-chain molecule (11) was also compared with the entire 564-residue sequence to identify additional internal homology. These searches were done by W. C. Barker of the National Biomedical Research Foundation, Washington, DC.

RESULTS AND DISCUSSION

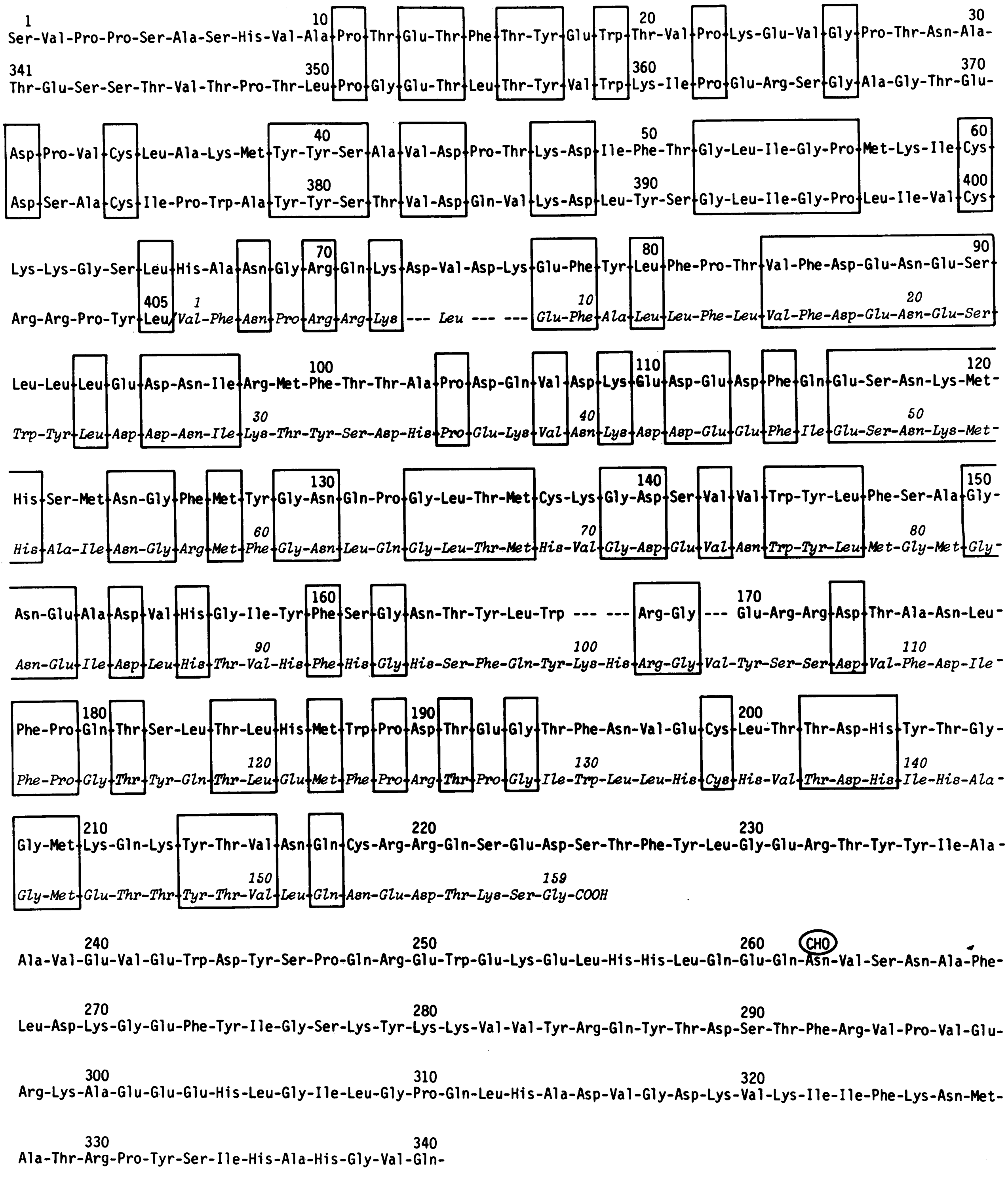
When the primary structures of the 50-kDal and 19-kDal fragments are aligned as in Fig. 1, two long regions of homologous sequences are readily identified, which suggest an internal duplication. We call the two regions homology segments; they are defined by aligning residues 1–224 of the 50-kDal fragment with residues 341–405 of the 50-kDal fragment immediately followed

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.

