J Mol Evol (1993) 36:41–66

Journal of
Molecular Evolution

Evolution of Protein Complexity: The Blue Copper-Containing Oxidases and Related Proteins

Lars G. Rydén[1] and Lois T. Hunt[2]

[1] Department of Biochemistry, Uppsala University, Uppsala Biomedical Center, Box 576, 751 23 Uppsala, Sweden
[2] National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road, NW, Washington DC 20007, USA

Summary. The blue copper proteins and their relatives have been compared by sequence alignments, by comparison of three-dimensional structures, and by construction of phylogenetic trees. The group contains proteins varying in size from 100 residues to over 2,300 residues in a single chain, containing from zero to nine copper atoms, and with a broad variation in function ranging from electron carrier proteins and oxidases to the blood coagulation factors V and VIII. Difference matrices show the sequence difference to be over 90% for many pairs in the group, yet alignment scores and other evidence suggest that they all evolved from a common ancestor. We have attempted to delineate how this evolution took place and in particular to define the mechanisms by which these proteins acquired an ever-increasing complexity in structure and function. We find evidence for six such mechanisms in this group of proteins: domain enlargement, in which a single domain increases in size from about 100 residues up to 210; domain duplication, which allows for a size increase from about 170 to about 1,000 residues; segment elongation, in which a small segment undergoes multiple successive duplications that can increase the chain size 50-fold; domain recruitment, in which a domain coded elsewhere in the genome is added on to the peptide chain; subunit formation, to form multisubunit proteins; and glycosylation, which in some cases doubles the size of the protein molecule. Size increase allows for the evolution of new catalytic properties, in particular the oxidase function, and for the formation of coagulation factors with multiple interaction sites and regulatory properties. The blood coagulation system is examined as an example in which a system of interacting proteins evolved by successive duplications of larger parts of the genome. The evolution of size, functionality, and diversity is compared with the general question of increase in size and complexity in biology.

Key words: Blue copper proteins — Blue oxidases — Coagulation factors — Discoidins — Domain evolution — Evolutionary mechanisms — Phylogeny

The blue copper proteins (for review see, e.g., Rydén 1988) form an extremely diverse group of proteins containing from 1 to 9 copper atoms per molecule and with about 100 to more than 1,000 residues in the single peptide chain. The blue color is due to binding of a so-called type 1 copper (Malkin and Malmström 1970), which has an extinction coefficient of about 5,000 at 610 nm and unusual spectral, magnetic, and redox properties. Recent reports on the structure of ascorbate oxidase (Messerschmidt et al. 1989; Messerschmidt and Huber 1990) and previous reports on cDNA sequences of the three blue oxidases laccase (Germann et al. 1988), ascorbate oxidase (Ohkawa et al. 1989), and ceruloplasmin (Koshinsky et al. 1986 for the cDNA

Offprint requests to: L.G. Rydén

Table 1. Blue copper-containing proteins and proteins related to them

Protein	Source	Peptide chain length	Copper (number)	Function
Small blue proteins				
Plastocyanin	Bacteria/plants	ca. 100	1	Electron transfer in photosynthesis
Azurin	Bacteria	ca. 130	1	Electron transfer
Pseudoazurin	Bacteria	ca. 100	1	Electron transfer
Amicyanin	Bacteria	ca. 100	1	Electron transfer
Phytocyanin	Plants	ca. 100	1	Unknown
Blue oxidases				
Ascorbate oxidase	Plants	540	4	Oxidation of ascorbate
Laccase	Plants/fungi	540	4	Lignin degradation
Ceruloplasmin	Vertebrate plasma	1,046	6	Unknown
Cytochrome c oxidase subunit II	Mitochondria	150	2	Oxidation of cytochrome c
Coagulation factors				
Factor V	Vertebrate plasma	2,196	1	Blood coagulation
Factor VIII	Vertebrate plasma	2,332	?	Blood coagulation

The two coagulation factors are homologous to the blue copper proteins over a part of the peptide chain. The homology with subunit II of cytochrome c oxidase is only tentative

sequence; and Takahashi et al. 1984, for the protein sequence), as well as on the two coagulation factors V (Jenny et al. 1987) and VIII (Wood et al. 1984) from human plasma that are related in part to ceruloplasmin, have disclosed a complex of homologous proteins widely different in size, structure, organization, and function. These proteins illustrate, perhaps more clearly and interestingly than in any other group, the various routes evolution has taken to design proteins with new functions and new levels of size and complexity by modifying an original small and simple structure. We have attempted to trace how this evolution took place by using methods common in the study of molecular evolution—in particular, construction of phylogenetic trees.

The proteins to be discussed are presented in Table 1. The small blue proteins (for review see Rydén 1984a) are found in bacteria and in plants. They all contain a single peptide chain of 100–140 residues that binds a single intensely blue type 1 copper. So far five clearly distinct families of small blue proteins have been discovered. The functions of these proteins, as far as they are known, are all related to electron transfer. Plastocyanin, which transfers electrons from photosystems PII to PI in photosynthesis, is best characterized. High-resolution three-dimensional structures are available for plastocyanin (Guss and Freeman 1983), azurin (Adman et al. 1978; Norris et al. 1983), pseudoazurin (Adman et al. 1989), and cucumber basic blue protein (Guss et al. 1988). The small blue proteins from nonphotosynthetic parts of plants, e.g., the cucumber basic blue protein and stellacyanin from the lacquer tree, form a group of more closely related families. We will call these the phytocyanins, as suggested earlier (Rydén 1984a). Finally, the ragweed allergen, a pollen protein, has been shown to be related to the group of small blue proteins (Hunt et al. 1985)—most closely to the phytocyanins—although it is neither blue nor contains copper.

Three different blue oxidases are known. Laccase has been isolated from the lacquer tree and several fungi (for review see Reinhammar 1984), ascorbate oxidase from cucumber and squash (reviewed by Mondovi and Avigliano 1984), and ceruloplasmin from vertebrate plasma (for review see Rydén 1984b). Laccase and ascorbate oxidase contain peptide chains of 540–570 residues, which form a dimer in ascorbate oxidase, whereas ceruloplasmin has a single peptide chain, which is 1,046 residues long in the human protein. The blue oxidases also contain, in addition to the type 1 copper, a trinuclear cluster as revealed by the structure of ascorbate oxidase (Messerschmidt et al. 1989). The cluster consists of a pair of magnetically coupled so-called type 3 coppers (Malkin and Malmström 1970) and a magnetically isolated type 2 copper, with properties more similar to those of copper in ordinary small-molecular-weight complexes. The blue oxidases have the ability to transfer four electrons from a reducing substrate to a molecule of oxygen, which thereby is reduced to water. They are, together with cytochrome c oxidase, the only enzymes known to catalyze this four-electron transfer reaction.

The blood coagulation factors V (Jenny et al. 1987) and VIII (Wood et al. 1984) have been shown to consist of single peptide chains of 2,196 and 2,332 residues, respectively. The two factors have essen-

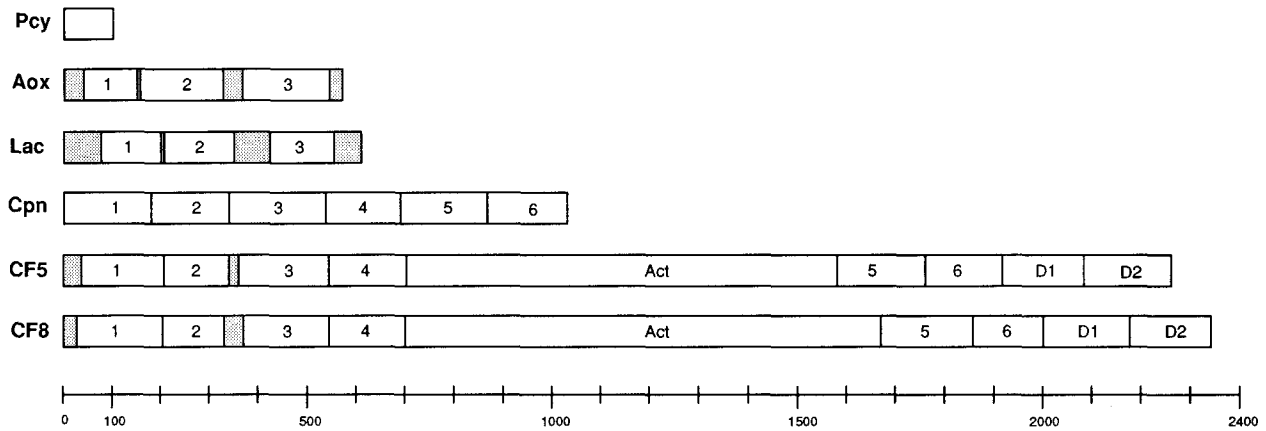


Fig. 1. Domain structure of a small blue protein plastocyanin (Pcy); the three blue oxidases plant ascorbate oxidase (Aox), fungal laccase (Lac), and human ceruloplasmin (Cpn); and the two human blood coagulation factors V (CF5) and VIII (CF8). Homologs of the copper-binding domains are numbered. In CF5 and CF8, Act is the activation domain and D1 and D2 are the discoidinlike domains. The lengths of the peptide chains are drawn to scale.

tially the same mosaic gene structure and evidently share a common ancestor. Parts of the sequences are 30–40% identical with the ceruloplasmin sequence and thus they belong to the group of proteins related to the blue oxidases. In plasma both factors occur as high-molecular-weight complexes with other proteins. They function in the blood coagulation cascade by binding factors Xa and IXa, respectively, to promote the proteolytic activity of these enzymes (Kane and Davies 1988). It appears that factors V and VIII are not blue and have no oxidase activity. However, factor V has been shown to contain one copper per mole, whereas the copper content of factor VIII is not known, although it is likely to be the same or higher.

The larger of the proteins mentioned all show internal homologies. Thus human ceruloplasmin is composed of three homology regions, each of about 340 residues, and some evidence for a duplication in these regions has indicated a possible sixfold repeat (Rydén 1982). The ascorbate oxidase structure revealed that this enzyme contains three homologous domains (Messerschmidt et al. 1989; Messerschmidt and Huber 1990). As laccase and ascorbate oxidase are comparatively closely related, it is evident that laccase also has this three-domain structure. The blood coagulation factors V and VIII were shown to be composed of a number of domains. In addition to the parts homologous to the entire ceruloplasmin sequence, which accounts for about 1,050 residues, a highly glycosylated stretch of about 900 residues is found between the second and third homology regions; finally, two repeats homologous to discoidin, a lectin produced by the cellular slime mold *Dictyostelium discoideum* (Poole et al. 1981), are added to the C-terminus. The domain structures of the proteins are depicted in Fig. 1.

The crystallographic structures of the small blue proteins (for plastocyanin, Guss and Freeman 1983; for azurin, Adman et al. 1978; and Norris et al. 1983; for pseudoazurin Adman et al. 1989; for cucumber basic blue protein, Guss et al. 1988) all feature a peptide chain folded in an eight-stranded double sandwich with a Greek key topology. The type 1 copper in each of these is coordinated in a flattened tetrahedron with one ligand from a strand 4 histidine, three ligands (a histidine, a cysteine, and a methionine) from the turn between strands 7 and 8, and possibly a carbonyl oxygen as a fifth ligand. The same folding is found in the three ascorbate oxidase domains although they are up to 100 residues longer than the small blue proteins. Furthermore, the type 1 copper in domain three in ascorbate oxidase is bound in the same corresponding position as in the small blue proteins. The similarities confirm the homology between the small blue proteins and the domains in the blue oxidases.

Direct calculations also support the assumption that the small blue proteins (Rydén and Lundgren 1976) and the domains in the blue oxidases are homologous (Rydén 1982). The proteins described thus form a large group of related proteins. Since many of these form families and superfamilies as defined in the PIR database (Dayhoff et al. 1983), the proteins mentioned form a group of related superfamilies.

Several other proteins have been suggested to belong to the group. In particular, these include subunit II of cytochrome c oxidase (Rydén and Lundgren 1976; Holm et al. 1987), which binds two copper ions in addition to a heme group; the superoxide dismutase (Rydén 1988), which binds a copper-zinc pair and has a peptide chain fold consisting of the same arrangement of eight beta-strands; and more recently lysyl oxidase (Trackman et al. 1990), in which we observed that a region consisting of

residues 279–340 has a suggestive similarity to the type 1 copper-binding region. Of these only the cytochrome oxidase subunit will be considered below.

Groups of related protein superfamilies are also formed by the immunoglobulins (Williams and Barclay 1988) and the serine proteases (Young et al. 1979; Patthy 1985). In these groups also phenomena similar to those seen in the blue protein group occur. A detailed analysis of the evolutionary events that led to the complexity has, however, not been done in any of these cases. It appears that the blue proteins, whose history extends over nearly the entire estimated time span for life on Earth, offer the best prospects for actually arriving at a credible description of an evolutionary history.

Sequences, Structures, and Methods

Sequence, Structures, and Designations. The amino acid sequences used have been obtained from the following sources: plastocyanin from poplar, *Populus nigra* var. *italica* (Guss and Freeman 1983); plastocyanin from *Anabaena variabilis* (Aitken 1975); azurin from *Pseudomonas aeruginosa* (Ambler 1971); amicyanin from *Methylobacterium extorquens* (Ambler and Tobari 1985); pseudoazurin from *Alcaligenes faecalis* (Hormel et al. 1986); cucumber basic blue protein (phytocyanin) from cucumber, *Cucumis sativus* (Murata et al. 1982); stellacyanin (phytocyanin) from lacquer tree, *Rhus vernicifera* (Bergman et al. 1977); ceruloplasmin from human, *Homo sapiens,* plasma (Takahashi et al. 1984); laccase from *Neurospora crassa* (Germann et al. 1988); laccase from the white rot fungus, *Phlebia radiata* (Saloheimo et al. 1991), whose sequence was kindly communicated before publication by Markku Saloheimo; ascorbate oxidase from cucumber, *Cucumis sativus* (Ohkawa et al. 1989); factor VIII from human plasma (Wood et al. 1984); factor V from human plasma (Jenny et al. 1987); discoidin from *Dictyostelium discoideum* (Poole et al. 1981); and fat globule membrane glycoprotein from mouse, *Mus musculus* (Stubbs et al. 1990).

The protein three-dimensional structures discussed are from the following sources: plastocyanin from poplar, *Populus nigra* var. *italica* (Guss and Freeman 1983); azurin from *Pseudomonas aeruginosa* (Adman et al. 1978); pseudoazurin from *Alcaligenes faecalis* (Adman et al. 1989); cucumber basic blue protein (phytocyanin) from cucumber, *Cucumis sativus* (Guss et al. 1988); and ascorbate oxidase from zucchini squash, *Cucurbita pepo condensa* (Messerschmidt et al. 1989; Messerschmidt and Huber 1990).

Proteins are designated by a three-character symbol as follows: plastocyanin, Pcy; amicyanin, Amy; azurin, Azu; pseudoazurin, Paz; phytocyanin, Phy (which includes stellacyanin, Sty, and cucumber basic blue protein, Cub); laccase, Lac; ascorbate oxidase, Aox; ceruloplasmin, Cpn; coagulation factor V, CF5; and coagulation factor VIII, CF8. Domain numbering is from N-terminal to C-terminal by arabic numerals, e.g., the most N-terminal domain in ceruloplasmin is designated Cpn1 and the most C-terminal Cpn6. In the mosaic proteins factors V and VIII the domains are designated to indicate their origin: the ceruloplasmin-related domains are called 1–6, the discoidin domains D1 and D2, and the middle activation domain Act.

Computational Methods. All sequences used were included in the Protein Sequence Database at the Protein Identification Resource (PIR) at the National Biomedical Research Foundation. The database as well as its software support is described in Barker et al. (1988 and 1990) and George et al. (1988).

