Protein Science (1997), 6:761–770. Cambridge University Press. Printed in the USA.
Copyright © 1997 The Protein Society

Structural comparison of cupredoxin domains: Domain recycling to construct proteins with novel functions

MICHAEL E.P. MURPHY,[1] PETER F. LINDLEY,[2] and ELINOR T. ADMAN
Department of Biological Structure, University of Washington, Box 357420, Seattle, Washington 98195-7420

(RECEIVED October 24, 1996; ACCEPTED January 28, 1997)

Abstract

The three-dimensional structures of the copper-containing enzymes ascorbate oxidase, ceruloplasmin, and nitrite reductase, comprised of multiple domains with a cupredoxin fold, are consistent with having evolved from a common ancestor. The presence or absence of copper sites has complicated ascertaining the structural and evolutionary relationship among these and related proteins. Simultaneous structural superposition of the enzyme domains and their known cupredoxin relatives shows clearly that there are at least six cupredoxin classes, and that the evolution of the conserved core of these domains is independent of the presence or absence of copper sites. Relationships among the variable loops in these structures show that the two-domain ancestor of the blue oxidases contained a trinuclear–copper interface but could not have functioned in a monomeric state. Comparison of the sequence of the copper-containing, iron-regulating protein. Ferrous transport (Fet3) from yeast to the structurally defined core and loop residues of the cupredoxins suggests specific residues that could be involved in the ferroxidase activity of Fet3.

Keywords: cupredoxin, domain phylogeny, structural comparison

Nature has often recycled protein folds, constructing a family of related proteins with different but related functions. Some well-known examples are the TIM barrel in glycolysis (Farber & Petsko, 1990) and the immunoglobulin fold in molecular recognition (Bork et al., 1994), each of which has resurfaced, often unexpectedly, in many structures. The cupredoxin fold is another example. Sequence alignments combined with structural and spectroscopic evidence for Type I blue copper sites[3] led to the conclusion that not only the cupredoxins but also blue oxidases probably evolved from a common ancestor (Ryden & Hunt, 1993). Indeed, the structures of two blue oxidases, ascorbate oxidase (AO) from plants, ceruloplasmin (CP) from mammalian plasma, and nitrite reductase (NIR), a bacterial denitrification enzyme (Godden et al., 1991), are built exclusively from cupredoxin-like domains. Only some of these domains bind Type I copper, and some are involved in binding non-blue mononuclear and trinuclear copper sites.

The cupredoxin fold, an eight-stranded Greek key β-barrel, was first observed in the small blue copper proteins, plastocyanin and azurin (Colman et al., 1978; Adman et al., 1978), and was subsequently observed in the related proteins pseudoazurin, amicyanin, and cucumber basic protein. These proteins show minimal sequence identity but are formed from a single domain bearing a Type I copper site and function as electron transfer agents (reviewed in Adman, 1991). The cupredoxin fold is distinct from other Greek key β-barrels such as superoxide dismutase, the immunoglobulins, and fibronectin domains. The latter contain only anti-parallel β-strands, whereas the first and third strands of cupredoxin domains form parallel connections to their respective β-sheets. There is always a discontinuity ("dog-leg") as strand two crosses to strand three from one sheet to the other.

The cupredoxin fold has been recruited as a subunit in two respiratory proteins. Subunit II of cytochrome oxidase has a cupredoxin fold but with a binuclear copper (CuA) site that transfers electrons from cytochrome c to the oxidase active site (Iwata et al., 1995; Tsukihara et al., 1996). The copperless cyoA domain of

Reprint requests to: Elinor T. Adman, Department of Biological Structure Box 357420, University of Washington, Seattle, WA 98195-7420 USA; e-mail: adman@u.washington.edu.
[1]Present address: Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada.
[2]CCLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England.
[3]Type(s) I, II, and III Cu refer to spectroscopically characterized copper centers in proteins. Type I Cu has very intense absorption at about 600 nm and narrow hyperfine coupling in EPR spectroscopy. The copper is generally coordinated by three strong ligands, Cys and two His, and one or two weaker ligands (Met and, in azurin, a carbonyl oxygen). Type II copper has much weaker absorption, broader hyperfine interactions, and is generally 4–5 coordinate with histidine and water–oxygen ligands. Type III is a pair of coppers antiferromagnetically coupled, usually coordinated by three histidines per copper and a bridging moiety (Adman, 1991). The recently characterized CuA site in cytochrome oxidase is also a pair of Cu atoms, bridged by two cysteines, and each also ligated by two histidines, and either a main chain carbonyl oxygen or a methionine. The trinuclear site in the blue oxidases combines a pair of Type III copper atoms and a Type II copper.

Escherichia coli quinol oxidase is also a cupredoxin fold (Wilmanns et al., 1995). This domain has been engineered to bind a binuclear copper (CuA) site by the introduction of six copper ligands (van der Oost et al., 1992). Both cytochrome oxidase and quinol oxidase appear to have recruited a single cupredoxin subunit as part of a large multi-subunit enzyme.

The structure of AO is constructed exclusively from three different cupredoxin domains, only one of which binds a Type I copper (Messerschmidt et al., 1989). A second, trinuclear, copper site is formed at the interface between domains 1 and 3. Based on sequence identity with AO, the structure of CP was predicted to have three repeats of two cupredoxin domains (Messerschmidt & Huber, 1990). Both Putnam (Dwulet & Putnam, 1981) and L. Ryden (Ryden, 1984a) predicted that CP has three repeats of two linked domains. When the structure and chemical sequence of NIR revealed that it consists of a trimer of two cupredoxin domains with homology to both AO and CP, it was suggested that the six domains of CP were arranged like the six domains of NIR (Fenderson et al., 1991). The crystal structure of CP (Zaitseva et al., 1996) has shown that the domain organization of these two distantly related proteins is indeed similar. However, the Type I binding sites are not in equivalent domains with respect to the pseudosymmetry elements (Fig. 1). The domain structure of AO may be thought of as three of the six domains of CP with the Type I copper sites in equivalent domains (Zaitseva et al., 1996).

Since these structures are so clearly related, we compare the domains in detail to define a common core and to assign functions to variable loops. This comparison allows us to comment on the domain evolution of this family and on a proposed model (Pan et al., 1995) of the blood coagulation protein, Factor VIII, and to suggest particular residues of the yeast ferrous transport (Fet3) protein (Askwith et al., 1994) likely to be involved in ferroxidase activity.