Abbreviation: kDal, kilodalton(s)

* Present address: Department of Medicine, Indiana University, Indianapolis, IN 46223.

† To whom reprint requests should be addressed.



```
1                                   10                     20                                          30
Ser-Val-Pro-Pro-Ser-Ala-Ser-His-Val-Ala-Pro-Thr-Glu-Thr-Phe-Thr-Tyr-Glu-Trp-Thr-Val-Pro-Lys-Glu-Val-Gly-Pro-Thr-Asn-Ala-
341                                 350                    360                                         370
Thr-Glu-Ser-Ser-Thr-Val-Thr-Pro-Thr-Leu-Pro-Gly-Glu-Thr-Leu-Thr-Tyr-Val-Trp-Lys-Ile-Pro-Glu-Arg-Ser-Gly-Ala-Gly-Thr-Glu-

                              40                                        50                                          60
Asp-Pro-Val-Cys-Leu-Ala-Lys-Met-Tyr-Tyr-Ser-Ala-Val-Asp-Pro-Thr-Lys-Asp-Ile-Phe-Thr-Gly-Leu-Ile-Gly-Pro-Met-Lys-Ile-Cys-
                              380                                       390                                         400
Asp-Ser-Ala-Cys-Ile-Pro-Trp-Ala-Tyr-Tyr-Ser-Thr-Val-Asp-Gln-Val-Lys-Asp-Leu-Tyr-Ser-Gly-Leu-Ile-Gly-Pro-Leu-Ile-Val-Cys-

                                 70                                           80                                90
Lys-Lys-Gly-Ser-Leu-His-Ala-Asn-Gly-Arg-Gln-Lys-Asp-Val-Asp-Lys-Glu-Phe-Tyr-Leu-Phe-Pro-Thr-Val-Phe-Asp-Glu-Asn-Glu-Ser-
                405 1                                          10                                   20
Arg-Arg-Pro-Tyr-Leu/Val-Phe-Asn-Pro-Arg-Arg-Lys --- Leu --- --- Glu-Phe-Ala-Leu-Leu-Phe-Leu-Val-Phe-Asp-Glu-Asn-Glu-Ser-

                                  100                                   110                                     120
Leu-Leu-Leu-Glu-Asp-Asn-Ile-Arg-Met-Phe-Thr-Thr-Ala-Pro-Asp-Gln-Val-Asp-Lys-Glu-Asp-Glu-Asp-Phe-Gln-Glu-Ser-Asn-Lys-Met-
                            30                                      40                                          50
Trp-Tyr-Leu-Asp-Asp-Asn-Ile-Lys-Thr-Tyr-Ser-Asp-His-Pro-Glu-Lys-Val-Asn-Lys-Asp-Asp-Glu-Glu-Phe-Ile-Glu-Ser-Asn-Lys-Met-

                                     130                                        140                                150
His-Ser-Met-Asn-Gly-Phe-Met-Tyr-Gly-Asn-Gln-Pro-Gly-Leu-Thr-Met-Cys-Lys-Gly-Asp-Ser-Val-Val-Trp-Tyr-Leu-Phe-Ser-Ala-Gly-
                                 60                                      70                                      80
His-Ala-Ile-Asn-Gly-Arg-Met-Phe-Gly-Asn-Leu-Gln-Gly-Leu-Thr-Met-His-Val-Gly-Asp-Glu-Val-Asn-Trp-Tyr-Leu-Met-Gly-Met-Gly-

                                       160                                                 170
Asn-Glu-Ala-Asp-Val-His-Gly-Ile-Tyr-Phe-Ser-Gly-Asn-Thr-Tyr-Leu-Trp --- --- Arg-Gly --- Glu-Arg-Arg-Asp-Thr-Ala-Asn-Leu-
                               90                                  100                                     110
Asn-Glu-Ile-Asp-Leu-His-Thr-Val-His-Phe-His-Gly-His-Ser-Phe-Gln-Tyr-Lys-His-Arg-Gly-Val-Tyr-Ser-Ser-Asp-Val-Phe-Asp-Ile-

        180                                  190                                        200
Phe-Pro-Gln-Thr-Ser-Leu-Thr-Leu-His-Met-Trp-Pro-Asp-Thr-Glu-Gly-Thr-Phe-Asn-Val-Glu-Cys-Leu-Thr-Thr-Asp-His-Tyr-Thr-Gly-
                            120                                     130                                     140
Phe-Pro-Gly-Thr-Tyr-Gln-Thr-Leu-Glu-Met-Phe-Pro-Arg-Thr-Pro-Gly-Ile-Trp-Leu-Leu-His-Cys-His-Val-Thr-Asp-His-Ile-His-Ala-

        210                                  220                                        230
Gly-Met-Lys-Gln-Lys-Tyr-Thr-Val-Asn-Gln-Cys-Arg-Arg-Gln-Ser-Glu-Asp-Ser-Thr-Phe-Tyr-Leu-Gly-Glu-Arg-Thr-Tyr-Tyr-Ile-Ala-
                            150                                159
Gly-Met-Glu-Thr-Thr-Tyr-Thr-Val-Leu-Gln-Asn-Glu-Asp-Thr-Lys-Ser-Gly-COOH

        240                                  250                                        260  CHO
Ala-Val-Glu-Val-Glu-Trp-Asp-Tyr-Ser-Pro-Gln-Arg-Glu-Trp-Glu-Lys-Glu-Leu-His-His-Leu-Gln-Glu-Gln-Asn-Val-Ser-Asn-Ala-Phe-

        270                                  280                                        290
Leu-Asp-Lys-Gly-Glu-Phe-Tyr-Ile-Gly-Ser-Lys-Tyr-Lys-Lys-Val-Val-Tyr-Arg-Gln-Tyr-Thr-Asp-Ser-Thr-Phe-Arg-Val-Pro-Val-Glu-

        300                                  310                                        320
Arg-Lys-Ala-Glu-Glu-Glu-His-Leu-Gly-Ile-Leu-Gly-Pro-Gln-Leu-His-Ala-Asp-Val-Gly-Asp-Lys-Val-Lys-Ile-Ile-Phe-Lys-Asn-Met-

        330                                  340
Ala-Thr-Arg-Pro-Tyr-Ser-Ile-His-Ala-His-Gly-Val-Gln-
```

