Heterogeneity of the covalent structure of the blue copper protein umecyanin from horseradish roots

GONZALEZ VAN DRIESCHE,¹ CHRISTOPHER DENNISON,²
A. GEOFFREY SYKES,² AND JOZEF VAN BEEUMEN¹

¹ Department of Biochemistry, Physiology and Microbiology, Laboratory of Biochemistry,
University of Gent, Gent, Belgium

² Department of Chemistry, University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom

(RECEIVED June 2, 1994; ACCEPTED November 4, 1994)

Abstract

The covalent structure of umecyanin has been determined by a combination of classical Edman degradation sequence analysis and plasma desorption, laser desorption, and electrospray ionization mass spectrometry. The preparation appeared to contain two isoforms having either a valine (75%) or an isoleucine (25%) residue at position 48. The polypeptide chain of 115 amino acids is strongly heterogeneous at its C-terminal end as a result of proteolytic cleavages at several places within the last 10 residues. The major fraction of the umecyanin preparation is only 106 residues long. The C-terminal tail 107–115 contains mainly alanine and glycine residues and a single hydroxyproline residue. In the native protein there is a disulfide bridge between Cys 91 and Cys 57, but in the apoprotein there is a disulfide shift that involves Cys 91 and one of the four copper binding residues (Cys 85). The three other ligand binding residues are His 44, His 90, and Gln 95. This tetrad of amino acids is the same as occurs in other type 1 copper proteins from plants such as cucumber peeling cupredoxin and lacquer tree stellacyanin. The umecyanin isoforms are glycoproteins with a glycan core having the same carbohydrate composition as that of horseradish peroxidase, a fact that is convincingly supported thanks to the high accuracy of the electrospray mass spectrometric technique. We suggest that the glycan may play a role in the association of the protein to the cellular membrane, but the precise functional role of umecyanin remains to be determined. We also discuss the evolutionary position of umecyanin in relation to the type 1 copper proteins in general.

Keywords: amino acid sequence; electrospray ionization mass spectrometry; laser desorption mass spectrometry; plasma desorption mass spectrometry; posttranslational modifications; umecyanin

The blue copper proteins constitute a very diverse group of proteins that may contain up to six copper atoms per polypeptide chain (Ryden & Björk, 1976; Marchesini & Kroneck, 1979; Ryden & Hunt, 1993). The characteristic intense blue color in the oxidized form is due to the absorbance at ≈ 600 nm (ϵ 's: $2\text{--}5 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$) resulting from an S(Cys) $\rightarrow$ Cu(II) charge transfer (Sykes, 1990; Adman, 1991). The proteins are isolated in single or multiple forms (Markossian et al., 1974; Ryden, 1984). Umecyanin is a type 1 blue copper protein from horseradish

roots (*Armoracia laphatifolia*). Umecyanin, first isolated as a co-product during the isolation of peroxidase from the peelings of horseradish roots (Paul & Stigbrand, 1970), has a molecular mass of about 14 kDa (Paul & Stigbrand, 1970) and an isoelectric point of 5.85 (Stigbrand, 1971). Its reduction potential was determined to be +283 mV (Stigbrand, 1972), which is comparable to that of azurins and pseudoazurins (Sykes, 1990). The reduction potentials of other related type 1 copper proteins are +370 mV for plastocyanin (Katoh et al., 1962), +184 mV for stellacyanin (Reinhammar, 1972), and +680 mV for rusticyanin (Ingledew & Cobley, 1980). The latter two values are the lowest and highest values, respectively, which have been determined for type 1 copper proteins. The functional role of umecyanin is unknown, but it should be noted that many small type 1 copper proteins have been isolated as co-products in the purification of large enzymes to which they appear to be functionally related. Two notable examples in this respect are stellacyanin to laccase (Aikazyán & Nalbandyan, 1979) and cucumber basic protein to indol-3-ethanol oxidase (Vickery, 1971).

Reprint requests to: Jozef Van Beeumen, Vakgroep Biochemie, Fysiologie en Microbiologie, Universiteit Gent, Ledeganckstraat 35, B-9000 Gent, Belgium; e-mail: jozef.vanbeeumen@rug.ac.be.

Abbreviations: ACN, acetonitrile; CBP, cucumber basic protein; CMC, carboxymethylcysteine; CPC, cucumber peeling cupredoxin; DTT, dithiothreitol; ESMS, electrospray ionization mass spectrometry; Hyp, hydroxyproline; MALDI-MS or LDMS, matrix assisted laser desorption ionization mass spectrometry; PDMS, plasma desorption mass spectrometry; PMP, 1-phenyl-3-methyl-5-pyrazolone; PTH, phenylthiohydantoin; TFA, trifluoroacetic acid.

CD experiments at neutral pH and EPR studies suggest a relationship between umecyanin and the azurins (Stigbrand & Sjöholm 1972; Chapman et al., 1986). On the other hand, the shift of the absorption maximum in the visible region from 610 nm (blue) at neutral pH to 580 nm (violet) at alkaline pH (Stigbrand, 1972; Stigbrand & Sjöholm 1972; Chapman et al., 1986) implies a similarity of umecyanin to stellacyanin for which a similar behavior has also been observed (Malmström et al., 1970).

The sequence of the first 88 residues of umecyanin has already been determined (Bergman, 1980) and exhibits 38% identity with the stellacyanin sequence (Bergman et al., 1977; Bergman, 1980). The amino acid composition of umecyanin shows that it contains four methionines (Stigbrand, 1971), and in the partial sequence these all occur before position 75. Therefore, like stellacyanin, which contains no methionine residues, the copper site in umecyanin is also believed to possess a different coordination as compared to most other type I blue copper proteins, which have a C-terminal methionine ligand. It has been suggested recently that in stellacyanin the fourth ligand is a glutamine (Fields et al., 1991; Thomann et al., 1991). An EXAFS study on umecyanin shows that there is one sulfur and two nitrogen ligands to the copper atom (Chapman et al., 1986), but the nature of the fourth ligand has remained uncertain.