A first general comparison of two sequences was done using the program DOTMATRIX, with the mutation data scoring matrix (Dayhoff et al. 1983; George et al. 1990). This program produces a graphic matrix display in which each dot indicates a short region of similarity between the two sequences. Extensive regions of similarity appear as diagonal lines; repeats appear as repetitive diagonal lines along one axis. Internal homologies were observed by comparison of a sequence to itself. Domain structures were in many cases noted as stretches of similarity interrupted by stretches without such similarity.

The alignment of any pair of sequences, either complete or a domain in a longer sequence, was based on DOTMATRIX comparisons and the known structural features of the proteins to be compared, in particular the positions of beta-strands and large loops. Pairwise alignments were made with the program ALIGN (Dayhoff et al. 1983), which is based on the Needleman and Wunsch algorithm (Needleman and Wunsch 1970). The program was run using the Mutation Data Matrix of 250 PAMs (accepted point mutations per hundred residues) for comparison of amino acid pairs. The parameter values chosen were normally a bias of six and a gap penalty of six or higher. The similarity of the two sequences was finally evaluated as an alignment score for the comparison of the two real sequences with the mean of a distribution of the scores from comparisons of 100 randomized sequences.

Alignments of three or more sequences were done as an extension of the pairwise alignments. The new sequence was aligned with one or more of the old and inserted into the grand alignment by manual adjustment and by minimizing the number of gaps and maximizing identities in all positions, with the aid of the alignment editing program ALNED. To aid in aligning sequences for which no three-dimensional structures were available, secondary structure predictions were made with the program CHOFAS (Kanehisa 1985), which is based on the method of Chou and Fasman (1978a,b). Finally, the program DISP was used to produce an alignment and to calculate a difference matrix. The parts in any two sequences where the differences in length were too large to allow a meaningful alignment (in particular between beta-strands 1 and 2, and 4 and 5) were excluded in the comparison, permitting a much better "signal-to-noise ratio" in the comparisons. Details for specific alignments are given in the Results section.

Phylogenetic trees (topologies or dendrograms) were calculated from the alignments using the program MATTOP (Barker et al. 1988), which is a modification of the exhaustive-search distance-matrix method first introduced by Fitch and Margoliash (1967). MATTOP generates a large number of topologies (up to 10,395 for eight branches) from a matrix of the estimated evolutionary distances between all pairs of sequences, computed from the pairwise percent differences by correcting for inferred superimposed and parallel mutations.

Distance-matrix methods are among the best for deriving topologies from distantly related sequences (where superimposed, parallel, and back-mutations have occurred), assuming that rates of evolution can vary on various branches (Astolfi et al. 1981; Barker and Dayhoff 1979; Blanken et al. 1982; Hasegawa et al. 1991; Saitou and Imanishi 1989; Sourdis and Nei 1988). The program calculates for each branch of the tree a distance in PAMs, using the Mutation Data Matrix of 250-PAM distance. A tree obtained from this calculation is in itself an unrooted topological relationship between sequences. The root, or the point of earliest time, must be estimated from other considerations. In principle, a sequence known to branch off first from all of the others could provide the trunk for the rest of the tree. However, sequences that diverged early are often so different that they cannot be

placed relative to the remaining topology with certainty. In the derivation of our trees we included an outgroup (i.e., a distantly related sequence, most often plastocyanin) to aid in selecting a position for the root, which was usually placed on the longer branch to the outgroup. Normally, no more than eight sequences can be accommodated by the MATTOP program. However, if some sequences are clustered on a single branch, the number of sequences treated can be increased correspondingly.

For each run the program calculated all possible trees and printed the 5–10 shortest trees—that is, with the minimal absolute value for total branch length expressed in PAMs (Barker et al. 1988). When trees with all positive branch lengths were obtained, trees with negative branch lengths were excluded from consideration. In one case, the best tree had one very short negative branch that was depicted as a region of uncertainty in the order of branching. The best tree is usually the shortest with all positive branches, having biological consistency; however, on occasion one must choose between the best biological tree and the shortest tree. (Here a difference in length of less than 3 PAMs was not considered significant.)

Results—Evolution of Copper-binding Domains

Alignments and Alignment Scores; Domain Structures

The alignment of the domains in all three oxidases is shown in Fig. 2a and the domain structures in Fig. 1. The domains in ascorbate oxidase were those defined by the three-dimensional structure as described by Messerschmidt et al. (1989). The laccase sequence is similar enough to the ascorbate oxidase to allow domain borders to be clearly seen. In ceruloplasmin three regions with internal homology are easily recognized; their size is approximately twice as large as the domains in ascorbate oxidase and laccase. The homology with ascorbate oxidase and laccase suggests that these regions consist of two domains. Indeed, extensive sequence similarities, including about 40% identities, are seen between the middle and the beginnings of the homology regions, which allows the putative domain borders to be identified. The positions of domain borders are further supported by the limited proteolysis pattern, as the points where the peptide chains are nicked coincide with these borders. The assumed six-domain structure of ceruloplasmin has thus been used in the alignment. The four ligands for type 1 copper are found in all copper-binding domains; however, it is clear from the alignment that no residue is conserved in all sequences.

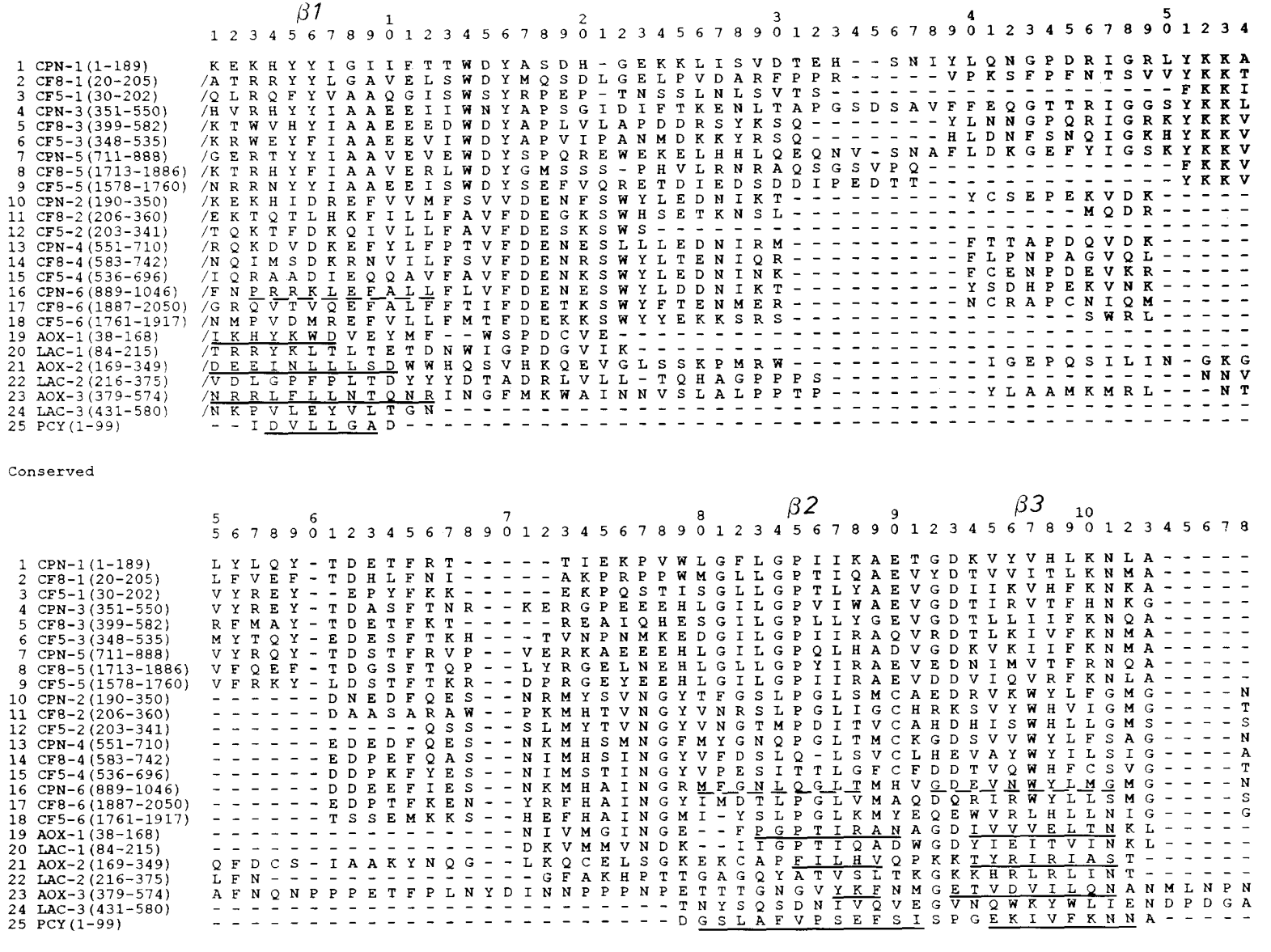
The sequences of coagulation factors V and VIII can easily be aligned along the entire ceruloplasmin sequence as shown previously (Wood et al. 1984; Jenny et al. 1987). The parts of the sequences of the two coagulation factors that appear within and after regions homologous with the ceruloplasmin sequence can easily be recognized. Thus the definition of domains in ceruloplasmin automatically defines domains in the two coagulation factors. Considerable care was taken in making the alignments, as tree-building is critically dependent on them. The alignments of the coagulation factors agree well with those previously published. In these three proteins, there is also a conserved disulfide bond in each of domains 1–5 but not in domain 6. (See Fig. 2a.)

Alignments of ascorbate oxidase, laccase, and ceruloplasmin differ from most previously published suggestions but are similar to that in Messerschmidt and Huber (1990) except for laccase domain 2; our alignment of this domain is based on DOTMATRIX and ALIGN comparisons combined with secondary structure predictions. DOTMATRIX comparison of ascorbate oxidase and laccase showed that the regions of closest similarity for the second domain were residues 240–325 and 265–350, respectively. For residues 240–340 of ascorbate oxidase (which includes beta-strands 2–8) and 265–367 of laccase the ALIGN score was 5.9. Comparison of ascorbate oxidase residues 169–346 (entire domain 2) with laccase residues 216–366 (our suggested domain 2) or laccase residues 263–386 (domain 2 according to Messerschmidt and Huber, 1990) gave ALIGN scores of 5.1 and 0.2, respectively. Beta-strand prediction was slightly better for our proposed domain.

In the alignment the beta-strands (as defined by Guss and Freeman 1983) in plastocyanin, added as a reference sequence, and in ascorbate oxidase (as defined by Messerschmidt et al. 1989) have been indicated. Proposed beta-strands for ceruloplasmin domain 6 (Rydén 1982) have also been indicated. Only strands in the double sandwich, referred to as beta-1 through beta-8, have been included. Strand 5 is not in a good conformation for sheet formation and is also of very different length in the proteins. It is apparent that areas of poor alignment are those between strands, in particular between beta-1 and beta-2, as well as before and after beta-5. These areas have not been included in the calculations of difference matrices, trees, and alignment scores, as indicated in figure legends and table headings. Likewise, residues known to be copper-binding in the C-terminal domains of ascorbate oxidase have been marked. The difference matrix for the alignment is given in Fig. 2b. Percent difference varies from 41 to 96%.

General Relationships Between the Proteins

In order to obtain a comparison of all proteins in the group we used the C-terminal domains in the oxidases, all of which contain a type 1 copper-binding site, and the corresponding domain (domain 6) in


β1
β2
β3
1 CPN-1(1-189)
2 CF8-1(20-205)
3 CF5-1(30-202)
4 CPN-3(351-550)
5 CF8-3(399-582)
6 CF5-3(348-535)
7 CPN-5(711-888)
8 CF8-5(1713-1886)
9 CF5-5(1578-1760)
10 CPN-2(190-350)
11 CF8-2(206-360)
12 CF5-2(203-341)
13 CPN-4(551-710)
14 CF8-4(583-742)
15 CF5-4(536-696)
16 CPN-6(889-1046)
17 CF8-6(1887-2050)
18 CF5-6(1761-1917)
19 AOX-1(38-168)
20 LAC-1(84-215)
21 AOX-2(169-349)
22 LAC-2(216-375)
23 AOX-3(379-574)
24 LAC-3(431-580)
25 PCY(1-99)
Conserved

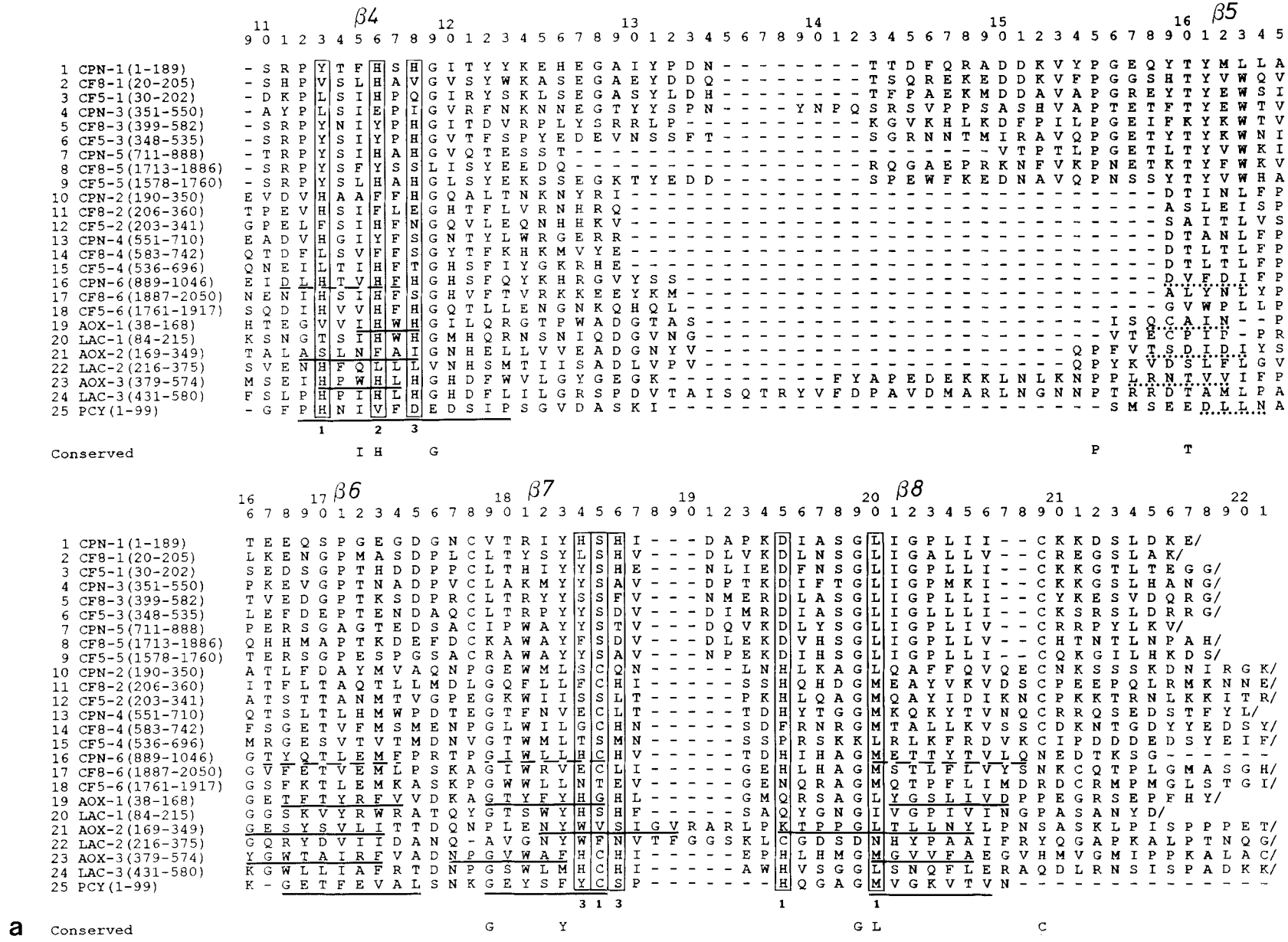


Fig. 2. Continued on page 48.