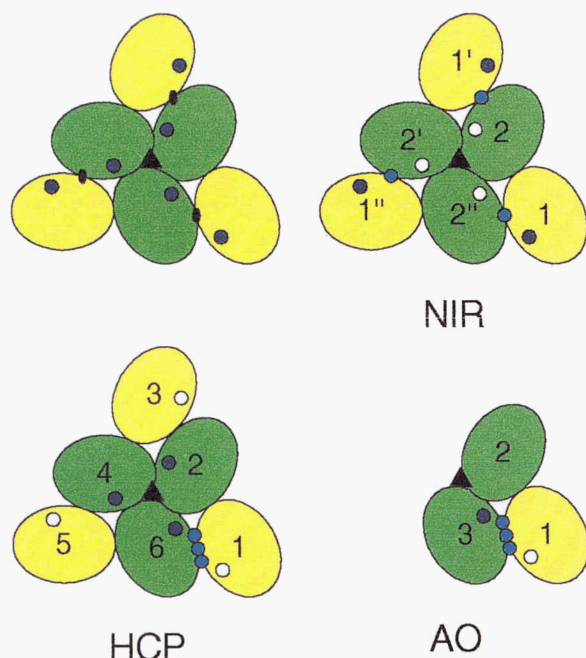


Fig. 1. Schematic diagrams of the domain organization and copper sites of ascorbate oxidase, ceruloplasmin, and nitrite reductase. The upper left diagram shows the location of potential Type I copper sites and of the (pseudo) three-fold (▲) and pseudo two-fold symmetry axes (♦) as if a six-domain structure were to be generated by symmetry from a single cupredoxin domain. For each of the structures diagrammed, the domains are numbered, the Type I "blue" copper sites are indicated by purple circles or white circles if empty, and Type II and Type III copper sites are drawn as blue circles. Unprimed numbers indicate covalently linked domains. The yellow domains correspond to the Class IV domains, the green to class V (see text and Fig. 3).

Results

Core definition and classification of the cupredoxin family

A total of 23 domains from 15 crystal structures were obtained from the Protein Data Bank (PDB) (Bernstein et al., 1977) or were provided by the authors (Table 1). Domains with a sequence identity greater than 50% to an existing domain were not included. The domains vary greatly in copper content and function. Of the 23 domains, 14 domains contain a Type I copper site, two contain dinuclear CuA sites, six provide ligands for inter-domain copper sites, and three are not involved in copper binding. A simultaneous superposition (see Methods) has defined a conserved core of 65 residues among these domains. The overall sequence identity and root-mean-square difference (RMSD) between all pairs is 18.5% and 2.05 Å, respectively.

The residues of the conserved core are organized into eight structurally conserved regions (SCRs) that include all eight strands of the β-barrel (Fig. 2B). The core also includes residues forming loops and sections of polypeptide chain without conserved main-chain H-bonds. The conserved core contains three of the four residues that can potentially form ligands to a bound Type I copper. There are no positions with absolutely conserved residues but there are nine residues conserved in more than half the domains. None of these conserved residues are found in the first three SCRs.

A total of 30 conserved (observed in more than half of the domains) main-chain H-bonds were found between residues of the conserved core. The two sheets that make up the β-barrel are linked by a single conserved H-bond between the side-chain car-

Table 1. *Cupredoxin domains superimposed*

Domains	Code	Source[a]
A. Single domain e- transfer proteins		
Poplar plastocyanin	pcy	PDB 1plc
Chlamydomonas reinhardtii plastocyanin	plt	PDB 2plt
Enteromorpha prolifera plastocyanin	ple	PDB 7pcy
Alcaligenes faecalis pseudoazurin	pza	PDB 1pza
Methylobacterium extorquens pseudoazurin	pmy	PDB 1pmy
Paracoccus denitrificans amicyanin	aan	PDB 1aan
Alcaligenes denitrificans azurin	aza	PDB 2aza
Pseudomonas aeuroginosa azurin	azu	PDB 4azu
Thiobacillus ferroxidans rusticyanin	ruc	R. Walter
B. Multi-domain enzymes[b]		
Alcaligenes faecalis nitrite reductase (2)	afn	PDB 2afn
Zucchini ascorbate oxidase (3)	aoz	PDB 1aoz
Human ceruloplasmin (6)	hcp	P. Lindley
Escherichia coli quinol oxidase (1)	cyw	PDB 1cyw
Bovine cytochrome oxidase (1)	bco	T. Tsukihara
Paracoccus denitrificans cytochrome oxidase (1)	pco	S. Iwata

[a]Coordinates obtained from the Protein Data Bank (PDB) are identified by their access code, otherwise the author providing coordinates is listed.
[b]Number of cupredoxin domains indicated in parentheses.

boxylate of residue 41 (Fig. 2) to the main-chain amide nitrogen of residue 38. This interaction stabilizes a tight turn between these residues and leaves the remaining carboxylate oxygen available for hydrogen bonding to the extended N-terminal loop in NIR, or the N-terminus itself in ceruloplasmin. Interestingly, the interactions formed with this residue perform the same function as a disulfide bond unique to azurins. Azurin, aoz2, cyoA, afn1, and hcp2 are the only cupredoxin domains with something other than glycine at position 40; azurin is the only one with cysteine in that position.

A structural similarity tree (Table 2, Fig. 3) derived from the RMSD between all pairs in the simultaneous superposition divides the cupredoxin domains into five classes. The first class consists of the single-domain electron transfer proteins pseudoazurin, plastocyanin, and amicyanin. The CuA domains of bovine and bacterial cytochrome oxidase and the related copperless domain of quinol oxidase make up the second class. The two available azurin structures form a third class distinct from the other single-domain electron transfer proteins in the first class. Class IV is made up of the first domains of NIR and AO and domains 1, 3, and 5 of CP. The last class contains the remaining domains of CP, domain 2 of NIR, domains 2 and 3 of AO, and, intriguingly, rusticyanin. A smaller common core can be defined without significantly altering the similarity tree if two additional distantly related cupredoxin domains, stellacyanin and cucumber blue protein, are included in the superposition. These two cupredoxins have been termed phytocyanins (Ryden 1984b) and represent a sixth cupredoxin domain class.

The segregation of the cupredoxin domains of NIR, AO, and CP into two different classes (IV and V) is surprising in that domain 1 of NIR, which contains a type I copper site, is clustered with four Type I copper absent domains of AO and CP, whereas domain 2 of NIR, which does not bind Type I copper, is clustered with four Type I copper containing domains of AO and CP. Thus the structural similarity of the cores of these domains does not correlate with copper content and therefore with electron transfer function. This division of these domains into two classes is consistent with the domain organization shown in Figure 1.

To further explore the relationship among the 11 domains of the enzymes AO, CP, and NIR, we superimposed them separately resulting in an alignment with 78 α-carbons (Fig. 2). An additional 11 conserved main chain H-bonds were also found, including two between a loop at the top of the barrel and position 56, a potential Type I copper ligand. A separate superposition of the five domains of Class IV and six domains of Class V using the same cutoff distance results in an additional 22 core residues in class IV (yellow) and 19 core residues in class V (green). The larger number of aligned residues supports the class identification in the original structural similarity tree.

Comparison of structurally unconserved segments

The eight conserved segments that make up the core structure are connected by seven structurally unconserved regions (SURs) of polypeptide chain (Figs. 2, 4). In the enzyme domains, half of these regions are twelve residues or less in length: SURs 3, 4, 6, and 7. The shortest region, SUR3, divides two conserved segments by a single residue insertion found only in azurin, where the loop is stabilized by a disulfide bridge. The absence of SUR3 in the enzyme domains results in the largest contiguous segment (17 positions) of conserved residues. SUR4 has been suggested to be part of a β-zipper that may nucleate folding of Greek key β-barrels (Hazes & Hol, 1992).