In this paper we present the complete amino acid sequence of umecyanin including the posttranslational modifications that it undergoes. A comparison is made between the sequence and those for other type I copper proteins, in particular to stellacyanin, cucumber peeling cupredoxin, and CBP, the three-dimensional structure of which is known (Guss et al., 1988).

Results

Sequence determination

The complete sequence of umecyanin (Fig. 1) was obtained by combining chemical sequence analysis of peptides from different enzymatic digests and measuring their precise masses by plasma or laser desorption and/or electrospray ionization mass spectrometry. The first 88 residues, with the exception of residue 76, had already been determined by Bergman (1980) to be: E D Y D V G G D M E W K R P S D P K F Y I T W A T G K T F R V G D E L F D F A A G M H D V A V V T K D A F D N C K K E N P I S H M T T P P V K I M L ? T T G P Q Y Y I C T V G. Although the completion of the umecyanin sequence required the confirmation of these data, we limit ourselves here to a description of the new sequence information.

Over the course of time, we used two batches of purified umecyanin to establish the sequence. The second batch was mainly used to establish the C-terminal sequence after the material from the first batch was used up (see further). SDS-PAGE of the latter batch revealed the presence of three major bands (D, E, and F in Fig. 2A), one minor band (A in Fig. 2A), and two faintly visible bands (B and C in Fig. 2A). The electrospray ionization mass spectrum of the native preparation, however, betrayed the presence of at least 15 molecular masses ranging from 11,986.6 Da to 19,712.7 Da (result not shown), suggesting that posttranslational modification(s) could be at the origin of this heterogeneity. Reversed-phase HPLC carried out as a control experiment of this apparent heterogeneity resulted in two major fractions (II and III) and one minor fraction (I) (Fig. 2B). The SDS-PAGE electropherogram of each of these fractions is given in Figure 2A.

Edman degradation of fraction III and each of the bands D, E, and F revealed the same N-terminal protein sequence as already been determined by Bergman (1980) (Table S1, Electronic Appendix). Fraction I had the totally different sequence S E Y V T D I S G N P. It was considered to be a contaminant protein present at most to 10% of the combined umecyanin species. The result for fraction II was as for fraction III, with a slight contamination of fraction I. In view of their identical N-terminal sequences, it was concluded that fractions II and III must contain different C-terminal polypeptide fragments of umecyanin, with fraction III containing the highest concentration of the complete polypeptide.

The best way to obtain useful new sequence information from position Lys 72 onward was digestion with Lys-C endoproteinase (Fig. 3A). Two of the resulting fractions, K6B and K7, extended the sequence up to position 94, with exception of the residues 76 and 91–92. In both the sequence analyses of K6B and K7, the fourth residue appeared as a blank during the run. It was also assumed that peptides K4 and K5B were the C-terminal peptides of umecyanin (9 residues long) because all the other peptides of the digest could be located in the region 1–59 of the protein.

Digesting the carboxymethylated apoprotein with Asp-N endoproteinase revealed the presence of peptides upon subsequent RP-HPLC separation (Fig. 4). Sequence analysis of peptide nD6 demonstrated an overlap of peptide K7 with peptide K4. It also identified residues 91, 92, 95, and 96 to be, respectively CMC, Arg, Gln, and Lys. Peptide nD2 was 7 residues long, confirming the sequence Asp 45–Lys 51. The peptide nD3, which eluted later, appeared to be a mixture of two very similar fragments, one being identical to Asp 38–His 44, the second differing from nD2 by having an isoleucine instead of valine as its fourth residue. The molecular mass of both fractions, determined by PDMS (Fig. 5), confirmed the presence of two different amino acids at the same position. This heterogeneity was confirmed by sequence and mass analysis of peptides obtained by cleavage of the carboxymethylated apoprotein with Arg-C proteinase (Fig. 6) and subdigesting the three largest peptides R7, R8, and R9 with Lys-C endoproteinase (Fig. 7). Quantitative sequence and mass analysis data for all the peptides are given in Tables S2–S6 (Electronic Appendix).

Type of N-glycosylation

The absence of a PTH-amino acid at position 76 upon sequence analysis of the appropriate peptides, K6B and K7, and the fact that a Thr was identified two residues further along this section, suggested that this position was taken by a glycosylated asparagine residue. This presumption was first confirmed by amino acid analysis of the peptides K6B and K7, which showed the presence of one more aspartic acid and of glucosamines (Table S7, Electronic Appendix). The differences between the calculated mass value of the peptides and the experimentally determined values further suggests that more than one sugar residue is bound to Asn. A carbohydrate compositional analysis following the reversed-phase HPLC method of Honda et al. (1989) was applied to umecyanin to allow comparison with those for the peroxidase from the same species. Although the column used was not able to separate carbohydrates with similar structural formulae, the separation was very convenient for structurally different carbohydrates such as hexose, hexosamine, pentose, and

Fig. 1. Amino acid sequence of the blue copper protein from horseradish root. The peptides were obtained by digestion of carboxymethylated apoprotein with Lys-C protease (K), Asp-N endoprotease (nD), and Arg-C protease (R). Peptides from a digest with Lys-C endoprotease of apoprotein are indicated by K'. The symbol ■ means that the run was deliberately stopped. The heterogeneity of some C-terminal peptides is indicated by vertical dashed lines.