Number of differences

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1 CPN-1(1-189)	.	52	51	55	50	48	50	53	45	78	82	77	79	82	78	79	84	77	66	72	83	83	76	75	82
2 CF8-1(20-205)	58	.	48	49	48	49	47	49	47	80	83	79	80	81	80	76	81	81	70	71	84	85	77	82	76
3 CF5-1(30-202)	57	53	.	43	46	47	48	52	46	84	81	78	82	82	76	84	82	82	70	69	86	84	81	81	80
4 CPN-3(351-550)	61	54	48	.	45	45	48	48	44	84	84	82	84	85	78	84	83	84	77	70	83	84	82	83	82
5 CF8-3(399-582)	56	53	51	50	.	37	48	50	49	83	85	83	82	86	81	84	82	80	77	73	85	84	83	82	81
6 CF5-3(348-535)	53	54	52	50	41	.	43	48	42	83	82	82	83	84	79	83	80	82	75	70	80	85	83	81	78
7 CPN-5(711-888)	56	52	53	53	53	48	.	48	38	75	81	77	82	82	78	79	79	79	72	72	81	84	78	82	78
8 CF8-5(1713-1886)	59	54	58	53	56	53	53	.	41	82	85	82	83	84	82	84	82	84	77	76	88	83	81	82	82
9 CF5-5(1578-1760)	49	52	51	48	54	46	42	45	.	84	87	80	86	85	79	80	78	79	75	70	84	84	78	80	78
10 CPN-2(190-350)	85	87	91	91	90	90	82	89	91	.	66	60	56	61	68	56	62	60	75	84	85	83	72	76	80
11 CF8-2(206-360)	89	90	88	91	92	89	88	92	95	73	.	67	65	66	75	62	63	68	79	82	82	85	74	76	79
12 CF5-2(203-341)	84	86	85	89	90	89	84	89	87	66	74	.	64	67	73	69	66	69	76	77	84	87	80	82	83
13 CPN-4(551-710)	86	87	89	91	89	90	89	90	93	62	71	70	.	66	70	53	59	68	74	81	82	83	77	74	76
14 CF8-4(538-742)	89	88	89	92	93	91	89	91	92	67	73	74	73	.	61	65	66	67	82	87	81	87	78	78	80
15 CF5-4(536-696)	85	87	83	85	88	86	85	89	86	75	82	80	77	67	.	73	72	76	78	83	82	86	73	77	80
16 CPN-6(889-1046)	86	83	91	91	91	90	86	91	87	62	68	76	58	71	80	.	56	56	76	82	85	85	69	73	78
17 CF8-6(1887-2050)	91	88	89	90	89	87	86	89	85	68	69	73	65	73	79	62	.	56	76	81	82	86	73	72	77
18 CF5-6(1761-1917)	84	88	89	91	87	89	86	91	86	66	75	76	75	74	84	62	62	.	76	78	86	80	74	75	80
19 AOX-1(38-168)	72	76	76	84	84	82	78	84	82	82	86	83	80	89	85	83	83	84	.	57	84	85	73	76	83
20 LAC-1(84-215)	78	77	75	76	79	76	78	83	76	91	89	84	88	95	90	89	88	86	64	.	81	79	75	78	82
21 AOX-2(169-349)	90	91	93	90	92	87	88	96	91	92	89	91	89	88	89	92	89	93	91	88	.	76	84	84	85
22 LAC-2(216-375)	90	92	91	91	91	92	91	90	91	90	92	95	90	95	93	92	93	87	92	86	83	.	82	86	87
23 AOX-3(379-574)	83	84	88	89	90	90	85	88	85	78	80	87	84	85	79	75	79	80	79	82	91	89	.	62	80
24 LAC-3(431-580)	82	89	88	90	89	88	89	89	87	83	83	89	80	85	84	79	78	82	83	85	91	93	67	.	81
25 PCY(1-99)	90	84	88	90	89	86	86	90	86	87	86	90	83	87	87	85	84	87	90	89	92	95	87	88	.

Percent difference

b

Fig. 2. Continued. **a** Alignment of sequences of domains in ascorbate oxidase, laccase, and ceruloplasmin and the parts of coagulation factors V and VIII homologous to ceruloplasmin. Abbreviations are the same as in Fig. 1. Domain boundaries are those defined by Messerschmidt et al. (1989) for ascorbate oxidase and as described in the text for laccase and ceruloplasmin. The domains of coagulation factors V and VIII are defined by homology to ceruloplasmin. Plastocyanin is included as a reference sequence. Known beta-strands are indicated by *lines* below the sequence for plastocyanin and ascorbate oxidase; *dashed lines* indicate suggested beta-strands for ceruloplasmin domain 6, and *dotted lines,* suggested beta-5 strands. Positions at which copper ligands occur in some of the domains are *boxed and numbered* according to the type of copper-binding. Conserved residues are those occurring in more than half of the sequences. Conserved disulfide bonds in ceruloplasmin and the coagulation factors are at positions 178–209 in domains 1, 3, and 5, and at positions 90–209 in domains 2 and 4. **b** Difference matrix calculated from the alignment. Regions in which no conserved beta-strands occur, in which the sequences vary in length, or in which most sequences have gaps are omitted from the calculations of the matrix. These are positions 1, 13–79, 103–108, 124–157, 188–194, and 208–221, or a total of 129 positions in the comparisons. The lower left half of the matrix shows percent difference; the upper right half shows number of differences.

the coagulation factors. These domains are assumed to be monophyletic and to give a correct phylogeny when compared. Plastocyanin from poplar was used as an outgroup. Alignments were those in Fig. 2a.

Only seven trees of all possible topologies (945) had all branches positive. The best biological tree, shown in Fig. 3, was fourth in absolute length (772 PAMs) and weighted variance; it was only 2 PAMs longer than the shortest tree. Coagulation factors V and VIII group together with ceruloplasmin on one branch, with very small distances between their points of divergence. The laccases and ascorbate oxidase diverge on another branch; again there is a very short internodal branch. These two main branches also correspond to animal and plant proteins, respectively. The base of the tree was arbitrarily placed on the long branch to the outgroup plastocyanin.

Evolution of the Oxidase Domain Structure

Trees for all domains in laccase and in ascorbate oxidase, and then all six domains together, were calculated with plastocyanin as an outgroup. No trees with all positive branches were obtained; two trees had one very short (4 PAMs) negative branch

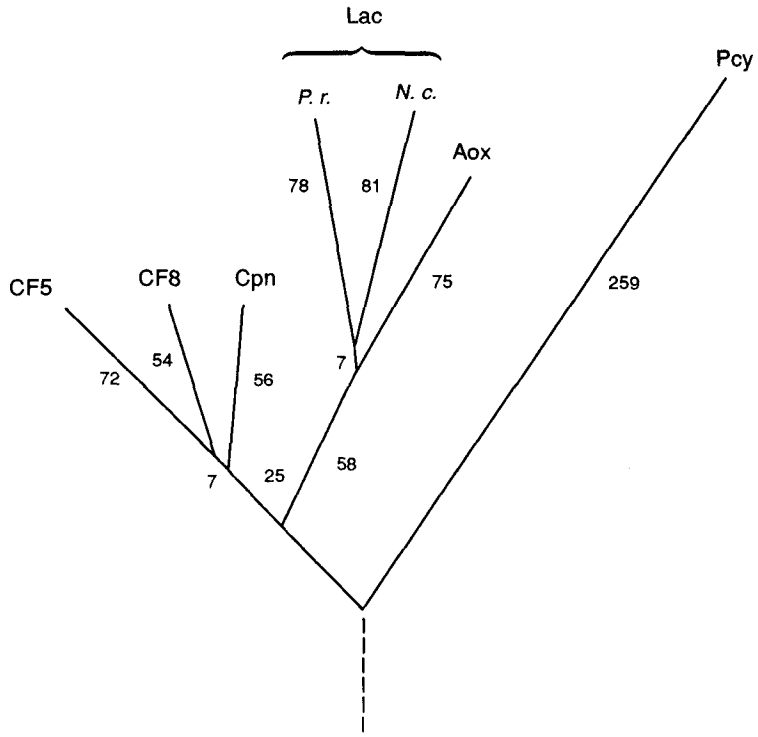


Fig. 3. Phylogenetic tree for the C-terminal type 1 copper-binding domains in the three blue oxidases, ascorbate oxidase, laccase, and ceruloplasmin, and the corresponding domains in coagulation factors V and VIII, using plastocyanin as an outgroup. Sequences are those in Fig. 2a with the addition of *Phlebia radiata* (*P.r.*) laccase, and positions omitted from calculating the tree are those given in Fig. 2b. Branch lengths are in PAMs and the suggested trunk is arbitrarily positioned, on this and all other trees. Scale: 0.77 mm = 2 PAMs.

and a difference in length of 2 PAMs. The negative branch, i.e., an uncertainty in the order of branching, arises from the great difference in branch lengths for the two middle domains as compared with the rest of the tree. The biologically best tree, with a length of 1,272 PAMs, is shown in Fig. 4a. It is evident that a structure closest to what is now the C-terminal copper-binding domain evolved first and then duplicated to form a double domain, similar to the N-terminal and C-terminal domains in today's proteins. Finally the middle domain was added to form the three-domain structure. This middle domain has diverged considerably from the other two. The two oxidases diverged after the three-domain structure had evolved.

Likewise a tree was calculated for the six ceruloplasmin domains, with the C-terminal domain of plastocyanin as an outgroup. The only all-positive tree (Fig. 4b), with a total branch length of 763 PAMs, shows the ancestral double domain, the ancestor of the so-called homology region, at the first node. Then the ancestor of the N-terminal double domain formed, and finally the middle double domain arose from duplication of the C-terminal double domain. The tree shown, as well as the combined trees of the 12 domains (not shown), proves that the common ancestor of the three oxidases was a double domain. We conclude that this double domain was an active enzyme.

Relationship to the Small Blue Proteins

To investigate the relationship of the blue oxidases to the small blue proteins, an alignment of the small blue proteins was constructed as shown in Fig. 5a. The beta-strands in the four known three-dimensional structures of plastocyanin, azurin, pseudoazurin, and a phytocyanin (the cucumber basic blue protein) were used to guide the alignment. The C-terminal domain of the ascorbate oxidase was included in accordance with Fig. 2a and used as an outgroup in the calculation of a phylogenetic tree. The difference matrix of the alignment is shown in Fig. 5b.

The shortest tree (Fig. 6) with all branches positive has an overall length of 1,228 PAMs. This 10-branch tree was first calculated as an eight-branch tree on which the two azurin and the two plastocyanin sequences were clustered. Then the shortest topology was run with all 10 branches. Amicyanin, pseudoazurin, and plastocyanin cluster together, with azurin at some distance on one side and the phytocyanins, including the pollen protein allergen Ra3, well removed from them on the other side. This makes a clear distinction between the plant proteins and the bacterial proteins, if plastocyanin as a chloroplast protein is counted among the bacterial ones. Ascorbate oxidase branches off from the bacterial side closest to azurin. The root of the

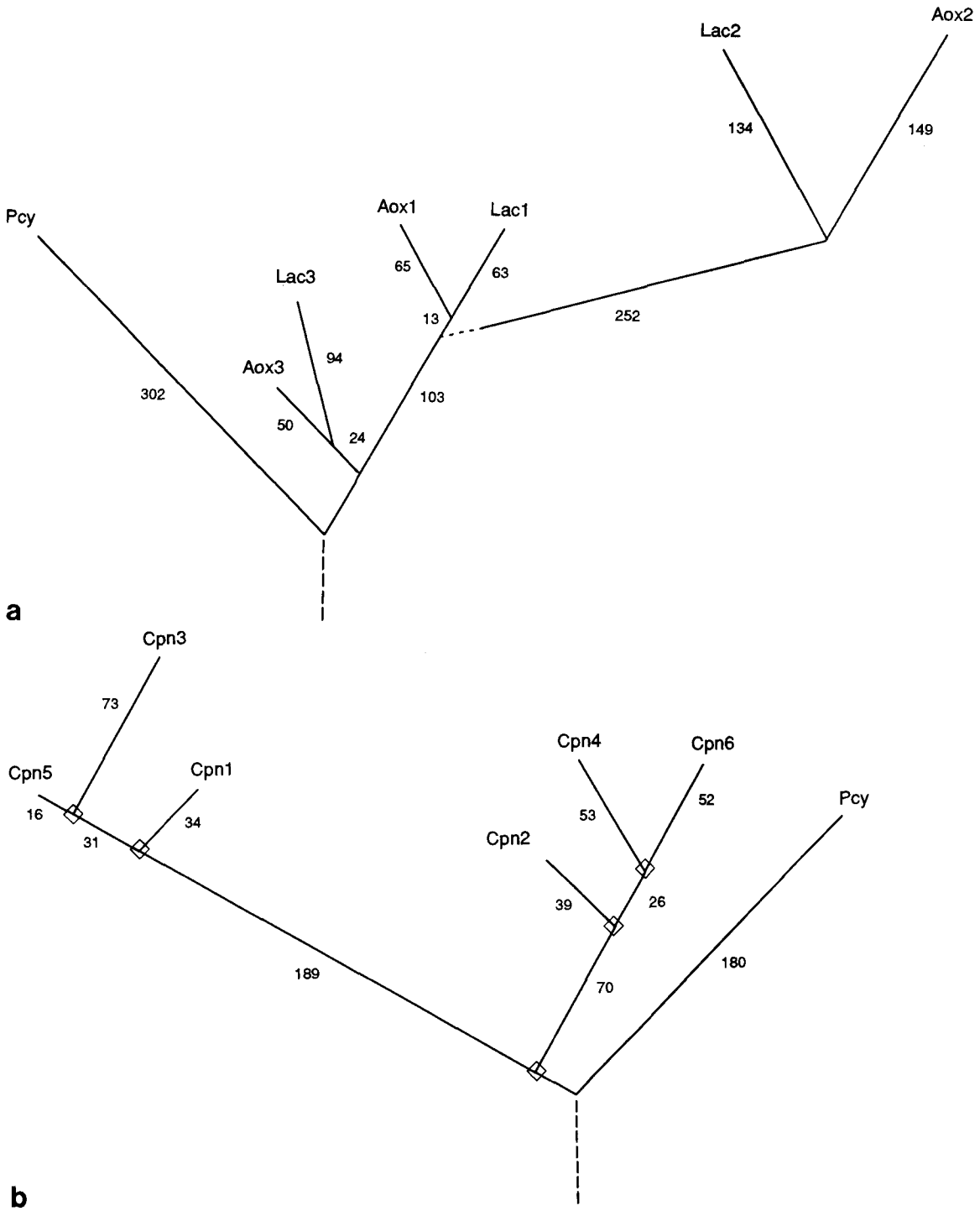
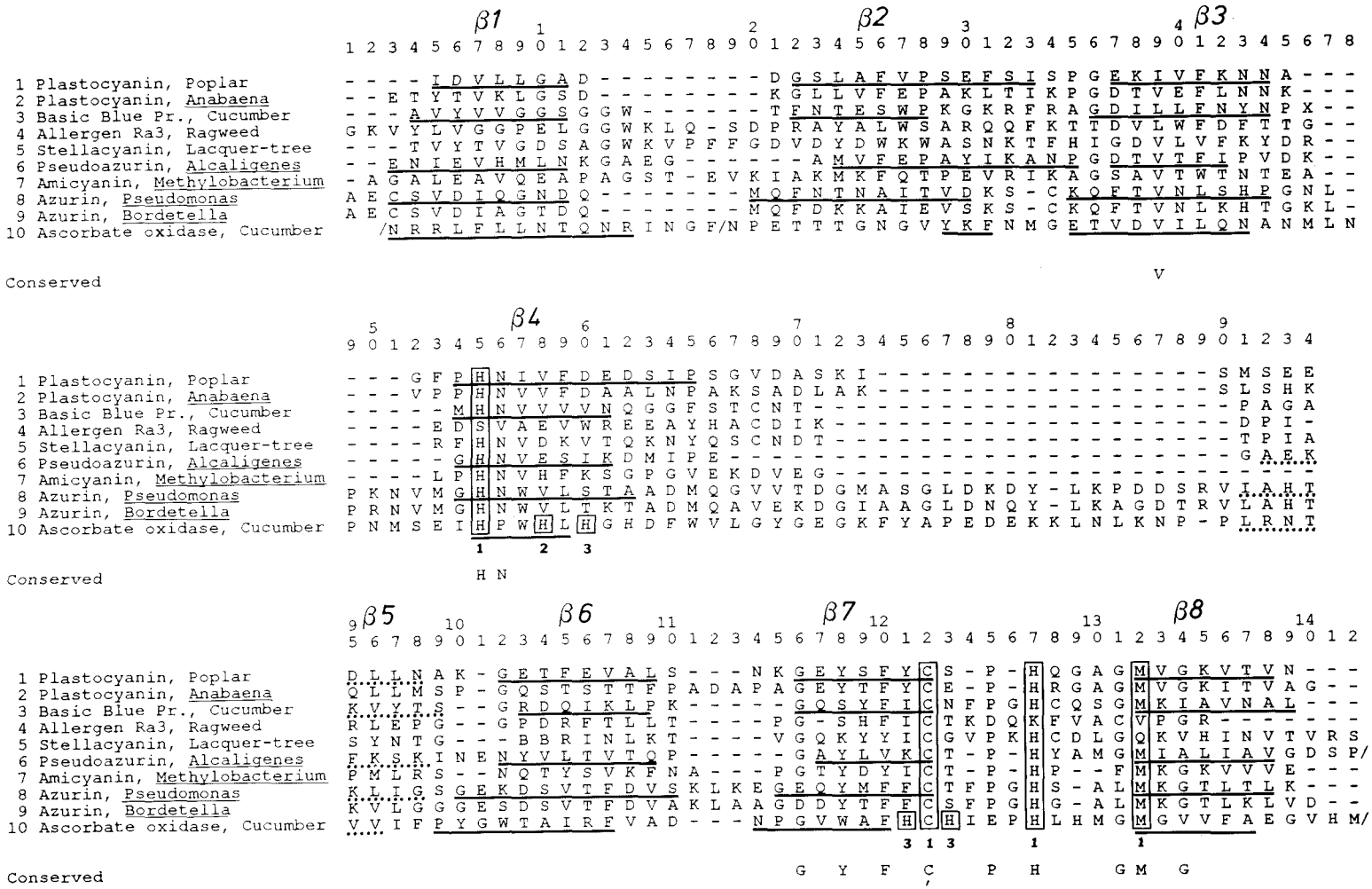