SUR6 is two to four residues in length, except in the class IV domains of CP, where it is 12 residues long. In these domains SUR6 forms a larger protrusion from the bottom of the β-barrel, and is packed against SUR5 of class V domains of CP, which is concomitantly shorter than in the other enzymes. Disulfides in CP not present in the other proteins stabilize these loops. SUR7 forms part of the Type I copper binding loop and is surprisingly unconserved, varying by eight residues in length, with no correlation with the presence or absence of copper. It is intriguing that SUR7 in aoz2 contains a histidine, which binds a non-functional copper external to where a Type I site might have been and might have resulted from a site that once was a Type I site.

Portions of the SURs that are most variable in length are each structurally conserved within the two subfamilies of enzyme domains (Figs. 2, 4). SUR1 contains β-strand extensions conserved only in class IV domains, and nine residues in SUR5 are conserved only in class V domains. SUR6 contains two structurally conserved residues in domain class V but is variable in class IV.

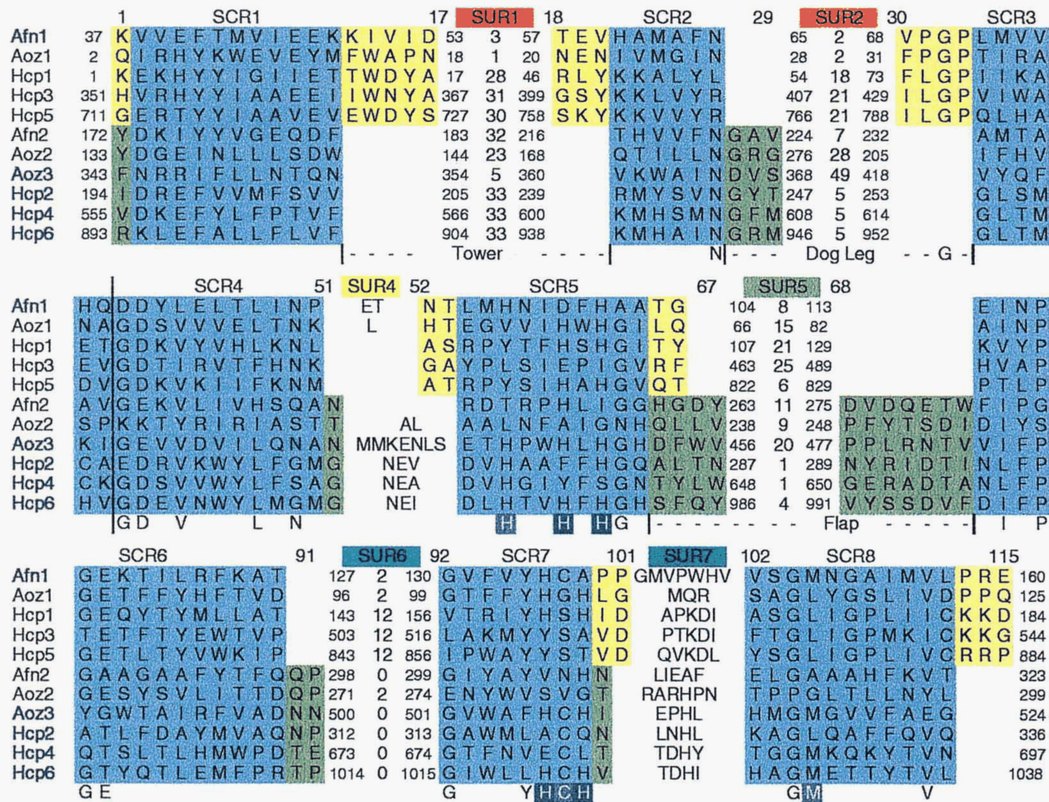
SUR1 (Figs. 2, 4), the tower, is quite long in most enzyme domains and provides domain–domain interactions, much like twist-ties between adjacent domains. SUR1 also contains residues important to defining access to the Type I sites in these enzymes. In domain 1 of NIR, this SUR contains a helix roughly parallel to the vector from the Type I copper to the active site copper. We have previously shown that negatively charged residues in this helix contribute to the interaction surface with the electron donor to this protein, pseudoazurin (Kukimoto et al., 1996). This same helix also forms the top of the active site pocket of NIR (pink in Fig. 4B). Similarly, Trp 163 in SUR1 in AO (Fig. 2) of domain 2 has been proposed to be involved in binding organic substrates near the Type I site found in domain 3 (Messerschmidt et al., 1992) and forms the top of the active site along with SUR1 of domain 3. In CP, SUR1 of domain 6 covers the Type I site of that domain and SUR1 of domain 5 occludes the trinuclear site. SUR4 also contributes to defining access to the Type I site in AO and CP, but from remote domains. For example, SUR4 of hcp2 restricts access to the Type I site of hcp6 and provides negatively charged residues likely to be involved in putative ferroxidase activity (Lindley et al., submitted).

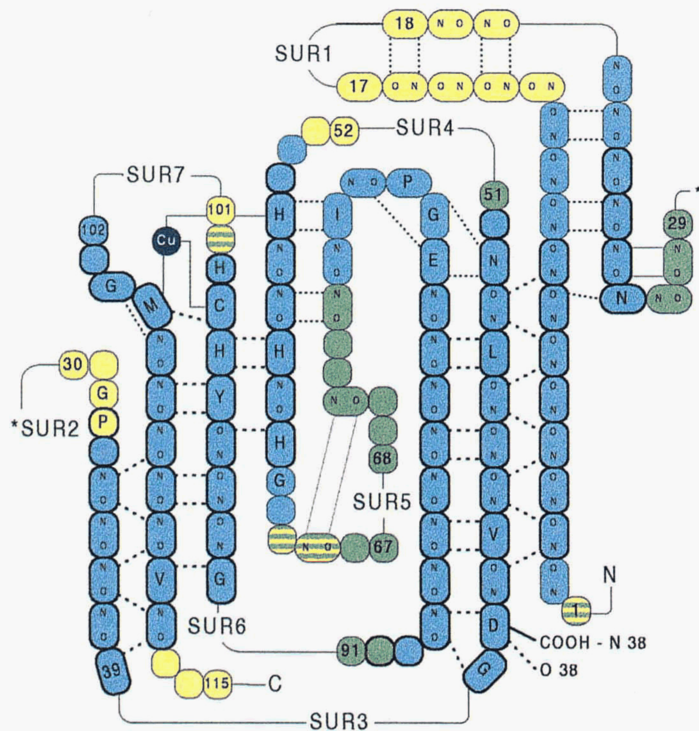
Only six structurally conserved residues of SCR2 separate SUR1 from SUR2 (Figs. 2, 4), which can be up to 49 residues in length. SUR2 is also known as the "dog-leg" because in many cupredoxins it is much shorter and connects the two short strands (SCR2 and SCR3), split across the barrel one to each sheet. In the enzyme domains, the non-regular structures of SUR2, which lie to the outside of the molecule, provide additional interactions between the odd–even domains and minimize surface roughness. The dog-leg is the longest in aoz3, partially occupying space that would have been filled by the dog-leg of a neighboring subunit as in the six-domain proteins. Interestingly, the space analogous to this loop of the odd-numbered domains of CP (see Fig. 4B: orange at the lower left of P) is filled by the N-terminal extension in NIR (Fig. 4B, gray at the lower left of NIR), which has shorter dog-legs.