Fig. 2. A: SDS-PAGE of native umecyanin. Positions of the molecular mass markers in the first lane are shown in the left margin: (from top to bottom) bovine serum albumin (66.2 kDa), ovalbumin (45.0 kDa), glyceraldehyde-3-phosphate dehydrogenase (36.0 kDa), carbonic anhydrase (31.0 kDa), trypsinogen (24.0 kDa), soybean trypsin inhibitor (21.5 kDa), lysozyme (14.5 kDa). Positions of the myoglobin fragment markers in lane 2 are at 17.0, 14.5, and 10.7 kDa, respectively. Lanes 3 and 4 contain native umecyanin from batches 1 and 2, respectively (see text); the different components are marked from A to F in the right margin. Lanes 5, 6, and 7 represent fractions I, II, and III obtained after reversed-phase HPLC of umecyanin batch 1. **B:** Reversed-phase HPLC separation of native umecyanin. Conditions are described in the Materials and methods.

deoxypentose. It is remarkable that the elution pattern of both analyses (Fig. 8) suggests the presence of the same carbohydrates, namely glucosamine, mannose, fucose, and xylose. The mass of fraction K7, determined by electrospray mass spectrometry (3,018.3 Da) was in agreement with the calculated mass of

3,017.4 Da, including an asparagine and a single residue of GlcNAc (203.19 Da) (Fig. 9A). The experimental mass of peptide K6B (3,985.2 Da) (Fig. 9B), in combination with the result from carbohydrate analysis (Fig. 8), is in perfect agreement with the sum of the masses of the amino acids supplemented with the carbohydrate structure found for horseradish peroxidase, (GlcNAc)₂(Man)₃(Fuc)(Xyl). The same conclusion was also obtained from the measured masses of the other peptides containing Asn 76, namely nD7 and nD8 (Table S3).

C-terminal sequence

Evidence for the C-terminal sequence was obtained by Lys-C cleavage of a second umecyanin preparation (batch 2) that was declustered but not chemically modified by carboxymethylation. It resulted in the generation of three different C-terminal peptides (K'4, K'5, and K'6) (Fig. 3B), each starting at Leu 97 but having different lengths (Fig. 1). From sequence analysis, K'4 was identical to peptide K4 isolated from the digest of the carboxymethylated sample (batch 1) of umecyanin. The second one (K'5) could be sequenced for four residues further than K'4. Electrospray mass analysis of both K'4 and K'5, however, revealed that the peptides were still longer than determined from the sequence analyses. Moreover, the mass spectra showed a series of peaks, corresponding to the cleavage of various positions between amino acids at position 105 and 110 (Table 1; Fig. 10). Peptide K'6 was the longest of the three C-terminal peptides. Edman degradation allowed the identification of 17 residues with the last of them having an unusual retention time of its PTH derivative. By separate comparison with PTH analysis on the reference product submitted to one Edman cycle, we were able to identify the 17th residue of K'6 as hydroxyproline. Due to the small amount of material remaining, and probably also because of the hydrophobic nature of peptide K'6, we could not confirm this assignment by either laser desorption or electrospray mass spectrometric analysis. Confirmation was obtained by both sequence and electrospray mass analysis of the low yield peptide R5 from the carboxymethylated umecyanin (Fig. 6). The electrospray spectrum also confirmed that the hydroxyproline residue is followed by a Gly-Ala sequence, but that two other peptide species were present that were, respectively, one and two residues shorter than the Val 93–Ala 115 sequence. Quantitative data for the K' peptides are given in Table S8 (Electronic Appendix). From all these data, it should be clear that the argument to state that peptides K'6 and R5 and the shortened peptides K'4 and K'5 are C-terminal peptides of the C-terminally ragged polypeptide chains of umecyanin, comes from the summed masses of the sequence-determined amino acids Glu 1–Gly 105 (11,687.3 Da, with residue 48 = Ile) supplemented with the mass of the largest glycan chain (glycan II = 1,171.1 Da).

Disulfide bridges in apo and native umecyanin

The mass of peptide K'8 (Ile 73–Lys 96) obtained from the digest of apo-umecyanin was found by ESMS to be 3,866.2 Da and by LDMS to be 3,866.5 Da (Fig. 11A). The calculated mass being 3,867.2 Da suggested that the two cysteine residues in this peptide, Cys 85 and Cys 91, might be linked by a disulfide bridge. This was confirmed to be indeed the case because Asp-N subdigestion did not result in two peptides but, on the contrary,

Fig. 3. Reversed-phase HPLC of peptides from digestion of, respectively, carboxymethylated (A) and apo-protein (B) with Lys-C endoproteinase. The peptides that correspond to those in A are mentioned in parentheses in the latter chromatogram. Conditions for the separation are described in the Materials and methods.

Fig. 4. Reversed-phase HPLC of peptides generated by digestion with Asp-N endoproteinase of carboxymethylated apo-protein. Conditions are described in the Materials and methods.

revealed an increase of mass of 22 Da as shown by MALDI-MS (Fig. 11B). Although this increase is close to the value for a Na⁺ ion, we rather think that the addition of a water molecule should be considered here as required for the cleavage of the Gly 88–Asp 89 bond. The difference of 4 Da may be attributed to the fact that the masses were measured using an external instead of an internal mass standard.