Fig. 4. Phylogenetic trees for the domains in the blue oxidases laccase, ascorbate oxidase, and ceruloplasmin using plastocyanin as an outgroup. Sequences are those in Fig. 2a, and positions omitted from calculating the tree are those given in Fig. 2b. **a** Tree for ascorbate oxidase and laccase domains. The *dashed line* at the junction of the 252-PAM branch indicates a small degree of uncertainty as to the divergence time. Scale: 0.75 mm = 3 PAMs. **b** Tree for the six ceruloplasmin domains; domains 2, 4, and 6 are the C-terminal-like domains and 1, 3, and 5 are the N-terminal-like domains. Boxes indicate duplication events. Scale: 0.75 mm = 2 PAMs.

tree was arbitrarily placed on the branch between azurins and the other small blue proteins, in accordance with a bacterial origin of the small blue proteins.

Evolution of the Coagulation Factors

The alignment of the ceruloplasmin-related domains in the blood coagulation factors V and VIII



a

Fig. 5. **a** Alignment of sequences for the small blue proteins plastocyanin, azurin, amicyanin, and pseudoazurin, the phytocyanins cucumber basic blue protein and stellacyanin, and the homologous plant protein ragweed Ra3 allergen. The ascorbate oxidase C-terminal copper-binding domain has been added as a reference sequence. Beta-strands are indicated by *underlining* for those proteins whose three-dimensional structures are shown, with *dotted lines* indicating suggested beta-5 strands. Conserved residues are those occurring in more than half of the sequences; only Cys at position 122 is present in all sequences. Positions 55, 122, 127, and 132, at which type 1 copper ligands occur in most sequences, are *boxed;* type 2 and 3 ligands at positions 58, 60, 121, and 123 in ascorbate oxidase are also *boxed.* **b** Difference matrix calculated from the alignment; positions omitted from the calculations are 1–3, 14–21, 46–52, 66–94, 100–101, 111–114, 124, 126, and 140–142, leaving a total of 84 positions in the comparisons. Continued on page 52.

Number of differences

		1	2	3	4	5	6	7	8	9	10
1	Plastocyanin, Poplar	.	42	66	74	76	63	61	67	67	73
2	Plastocyanin, Anabaena	51	.	63	72	69	57	59	65	67	73
3	Basic Blue Pr., Cucumber	79	75	.	68	57	67	66	66	69	75
4	Allergen Ra3, Ragweed	88	86	81	.	72	75	74	78	78	78
5	Stellacyanin, Lacquer-tree	90	82	68	86	.	69	71	72	71	76
6	Pseudoazurin, Alcaligenes	75	68	82	89	82	.	65	72	72	74
7	Amicyanin, Methylobacterium	73	70	79	88	85	77	.	70	70	75
8	Azurin, Pseudomonas	81	78	79	93	86	86	84	.	24	70
9	Azurin, Bordetella	81	81	82	93	85	86	84	30	.	72
10	Ascorbate oxidase, Cucumber	87	87	89	93	90	88	89	83	86	.

Percent difference

b

Fig. 5. Continued.

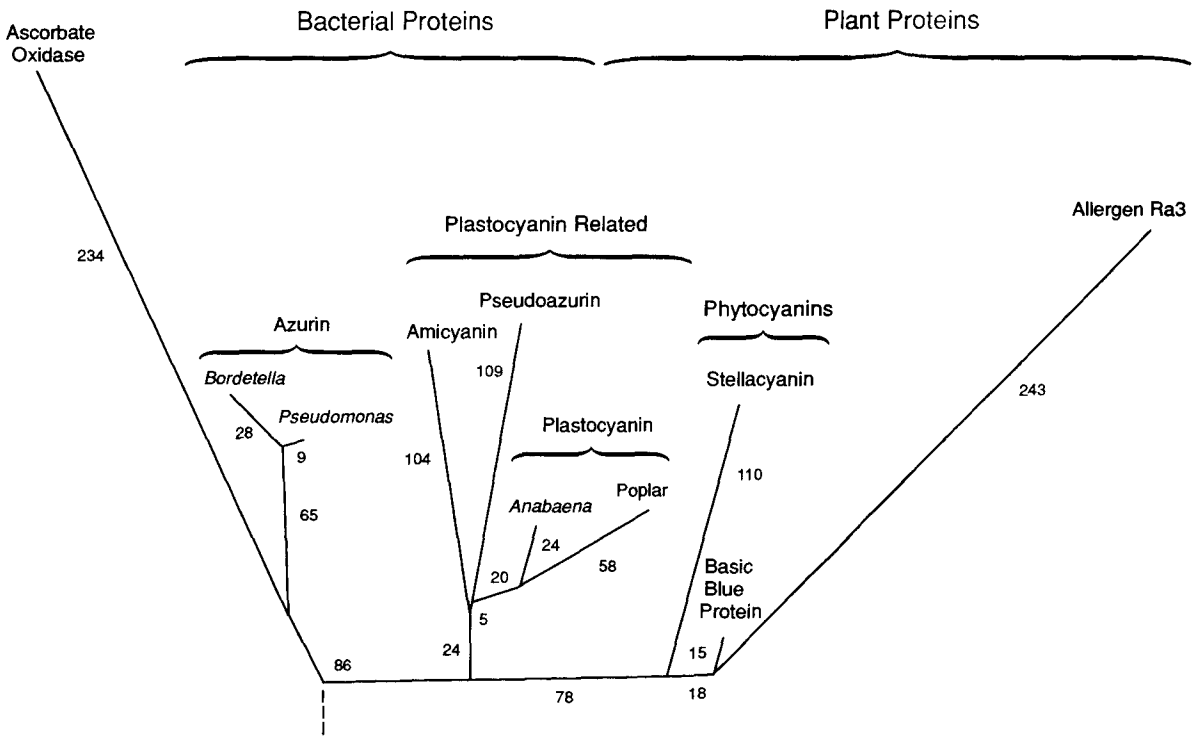


Fig. 6. Phylogenetic tree for the small blue proteins. Sequences are those in Fig. 5a, and positions omitted in calculating the tree are those indicated in Fig. 5b. Ascorbate oxidase C-terminal domain is used as an outgroup, showing the point at which the oxidases branched off. The trunk of the tree has arbitrarily been placed on the branch between the azurins and the other small blue proteins. Scale: 0.68 mm = 2 PAMs.

has already been given in Fig. 2a. The difference matrix (Fig. 2b) shows that in these three proteins domains 1, 3, and 5 are about equally similar as are domains 2, 4, and 6, but there is more than 80% difference between the two sets of domains. For domain 6, the close relationship is also shown by the tree of the C-terminal domains (Fig. 3).

Figure 7a and b shows unrooted trees calculated for the first and second halves of the double domain in factors V and VIII. The trees suggest that the two factors had a common ancestor with the domain structure we see in them today. It also suggests that the two factors diverged soon after ceruloplasmin obtained its six-domain form. It does not show how the additional domains in these proteins were acquired but of course suggests that this happened after the divergence from ceruloplasmin, since the overall domain structure is the same in factors V and VIII.

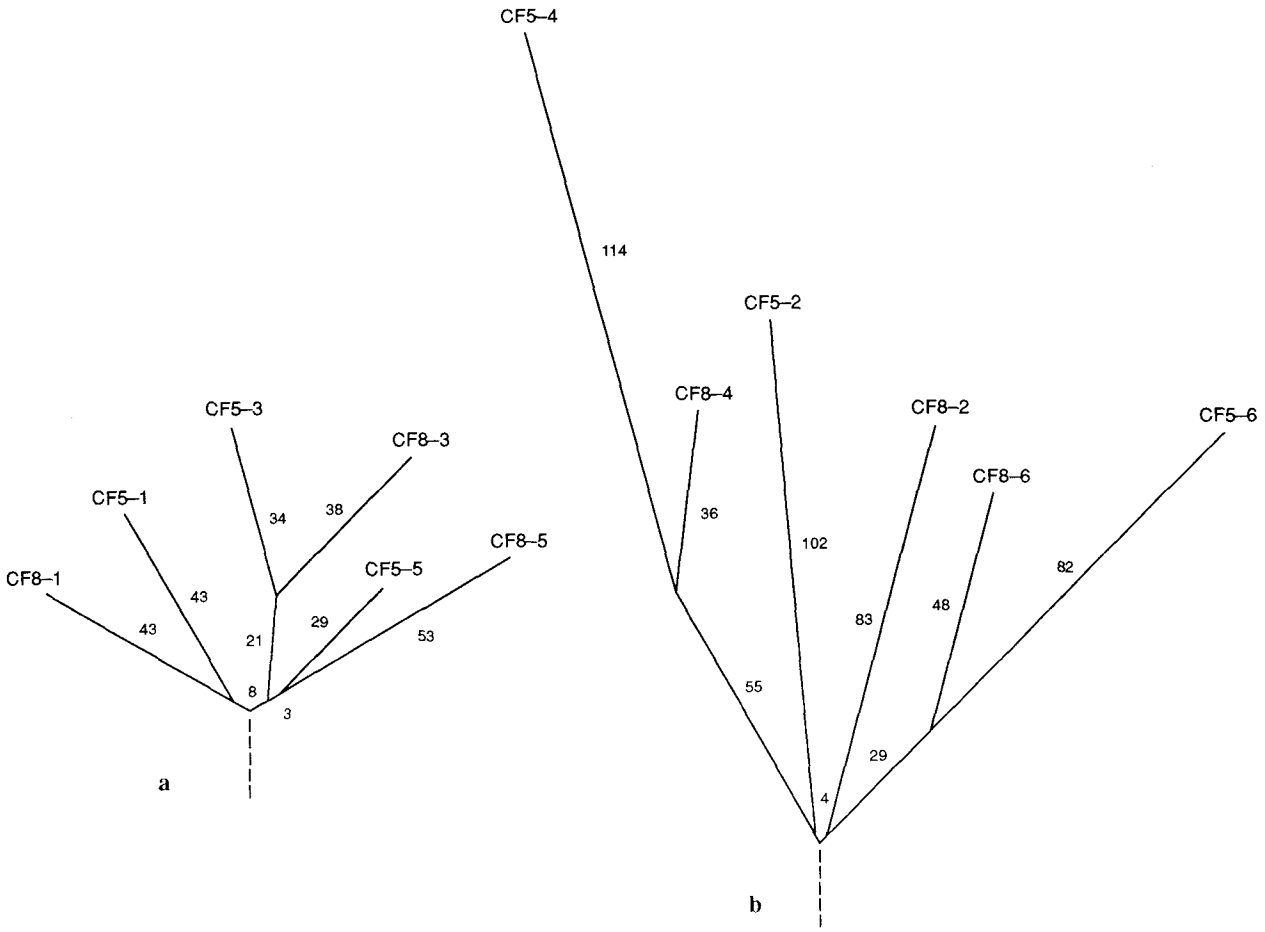


Fig. 7. Phylogenetic trees for the domains in coagulation factors V and VIII. Sequences are those of Fig. 2a, and omitted positions are those indicated in Fig. 2b. **a** N-terminal-like domains 1, 3, and 5. **b** C-terminal-like domains 2, 4, and 6. Scale: 0.66 mm = 1 PAM.

Time of Events

Dates assigned to the events we are discussing necessarily have to be provisional. The earliest event is the divergence of plastocyanin and azurin lines. This corresponds to the divergence of the cyanobacteria from other Eubacteria and has been suggested to have occurred some 3,000 million years ago. (See discussion and references in Rydén and Lundgren 1980.) Chloroplasts are thought to have originated from multiple cyanobacterial symbioses with eukaryotes (Knoll 1992 and references therein), the earliest of which may have occurred before 2.1 billion years ago (Han and Runnegar 1992).

The divergence of plant and animal lines is believed to have occurred about 1,400 million years ago (see, e.g., Schopf 1978; Knoll 1992) and would correspond to the divergence of the ascorbate oxidase–laccase branch from the ceruloplasmin branch. The divergence of the two coagulation factors from each other and from ceruloplasmin would, by inference and using the tree in Fig. 3, be approximately dated to about 500–800 million years ago, assuming that the rate of accumulation of mutations is uniform. This agrees reasonably well with dates calculated from the phylogeny of the serine protease domains in those coagulation factors having this domain, assuming that coagulation factors IX and X evolved together with factors V and VIII with which they interact, as is discussed below. The divergence of coagulation factors IX and X occurred at a time distance about seven times that of the divergence of the bovine and human lines, which is around 70 million years ago (e.g., Goodman 1982). This indicates that factors IX and X, and also factors V and VIII, diverged from each other about 500 million years ago.