FIG. 1. The entire 564 residues of the 50-kDal and 19-kDal fragments of human ceruloplasmin aligned for maximum homology. The 405-residue sequence of the 50-kDal fragment (6) is shown in gothic type and is followed by the 159-residue sequence of the 19-kDal fragment (7–9) in italic type. The shilling mark between the two sequences indicates that no overlap has yet been established. The boxed pairs of residues are identical; many other pairs are homologous (see text). The glucosamine oligosaccharide at Asn-262 is designated CHO.

by the 159 residues of the 19-kDal fragment. Both homology segments contain 224 amino acids, and each has a three-residue gap to maximize the homology. In this alignment there are no matching sequences for an internal part of the 50-kDal fragment composed of residues 225–340. However, this segment is probably homologous to the amino-terminal sequence of the intact ceruloplasmin molecule (see later).

Many short stretches of identical sequence are present in the two homology segments. Of these, the most prominent are the seven-residue sequence Val-Phe-Asp-Glu-Asn-Glu-Ser and the six-residue sequence Glu-Ser-Asn-Lys-Met-His, which are present in both the 50-kDal and the 19-kDal fragments (see positions 84–90 and 116–121, respectively, in the sequence of the 50-kDal fragment). In addition, there is one five-residue repeat within the 50-kDal fragment (Gly-Leu-Ile-Gly-Pro) and one four-residue sequence (Gly-Leu-Thr-Met) that is present in both fragments. As further evidence that the homology is not due to chance, there are six pairs of identical three-residue sequences; one pair is a repeat within the 50-kDal fragment, and five occur on alignment of the two fragments. Finally, there are numerous instances in which the two homology segments are identical at one or two positions. Altogether, the two homology segments are identical at 97 positions out of 224 pairs compared and thus have 43.3% identity. This is almost exactly the degree of identity between the α and β chains of human hemoglobin and of the κ and λ chains of human immunoglobulins. Duplication of an ancestral gene has been proposed as the mechanism for evolutionary divergence for superfamilies of proteins such as the globins and immunoglobulins (3). We postulate that internal duplication was the mechanism for evolutionary expansion of the blue multicopper oxidases such as ceruloplasmin.