SUR5 can also fill gaps between domains and forms the "flap" in azurin. The longer SUR5 in AO and CP (odd domains, Class IV) blocks access to the active site channel to the trinuclear site relative to NIR, which has a much smaller flap in domain 1. SUR5 from even domains (Class V) forms part of the back side of the active site pocket and also provides the interactions near the three-fold axis. Because of this constraint (and the above mentioned interaction with SUR6),

SUR5s are generally shorter and more similar to each other in the even- than in the odd-numbered domains.

The domain linker in NIR is 16 residues long and provides stabilizing interactions in addition to filling the cleft between the two domains. The first linker in AO is 14 residues in length, following a path similar to NIR. The linker between aoz2 and aoz3 is composed of three sections, two proline-rich regions flanking a 16-residue long α-helix. The linkers are also 13 to 15 residues long in CP; however, the center portion is directed toward the trimer axis rather than directed outward.





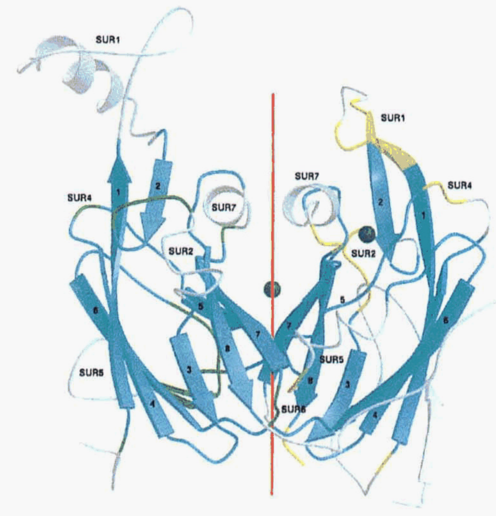


Table 2. *Structure and sequence comparison table for domains of AO, CP, and NIR*[a]

	afn1	aoz1	hcp1	hcp3	hcp5	afn2	aoz3	hcp2	hcp4	hcp6	aoz2
afn1	—	26.2	20.0	13.8	16.9	12.3	26.2	23.1	24.6	29.2	7.7
aoz1	0.69	—	35.4	32.3	30.8	20.0	26.2	23.1	23.1	23.1	6.2
hcp1	0.98	0.87	—	40.0	47.7	18.5	18.5	12.3	15.4	16.9	7.7
hcp3	1.03	0.91	0.46	—	53.8	16.9	9.2	12.3	10.8	10.8	6.2
hcp5	0.97	0.99	0.54	0.64	—	21.5	15.4	18.5	16.9	20.0	4.6
afn2	0.89	1.06	1.18	1.31	1.20	—	21.5	18.5	15.4	18.5	6.2
aoz3	1.10	1.15	1.38	1.53	1.45	0.86	—	21.5	18.5	29.2	7.7
hcp2	1.01	1.15	1.33	1.41	1.38	1.04	1.14	—	35.4	40.0	3.1
hcp4	0.97	1.13	1.33	1.42	1.38	0.90	1.08	0.48	—	43.1	12.3
hcp6	0.95	1.12	1.32	1.41	1.39	0.94	1.06	0.56	0.41	—	9.2
aoz2	1.07	1.19	1.32	1.44	1.34	1.16	1.29	1.14	1.08	1.06	—

[a]The table refers to the superposition of 65 α-carbons and is a subset of the complete matrix provided in the electronic appendix. The lower left of the table lists the RMSD in Ångstroms and the upper right is the percent identity of the sequences of the superimposed residues. The domain codes are described in Table 1. The complete table for the 23 domains is included in the Electronic Appendix.

Active sites

The active sites of NIR, AO, and CP are formed at the interface between two domains, one class IV and one class V (Fig. 2). The trinuclear copper sites of AO and CP are liganded by eight histidines, four from each domain related by a pseudo-two-fold axis (Messerschmidt et al., 1989; Zaitseva et al., 1996). Interestingly, these four histidine ligands are found within the conserved core (Fig. 2; positions 59, 61, 97, and 99, corresponding to NIR residues D98, H100, H255, I257, and H135, A137, V304, and H306). In NIR, only four of the eight histidines are present; three are ligands forming a mononuclear copper site and the fourth is located nearby but is not liganded to the copper. Each domain provides two of the four histidines but none is related by the pseudo–two-fold axis. Amino acid replacements at two pseudo-symmetry related positions, Asp 98 (domain 1) and Ile 257 (domain 2), and the fourth histidine, His 255 (domain 2), are likely involved in the mechanism of nitrite reduction (Murphy et al., 1995; Adman et al., 1995). The remaining two substitutions, Ala 137 (domain 1) and Val 304 (domain 2), form part of the active-site cavity but do not appear to play a direct role in catalysis.

The six-domain structures of NIR and CP each possess three pseudo–two-fold symmetry axes relating the conserved cores of the domains forming the active sites. For both proteins, rotation is within 4° of 180° and the translational component is less than 0.6 Å. The α-carbons of the eight ligands of the trinuclear site of AO are also related by two-fold symmetry; however, the conserved cores are related by a 165° rotation resulting in the strands bearing the histidine ligands being distorted relative to the rest of the β-barrel as compared to NIR and CP. This distortion in AO is likely due to formation of a compact structure using only three domains instead of six.

The presence of a local two-fold axis nearly parallel to a three-fold axis imposes an interesting constraint on the relative orientation of the two domains in the monomer of NIR. The barrel axes of each domain must be 60° apart if the two-fold is parallel to the three-fold as can be seen in Fig. 1. As the two-fold is no longer parallel to the three-fold, the actual angle increases. The actual angle in NIR and CP is 72° and in AO it is 64°. This, in turn, implies that not only does there need to be gene duplication of a cupredoxin ancestor, but, to form a functional enzyme, there needs to be a linker that will allow the two domains to be oriented correctly.