In order to prove that this is also the case in the native umecyanin, we also digested 7 nmol of the carboxymethylated (but not reduced) umecyanin with Asp-N endoproteinase. Sequence analysis of the peak GF-1 obtained after gel filtration of the generated fraction (Fig. 12) revealed the presence of the peptides D55–G88 and D89–A115 with a blank analysis at Edman cycle 3 (Table S9, Electronic Appendix). Because mass analysis on the fraction GF-1 failed, a subdigest with Lys-C endoproteinase was carried out with subsequent separation of the peptides (Fig. 13). Although one of these fractions (GF-1A) eluted as a sharp single peak, it again revealed the presence of two peptides with a blank at Edman cycle 3 (Table S9). Combined with the mass analyses by PDMS (Fig. 14A), it could be unambiguously concluded that the peptides D55–K58 (calc. mass 478.5 Da) and D89–K96 (calc. mass 942.0 Da) were linked by a disulfide bridge, resulting in a molecular species with a theoretical mass of 1,418.5 Da. This is within experimental error of the measured mass of 1,418.1 Da. The presence of the 941.8-Da peak in the non-reduced sample GF-1A (Fig. 14A), together with the peak at 478.5 Da, is very likely due to fragmentation. Fragment ions due to the cleavage of the disulfide bond have often been observed in the PDMS analyses of disulfide-bound peptides (Sørensen et al., 1990). Confirmation of the existence of this disulfide bridge was obtained by reduction of the sample with DTT and mass analysis by PDMS, revealing the increase of a fragment of 942 Da (Fig. 14B).

From the experiments on the native protein, it thus appears that the disulfide bridge detected in peptide K'8 must be due to

Fig. 5. Plasma desorption mass spectrum of peptide nD2 (A) (MH^+ : 731.9 Da) and nD3 (B) obtained by Asp-N digestion on carboxymethylated apoprotein. The mass of 765.7 Da corresponds to the calculated mass of fragment nD3A: Asp 38–His 44 (MH^+ : 764.8 Da).

thiol–disulfide exchange. Such an exchange is known to be possible at pH 8, the pH value at which the digest of the reduced apoprotein with Lys-C endoproteinase has been carried out (Engeseth et al., 1984). A disulfide interchange will not have been possible upon Asp-N digestion of the carboxymethylated native protein because this derivatization has been carried out at pH 7.

Discussion

Heterogeneity of the primary structure of umecyanin

The protein preparation known as umecyanin that has been isolated from horseradish roots appears to contain two isoforms

Table 1. Degraded forms of the C-terminal region of umecyanin, determined by a combination of electrospray mass spectrometry and N-terminal sequence analysis^a

Peak no.	Structure	Measured mass (Da)	Calculated mass (Da)
K4	97 104 L–V V G A	772.3	771.9
	L–V V G A G	828.4	829.0
	L–V V G A G G	885.3	886.0
R4	93 105 V–V V G A G	1,240.9	1,241.4
	V–V V G A G G	1,297.3	1,298.5
R5	V–V V G A G G A G G G A T Hyp	1,825.1	1,827.0
	V–V V G A G G A G G G A T Hyp G	1,883.4	1,884.1
	V–V V G A G G A G G G A T Hyp G A	1,954.0	1,955.2
K'4	97 104 L–V V G A	771.9	771.9
	L–V V G A G	828.6	829.0
	L–V V G A G G	885.8	886.0
K'5	L–V V G A G G A	956.5	957.1
	L–V V G A G G A G	1,013.6	1,014.1
	L–V V G A G G A G G	1,070.6	1,071.3
	L–V V G A G G A G G G	1,127.7	1,128.3
K'6	L–V V G A G G A G G G A T Hyp G A	–	1,542.7

^a K4, R4, R5 are isolated from digestion on carboxymethylated apoprotein with, respectively, Lys-C (Fig. 3A) and Arg-C protease (Fig. 6). C-terminal fragments from the Lys-C digestion on the second batch of umecyanin (K'4, K'5, and K'6, Fig. 3B) are also isolated as single HPLC peaks.

Fig. 6. HPLC separation of peptides obtained by Arg-C endoproteinase digestion on carboxymethylated apoprotein.

differing by only one amino acid. Each isoform has a polypeptide chain containing 115 residues, but one isoform has a valine and the other isoform an isoleucine residue at position 48. However, the major fraction of the preparation contains only 106 amino acids (starting from the N-terminus) as appears from the electrospray spectrum of HPLC fractions II and III of batch 1 (see also further). The initial yields of the peptides R4 versus R5 and K4 versus K5 indicate that the amounts of this fraction may be as large as 70–85%. The calculated amino acid composition of umecyanin agrees rather well with those obtained by Stigbrand (1971) and Bergman (1980) after recalculation of their data, taking into account a total number of 104 amino acids (two tryptophan residues not included) instead of the 122 and 120 amino acids that they respectively took into account (Table 2). The major fraction of umecyanin contains 15 acidic and 11 basic residues, 3 cysteines, 4 methionines, and 3 histidines. The polypeptide chain also contains 1 hydroxyproline (position 113) and 1 glycan binding asparagine (position 76). We were unable to detect Asn 76 as an unmodified residue in any of the peptides that were generated.

The heterogeneity in the primary structure of umecyanin also relates to the fact that polypeptide chains have been detected of different lengths, all of them resulting from extensive cleavage at the C-terminal end. Evidence for cleavage of peptide bonds

Fig. 7. HPLC separation of peptides obtained by Lys-C protease subdigestion on the three large Arg-C peptides: R7 (A), R8 (B), and R9 (C).

Ala 104–Gly 105, Ala 107–Gly 108, Gly 108–Gly 109, Gly 109–Gly 110, Hyp 113–Gly 114, and Gly 114–Ala 115 was also obtained from electrospray mass analysis of peptides K4, K'4, K'5, R4, and R5. Without the use of this high precision mass spectrometric technique most of these cleavage sites would very likely not have been detected. We will further refer to the polypeptide chain of 106 amino acids as "isoform I" (valine) and "isoform II" (isoleucine), respectively.