Discussion—Mechanisms for Evolution of Protein Complexity

The evolutionary events investigated in the group of proteins related to the blue proteins exhibit the wide spectrum of mechanisms that have been observed so far in protein evolution. Below we will examine these mechanisms for this group of proteins by an-

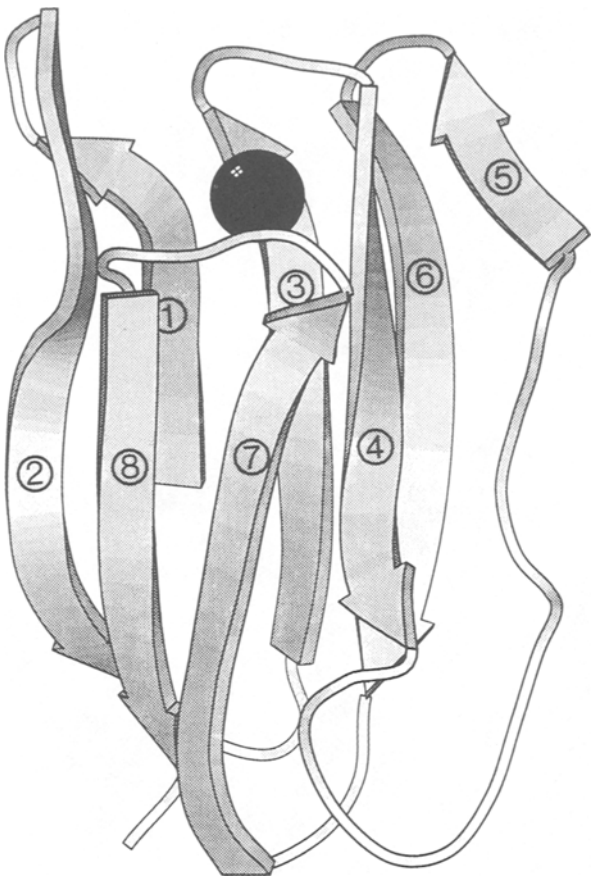


Fig. 8. Structure of poplar plastocyanin showing the beta-sheet structure. The coordinates for oxidized plastocyanin from poplar, as determined by Guss and Freeman (1983), were obtained from the Brookhaven Protein Data Bank. The program MolScript (Kraulis 1991) was used to draw the figure. The molecule essentially forms a double sandwich in which the lower part of strand 2 and strands 8, 7, and 4 form a sheet in the foreground, whereas the upper part of strand 2 and strands 1, 3, and 6 form a sheet in the background. Strand 5, which is mostly irregular, does not belong to either sheet. The copper as shown has its van der Waals radius.

alyzing critically the existing data on sequence and structure as well as some new data, in particular for the coagulation factors.

Size Increase

Domain Enlargement

A direct route to size increase in a globular protein is the insertion of residues in the peptide chain in the domain. To study this possibility all domains with a known structure—that is, four small blue proteins and the three ascorbate oxidase domains, as well as two proteins (amicyanin and stellacyanin) for which alignments could confidently be made to proteins with known structures—were examined. Figure 8 shows a diagrammatic representation of the 3-D structure of poplar plastocyanin, as a reference molecule. Table 2 shows that domain size in the group varies from 97 to 210 residues. Even in the group of small blue proteins the peptide chain varies in length from the minimum of 97 in cucumber basic blue protein to 129 in azurin. The enlargements are due mainly to added length in the loops between strands. The total number of residues in strands varies between 59 and 73, with a mean of 64. The largest variations occur in the loop between beta-4 and beta-5, where azurin has 27 residues that form a large flap that even contains a short helix, whereas pseudoazurin has only a six-residue-long turn. An equally long C-terminal extension is present in pseudoazurin, where it forms a ninth strand. A similar C-terminal extension is present in the phytocyanin from horseradish (Bergman 1980).

The eight strands in each of the three domains of ascorbate oxidase contain a total of 54–63 residues (Messerschmidt et al. 1989), the decrease compared with the small blue proteins mainly being due to a shorter second and fourth strand. The number of residues outside the double sandwich has thus increased from an average of 45 (span 34–60) to 125 (span 81–147) residues. Again a large part of the increase is due to the presence of C-terminal extensions. Two of these are interdomain sequences and could of course have been defined differently. But the dramatic difference compared with the small blue proteins is a large excursion of the chain between beta-1 and beta-2 (22, 60, and 63 residues in the three domains). This new part of the chain forms additional strands that are added on to the existing sheets. There are also several new helices, in particular in the loop between strands 4 and 5 as in azurin. It is clear that the copper-binding strands beta-7 and beta-8 vary little.

The definition of strand used by the various authors varies slightly and makes an exact comparison of the seven structures difficult. It is, however, not clear what a good definition for the purpose of comparing the proteins would be. The possibility of forming hydrogen bonds with a neighboring strand would not be the best criterion when homology forms the basis for the comparison. A rough examination of alternative definitions suggests that these uncertainties do not affect the conclusions to any important extent.

The added residues on the oxidase domains compared with the small blue proteins do not contribute to the hydrophobic core of the domains. This fact might set an upper limit to the increase in protein size by domain enlargement. The changes observed, however, illustrate the possibilities a protein has to change its secondary structure.

Table 2. Distribution of residues within and outside the basic eight-stranded beta-sheet fold in domains of type 1 copper proteins[a]

Protein/ domain	N-ter	β-1	1–2	β-2	2–3	β-3	3–4	β-4	4–5	β-5	5–6	β-6	6–7	β-7	7–8	β-8	C-ter	Total	Total β-sheet
Pcy	0	7	2	12	3	8	3	12	13	4	2	8	3	7	7	7	1	99	65
Amy	1	9	11	10	3	8	3	10	9	3	1	8	3	7	5	7	1	99	62
Azu	2	9	1	11	3	10	8	9	27	8	1	11	4	8	8	7	2	129	73
Paz	0	9	6	12	1	6	4	8	6	7	2	9	2	6	7	7	31	123	63
Cub	0	8	4	7	7	9	2	8	13	5	1	7	1	7	9	8	0	96	59
Sty	0	8	6	12	7	9	2	8	14	4	1	7	2	7	9	8	3	107	63
Aox1	4	7	22	8	3	8	8	4	17	5	3	7	4	7	10	7	11	135	54
Aox2	—	10	60	5	4	8	4	7	19	6	2	8	8	8	5	11	45	210	63
Aox3	—	12	63	3	3	9	11	5	30	6	2	9	3	7	11	6	24	204	57

[a] Protein names were abbreviated as follows: Pcy, plastocyanin; Amy, amicyanin; Azu, azurin; Paz, pseudoazurin; Cub, cucumber basic blue protein; Sty, stellacyanin; Aox1, Aox2, and Aox3, the N-terminal, middle, and C-terminal domains in ascorbate oxidase. The assignment of residues to the beta-strands was based on the published structures in the case of plastocyanin, azurin, pseudoazurin, cucumber basic blue protein, and the ascorbate oxidase domains. For ascorbate oxidase the structure assignments given by Messerschmidt et al. (1989) were used with the exception that (after visual inspection) residues 261–268 instead of 267–268 were assigned to beta-strand 6 in the second domain. The identification of beta-strands as corresponding to beta-1 through beta-8 in plastocyanin was done by visual comparison with the plastocyanin structure. For amicyanin, assignments were based on the alignment with plastocyanin and for stellacyanin they were based on the alignment with cucumber basic blue protein

Domain Duplication

By duplication of a single type of domain, the proteins in the group have increased in size from the single domain of about 100 residues to the three-domain structures in the plant and fungal oxidases and the six-domain structure of ceruloplasmin that has more than 1,000 residues. The sequence of events leading up to these larger proteins, which have been described above, are summarized in Fig. 9.

The trees in Fig. 9 and in Fig. 4 suggest that the small blue protein ancestor increased in size by domain enlargement before the first duplication. Obviously then, the domain duplication in this instance made the evolution of an oxidase active site possible, because the active-site coppers in all three enzymes contain ligands from two different domains. The common ancestral oxidase consisted of only two domains, corresponding to the double domain in ceruloplasmin. The advantage for increasing the size beyond this stage is not clear. A consideration relevant for plasma proteins is that size is important to avoid clearing from the circulation by the kidneys. In fact almost all plasma proteins are larger than albumin, which with its 587 residues is on the same order of size as the four-domain structure of the ceruloplasmin ancestor. For several plasma proteins, such as the immunoglobulins, plasmin, and some coagulation factors, it has been shown that multiple domains are involved in additional functions and contain new binding sites. This might also be a reason for the large size of ceruloplasmin. Indeed, the binding of ceruloplasmin to a receptor has been described (Omoto and Tavassoli 1990).

After the duplication of the domains beyond the two-domain stage, copper-binding sites were lost. In the N-terminal domain in all three oxidases the type 1 copper site has disappeared, as shown by the structure of ascorbate oxidase; this is consistent with chemical analysis and with the mutation of the copper-binding residues in the N-terminal domain (Fig. 2a). In ceruloplasmin the type 1 site is still present in domains 2 and 4, which are most closely related to the C-terminal catalytic domain. The retention of copper sites in these two cases might reflect the fact that this copper is buried in the interior of the molecule and is not lost unless a considerable change in structure occurs.

Domain Requisition—the Discoidin Domains

In coagulation factors V and VIII, two domains not related to ceruloplasmin but to discoidin are present at the carboxyl end, as first reported for factor VIII by Vehar et al. (1984) and Wood et al. (1984) and later for factor V by Kane and Davie (1986). We will call this mechanism, by which an exogenous type of domain is incorporated into a protein molecule from elsewhere in the genome, domain requisition.

Discoidins I and II, two families of lectins, are among several cell-surface proteins produced by the cellular slime mold *Dictyostelium discoideum* during the aggregation phase of the life cycle (Poole et al. 1981). Of the four genes of the discoidin I family, two have been completely sequenced and the other two partially sequenced. The protein sequences are highly conserved (less than 2% difference), but probably no two are identical. It has been proposed that the tetrameric discoidin I molecule, which has a high affinity for galactose and modified

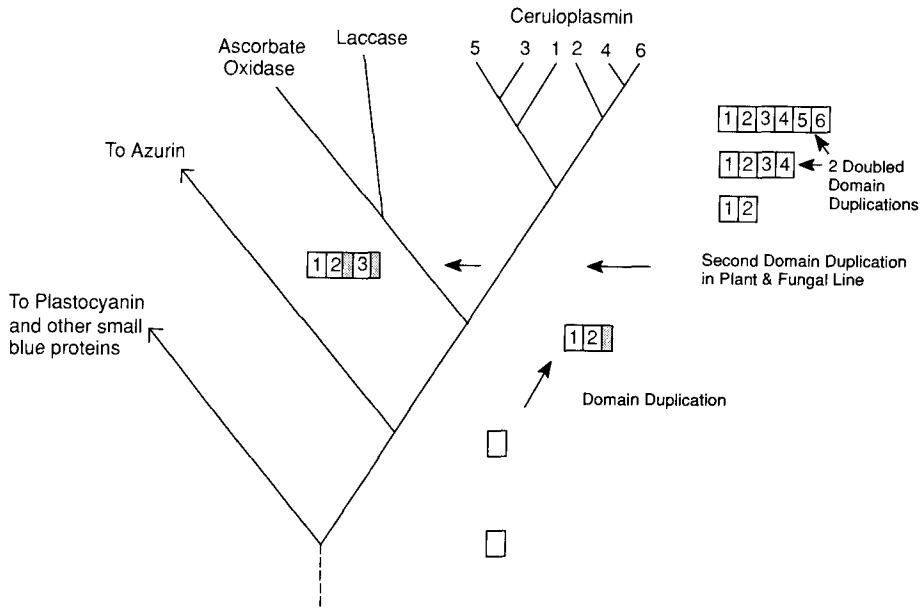


Fig. 9. Evolution of the domain structure in the blue oxidases based on the phylogenetic trees of Figs. 3, 4, and 6. *Shaded areas in boxes* represent C-terminal extensions of domains.

galactose residues, is an integral component in the cell–cell cohesion process. (See Poole et al. 1981 and references therein.)

In the coagulation factors, each of the discoidin-like domains, approximately 160 residues in length, corresponds to the amino-terminal 158 residues (approximately two-thirds) of a discoidin I sequence. More recently a similar pair of domains has been reported in a 51.5K glycoprotein (MGP) present in the milk-fat-globule membrane produced by mouse mammary epithelial cells (Stubbs et al. 1990). These domains are closer to those in factors V and VIII than to those in *Dictyostelium*. An alignment derived from all pairwise ALIGN comparisons of the factor V and factor VIII domains, the milk glycoprotein, and discoidin I-C is shown in Fig. 10a. It differs in some minor details from that of Kane and Davie (1986) and that of Stubbs et al. (1990). All sequences except the second factor VIII domain contain one NXS/T potential glycosylation site. There are 11 invariant sites, 12 positions at which aromatic residues predominate, and several short regions where basic residues cluster. Identities and similarities are clustered in five regions, the longest being at positions 121–164. The discoidin I sequence contains an R-G-D motif (alignment position 86–88) capable of binding to a specific cell-surface receptor; the discoidinlike domains in the coagulation factors lack this peptide (Kane and Davie 1986).

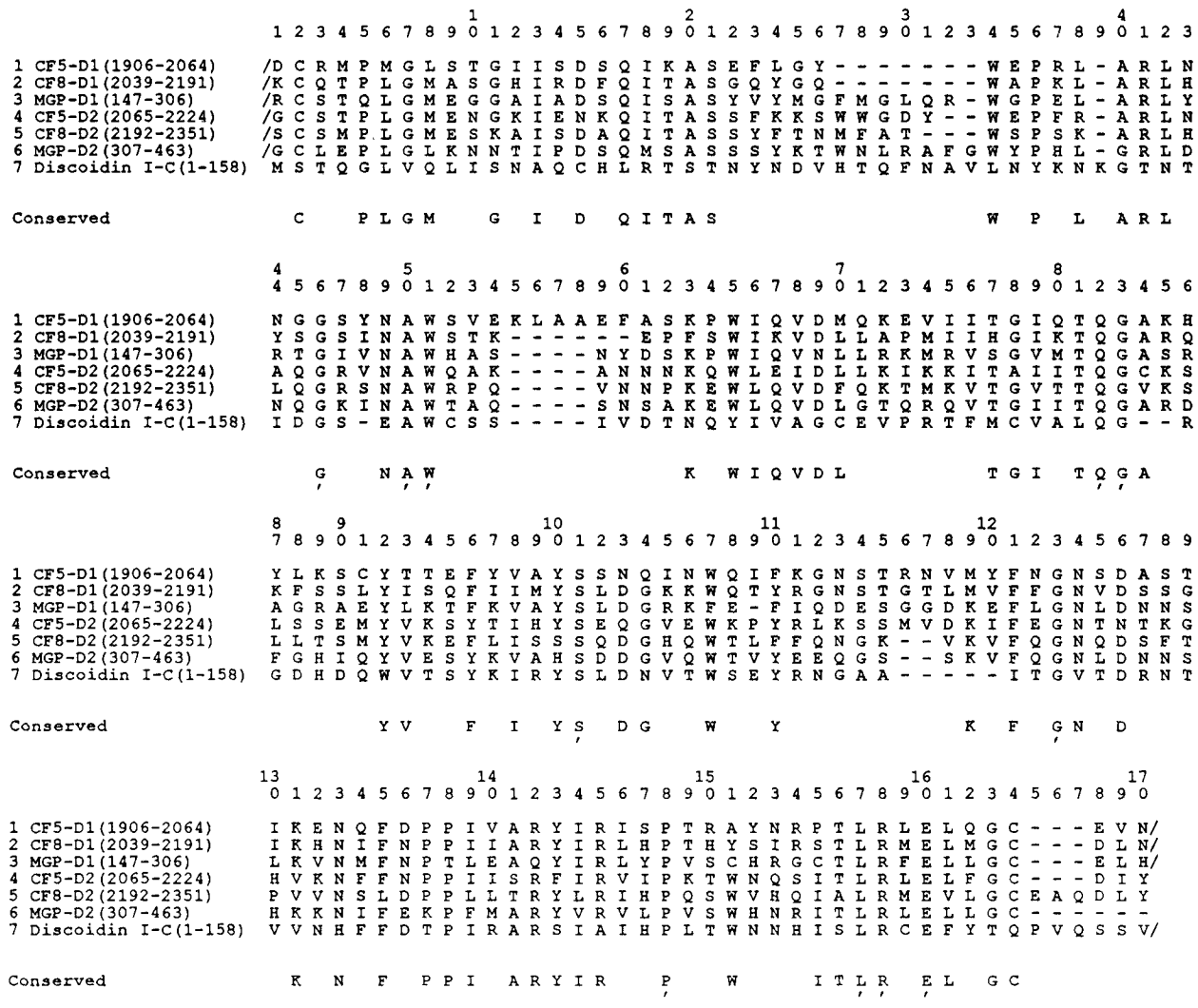
The discoidinlike domains in factor V and factor VIII are only very distantly related to discoidin I, as shown by the values in Fig. 10b and the tree in Fig. 10c. In the alignment (Fig. 10a) the domains are considered individually in order to emphasize the conservation of residues; while the topology (Fig. 10c) reflects the alignment, the branching pattern may not quite correspond to evolutionary events. The time of divergence for lines leading to cellular slime molds and to animals may have been between 1.5 and 2 billion years ago (Knoll 1992). We suggest that at some point since then, in the animal line at least, the DNA for this domain doubled; the doubled domain would then be available for incorporation into other genes by exon shuffling. It appears that this occurred in the ancestral coagulation factor gene after its divergence from ceruloplasmin, but before the gene duplication giving rise to separate genes for the two factors.