Fig. 2. (*facing page*) **A:** Derived sequence alignment from the superposition of 11 enzyme domains (cyan). The additional superimposed residues in the alignment of the two subfamilies, IV and V, are shaded yellow and green, respectively. Residues found in more than half the sequences are listed below as well as Type I ligands (blue boxes) and Type II/III ligands (dark green boxes). Structure codes are as in Figure 3. The size of insertions between conserved blocks is indicated in large font and the residues numbers at the block edges are added in small font. The colors of the SUR labels correspond to loop colors in Figures 4A, B. The complete alignment of 23 domains is given in the electronic appendix. **B:** A diagram of the conserved core of the cupredoxin domains using the same color scheme as in **A**. A bold outline indicates the subcore conserved in all 23 cupredoxin domains listed in Figure 3. H-bonds are included if present in more than half the structures included in the superposition. The seven structurally unconserved regions (SUR) are labeled. Note SUR3 is indicated by a vertical bar being of zero length in the enzyme domains. The beginning and end of each conserved segment is numbered by residue position as in **A**. **C:** Domains 1 and 2″ of NIR related by a pseudo two-fold axis (red) and color-coded as in **A**. The conserved segments are numbered, the SURs are labeled, and the two copper atoms are drawn as green spheres. Note that the SCRs are not exclusively secondary structure elements: SCR3/4 contains a turn, as does SCR6. Figures 2C, 4, 5 were produced using MOLSCRIPT (Kraulis, 1991) as extended by Robert Esnouf and Raster3D (Merritt & Murphy, 1994).

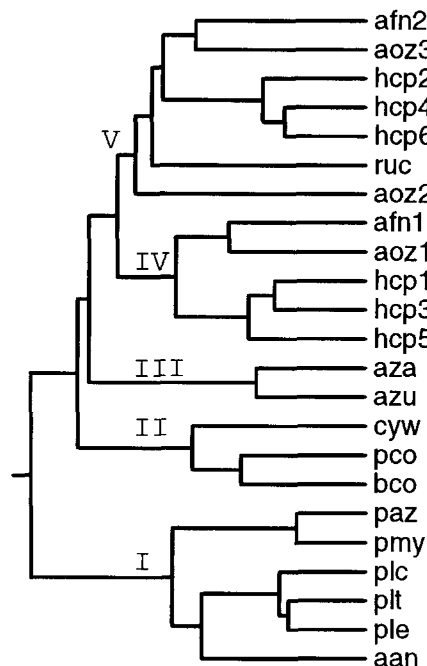


Fig. 3. Structural similarity tree derived from a superposition of 23 cupredoxin domains. The domain codes are as listed in Table 1 followed by a domain number in structures with multiple cupredoxin domains. The five cupredoxin classes defined by this method are indicated by roman numerals at the nodes of the tree containing the domains of the indicated class.

Discussion

Cupredoxin phylogeny

Trees constructed from 3D structure alignments have been compared to trees derived from sequence alignments of cytochromes *c*, globins, and the aspartic proteinases (Johnson et al., 1990). In each case the phylogenetic trees based on sequence are consistent with those constructed from pairwise rigid body superposition of the 3D structures.

Phylogenetic trees have been derived from sequence alignments of CP, AO, azurin, plastocyanin, pseudoazurin, and amicyanin, as well as from domains for which 3D structures are not available: factor V, factor VIII, laccase, and ragweed allergen (Ryden & Hunt, 1993). In general, these sequence-based phylogenetic trees are consistent with the structural similarity tree in Figure 3. In both trees, pseudoazurin, plastocyanin, and amicyanin are clustered together distinctly from azurin. The trees also show that, of these cupredoxins, azurin is most similar to the blue oxidase domains (rusticyanin was not considered). The sequence alignment used to construct the phylogenetic tree of cupredoxins with domain 3 of AO as the outgroup differs in many regards to the alignment we infer from the structural alignment. In particular, of the 84 sequence positions used to construct the sequence-based phylogenetic tree, only 24 positions in three regions that contain a conserved copper ligand are identical with the structural alignment of 65 residues for the five domains in common. Most of the differences are a result of a two-residue shift that retains the matching of buried hydrophobic residues. The goals of sequence- and structure-based alignments are different and the resulting alignments need not agree. For example, it is possible that a segment of sequence has threaded differently onto a similar polypeptide backbone due to mutation. The resulting sequence alignment would then be displaced from the corresponding structural alignment. Despite these differences, both methods yield similar trees.

Using a sequence-based phylogenetic tree, Ryden proposed a model for the evolution of the blue oxidase family (Ryden & Hunt, 1993). In this model, the class V domain evolved first, followed by gene duplication to produce the class IV domain. After the formation of the double domain, the trinuclear copper site was formed by the addition of ligands to both domains, yielding a functional protein. This two-domain structure gave rise to the six-domain ceruloplasmin structure through two duplications of the double domain followed by the loss of the blue copper sites in the class IV domains. For AO, a single domain duplication from the double-domain protein created the second domain that has since diverged from the other two.

We believe that the protein could not have functioned as a monomer of two linked domains. From inspection of the available structures, it seems unlikely that the ancestral two-domain structure existed as a monomer with the active site at the interface of the two domains because the relative orientation necessary for function is maintained by either the whole six-domain structure, or the second domain of AO plus the intervening linker that effectively replaces the other three domains of CP and NIR.

A model alternative to that of Ryden and Hunt for the evolution of AO suggests that three domains from an ancestral six-domain structure are replaced with the short linker described above rather than with the addition of a single domain to a two-domain monomer. Consistent with this, the structural core alignment places both aoz2 and aoz3 in class V. SUR1 and SUR5 of aoz2 appear to be more similar to SURs of this class than to those of class IV, suggesting that these similarities evolved from a common ancestor. This is also likely to be the case for the fungal laccases (Messerschmidt & Huber, 1990), and possibly the related phenoxazine synthase (Freeman et al., 1993) and bilirubin oxidase (Koikeda et al., 1993), all also likely to be three-domain proteins.

Our model and Ryden and Hunt's both require the loss of Type I copper sites. However, the copper site is lost in the class V domain of NIR instead of the class IV domain as in the blue oxidase family. Retention of a Type I copper site in the outer domain of NIR is needed for access to it by a large substrate, its electron transfer partner, pseudoazurin. The various SURs of CP and AO restrict access to and define binding sites for smaller substrates, such as ascorbate, iron (as described above), or the various substrates of the laccase-like family, by retaining the less accessible interior copper.

It seems most likely that oxidase function evolved from an NIR-like trimer, adding four additional mutations to generate the eight histidines required for oxidase function as AO and CP evolved. After CP evolved, two of the three oxidase sites were lost. Vestigial histidine residues 816 and 818 at the hcp4–hcp5 interface of CP are consistent with this as is His 280 at the hcp2–hcp3 interface.

Inspection of the structure-based tree suggests that rusticyanin is a close relative of the class V domains of the blue oxidases and nitrite reductase and may resemble the common ancestor of the class IV and V domains of NIR and the blue oxidases. It is intriguing that both rusticyanin and ceruloplasmin have been ascribed ferroxidase activity. The structural tree also suggests that subunit II of cytochrome oxidase and the quinol oxidase domain are more closely related to the blue oxidase–NIR family than to the rest of the cupredoxin family.