Not only are 43% of the paired amino acids identical in the two 224-residue sequences compared in Fig. 1, but an additional 13% are homologous in the sense that they have similar physical parameters and structural characteristics. There are frequent interchanges within the groups of hydrophobic, aromatic, basic, acidic, polar, or neutral amino acids. For example, there are six instances of the pair serine/threonine, four of aspartic acid/glutamic acid, four of tyrosine/phenylalanine, and nine involving combinations of valine, leucine, or isoleucine.

Computer Analysis of the Statistical Significance of Internal Homology. The very high degree of internal homology in a continuous alignment of the sequences of the 50-kDal and 19-kDal fragments was confirmed by W. C. Barker in an intrasequence computer search using the program RELATE. In a comparison of all possible sequence lengths of 25 residues, the top 164 scores were for alignments that were separated by either 340 or 337 residues. The difference of three residues accords with the gaps we inserted in both the 50-kDal and the 19-kDal fragments (Fig. 1). Thus, the computer search with the RELATE program showed that a long stretch beginning with residue 1 will align with residue 341, as we had already determined by visual inspection. The distance of the real sequence comparison from the average of 100 randomized sequence comparisons = 25.46 SD. This is the highest score for duplication in protein sequence yet reported except for the well-known internal duplication in the α-2 chain of human haptoglobin and the 5-fold replication of a segment length of 79 residues in human plasminogen (2, 5).

Location of Free Sulfhydryl Groups. Ceruloplasmin is reported to contain three free sulfhydryl groups (12). Recently, Rydén and Lundgren (13) reported the sequence of two short peptides designated Cer 1 and Cer 2, which contain cysteine. These cysteines are found in the 19-kDal and 50-kDal fragments, respectively, and are located at positions 134 of the 19-kDal fragment (Cer 1) and 199 of the 50-kDal fragment (Cer 2). The sequences around these cysteine residues are homologous and show great similarity to the active site amino acid residues reported for azurin (see below).

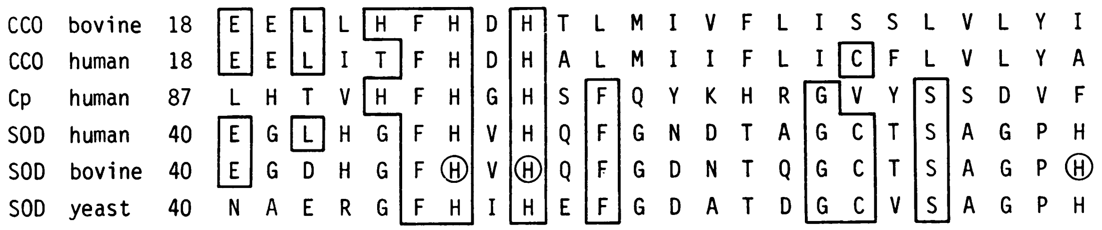
Disulfide Bridging Pattern. Because cleavage experiments with plasmin gave the same size pattern for all fragments both before and after reduction, it has been assumed that all disulfide bridges are within each fragment (14). From this and from the striking homology of the half-cystine residues in the third and fourth lines of Fig. 1, we predict that the three disulfide bridges within the 50-kDal fragment are as follows: Cys-34 to Cys-60, Cys-137 to Cys-218, and Cys-374 to Cys-400. This bridging pattern would provide the most parsimonious interpretation of the data, but it has yet to be confirmed by the actual isolation of the bridge peptides.