Segment Elongation—the Activation Domains

Both coagulation factors contain a region about 900 residues long between the second and third ceruloplasmin-related double domain (domains 4 and 5). This region has cleavage sites in the amino and carboxyl ends that, when cleaved by thrombin, factor IX, or factor X, will generate active factors V and VIII, respectively, while the region is dissociated from the rest of the molecule. Since this is the classical way for an activation peptide to function we will call this region the activation domain.

In a general sense the regions in the two factors are similar (Kane et al. 1987): they are of similar length; they both contain a large number, about 20, of potential glycosylation sites with the sequence NXS/T; they both have a predominance of acidic residues in the N-terminal part, followed by a basic region, and are quite acidic in their C-terminal part. They also both have conspicuous PEST regions (Rogers et al. 1986), known to make the region sensitive to proteolysis.

When comparing the two sequences, close simi-



a

ALIGN Scores

	CF5 D1	CF8 D1	MGP D1	MGP D2	CF5 D2	CF8 D2	DISC I-C
CF5 D1	—	17	19	17	24	20	5
CF8 D1	52	—	19	18	21	22	9
MGP D1	61	57	—	18	18	16	5
MGP D2	61	61	56	—	21	21	8
CF5 D2	63	58	64	57	—	19	6
CF8 D2	60	59	61	58	56	—	6
DISC I-C	83	79	83	79	78	81	—

b Percent Difference

Fig. 10. **a** Alignment of discoidin domains found in the *Dictyostelium* discoidins, in coagulations factors V and VIII, and in the milk-fat-globule membrane protein. Conserved residues are those occurring in more than half of the sequences; totally conserved residues are marked. **b** Difference matrix calculated from the alignment. In the lower left half are shown percent differences; positions omitted in calculating the matrix and tree are 27–33, 55–58, and 165–167. In the upper right half are shown the alignment scores, calculated with a bias of 6 and a gap penalty of 6 or 10 (with discoidin I-C). Continued on page 58.

larities were not found. However, there is, as has been pointed out by Jenny et al. (1987) and Kane et al. (1987), clear evidence for internal repeats in factor V. There is a nine-residue peptide that is repeated very closely 31 times and less closely up to 15 additional times. Dot-matrix plots (Fig. 11a and 11b) of a part of the activation domain in the two proteins demonstrate this dramatically. This nine-residue peptide cannot be seen in the corresponding region in factor VIII. In factor VIII there is a weak indication of a repeat of a 20-residue sequence, as shown by the regular appearance of the glycosyla-

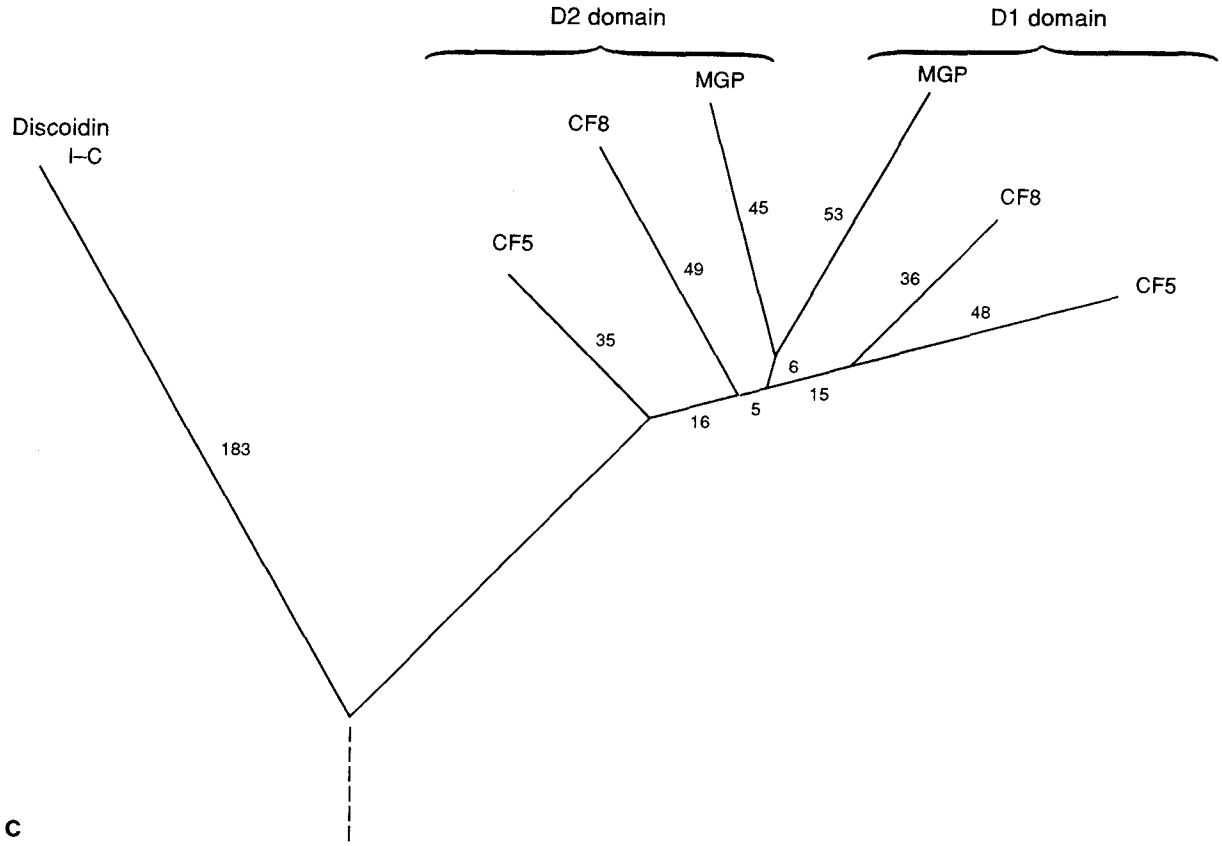


Fig. 10. Continued. **c** Tree of the mammalian discoidinlike domains and discoidin I-C. Scale: 0.75 = 1 PAM.

tion site. The nine-residue peptides, with a consensus sequence of T/N/PLSPDLSQT, also have a structure compatible with a possible origin from a duplication of a four-residue peptide L-S-P/Q-D/T.

The mechanism for size increase in this part, which at least in factor V is due to repeated duplications of a short peptide segment, we will call segment elongation. It appears that the glycosylated regions have developed into large areas of the factor V and factor VIII structure independently of each other. If this is so the elongation of an originally shorter stretch occurred independently. It is also likely that selection pressure for the sequences in this part is weak and that sequences change rapidly, and thus their evolutionary history is more difficult to trace. We found no evidence for an origin of the shorter peptides elsewhere in the ceruloplasmin-related domains or the discoidin domains. The possibility exists that the activation domain originated from a change in the exon–intron border of the ceruloplasmin gene, a suggestion that will be possible to investigate when its genomic DNA sequence becomes available. In this context it should be noted that in factor V the entire activation domain is contained in a single unusually large exon. This is also consistent with the suggested mechanism of segment elongation. The structure of the gene for factor VIII is not yet known.

Subunit Aggregation

Subunit structure formation is not an important mechanism for size increase in the group of proteins discussed. Ascorbate oxidase, however, associates as a two-subunit protein, in which one metal ion, most likely copper, bridges the two subunits.

Weak evidence exists that subunit II of mitochondrial cytochrome c oxidase is related to the blue proteins (Rydén and Lundgren 1980; Holm et al. 1987; Saraste 1990). Mitochondrial cytochrome c oxidase consists of three membrane-bound and mitochondrially coded subunits and a variable number of soluble subunits. It has a complex function, as it both transfers four electrons from cytochrome c to oxygen and pumps protons through the mitochondrial membrane when doing so. Subunit II contains two copper-binding sites that are spectroscopically quite distinct from the site in the blue proteins. There is some weak sequence similarity between subunit II and azurin. For a more reliable comparison the detailed structure of cytochrome c oxidase will be needed. Cytochrome c oxidase is an outstanding example of the role of subunit aggregation

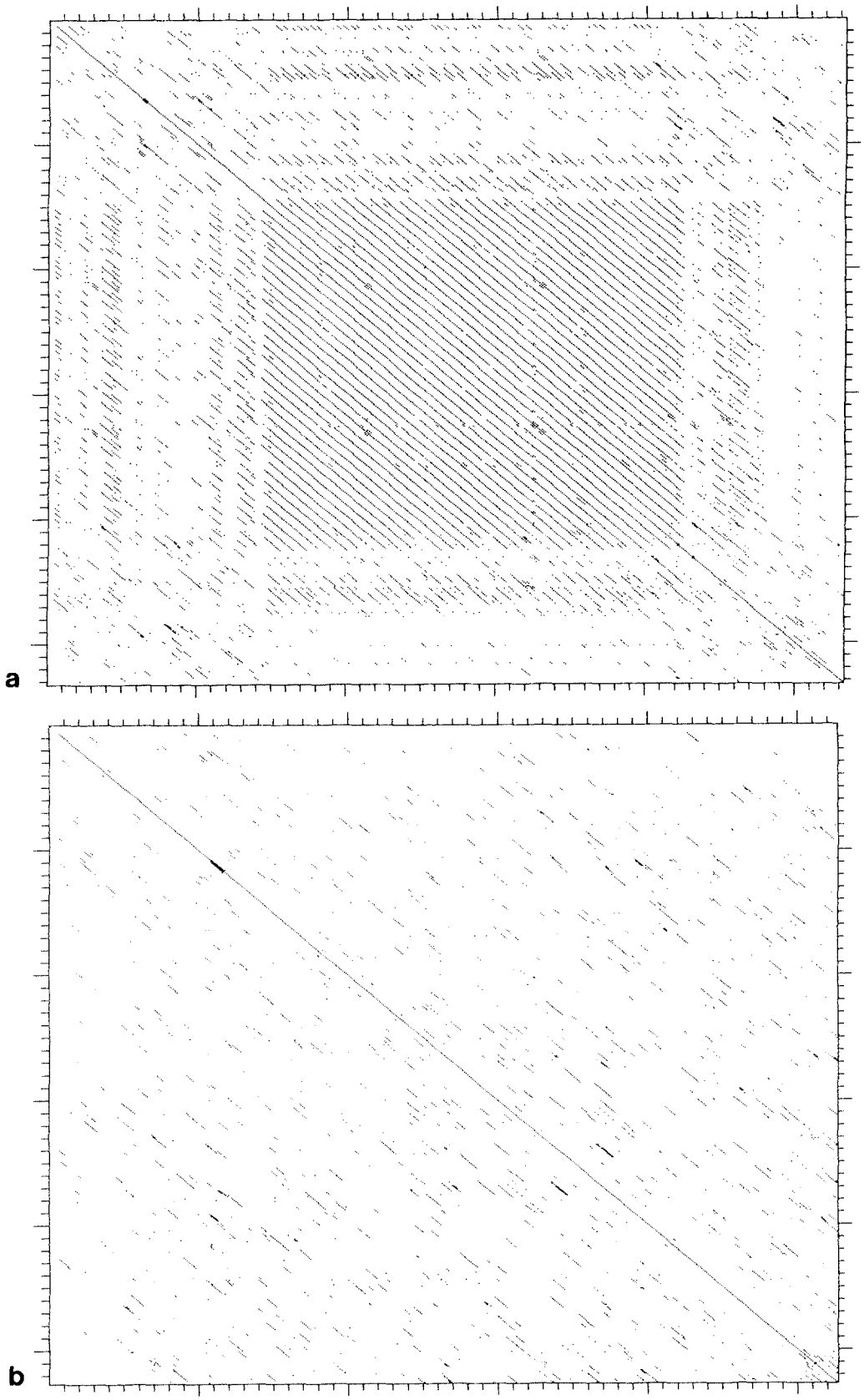


Fig. 11. Dot-matrix plots for the comparison of coagulation factors V and VIII with themselves to show internal homologies. **a** Matrix for factor V positions 1,040–1,075, using a window size of 9 and a minimum score of 10. Internal repeats, especially the nine-residue repeats, are shown as *diagonal lines* parallel to the main diagonal. **b** Matrix for factor VIII positions 700–1,230, using a window size of 12 and a minimum score of 12. Several areas with a few short repeats are seen.

in evolving both size and functional complexity in a group of proteins.

Glycosylation

For several of the proteins in the group, glycosylation makes an important contribution to increase in molecular size. Among the small blue proteins only the plant-derived phytocyanins are glycosylated. Some of these, however, contain up to 50% carbohydrate. Stellacyanin contains 50% carbohydrate (Bergman et al. 1977), as does umecyanin (Bergman 1980). The role of the carbohydrate in these proteins is not well known, although it has been shown that it contributes to protection against proteolysis.

The larger oxidases and the coagulation factors are all glycosylated. A few glycosylation sites exist in the domains of the oxidases and in ceruloplasmin; these have been examined in some detail (Rydén 1984b). These account for around 10% by weight of the proteins. In general, they are not conserved, and most occur in loops between beta-strands, i.e., at the surface. A more impressive amount of glycosylation is found in the activation domain in coagulation factors V and VIII. About 20 potential NXS/T glycosylation sites, as well as O-linked carbohydrate in factor V (Kane et al. 1987), are present, which would permit 30% or more carbohydrate. Again the carbohydrate might protect this region against proteolysis until after the removal of the domain in activation.

Functional Diversification

Catalytic Properties

All of the small blue proteins have the capacity to transfer one electron from a donor to an acceptor. The redox potential varies considerably from 180 mV for stellacyanin to the unusually high 345–390 mV for plastocyanins (for review see Rydén 1984a), thereby conferring a variation in functional properties and adaptation to the varying properties of acceptors and donors in the different cells. Plastocyanin in addition has the capacity to bind to the thylakoid membrane, by a highly negatively charged spot on the surface.

The blue copper proteins demonstrate an unusually dramatic case of the evolution of a complex enzymatic function from a simple origin. As the blue oxidases evolved from the small blue proteins, three new copper-binding sites formed close to the already-existing type 1 copper site. This involved the formation of eight new ligands in the protein, four of these being in the same domain as the type 1 copper, while four are found in the N-terminal domain. According to the evolutionary trees, these four copper sites were present in the two-domain ancestral oxidase. In ceruloplasmin the three most recently acquired sites were again lost in the two new double domains after triplication occurred, while the type 1 site retained its copper. This scheme of events accounts for the six coppers observed in ceruloplasmin.

The blue oxidases have the capacity to catalyze the four-electron transfer to oxygen from a reducing substrate. In laccase and ceruloplasmin the uptake of electrons is similar to that in the small blue proteins, i.e., a one-electron uptake to type 1 copper converting the reducing substrate into a free radical. Oxygen is reduced by transfer of pairs of electrons from the "new" trinuclear copper center, which also is oxygen-binding. Several different types of substrate specificity have evolved within the group. While ascorbate oxidase has a remarkably narrow specificity for ascorbate, ceruloplasmin and laccase have no proven capacity to bind the reducing substrate at all and thus do not form an enzyme–substrate complex, with the possible exception that ceruloplasmin might bind iron ions. These enzymes seem to be able to reduce all compounds that have a redox potential lower than the type 1 copper.