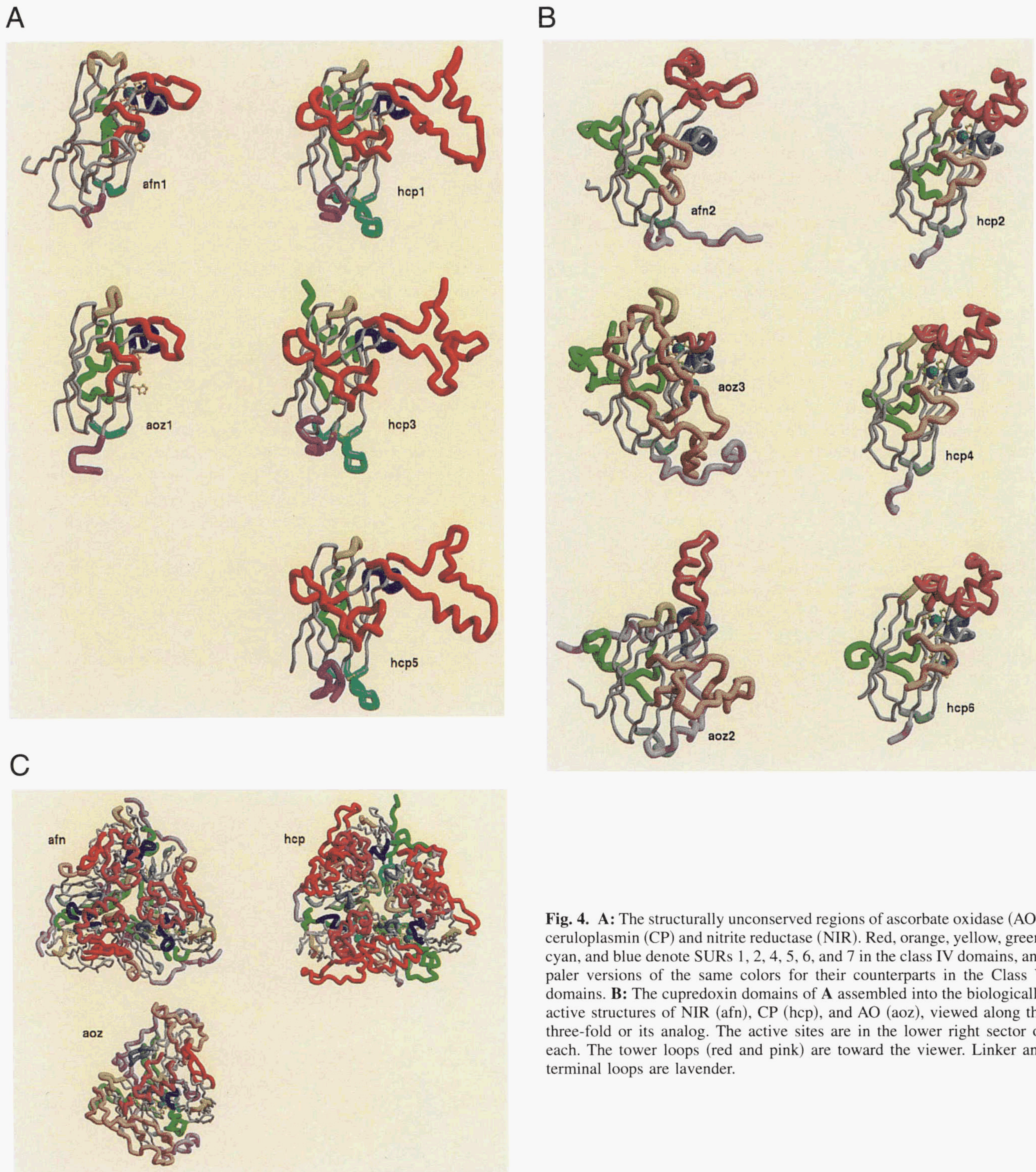


Fig. 4. A: The structurally unconserved regions of ascorbate oxidase (AO), ceruloplasmin (CP) and nitrite reductase (NIR). Red, orange, yellow, green, cyan, and blue denote SURs 1, 2, 4, 5, 6, and 7 in the class IV domains, and paler versions of the same colors for their counterparts in the Class V domains. **B:** The cupredoxin domains of **A** assembled into the biologically active structures of NIR (afn), CP (hcp), and AO (aoz), viewed along the three-fold or its analog. The active sites are in the lower right sector of each. The tower loops (red and pink) are toward the viewer. Linker and terminal loops are lavender.

Implications for coagulation factor VIII and yeast Fet3

A six-domain model for the blood coagulation protein factor VIII was based on the structural similarity of AO and NIR, and the sequence similarity of CP and factor VIII (Pan et al., 1995). The sequence alignments proposed in that work are largely correct, although forcing the Type I copper ligands to align with the Type I copper in the outer domains of NIR resulted in an incorrect domain alignment. The alignment is much more likely to be according to the published ceruloplasmin structure, not available at the time of the prediction. Interestingly, the now re-defined inter-domain interactions for factor VIII suggest vestigial histidine residues at three interfaces: at the domain 2–3 interface, there is FF**H**···FC**H**/YP**H**···SSF; at the domain 4–5 interface, FFS···GC**H**/

YSS···FSD, and at the domain 6–1 interface, **HFS**···ECL/**HAV**···**LSH**, giving rise to two interfaces with enough histidines to form type II sites and supporting the idea that trinuclear copper sites have been lost at these interfaces. It has been suggested that copper stabilizes active factor VIII (Bihoreau et al., 1994). Inasmuch as NIR is stable even in the absence of Type II copper (Adman et al., 1995) at these interfaces, it is difficult to predict the role of copper in factor VIII stability or function. Two of the even-numbered domains also provide Type I copper ligands; the third provides two of four required ligands (L···**C**···F···**M**).

The Fet3 gene from yeast has been shown to be required for iron uptake and to code for a protein with sequence similarity to the three-domain blue oxidases (Askwith et al., 1994), apparent even from the partial sequence reported at first. The protein was later shown to be a membrane-bound cell surface ferroxidase (DeSilva et al., 1995), providing a link for copper-regulated iron transport. Visual alignment (Fig. 5) of the Fet3 sequence with our structurally based core reveals several features that lend support to the proposed alignment. All 14 putative glycosylation sites are located on loops or external residues of SCRs. Residues believed to be important in ferroxidase activity in CP are in SUR1 of domain 6 (E932, E935) and SUR4 of domain 2 (E272, D230). By analogy one would expect to find negatively charged residues in the corresponding regions of Fet3, particularly in domain 3 SUR1 and domain 2 of SUR4. Indeed, several candidate residues can be found within these regions: E329, E334, and D336 of domain 3, SUR1; E227, D228, and E230 of domain 2, SUR4. Mutation of these residues would test their role in ferroxidase activity. Also consistent with putative ferroxidase activity, the "tower loop" of domain 2 is considerably shortened relative to that in AO, leaving SUR4 of domain 2 more surface accessible as it is in CP (see Fig. 4B). SUR4 of domain 3 and SUR5 are both considerably longer than their counterparts in AO. These portions of the structure would be adjacent to each other in our putative model of Fet3 and possibly interact and stabilize each other.