Types of Copper in Ceruloplasmin and Other Copper-Containing Proteins. Although copper may be bound to proteins in several forms, all blue proteins such as ceruloplasmin owe their intense color to the so-called type 1 Cu^{2+} ions. Unlike the small blue electron-transfer proteins such as the azurins and plastocyanins, which contain a single type 1 copper and have molecular weights of only 10,500 to 14,000, ceruloplasmin is a large multicopper oxidase and contains six copper ions, generally given as two type 1, one type 2, and three type 3 (15, 16). Type 1 copper, often called "blue copper," has strong absorbance around 600 nm, is detectable by electron paramagnetic resonance (EPR), and is characteristic of all blue proteins. Type 2 Cu^{2+}, often called "nonblue copper," because of its weak absorption at 600 nm, is also detectable by EPR; it is present not only in other large multicopper blue oxidases such as laccase and cytochrome oxidase but also in other nonblue oxidases such as monoamine oxidase and galactose oxidase. The EPR-nondectable type 3 Cu^{2+} ions are thought to constitute antiferromagnetically coupled binuclear metal centers and are present in blue oxidases that can catalyze the reduction of dioxygen to two molecules of water, e.g., ceruloplasmin, laccase, and cytochrome oxidase (16). Recently, another EPR signal from Cu^{2+} has been discovered in the latter proteins, which is similar to that in other metalloproteins containing Cu^{2+} in binuclear centers, e.g. superoxide dismutase (17).

Homology at the Binding Site for Type 1 Copper. Because crystallographic structures and amino acid sequences are established for some of the small blue proteins and some of the large multicopper oxidases, we have searched our sequence data for ceruloplasmin for segments homologous to the known copper-binding sites in these proteins. We have shown (7–9, 14) that the carboxy-terminal sequence of the 19-kDal fragment

STN		G	Q	K	Y	Y	I	C	G	V	P	K	H	C	D	L	G	Q	K
CCO-B	194	G	Q	C	S	E	I	C	G	S	N	-	H	S	F	-	-	M	P
CCO-H	194	G	Q	C	S	E	I	C	G	A	N	-	H	S	F	-	-	M	P
Pl	83	G	E	Y	T	F	Y	Ⓒ	E	-	P	-	Ⓗ	R	G	A	G	Ⓜ	V
Az	106	E	Q	Y	M	F	F	Ⓒ	T	F	P	G	Ⓗ	S	-	A	L	Ⓜ	K
Cp 19 kDal	128	G	I	W	L	L	H	C	H	V	T	D	H	I	H	A	G	M	E
Cp 50 kDal	193	G	T	F	N	V	E	C	L	T	T	D	H	Y	T	G	G	M	K
Laccase						/L	H	C	H	I	B	F/							

FIG. 2. Homology of blue oxidases at the binding site for type 1 copper in azurin (Az) and plastocyanin (Pl). Only residues identical with those in both the 19-kDal and the 50-kDal fragments of human ceruloplasmin (Cp) are boxed. The cysteine, histidine, and methionine are known ligands to blue type 1 Cu^{2+} in azurin (18) and plastocyanin (19). In these proteins, these ligands are close to the carboxy terminus, as they also are in stellacyanin (STN), human and bovine cytochrome oxidase polypeptide II (CCO-H and CCO-B) and the 19-kDal fragment of ceruloplasmin. The sources of sequence data are given in the text. The one-letter notation for amino acids is given in ref. 20. For each sequence the position number is given for the first amino acid shown.



CCO	bovine	18	E	E	L	L	H	F	H	D	H	T	L	M	I	V	F	L	I	S	S	L	V	L	Y	I
CCO	human	18	E	E	L	I	T	F	H	D	H	A	L	M	I	I	F	L	I	C	F	L	V	L	Y	A
Cp	human	87	L	H	T	V	H	F	H	G	H	S	F	Q	Y	K	H	R	G	V	Y	S	S	D	V	F
SOD	human	40	E	G	L	H	G	F	H	V	H	Q	F	G	N	D	T	A	G	C	T	S	A	G	P	H
SOD	bovine	40	E	G	D	H	G	F	Ⓗ	V	Ⓗ	Q	F	G	D	N	T	Q	G	C	T	S	A	G	P	Ⓗ
SOD	yeast	40	N	A	E	R	G	F	H	I	H	E	F	G	D	A	T	D	G	C	V	S	A	G	P	H