Umecyanin is glycosylated. Although we did not carry out experiments to prove the exact sequence of the glycan core, we

Fig. 8. Reversed-phase HPLC separation of PMP-derivatized sugars from native umecyanin. Positions of the peaks are given (from left to right): 1, PMP; 2, glucosamine; 3, artifact; 4, mannose; 5 and 6, artifacts; 7, xylose; 8, fucose. The insert in this chromatogram represents the separation of PMP-derivatized sugars from horseradish peroxidase. The order of the peaks is the same as from umecyanin. Conditions of both separations are given in the Materials and methods.

found by reversed-phase HPLC analysis that the glycan core is composed of the same carbohydrate residues as that of the peroxidase isolated from the same species (Clarke & Shannon, 1976; Green & Oliver, 1991), namely 2 (and 1) GlcNAc, 3 mannose,

Table 2. Amino acid composition of the major peptide chain of umecyanin (104 amino acids, 2 tryptophan residues not included)^a

	Literature data				Our data
	Stigbrand (1971)		Bergman (1980)		
Total	121.9 ^b	104.0 ^c	119.9 ^b	104.1 ^c	104 ^d
Asp	17.3	14.8	17.1	14.8	10
Asn	0.0	0.0	0.0	0.0	4
Thr	10.6	9.1	10.8	9.4	9
Ser	4.7	4.0	3.6	3.1	3
Glu	8.3	7.1	8.4	7.3	5
Gln	0.0	0.0	0.0	0.0	2
Pro	3.7	5.7	6.8	5.9	6
Gly	13.4	11.4	13.7	11.9	11
Ala	8.0	6.8	7.9	6.9	6
Cys	3.0	2.6	2.7	2.3	3
Val	10.4	8.9	10.5	9.2	10
Met	4.2	3.6	4.4	3.8	4
Ile	6.2	5.3	6.0	5.2	5
Leu	3.8	3.2	3.4	3.0	3
Tyr	3.8	3.2	4.0	3.5	4
Phe	4.6	3.9	4.9	4.2	5
Lys	10.0	8.5	9.9	8.6	8
His	3.5	3.0	2.9	2.5	3
Arg	3.4	2.9	2.9	2.5	3

^a All values are expressed in mol/mol of protein.

^b Values obtained from literature data.

^c Values recalculated from literature data.

^d Values obtained from our sequence.

1 fucose, and 1 xylose. The glycan structure, however, is the third cause of heterogeneity in the umecyanin preparation. This could be demonstrated by ES mass analysis of the two major fractions (II and III) obtained after reversed-phase HPLC of the native umecyanin (Fig. 2B). Both of these fractions have the same N-terminal sequence that corresponds to the N-terminal region of the polypeptide chain. The ES spectrum of fraction II (Fig. 15A), however, shows the presence of three components and so does the spectrum of fraction III (Fig. 15B). In Table 3 we have summarized the several different molecular species that we think, on the basis of the high accuracy of the electrospray mass method (Fenn et al., 1990), occur in the umecyanin preparation. The smallest molecular weight of 11,961.75 fits the 106-residue polypeptide isoform I with a singly bound N-acetylglucosamine and two of the methionine residues being oxidized to methionine sulfoxide. Because of the mass difference between valine and isoleucine being 14.03 Da, it may also represent isoform II, with only one of the methionines being oxidized. All other measured masses represent polypeptide chains with an elaborate glycan core of composition (GlcNAc)₂(Man)₃(Fuc)(Xyl). The largest of the masses (12,935.54 Da) fits the structure of isoform II with only one oxidized methionine, or alternatively, of isoform I having two methionine sulfoxide residues. The value of 12,915.16 Da is only one of the six measured masses that fits a molecular species without a modified methionine. It corresponds to isoform II with the full glycan core.

Two of the three cysteine residues in the umecyanin isoforms (Cys 57 and Cys 91) are involved in a disulfide bridge. This could be proven unambiguously by the fact that a digest of a carboxymethylated nonreduced sample of umecyanin with Asp-N endoprotease produced an Asp 55–Gly 88 and Asp 89–Lys96 fraction. Further proof came from subsequent subdigest with Lys-C protease. The latter cleavage resulted in a fraction of molecular mass of 1,418.1 Da, which corresponds to the disulfide forms of the peptides Asp 55–Lys 58 and Asp 89–Lys 96. It should be noted that peptide K'8, obtained from reduced carboxymethylated protein, reveals the presence of a disulfide

Fig. 9. Electrospray ionization mass spectrum of peptide K7 (A) and K6B (B) from digestion of the carboxymethylated apoprotein with Lys-C endoproteinase. "Na" and "K" represent sodium and potassium adducts to the major peaks. "CM" represents a carboxymethyl-adduct, very likely (not proven) of a histidine residue. The number following the capital letter at the top of each peak represents the number of positive charges for that particular m/z peak. **A:** Measured mass of 3,033.3 Da of component B corresponds to the methionine sulfoxide from component A. The measured difference between component C (3,163.9 Da) and component A (3,018.3 Da) amounts to 145.6 Da, which is in agreement with the mass of fucose (146.1 Da). **B:** Mass difference between measured components B (4,188.1 Da) and A (3,985.2 Da) represents an extra *N*-acetylglucosamine residue in component B.

bridge between Cys 85 and Cys 91. Cys 85, as further discussed, is the most likely candidate as one of the four Cu ligands in the native umecyanin isoforms. The origin of this disulfide bridge probably resides in thiol-disulfide interchange at the alkaline pH of the digest.