Methods

An initial list of equivalent residues was constructed from a limited number of manual pairwise superpositions or using the pairwise distance matrix method in WHATIF (Vriend, 1990). Simultaneous superposition of α-carbons was accomplished using the program POLYPOSE (Diamond, 1992). Additional equivalent residues were identified using the program CLUSTERID (detailed description to be published elsewhere). CLUSTERID searches an initial structural alignment for additional clusters of a-carbons, one from each structure, with a root-mean-square (RMS) from the average less than a specified initial cutoff. The program accepts these new clusters only if they are part of a contiguous segment of at least four clusters. The enlarged structural alignment is used to obtain an improved superposition and the search is repeated using a larger cutoff value until a specified maximum cutoff is reached and no new clusters are found.

The alignment resulting from this procedure was found to be sensitive to the starting conditions and the cutoff parameters resulting in several different alignments with similar overall RMS distances. Similar results have been recently reported when comparing alignment methods (Godzik, 1996). In this case, a unique

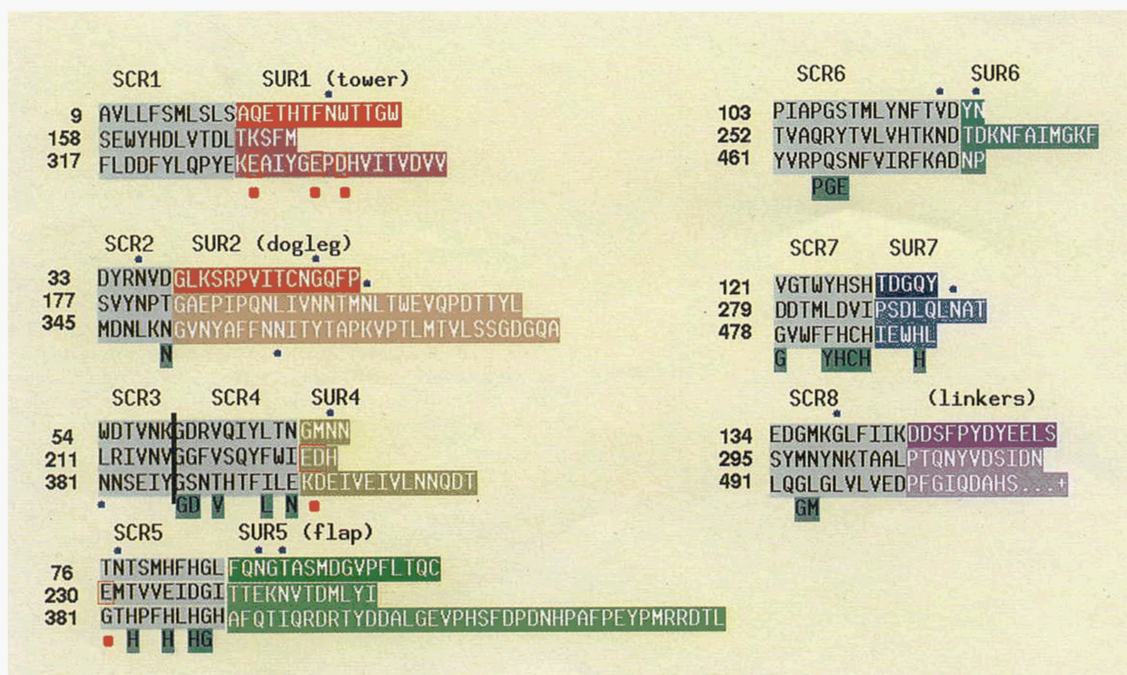


Fig. 5. Sequence of Fet3 arranged in three domains and divided according to proposed structurally conserved regions (SCRs) and structurally unconserved regions (SURs) based on sequence similarity to those regions identified in NIR, AO, and CP (Fig. 2). Owing to low sequence identity, the exact alignment of each SCR is not likely to be correct; however, the proposed regions reveal surprising consistency. The red dots flag three residues in SUR1 of domain 3, two in SUR4 of domain 2, and one in SCR5 of domain 2, which we propose to be involved in ferroxidase activity. Blue dots flag putative glycosylation sites. The colors of SURs correspond to those in Figure 4. Residues emphasized in line underneath are those that are highlighted in Figure 2.

alignment was obtained by adding the constraint that main-chain H-bonds be conserved in every cluster segment. This was accomplished by manually inspecting main-chain H-bonding patterns of the domains included in the superposition. The list of initial equivalenced residues was updated to maximize the number of conserved main-chain H-bonds and the alignment procedure using CLUSTERID was repeated. The added constraint of matching main-chain H-bonds resulted in an alignment more robust to changes in cutoff parameters. Including this added constraint results in an alignment with a slightly higher overall RMSD between pairs.

CLUSTERID produces a matrix of pairwise RMSDs between α-carbons. This matrix was used to produce the structural similarity tree in Figure 3 using the UPGMA method of the PHYLIP package (Felsenstein, 1989). The programs in the PHYLIP package are designed to analyze sequence comparisons and have not been widely exploited for comparison of three-dimensional structures as described here. Normally, the distance matrix is obtained by calculating the distance between the amino acid sequences of all pairs using a matrix describing their similarity. In the present study, the PHYLIP package is used to identify clusters of similar domains based on α-carbon distance and to represent these clusters as trees.

An initial tree was constructed from an alignment of the 11 domains from AO, CP, and NIR. The addition of more distantly related domains such as the cupredoxins reduced the size of the common core but produced a tree with greater discrimination between clusters of structures, including those that contain the 11 original domains. A final cutoff of 2.5 Å was used in the alignment of the 23 cupredoxin domains, which gave an overall RMS distance to the average structure of 1.42 Å and a maximal distance of 4.94 Å (Table 2). A cutoff of 2.5 Å was used to obtain a structural similarity tree with the highest discrimination of classes and maximizing the number of α-carbon clusters. Increasing the number of domains included in the superposition appears to improve the quality of similarity trees. When the same alignment procedure was applied to the 11 similar domains found in the NIR and the blue oxidases and each group of class IV and V domains, a cutoff of 2.0 Å was used. The lower cutoff was useful for defining a core of only highly conserved residues, and helps keep the alignment consistent between class IV and V domains.

Acknowledgments

Hartmut Michel (Max-Planck-Institut für Biophysik), Menachem Shoham (Case Western Reserve University), Tomitake Tsukihara (Osaka University), David Eisenberg (UCLA), and their colleagues provided atomic coordinates prior to PDB release. The authors thank Albrecht Messerschmidt for helpful early discussions. Part of this work was done in the laboratory of Professor Masaru Tanokura at the University of Tokyo. This work was supported by NIH Grant GM31770 (to E.T.A.) and an MRC of Canada Fellowship (to M.E.P.M.).

References

Adman ET. 1991. Copper protein structures. *Adv Prot Chem 42*:145–197.

Adman ET, Godden JW, Turley S. 1995. The structure of copper-nitrite reductase from *Achromobacter cycloclastes* at five pH values, with NO_2^- bound and with Type II copper depleted. *J Biol Chem 270*:27458-27474.

Adman ET, Stenkamp RE, Sieker LC, Jensen LH. 1978. A crystallographic model for azurin at 3 Å resolution. *J Mol Biol 123*:35–47.