FIG. 3. Homology of multicopper oxidases at the binding site for nonblue copper in bovine superoxide dismutase (SOD). Residues identical in two or more enzymes of different specificity are boxed. Histidines at the copper-binding site of bovine superoxide dismutase (23, 24) are enclosed in circles. Sources of sequence data: cytochrome oxidase polypeptide II (CCO), human (25), bovine (26); human ceruloplasmin (Cp) 19-kDal fragment (7); superoxide dismutase, human (27), bovine (24), yeast (28). For each sequence the position number is given for the first amino acid shown.

of ceruloplasmin has a cysteine, histidine, and methionine in positions homologous to these amino acids in the azurins and plastocyanins, and we proposed that the three amino acids were involved in binding a type 1 copper ion. The 50-kDal fragment has a cysteine, histidine, and methionine in homologous positions (Fig. 2). Crystallographic structures of azurin (18) and plastocyanin (19) confirmed that the combining site did involve these three amino acids as ligands, and that in azurin the fourth ligand was a histidine which was 66 residues amino-terminal to the cysteine. Significantly, in the 19-kDal fragment there is a histidine (His-69) that is 65 residues amino-terminal to the cysteine residue (Cys-134), but in the 50-kDal fragment this is replaced by cysteine (Cys-137) (Fig. 1). There are also homologous cysteine, histidine, and methionine residues in human and bovine cytochrome oxidase. Stellacyanin (21) and fungal laccase B (22) are also homologous to ceruloplasmin in this region; indeed, the 19-kDal fragment of ceruloplasmin and laccase have the identical sequence Leu-His-Cys-His. We propose, therefore, that the binding site for type 1 copper is similar in the small blue electron-transfer proteins and in the large multicopper oxidases and that all these enzymes may have evolved from the same ancestral gene that coded for a small blue protein having electron-transfer or oxidase function.

Homology of Ceruloplasmin and Other Copper Oxidases. A different sequence in the 19-kDal fragment of ceruloplasmin is homologous to a known site for binding nonblue copper in superoxide dismutase (23, 24) and also to other sequences in cytochrome oxidase (25, 26) (Fig. 3). Comparison by the computer program RELATE of the sequence of the 19-kDal fragment of ceruloplasmin to that of bovine superoxide dismutase gave the highest score to the 15-residue sequence of positions 92–106 in the 19-kDal fragment and positions 43–57 of bovine superoxide dismutase. Crystallographic study of bovine superoxide dismutase (23) has shown that His-46, His-48, and His-63 are ligands to copper. Although the 19-kDal fragment has histidines homologous only to His-46 and His-48, it has three other nearby histidines, any one of which might serve as a ligand equivalent to His-63. These two histidines are also in homologous positions in human and bovine cytochrome oxidase polypeptide II, but they are absent in the 50-kDal fragment (Fig. 1).

These results together with those described above for the binding site for type 1 copper suggest that the 19-kDal fragment is related in structure and function to other copper oxidases and electron-transfer proteins and that these proteins share a common evolutionary origin and were derived at least in part from