Comparison to other small blue cupredoxines

Umecyanin belongs to the family of small blue copper proteins (Stigbrand et al., 1971). In Figure 16A, the amino acid sequence of isoform I of umecyanin is aligned with those of small blue copper proteins stellacyanin (Bergman et al., 1977), CBP (Murata et al., 1982), rusticyanin (Ronk et al., 1991), and CPC

(Mann et al., 1992). The alignment is based on the known three-dimensional structure of the cucumber basic copper protein, the structure of which has recently been resolved by a novel X-ray diffraction technique (Guss et al., 1988). In summary, strands 3 and 6 form part I of a sandwich-like structure, whereas strands 7 and 8 constitute the second part of the sandwich. Both halves are oriented face to face. The antiparallel β -strands 4 and 5 are oriented perpendicularly to the sandwich structure and enclose the copper atom. The α -carbon backbone of CBP is shown in Figure 16B. The secondary structure elements (seven β -strands and one α -helix) are indicated in Figure 16A under the sequence of CBP. The four copper ligands are located, respectively, as follows (sequence alignment numbering of Fig. 16A): Cys 96 in

Fig. 10. Electrospray mass spectrum of peptides K'4 (A) and K'5 (B). Both peptides are obtained by digest of the apo umecyanin with Lys-C protease. Numbers present the MH^+ values.

Fig. 11. Laser desorption mass spectrum of peptide K'8 (Fig. 3B) before (A) and after (B) subdigestion with Asp-N endoprotease. The used matrix was α -cyanosinapinic acid. Numbers represent the MH^+ values.

Fig. 12. Gel filtration of peptides generated by a digest of the carboxymethylated but not reduced umecyanin with Asp-N endoproteinase. Only fraction GF-1 was investigated by sequence and mass analysis to elucidate the location of the disulfide bridge.

strand 7, His 50 in strand 4, Met 106 in strand 8, and His 101 at the end of the β -turn F98–H100. Three of these ligands are conserved between stellacyanin and CBP, but Met 106 appears to be replaced by a glutamine in stellacyanin. A model of stellacyanin has recently been generated using molecular modeling techniques (Fields et al., 1991). In this work, the structure of CBP, whose amino acid sequence is 45% homologous to that

Fig. 13. Reversed-phase HPLC separation of peptides generated by Lys-C subdigestion on fraction GF-1, obtained by gel filtration of carboxymethylated but not reduced umecyanin with Asp-N protease digestion (Fig. 12). Only fraction GF-1A was further investigated to determine the precise location of the disulfide bridge by mass and sequence analysis. Conditions of the separation are given in the Materials and methods.

Table 3. Proposal for the structure of the different molecular species observed upon ESMS analysis of umecyanin preparation II and III (Fig. 2B)

					Proposed modification	
		Measured (Da)	Calculated with glycan I ^a	Calculated with glycan II ^b	Difference (Da)	Type
Iso-I (Val)						
II	A	11,986.59	11,933.50		53.09	3 MetSO
	B	12,872.01		12,901.37	−29.36	−Gly + 2 MetSO
	C	12,935.54		12,901.37	34.17	2 MetSO
III	A	11,961.75	11,933.5		28.25	2 MetSO
	B	12,861.29		12,901.37	−40.1	−Gly + MetSO
	C	12,915.16		12,901.37	13.79	MetSO
Iso-II (Ile)						
II	A	11,986.59	11,947.52		39.06	2 MetSO
	B	12,872.01		12,915.40	−43.39	−Gly + MetSO
	C	12,935.54		12,915.40	20.14	MetSO
III	A	11,961.75	11,947.53		14.22	MetSO
	B	12,861.29		12,915.40	−54.11	−Gly
	C	12,915.16		12,915.40	−0.24	

^a The proposal is based on the evidence that the major polypeptide contains 106 amino acids (Glu 1–Gly 106), that no copper atom is present any more, and that there is one disulfide bridge (Cys 91–Cys 57). The measured masses obtained for the contaminating fraction I (18,442.8, 19,618.9, and 19,712.7 Da) and of the sodium and potassium adducts of the major peaks of fractions II and III are not included in the table. The calculated mass of isoform I without glycan bound is 11,730.31 Da and isoform II is 11,744.34.

^b Glycan I = GlcNAc (203.19 Da).

^c Glycan II = GlcNAc₂, fucose, mannose₃, xylose (1,171.06 Da).

Fig. 14. Plasma desorption mass spectra of peptides GF-1A obtained by subdigestion of fraction GF-1, as described in Figure 13, with Lys-C endoprotease before (A) and after (B) reduction with DTT. The peaks of 710.2 and 713.4 Da correspond to the double charged molecule of, respectively, 1,419.1 and 1,419.9 Da. Numbers present the MH^+ values.

of stellacyanin, was used as a starting point. The model shows that the fourth copper ligand is indeed a Gln 97 (sequence numbering) instead of Met 89 in CBP. This is further supported by pulsed ENDOR studies (Thomann et al., 1991) and by the fact that the oxidized Met121Gln *Alcaligenes denitrificans* azurin mutant possesses spectroscopic properties similar to stellacya-

nin (Romero et al., 1993). It is apparent from the alignments in Figure 16A that the fourth ligand in umecyanin is likewise a glutamine. It has been suggested to be the case also in CPC (Mann et al., 1992).