Askwith C, Elde D, Van Ho A, Bernard PS, Li L, Davis-Kaplan S, Sipe DM, Kaplan J. 1994. The FET3 gene of *S. cerevisiae* encodes a multicopper oxidase required for ferrous iron uptake. *Cell 76*:403–410.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Bruce MD, Rodgers JR, Kennard O, Shimanouchi T, Tasmui M. 1977. The protein databank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Bihoreau N, Pin S, de Kersabiec AM, Vidot F, Fontaine-Aupart MP. 1994. Copper-atom identification in the active and inactive forms of plasma-derived FVIII and recombinant FVIII-delta II. *Eur J of Biochem 222*:41–48.

Bork P, Holm L, Sander C. 1994. The immunoglobulin fold, structural classification, sequence patterns and common core. *J Mol Biol 242*:309–320.

Colman PM, Freeman HC, Guss JM, Murata M, Norris VM, Ramshaw JAM, Venkatappa MP. 1978. X-ray crystallographic structure analysis of plastocyanin at 2.7 Å resolution. *Nature 272*:319–324.

DeSilva DM, Askwith CC, Eide D, Kaplan J. 1995. The gene product required for high affinity iron transport in yeast is a cell surface ferroxidase. *J Biol Chem 270*:1098–1101.

Diamond R. 1992. On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci 1*:1279–1287.

Dwulet FE, Putnam FW. 1981. Internal duplication and evolution of human ceruloplasmin. *Proc Natl Acad Sci USA 78*:2805–2809.

Farber GK, Petsko GA. 1990. The evolution of the α/β barrel enzymes. *Trends Biochem Sci 15*:228–234.

Felsenstein, J. 1989. PHYLIP—Phylogeny inference package (version 3.2). *Cladistics 5*:165–166.

Fenderson FF, Kumar S, Liu M-Y, Payne WJ, LeGall J. 1991. Amino acid sequence of nitrite reductase: A copper protein from *Achromobacter cycloclastes*. *Biochemistry 30*:7180–7185.

Freeman JC, Nayar PG, Begley TP, Villafrance JJ. 1993. Stoichiometry and spectroscopic identity of the copper centers in phenoxazinone synthase: A new addition to the blue copper oxidase family. *Biochemistry 32*:4826–4830.

Godden JW, Turley S, Teller DC, Adman ET, Liu MY, Payne WJ, LeGall J. 1991. The 2.3 Ångstrom X-ray structure of nitrite reductase from *Achromobacter cycloclastes*. *Science 253*:438–442.

Godzik A. 1996. The structural alignment between two proteins: Is there a unique answer? *Protein Sci 5*:1325–1338.

Hazes B, Hol WGJ. 1992. Comparison of hemocyanin β-barrel with other Greek key β-barrels: Possible importance of the β-zipper in proteins structure and folding. *Proteins Struct Func Genet 12*:278–298.

Iwata S, Ostermeier C, Ludwig B, Michel H. 1995. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature 376*:660–669.

Johnson MS, Sali A, Blundell TL. 1990. Phylogenetic relationships from three-dimensional protein structures. *Methods in Enzymol 183*:670–690.

Koikeda S, Ando K, Kaji H, Inoue T, Murao S, Takeuchi K, Samejima T. 1993. Molecular cloning of the gene for bilirubin oxidase from and its expression in yeast. *J Biol Chem 268*:18801–18809.

Kraulis P. 1991. MOLSCRIPT, a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst 24*:946–950.

Kukimoto M, Nishiyama M, Tanokura M, Adman ET, Horinouchi S. 1996. Studies on protein–protein interaction between copper-containing nitrite reductase and pseudoazurin from *Alcaligenes faecalis* S-6. *J Biol Chem 271*:13680–13683.

Lindley PF, Card G, Zaitseva I, Zaitsev V, Reinhammer B, Solin-Lindgren E, Yoshida K. An X-ray structural analysis of the ferroxidase activity of human ceruloplasmin. *J Biol Inorg Chem.* Submitted.

Merritt EA, Murphy MEP. 1994. Raster3D (version 2.0). A program for photorealistic molecular graphics. *Acta Crystallogr D50*:869–873.

Messerschmidt A, Huber R. 1990. The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin, modelling and structural relationships. *Eur J of Biochem 187*:341–352.

Messerschmidt A, Ladenstein R, Huber R, Avigliano L, Petruzzelli R, Rossi A, Finazzi-Agro A. 1992. Refined crystal structure of ascorbate oxidase at 1.9 Å resolution. *J Mol Biol 224*:179–205.

Messerschmidt A, Rossi A, Ladenstein R, Huber R, Bolognesi M, Gatti G, Marchesini A, Petruzzeli R, Finazzi-Agro A. 1989. X-ray crystal structure of the blue oxidase ascorbate oxidase from zucchini. *J Mol Biol 206*:513–529.

Murphy MEP, Turley S, Kukimoto M, Nishiyama M, Horinouchi S, Sasaki H, Tanokura M, Adman ET. 1995. Structure of *Alcaligenes faecalis* nitrite reductase and a copper site mutant, M150E, that contains zinc. *Biochemistry 34*:12107–12117.

Pan Y, DeFay T, Gitschier J, Cohen FE. 1995. Proposed structure of the A domains of factor VIII by homology modelling. *Nat Struct Biol 2*:740–744.

Ryden L. 1984a. Ceruloplasmin. In: Lontie R, ed. *Copper proteins and copper enzymes III*. Boca Raton, Florida: CRC Press. pp 37–100.

Ryden L. 1984b. Structure and evolution of small blue proteins. In: Lontie R, ed. *Copper proteins and copper enzymes I*. Boca Raton, Florida: CRC Press. pp 157–183.

Ryden LG, Hunt LT. 1993. Evolution of protein complexity: The blue copper-containing oxidases and related proteins. *J Mol Evol 36*:41–66.

Tsukihara T, Aoyama H, Yamashita H, Tomizaka T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. 1996. The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science 272*:1136–1144.

van der Oost J, Lappalainen P, Musacchio A, Warne A, Lemieux L, Rumbley J, Gennis RB, Asa R, Pascher T, Malmstrom BG, Saraste M. 1992. Restoration of a lost metal-binding site: Construction of two different copper sites into a subunit of the cytochrome quinol oxidase complex. *EMBO J 11*:3209–3217.

Vriend G. 1990. WHATIF: A molecular modelling and drug design program. *J Mol Graph 8*:52–56.

Wilmanns M, Lappalainen P, Kelly M, Sauer-Eriksson E, Saraste M. 1995. Crystal structure of the membrane-exposed domain from a respiratory quinol oxidase with an engineered dinuclear copper center. *Proc Nat Acad Sci USA 92*:11955–11959.

Zaitseva I, Zaitsev V, Card G, Moshkov K, Bax B, Ralph A, Lindley P. 1996. The X-ray structure of human serum ceruloplasmin at 3.1 Å: Nature of the copper centres. *J Biol Inorg Chem 1*:15–23.