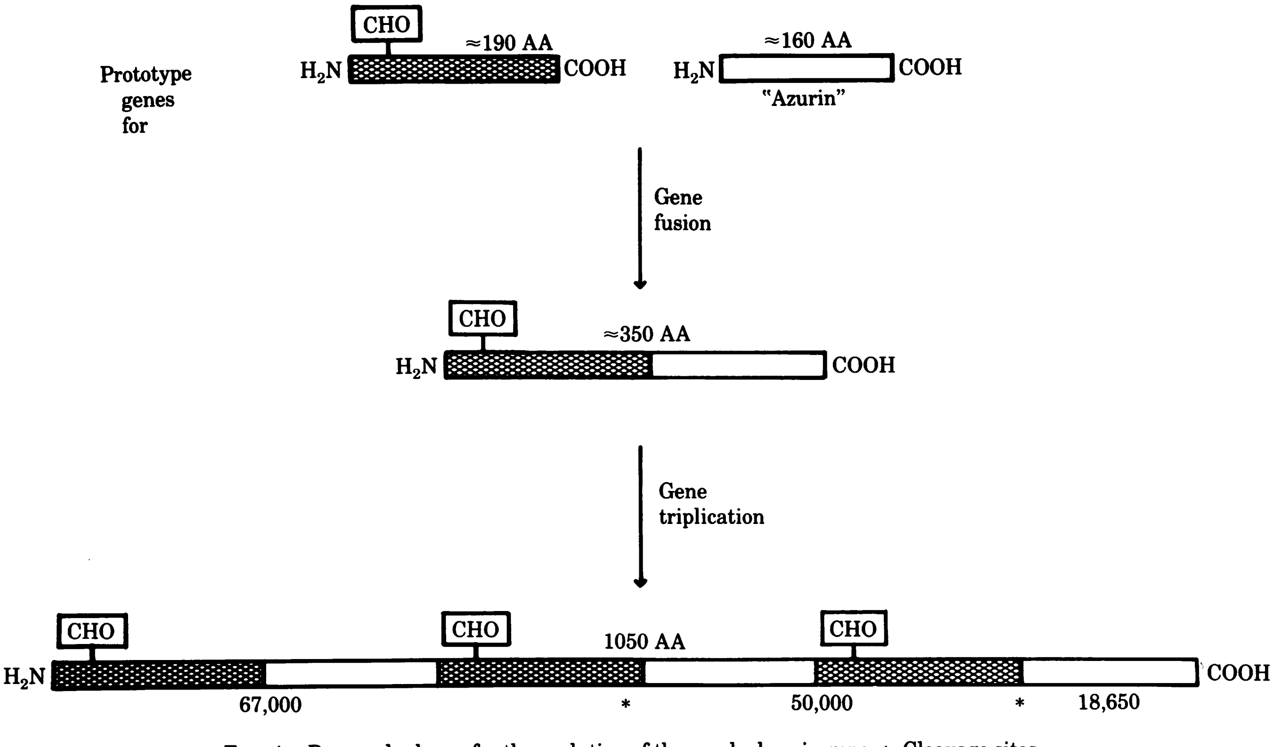


FIG. 4. Proposed scheme for the evolution of the ceruloplasmin gene. *, Cleavage sites.

the same ancestral gene. This is the gene referred to as azurin in Fig. 4.

Homology to the 67-kDal Fragment. Extensive but incomplete sequence data we have obtained for the amino-terminal 67-kDal fragment indicate that it has strong homology to the 564 residues in the carboxy-terminal half of ceruloplasmin. For example, the sequence of the first seven residues of the whole ceruloplasmin molecule is Lys-Glu-Lys-His-Tyr-Tyr-Ile- (11); this is homologous to residues 230–236 of the 50-kDal fragment and thus in Fig. 1 it fits best just after the end of the 19-kDal fragment. However, because Lys-Glu-Lys-His-Tyr-Tyr-Ile- is actually the amino terminus of the whole molecule, we propose that ceruloplasmin initiates with a segment of about 23 kDal that is related to an ancestral gene corresponding to the unmatched section of the 50-kDal fragment in Fig. 1.

Evolution of the Ceruloplasmin Gene. Fig. 4 presents a model for the evolution of the gene for ceruloplasmin that is based both on the internal homology of this protein (Fig. 1) and on its homology to copper-binding sites in other proteins (Figs. 2 and 3). In this model a primordial gene coding for an azurin-type protein of about 160 amino acids was fused to a gene coding for a protein of unidentified function and corresponding approximately to the 180 residues of sequence from positions 225 to 405 in the 50-kDal fragment. Tandem triplication of the ancestral fused gene could give rise to the present-day gene for ceruloplasmin. This model predicts that the complete sequence of human ceruloplasmin will exhibit a three-fold repeat pattern of two alternating structures, one of which is homologous to azurin and other blue proteins. If two copper ions are present in each azurin or blue oxidase subunit, the full complement of copper in ceruloplasmin would be provided for. This model also provides a mechanism for generating high molecular weight, multifunctional proteins from smaller preexisting molecules. This process probably occurred when vertebrates developed a closed vascular system and a urogenital system, because plasma proteins must have a molecular weight of about 60,000 to avoid renal excretion.