The color change (Stigbrand, 1972; Stigbrand & Sjöholm, 1972; Chapman et al., 1986) that has been observed for stella-

Fig. 15. Electrospray ionization mass spectrum of the fractions II and III (Fig. 2B) of umecyanin. The number following the capital letter at the top of each peak represents the number of positive charges for that particular m/z peak.

cyanin and umecyanin upon increasing the pH has been attributed in stellacyanin to a change in the way the glutamine coordinates, namely via the side-chain oxygen at neutral pH and via the deprotonated side-chain amide at higher pH values (Fields et al., 1991). This interpretation seems unlikely because in both cases the color changes are not accompanied by changes in EPR spectra, which are normally very sensitive to changes at the active site (Malmström et al., 1970; Chapman et al., 1986). However, changes in the EPR spectrum of oxidized stellacyanin are observed at pH values ≥ 11 . Moreover, evidence for a nitrogen atom donor at the fourth coordination site in stellacyanin has been obtained at pH 11.0 using pulsed ENDOR (Thomann et al., 1991). The color changes observed at the lower pH values (8–10) have therefore to be explained in some other way. The protonation/deprotonation of a lysine residue in close proximity is one possibility. This has been suggested as the cause of similar effects observed for CBP (Sakurai et al., 1990), and for the Met44Lys mutant of *Pseudomonas aeruginosa* azurin (Van de Kamp et al., 1993), both of which possess a methionine as the fourth ligand.

The coordination of the copper at the active site of stellacyanin by the side-chain amide oxygen of a glutamine would be expected to stabilize the Cu(II) form of the protein. It has been suggested (Fields et al., 1991; Romero et al., 1993) that this is the reason why stellacyanin has the lowest reduction potential of the small blue copper proteins (+184 mV). However, the reduction potential of umecyanin is 283 mV and so the effect of the coordinating oxygen atom cannot be as great in this case. This may be due to a longer Cu–O bond in umecyanin. We note that umecyanin has an axial EPR spectrum and a low absorption at ≈ 450 nm in the visible spectrum, whereas stellacyanin and CBP both have rhombic X-band EPR spectra and greater absorption at ≈ 450 nm in their visible spectra. These two features of small blue copper proteins are thought to depend on the relative position of the copper compared to the axial ligand (Har et al., 1993; Lu et al., 1993; Romero et al., 1993). The properties exhibited by stellacyanin and CBP are thought to be indicative of a copper atom more out of the plane of the equatorial ligands and closer to the axial ligands.

Alternatively, an explanation may relate to the observation that an Asn residue follows the N-terminal histidine ligand in the sequences of most small blue copper proteins (Adman, 1991), and in umecyanin this is replaced by an Asp. The Asn residue is known to be involved in the stabilization of the copper binding site (Hoitink & Canters, 1992; Pascher et al., 1993). The replacement of Asn by Asp introduces a negative charge at the copper binding site and will lead to an increase in the reduction potential of umecyanin. The same replacement in the Asn47Asp mutant of *P. aeruginosa* azurin leads to a 23-mV increase in the reduction potential compared to the wild-type protein at pH 7.0 (Pascher et al., 1993). It is of interest to note that rusticyanin, which has the highest reduction potential among the single small blue copper proteins (680 mV) has a valine residue in this position.

Umecyanin has a cysteine residue neighboring its C-terminal histidine ligand (His 101, Fig. 16A numbering), as is also the case for CPC, stellacyanin, and CBP. It has been shown in native stellacyanin that this cysteine is involved in a disulfide bridge with cysteine 65 (Engeseth et al., 1984). A similar disulfide bridge exists also in the isoforms of native umecyanin between Cys 65 and Cys 102. However, in the apoprotein, a disulfide

bridge is formed between Cys 96 and Cys 102 (Fig. 16A numbering). A disulfide switch is also observed in apo-stellacyanin and has been suggested to be initiated by an electron transfer between Cys 65 and Cys 96, which in the apoprotein is not any longer one of the four copper ligands (Engeseth et al., 1984). Such an electron transfer presumably occurs under the alkaline condition used during the proteolytic digests of the protein (Engeseth et al., 1984).

In the majority of small blue copper proteins, there are four intervening residues between the C-terminal Cys and His ligands (Cys 96 and His 101). In the CBP structure, this region is known to give a type II reverse β -turn, consistent with a Gly-Pro sequence. The corresponding β -turn in stellacyanin has a more type I character, but in umecyanin, the proline residue is replaced by a glycine and is followed by an aspartic acid. The situation is therefore like that in CPC, where the glycine is followed by a threonine. It is likely that a Gly-X sequence results in the formation of a more enclosed β -bend. The spacings between the C-terminal ligands in small blue copper proteins have been discussed recently (Dennison et al., 1994), and it appears that those small blue copper proteins exhibiting an active site protonation in the accessible range of pH (plastocyanin, amicyanin, and pseudoazurin) only have two intervening residues between the Cys and His ligands (Fig. 16A).

Another remarkable feature of the umecyanin sequence is that it contains, in comparison to those of rusticyanin, stellacyanin, and CBP, as well as those for the plastocyanins and azurins, a C-terminal extension of 10–14 residues. CPC also has an unusual C-terminal tail, but its amino acid sequence is totally different from that of umecyanin (Mann et al., 1992). Also, pseudoazurin has a C-terminal extension of which the total length is even 17 amino acids longer than for umecyanin. The tail in the latter sequence contains predominantly glycine and alanine residues, and a hydroxyproline two positions before the C-terminal alanine residue, whereas that in CPC is mainly constituted of proline and serine residues and is similar to the plant cell wall proteins known as extensins (Mann et al., 1992). Another protein family involved in the architecture of plants cell walls is the arabinogalactans, which are rich in hydroxyproline, serine, alanine, and glycine residues (McNeil et al., 1984). It thus appears that the C-terminal tail of umecyanin is more closely related to the arabinogalactans. The exact role of this part of umecyanin is unknown. The fact that several cleavage sites are detected in this region may be due to the special nature of the succeeding Ala and Gly residues. The partial proteolytic cleavages explain why the yield of intact C-terminal Lys-C or Arg-C peptides is very low compared to the peptides obtained from the polypeptide preceding them. An additional factor, although less dominant, may be that the hydrophobic nature of the C-terminal region results in adsorption losses on the reversed-phase column used for their isolation. It is likely that the proteolytic cleavage of the C-terminal region only occurs during the initial step of the purification of umecyanin, and that this tail anchors the protein to the cell wall in vivo. Extensive study on the precise location of the protein in vivo will be necessary to confirm this possibility.