The two coagulation factors V and VIII have not been shown to have any catalytic properties. Again the copper sites have been lost during evolution from ceruloplasmin. Factor V has only a single copper, presumably in a type 1 copper site in the last domain (domain 6); sequence comparison suggests that factor VIII retains copper at two of the type 1 sites.

Binding Properties

The small blue proteins interact with several substances in connection with electron transfer, as mentioned above. The pollen protein allergen Ra3 has lost copper entirely and its function depends on its capacity to bind to specific receptors on the pistil when the pollen germination tube is penetrating the pistil. It is thus an example wherein function is entirely connected with binding. The binding properties of the blue oxidases, as has been mentioned above, are part of their enzymatic properties, although it should be added that the subject is not well studied and a recent report on a ceruloplasmin receptor has been published (Omoto and Tavassoli 1990).

The functions of coagulation factors V and VIII, as has been said, depend on their capacity to interact with several other macromolecules, in particular the proteases (factor IXa and Xa) they activate, as well as von Willebrand factor and protein C, which

latter protein acts as an inactivator of factors V and VIII. The locations of the binding sites are not known, but they must reside in the ceruloplasmin-related domains and/or the discoidin domains, because the activation domain is dissociated during activation.

The lectinlike activity of the discoidin domain ancestor suggests that these domains are responsible for binding. However, there is no solid experimental evidence concerning such a possible functional role (such as a lectinlike activity). In the discoidins, even single-residue mutations may reduce or abolish function (Poole et al. 1981); therefore, any role of the discoidinlike domains in the two factors may be only remotely similar to lectins or may be entirely new. However, the complex of factor VIII with von Willebrand factor possesses several activities including platelet adhesion-promoting activity (Vehar et al. 1984); one is tempted to speculate that the discoidinlike domains could be involved in the complex formation or in the platelet adhesion-promoting activity.

Regulatory Properties

Coagulation factors V and VIII are the only proteins in the group that have developed regulatory properties. They are activated in conjunction with the coagulation cascade, as the activation domain is released after proteolytic cleavage by factor IX or X or thrombin. The evolution of this domain has been dealt with above. In several other groups of homologous proteins regulatory properties are of much greater relevance.

Evolution of a Protein System: the Coagulation Cascade

Coagulation factors V and VIII are part of the system of proteins that form the coagulation cascade (Hedner and Davie 1989). The homologies within this system allow us to suggest in some detail how this complex system evolved and where the two ceruloplasmin-related factors became part of the system.

Factors V and VIII are found in two consecutive steps in the cascade as they form complexes with the two proteases factors IX and X. These two in turn are similar in their general architecture. They consist of peptide chains of 476 and 461 residues, respectively. Both have a serine protease domain (residues 223–476 and 227–461) preceded by three shorter duplicated regions containing the EGF-like domains of types A and B. Also the coagulation factors XI and XII, as well as prothrombin, protein C, and factor VII, contain the serine protease domain, and the evolution of all these factors can be related.

An alignment and a difference matrix of the serine protease domain of the proteins mentioned were constructed (Fig. 12a and 12b) and an evolutionary tree was calculated (Fig. 13). The proteins are all monophyletic. The tree shows them to have diverged from each other during a relatively short time span. The most recent event in the tree is the duplication forming factors IX and X from a common ancestor, which we will call factor IX–X. Because factors V and VIII likewise developed from a common ancestor, we suggest that the entire system of factors V–X and VIII–IX is the result of a single duplication. This suggestion is supported by the observation that the genes for factors VIII and IX are located close to each other on the human X chromosome at positions Xq28 and Xq26.3–27.1, respectively, suggesting that the part of the chromosome with these two genes duplicated.

The divergence of prothrombin from the ancestral factor IX–X is the next-most-recent event in the tree, assuming mutation rates to be constant. The event before that seems to be the divergence of the two earlier factors in the cascade—that is, factors XII and XI. These are single-chain proteins of 602 and 625 residues, respectively, each with a serine protease domain (positions 360–602 and 384–618), preceded by several repeats of smaller domains that are different in the two proteins.

The information in the phylogenetic tree (Fig. 13) can be represented as a series of pictures of the coagulation cascade at different stages of evolution (Fig. 14). Each duplication in the tree gives the system one more amplification step. Counting the five consecutive steps that involve the factors discussed so far, there are four consecutive duplications. Each of the serine proteases that evolved in this way achieved additional domains. After the third duplication the ancestral factor V–VIII achieved a role in the system; it then evolved into the two separate proteins after the fourth duplication, which also involved the ancestral factor IX–X.

The tree includes both human and bovine sequences. The suggested date for the divergence of these lines is 70 million years ago, which then can be used to date the other events in the tree, assuming the mutation rates to be similar throughout the time considered. It thus suggests that the duplication of the ancestral factor V–VIII occurred close to 500 million years ago.

The evolution of the coagulation cascade is an instance of the evolution of a whole system of interactive proteins, each with new catalytic and/or regulatory properties. The ceruloplasmin-derived proteins became a part of this, in a way we do not know, and then developed with it.

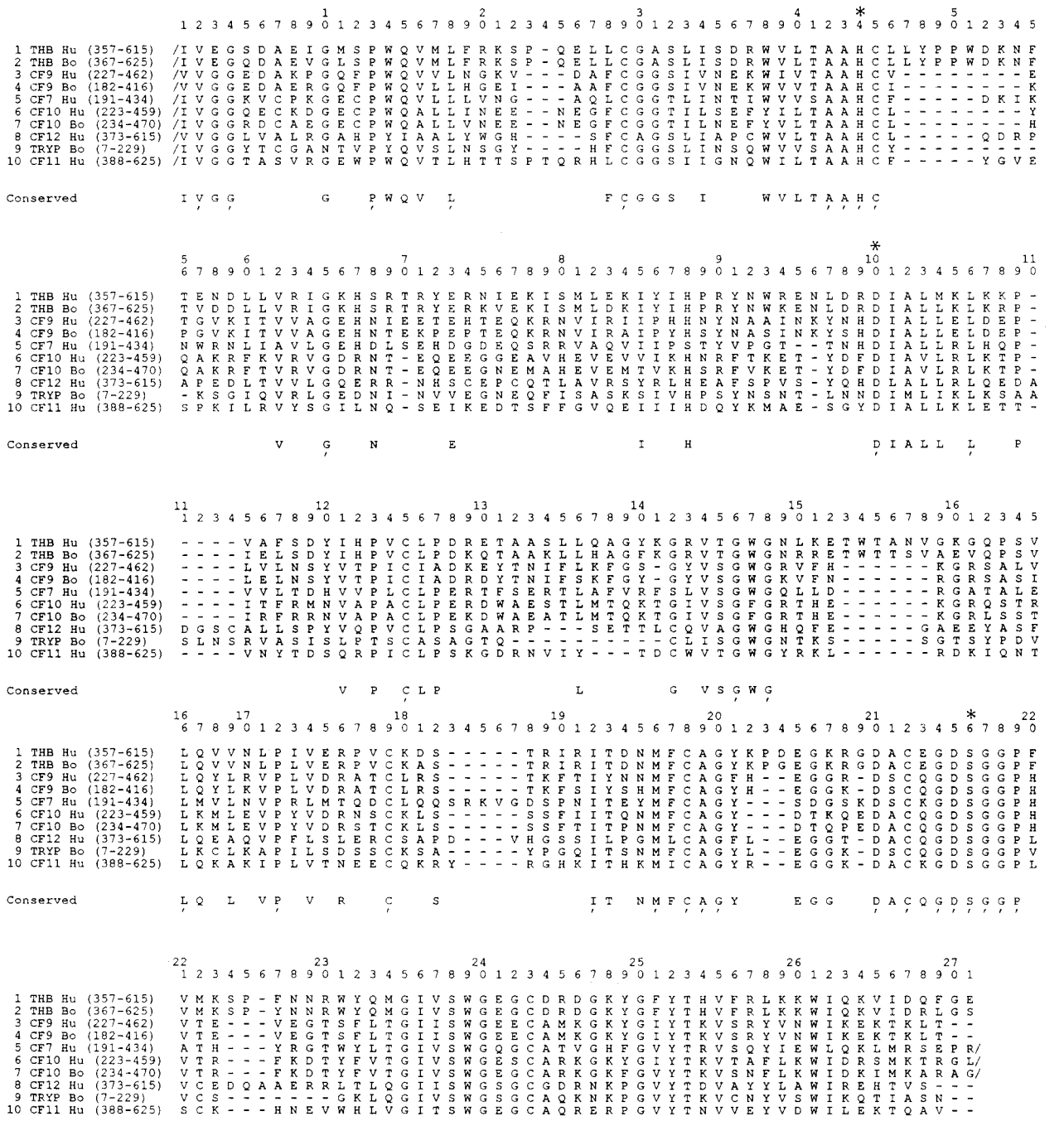


Fig. 12. **a** Alignment of serine protease domains in human and bovine coagulation factors VII, IX, X, XI, and XII and prothrombin (THB), with bovine trypsin (TRYP) as a reference sequence. The three active-site residues are indicated by *asterisks*. Conserved residues are indicated as in Fig. 10a. Continued on page 63.

Conclusions

In this paper we have attempted to bring together and analyze the different kinds of data available for the blue copper proteins and related proteins in order to trace their evolutionary history through at least three separate main lines—the small blue proteins, the blue oxidases, and two coagulation factors. In doing so we have defined the operation of several mechanisms that led to increased size, increased complexity of structure, multifunctionality, and diversity in these proteins. These mechanisms can be related both to organismal evolution and to the evolution of other groups of proteins.

Number of differences

	1	2	3	4	5	6	7	8	9	10
1 THB Hu (357-615)	.	29	146	147	155	141	142	159	152	148
2 THB Bo (367-625)	12	.	143	145	155	143	142	154	155	146
3 CF9 Hu (227-462)	62	60	.	36	139	136	134	148	150	152
4 CF9 Bo (182-416)	62	61	15	.	137	138	133	145	150	149
5 CF7 Hu (191-434)	65	65	59	58	.	148	144	161	154	154
6 CF10 Hu (223-459)	59	60	57	58	63	.	38	165	158	161
7 CF10 Bo (234-470)	60	60	57	56	61	16	.	159	153	158
8 CF12 Hu (373-615)	67	65	63	62	68	70	67	.	147	147
9 TRYP Bo (7-229)	64	65	64	64	65	67	65	64	.	149
10 CF11 Hu (388-625)	62	62	64	63	65	68	67	63	64	.

Percent difference

b

Fig. 12. Continued. **b** Difference matrix calculated from the alignment; positions omitted in calculating the matrix and tree are 23–24, 47–54, 96, 110–114, 153–158, 183–187, 203–204, 224–226, and 270–271.

In organismal evolution the phenomena we have described have their exact counterparts: living organisms have evolved from the simple small bacteria to the large complex multicellular animals and plants. Even more so: each separate group of animals and plants in turn evolved to even larger species, increased functional complexity, and increased diversity—i.e., an increased number of forms or species in the group (Bonner 1988). However, even if we accept that protein evolution is a prerequisite for organismal evolution, the relationship between the two is not straightforward. Organisms need not have larger proteins to be larger, it is not clear if they require multifunctional proteins to themselves be functionally complex, and they certainly do not need to have many different proteins in order to evolve many different forms.

Nor is there a parallelism in terms of selection forces. By size increase organisms can escape to new niches and evade predators, by multifunctionality they become more independent of their environment, and by developing diverse forms they exploit and create an ever-increasing number of different niches. The only obvious parallel phenomenon is that new organisms form the environment to which others must adapt. Also, new proteins form a molecular environment to which new—paralogous—proteins must adapt.

Still, it is obvious that the reason for the observed evolution of protein complexity must be found on the organismal level. What then are the selection pressures? A possible connection between protein and organismal evolution—not only for the copper proteins—is that proteins, by being larger, become more functionally complex and have more interactions with other molecules in their environment. This in turn creates new possibilities for diversity. It might be that the only selection force on the organismal level that drives protein evolution in the direction we have seen is the requirement for functional complexity.

Many groups of proteins exhibit the phenomena we have described. The serine proteases, in particular the coagulation and complement factors, form a well-studied example, as do the immunoglobulins and related proteins. However, the copper proteins might be unusual in several respects. They describe an evolutionary history that spans nearly the entire period of life on Earth and involves prokaryotes, fungi, plants, and animals, whereas most of the other groups studied are confined to evolution of animal life. Furthermore, among the small blue proteins we find the nonenzyme ancestor of an unusually complex group of enzymes—i.e., the blue oxidases—which in addition has an unusually complex active site structure. Subunit II of cytochrome oxidase would be a second case if the proposed homology to azurin is confirmed. The small blue proteins make up one of three different types of ancient electron-transfer molecules, the other two being cytochromes and ferredoxins. Although many enzymes contain a heme group or nonheme iron sulphur centers, there is not so far to our knowledge any demonstrated homology to these other electron-transfer proteins. The observed homology between ferredoxins and the noncatalytic subunit in the two homologous enzymes succinate dehydrogenase and fumarate reductase from *Escherichia coli* (Cole et al. 1982; Darlison and Guest 1984) is not entirely comparable.

From the time of their appearance, the blue copper proteins have been connected to the development of the basic conditions of life: plastocyanin in oxygenic photosynthesis and the development of an oxygen-containing atmosphere, subunit II of cytochrome oxidase in the evolution of oxygen respiration. They are a striking demonstration of how the evolution of proteins changed the direction of the evolution of life on Earth.

Acknowledgments. L.G.R. wishes to thank Drs. Per Kraulis and Ylva Lindquist of the Department of Molecular Biology, Uppsala University, and the Swedish University of Agricultural Sciences for the use of the MolScript program and for help in using the program, respectively. We are grateful to Dr. Markku Saloheimo, VTT, Biotechnical Laboratory, Espoo, Finland, for unpublished sequence data for *Phlebia radiata* laccase. We thank Dr. Winona C. Barker for her critical reading of the manuscript and for helpful discussions and Drs. Lai-Su Yeh and Geetha Y. Srinivasarao for assistance with computer analyses. We also appreciate the valuable assistance of Kathryn E. Sidman in designing most of the illustrations and of James K. Bair and Angela K. Murakami in producing the manuscript. This study was supported by a grant from the Natural Science Research Council in Sweden to L.R.; L.T.H. was partially supported by grants LM05206 and GM45838 from the National Institutes of Health.

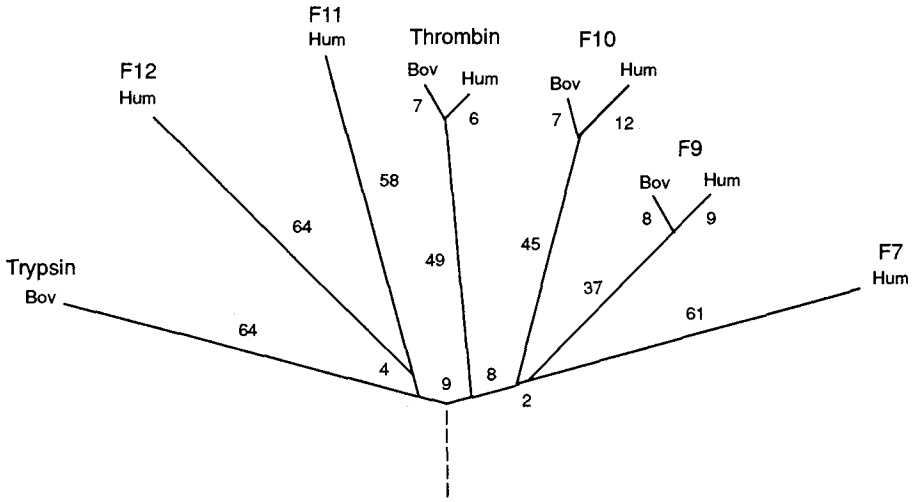


Fig. 13. Tree for the protease domains of the coagulation factors of Fig. 12, and bovine trypsin. The trunk was placed arbitrarily on the longest internodal branch. Scale: 0.74 mm = 1 PAM.