We thank P. H. Davidson, S. A. Dorwin, J. A. Dwulet, R. L. Fox, J. Madison, G. von Furstenburg, and A. Galen for their excellent technical assistance and Dr. Winona C. Barker for conducting the computer searches. This work was supported by National Institutes of Health Grant AM 19221.

1. Wu, T. T., Fitch, W. W. & Margoliash, E. (1974) *Annu. Rev. Biochem.* **43**, 539–566.
2. Dayhoff, M. O., ed. (1978) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 3, pp. 359–362.
3. Dayhoff, M. O., McLaughlin, P. J., Barker, W. C. & Hunt, L. T. (1975) *Naturwissenschaften* **62**, 154–161.
4. Bogardt, R. A. (1977) Dissertation (Indiana Univ., Bloomington, IN).
5. Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. (1978) *J. Mol. Evol.* **10**, 265–281.
6. Dwulet, F. E. & Putnam, F. W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 790–794.
7. Kingston, I. B., Kingston, B. L. & Putnam, F. W. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1668–1672.
8. Kingston, I. B., Kingston, B. L. & Putnam, F. W. (1980) *J. Biol. Chem.* **255**, 2878–2885.
9. Kingston, I. B., Kingston, B. L. & Putnam, F. W. (1980) *J. Biol. Chem.* **255**, 2886–2896.
10. Dwulet, F. E. & Putnam, F. W. (1980) *Protides Biol. Fluids Proc. Colloq.* **28**, 83–86.
11. Noyer, M., Dwulet, F. E., Hao, Y. L. & Putnam, F. W. (1980) *Anal. Biochem.* **102**, 450–458.
12. Egorov, T. A., Svenson, A., Rydén, L. & Carlsson, J. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3029–3033.
13. Rydén, L. & Lundgren, J.-O. (1979) *Biochimie* **61**, 781–790.
14. Kingston, I. B., Kingston, B. L. & Putnam, F. W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5377–5381.
15. Frieden, E. (1980) *Excerpta Med.*, Ciba Foundation Symposium 79, 93–124.
16. Malmström, B. G. (1978) in *New Trends in Bio-Inorganic Chemistry*, ed. Williams, R. J. P. & Da Silva, J. R. R. F. (Academic, New York), pp. 59–77.
17. Reinhammar, B., Malkin, R., Jensen, P., Karlsson, B., Andréasson, L.-E., Aasa, R., Vänngård, T. & Malmström, B. G. (1980) *J. Biol. Chem.* **255**, 5000–5003.
18. Adman, E. T., Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1980) *J. Mol. Biol.* **123**, 35–47.
19. Colman, P. M., Freeman, H. C., Guss, J. M., Murata, M., Norris, V. A., Ramshaw, J. A. M. & Venkatappa, M. P. (1978) *Nature (London)* **272**, 319–324.
20. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *J. Biol. Chem.* **243**, 3557–3559.
21. Bergman, C., Gandvik, E.-K., Nyman, P. O. & Strid, L. (1977) *Biochem. Biophys. Res. Commun.* **77**, 1052–1059.
22. Briving, C., Gandvik, E. K. & Nyman, P. O. (1980) *Biochem. Biophys. Res. Commun.* **93**, 454–461.
23. Richardson, J. S., Thomas, K. A., Rubin, B. H. & Richardson, D.C. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1349–1353.
24. Steinman, H. M., Naik, V. R., Abernethy, J. L. & Hill, R. L. (1974) *J. Biol. Chem.* **249**, 7326–7338.
25. Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) *Nature (London)* **282**, 189–194.
26. Steffens, G. J. & Buse, G. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 613–619.
27. Jabusch, J. R., Farb, D. L., Kerschensteiner, D. A. & Deutsch, H. F. (1980) *Biochemistry* **19**, 2310–2316.
28. Steinman, H. (1980) *J. Biol. Chem.* **255**, 6758–6765.