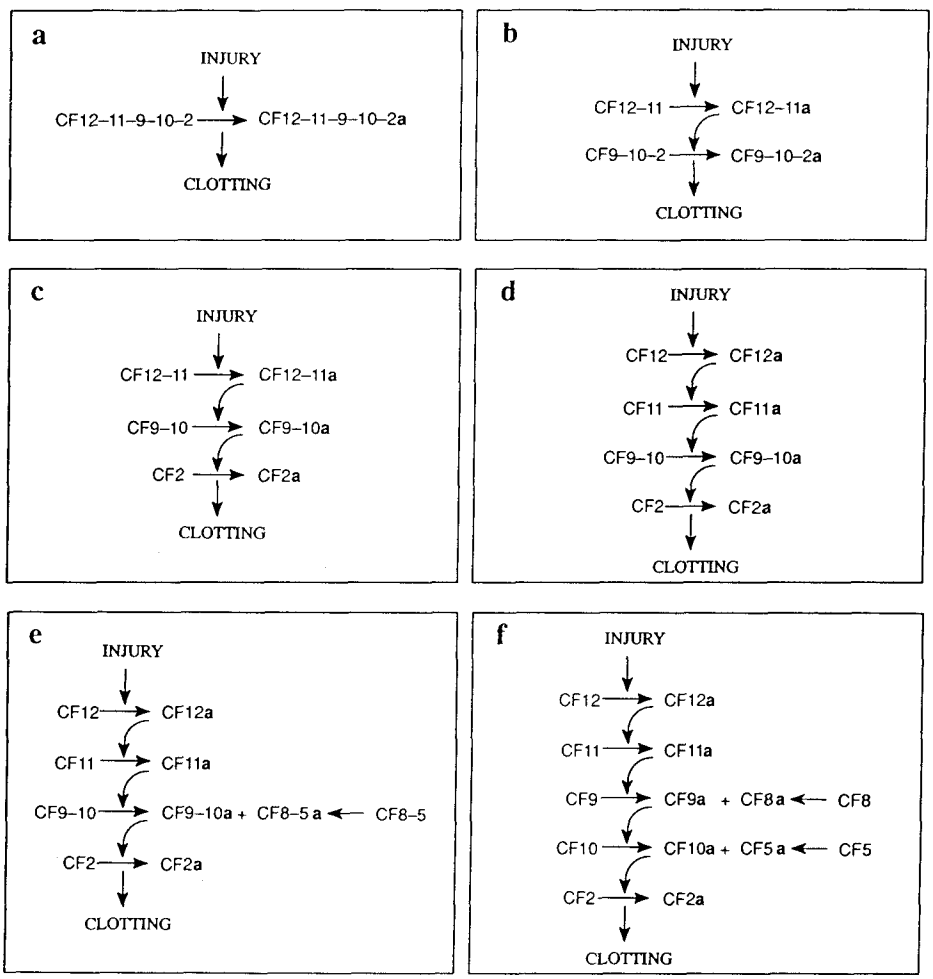


Fig. 14. Suggested route in six steps, **a–f**, for the evolution of the extrinsic pathway of blood coagulation as based on the phylogeny of coagulation factors V, VIII, IX, X, XI, and XII. Factors marked by two or more Arabic numerals are the ancestors; thus CF8-5 is the ancestor of factor V and factor VIII. The activated form of each factor or ancestral factor is indicated by a *lowercase a*.

References

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

Adman ET, Turley S, Bramson R, Petratos K, Banner D, Tsernoglou D, Beppu T, Watanabe H (1989) A 2.0-Å structure of the blue copper protein (cupredoxin) from *Alcaligenes faecalis* S-6. J Biol Chem 264:87–99

Aitken A (1975) Prokaryote-eukaryote relationships and the amino acid sequence of plastocyanin from *Anabaena variabilis*. Biochem J 149:675–683

Ambler RP (1971) Sequence data acquisition for the study of phylogeny. In: Previero A, Pechere J-F, Coletti-Previero

phylogeny. In: Previero A, Pechere J-F, Coletti-Previero M-A (eds) Recent developments in the chemical study of protein structures. INSERM, Paris, pp 289–305

Ambler RP, Tobari J (1985) The primary structures of *Pseudomonas* AM1 amicyanin and pseudoazurin. Two new sequence classes of blue copper proteins. Biochem J 232:451–457

Astolfi P, Kidd KK, Cavalli-Sforza LL (1981) A comparison of methods for reconstructing evolutionary trees. Syst Zool 30: 156–169

Barker WC, Dayhoff MO (1979) Evolutionary trees from protein sequences: probability of correct reconstruction of four-branch topologies. Biophysical J 25:158a

Barker WC, Hunt LT, George DG (1988) Homology and similarity in amino acid sequences: detection, analysis, and interpretation. In: Silverman BD (ed) Computer simulation of carcinogenic processes. CRC Press, Boca Raton, FL, pp 1–36

Barker WC, George DG, Hunt LT (1990) Protein sequence database. In: Doolittle RF (ed) Methods in enzymology, Vol 183. Academic Press, New York, pp 31–49

Bergman C, Gandvik E-K, Nyman PO, Strid L (1977) The amino acid sequence of stellacyanin from the lacquer tree. Biochem Biophys Res Commun 77:1052–1059

Bergman C (1980) Amino acid sequence studies of two small blue proteins: stellacyanin and umecyanin. PhD thesis (Department of Biochemistry and Biophysics), University of Göteborg, Sweden

Blanken RL, Klotz LC, Hinnebusch AG (1982) Computer comparison of new and existing criteria for constructing evolutionary trees from sequence data. J Mol Evol 19:9–19

Bonner JT (1988) The evolution of complexity by means of natural selection. Princeton University Press, Princeton, NJ

Chou PY, Fasman GD (1978a) Empirical predictions of protein conformation. Annu Rev Biochem 47:251–276

Chou PY, Fasman GD (1978b) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol 47:45–148

Cole ST, Grundstrom T, Jaurin B, Robinson JJ, Weiner JH (1982) Location and nucleotide sequence of frdB, the gene coding for the iron-sulphur protein subunit of the fumarate reductase of *Escherichia coli*. Eur J Biochem 126:211–216

Darlison MG, Guest JR (1984) Nucleotide sequence encoding the iron-sulphur protein subunit of the succinate dehydrogenase of *Escherichia coli*. Biochem J 223:507–517

Dayhoff MO, Barker WC, Hunt LT (1983) Establishing homologies in protein sequences. In: Hirs CHW, Timasheff SN (eds) Methods in enzymology, vol 91. Academic Press, New York, pp 524–545

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees: a method based on mutation distances as estimated from cytochrome c sequences is of general applicability. Science 155:279–284

George DG, Hunt LT, Barker WC (1988) The Protein Identification Resource (PIR) software. In: Lesk AM (ed) Computational molecular biology: sources and methods for sequence analysis. Oxford University Press, Oxford, pp 100–115

George DG, Barker WC, Hunt LT (1990) Mutation data matrix and its uses. In: Doolittle RF (ed) Methods in enzymology, vol 183. Academic Press, New York, pp 333–351

Germann UA, Müller G, Hunziker PE, Lerch K (1988) Characterization of two allelic forms of *Neurospora crassa* laccase. Amino- and carboxyl-terminal processing of a precursor. J Biol Chem 263:885–896

Goodman M, Romero-Herrera AE, Dene H, Czelusniak J, Tashian RE (1982) Amino acid sequence evidence on the phylogeny of primates and other eutherians. In: Goodman M (ed) Macromolecular sequences in systematic and evolutionary biology. Plenum Press, New York, pp 115–191

Guss JM, Freeman HC (1983) Structure of oxidized poplar plastocyanin at 1.6 Å resolution. J Mol Biol 169:521–563

Guss JM, Merritt EA, Phizackerley RP, Hedman B, Murata M, Hodgson KO, Freeman HC (1988) Phase determination by multiple-wavelength x-ray diffraction: crystal structure of a basic "blue" copper protein from cucumbers. Science 241: 806–811

Han T-M, Runnegar B (1992) Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee Iron-Formation, Michigan. Science 257:232–235

Hasegawa M, Kishino H, Siatou N (1991) On the maximum likelihood method in molecular phylogenetics. J Mol Evol 32: 443–445

Hedner U, Davie EW (1989) Introduction to hemostasis and the vitamin K-dependent coagulation factors. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) Metabolic basis of inherited disease, ed 6, vol 2. McGraw-Hill, New York, pp 2107–2134

Holm L, Saraste M, Wilkström M (1987) Structural models of the redox centres in cytochrome oxidase. EMBO J 6:2819–2823

Hormel S, Adman E, Walsh KA, Beppu T, Titani K (1986) The amino acid sequence of the blue copper protein of *Alcaligenes faecalis*. FEBS Lett 197:301–304

Hunt LT, George DG, Yeh L-SL (1985) Ragweed allergen Ra3: relationship to some type I copper-binding proteins. J Mol Evol 21:126–132

Jenny RJ, Pittman DD, Toole JJ, Kriz RW, Aldape RA, Hewick RM, Kaufman RJ, Mann KG (1987) Complete cDNA and derived amino acid sequence of human factor V. Proc Natl Acad Sci USA 84:4846–4850

Kane WH, Davie EW (1986) Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin. Proc Natl Acad Sci USA 83: 6800–6804

Kane WH, Davie EW (1988) Blood coagulation factors V and VIII: structural and functional similarities and their relationship to hemorrhagic and thrombotic disorders. Blood 71:539–555

Kane WH, Ichinose A, Hagen FS, Davie EW (1987) Cloning of cDNAs coding for the heavy chain region and connecting region of human factor V, a blood coagulation factor with four types of internal repeats. Biochemistry 26:6508–6514

Kanehisa M (1985) IDEAS. Integrated database and extended analysis system for nucleic acids and proteins. User manual. Laboratory of Mathematical Biology, NIH, Bethesda, MD

Knoll AH (1992) The early evolution of eukaryotes: a geological perspective. Science 256:622–627

Koschinsky ML, Funk WD, van Oost BA, MacGillivray RTA (1986) Complete cDNA sequence of human preceruloplasmin. Proc Natl Acad Sci USA 83:5086–5090

Kraulis PJ (1991) MolScript—a program to produce both detailed and schematic plots of protein structure. J Appl Crystallogr 24:946–950

Malkin R, Malmström BG (1970) The state and function of copper in biological systems. Adv Enzymol 33:177–243

Messerschmidt A, Rossi A, Ladenstein R, Huber R, Bolognesi M, Gatti G, Marchesini A, Petruzzelli R, Finazzi-Agró A (1989) X-ray crystal structure of the blue oxidase ascorbate oxidase from zucchini. Analysis of the polypeptide fold and a model of the copper sites and ligands. J Mol Biol 206:513–529

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

Mondovi B, Avigliano L (1984) Ascorbate oxidase. In: Lontie R

(ed) Copper proteins and copper enzymes, vol 3. CRC Press, Boca Raton, FL pp 101–118

Murata M, Begg GS, Lambrou F, Leslie B, Simpson RJ, Freeman HC, Morgan FJ (1982) Amino acid sequence of a basic blue protein from cucumber seedlings. Proc Natl Acad Sci USA 79:6434–6437

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Norris GE, Anderson BF, Baker EN (1983) Structure of azurin from *Alcaligenes denitrificans* at 2.5 Å resolution. J Mol Biol 165:501–521

Ohkawa J, Okada N, Shinmyo A, Takano M (1989) Primary structure of cucumber (*Cucumis sativus*) ascorbate oxidase deduced from cDNA sequence: homology with blue copper proteins and tissue-specific expression. Proc Natl Acad Sci USA 86:1239–1243

Omoto E, Tavassoli M (1990) Purification and partial characterization of ceruloplasmin receptors from rat liver endothelium. Arch Biochem Biophys 282:34–38

Patthy L (1985) Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. Cell 41:657–663

Poole S, Firtel RA, Lamar E, Rowekamp W (1981) Sequence and expression of the discoidin I gene family in *Dictyostelium discoideum*. J Mol Biol 153:273–289

Reinhammar B (1984) Laccase. In: Lontie R (ed) Copper proteins and copper enzymes, vol 3. CRC Press, Boca Raton, FL, pp 1–35

Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234:364–368

Rydén L (1982) Model of the active site in the blue oxidases based on the ceruloplasmin-plastocyanin homology. Proc Natl Acad Sci USA 79:6767–6771

Rydén L (1984a) Structure and evolution of the small blue proteins. In: Lontie R (ed) Copper proteins and copper enzymes, vol 1. CRC Press, Boca Raton, FL, pp 183–214

Rydén L (1984b) Ceruloplasmin. In: Lontie R (ed) Copper proteins and copper enzymes, vol 3. CRC Press, Boca Raton, FL, pp 37–100

Rydén L (1988) Evolution of blue copper proteins. In: King TP, Mason H (eds) Oxidases and related redox systems. Alan R Liss, New York, pp 349–366

Rydén L, Lundgren J-O (1976) Homology relationships among the small blue proteins. Nature 261:344–346

Rydén L, Lundgren J-O (1979) On the evolution of blue proteins. Biochimie 61:781–790

Rydén L, Lundgren J-O (1980) The relationship of the small blue proteins with the copper-containing oxidases. In: Peeters H (ed) Protides of the biological fluids, vol 28. Pergamon Press, Oxford, pp 87–90

Saitou N, Imanishi T (1989) Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol Biol Evol 6:514–525

Saloheimo M, Niku-Paavola M-L, Knowles JKC (1991) Isolation and structural analysis of the laccase gene from the lignin-degrading fungus *Phlebia radiata*. J Gen Microbiol 137:1537–1544

Saraste M (1990) Structural features of cytochrome oxidase. Q Rev Biophys 23:331–366

Schopf JW (1978) The evolution of the earliest cells. Sci Am 239:111–140

Sourdis J, Nei M (1988) Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. Mol Biol Evol 5:298–311

Stubbs JD, Lekutis C, Singer KL, Bui A, Yuzuki D, Srinivasan U, Parry G (1990) cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like domains linked to factor VIII-like sequences. Proc Natl Acad Sci USA 87:8417–8421

Takahashi N, Ortel TL, Putnam FW (1984) Single-chain structure of human ceruloplasmin: the complete amino acid sequence of the whole molecule. Proc Natl Acad Sci USA 81: 390–394

Trackman PC, Pratt AM, Wolanski A, Tang S-S, Offner GD, Troxler RF, Kagan HM (1990) Cloning of rat aorta lysyl oxidase cDNA: complete codons and predicted amino acid sequence. Biochemistry 29:4863–4870

Vehar GA, Keyt B, Eaton D, Rodriguez H, O'Brien DP, Rotblat F, Oppermann H, Keck R, Wood WI, Harkins RN, Tuddenham EGD, Lawn RM, Capon DJ (1984) Structure of human factor VIII. Nature 312:337–342

Williams AF, Barclay AN (1988) The immunoglobulin superfamily: domains for cell surface recognition. Ann Rev Immunol 6:381–405

Wood WI, Capon DJ, Simonsen CC, Eaton DL, Gitschier J, Keyt B, Seeburg PH, Smith DH, Hollingshead P, Wion KL, Delwart E, Tuddenham EGD, Vehar GA, Lawn RM (1984) Expression of active human factor VIII from recombinant DNA clones. Nature 312:330–337

Young CL, Barker WC, Tomaselli CM, Dayhoff MO (1979) Serine proteases. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington, DC, pp 73–93

Received March 5, 1992/Revised and accepted August 7, 1992