Evolutionary position of umecyanin

On the basis of the alignment in Figure 16A, it appears that umecyanin has the highest sequence similarity to CPC, namely 46.7%. In this calculation, the C-terminal regions were omit-

ted (107–130) because they do not occur in the other small blue copper proteins. There is about 8.4% less similarity with stellacyanin and a further 8.1 less with CBP. There is very little similarity to the prokaryotic azurins and to the plastocyanins, which occur both in cyanobacteria and in the chloroplast of plants and algae (Ambler, 1974; Ramshaw & Felton, 1982): with azurin from *A. denitrificans* (Ambler, 1974; Baker, 1988) only 13 out of the 129 comparable residues are identical and with plastocyanin from cucumber (Ramshaw & Felton, 1982), this is 13 out of a total of 99 residues.

The prokaryotic small blue copper protein that has the highest sequence similarity with umecyanin is rusticyanin from *Thiobacillus ferrooxidans* (23.6%) and this is why its sequence has been included in Figure 16A. Rusticyanin has two remarkable internal insertions of, respectively, 10 and 50 residues when compared to the small blue copper protein from plant sources. Because such similar extensions have not yet been found in other prokaryotic members of this class of proteins, it is possible that they originate by gene insertion or gene fusion at a more recent stage of bacterial evolution, rather than by deletion from an ancestral longer gene. In Figure 17, an evolutionary tree is proposed that takes into account the percentage sequence similarities as well as the phenomenon of disulfide bridge formation between Cys 65 and Cys 102, the mutation of the fourth copper ligand methionine to glutamine, and the C-terminal extension in umecyanin and CPC. The amino acid sequence determined herein adds to the general concept that all the small blue copper proteins originate from a common ancestral gene (Ryden & Hunt, 1993). Stellacyanin and cucumber basic (blue) protein have been suggested to be members of the family of phytocyanins (Ryden & Hunt, 1993). Figure 17 shows that umecyanin and CPC, which also occur in the non-green parts of higher plants, can be considered to be part of this family.

Compared to stellacyanin, umecyanin has a rather low carbohydrate content (3.7% versus 40%) (Malmström et al., 1970; Stigbrand, 1971). The glycan core in umecyanin has the same composition and molecular mass as that of horseradish peroxidase (Clarke & Shannon, 1976; Green & Oliver, 1991), and it can be presumed that their sequences are identical: (GlcNAc-fucose)-GlcNAc-mannose₃-xylose. The glycan is bound near the end of β -strand 6. Assuming a similar three-dimensional structure, this is in the neighborhood of two of the three glycan cores in stellacyanin (Fig. 16B). Glycosylation is known to play a role either in the correct folding of a protein, in protein transport, or in the biological activity itself (Paulson, 1990). Because CBP appears to be nonglycosylated, the first possible role is not relevant to the copper proteins. Protein targeting and a role in the association with larger proteins may be involved, but there is at present no experimental evidence for this assumption. It has been proposed (Mann et al., 1992) that the glycosylation of CPC, which occurs near the unusual C-terminal end of the protein (Fig. 16B), protects this part of the protein against proteolytic cleavage. Alternatively, according to these authors, the glycosylation should be seen as a supporting structural element that keeps the extracellular N-terminal region 1–115 (Fig. 16B) away from the C-terminal region so as to ensure that the latter can easily interact with the components of the cell wall. Although the situation *in vivo* may be different from that *in vitro*, the first of these two possibilities appears less evident given the extensive proteolytic cleavage that we have detected in our preparation of umecyanin. On the other hand, one may wonder

whether the absence of glycosylation in the CBP is not at the basis of the fact that this protein lacks a C-terminal tail like the one in umecyanin and CPC. We suggest that an ancestral form of CBP might have been glycosylated at the same site as in stellacyanin (Asn 66-Asp-Thr 68), but that the glycosylation may have been lost through mutation leaving an Asn Thr sequence instead of the required Asn-X-Ser/Thr sequence.

As said earlier, the physiological role of umecyanin is not yet known. The protein was isolated as a co-product during the isolation of peroxidase from the peelings of horseradish roots (Paul & Stigbrand, 1970). Because two of the three cysteines in umecyanin are linked in a disulfide bridge, whereas the third acts as one of the four copper ligands, it is likely that the association rather occurs through hydrophobic interaction. Umecyanin cannot replace laccase in the reduction of oxygen by ascorbic acid (Paul & Stigbrand, 1970), but like other small blue copper proteins, it can be used as an electron donor instead of vitamin C if laccase is present (Stigbrand, 1971).

Materials and methods

Isolation and purification of umecyanin from horseradish roots

Umecyanin was isolated using a modified version of a method from the literature (Paul & Stigbrand, 1970). The whole horseradish root was used in this preparation and was homogenized in a Waring blender. The homogenized product was filtered and centrifuged and the protein precipitation steps were as in the original procedure.

The final stage of purification was carried out on a Pharmacia FPLC system using a Mono-Q column.

Electrophoresis and blotting

SDS-PAGE was performed in a Hoefer Mighty Small SE250 electrophoresis unit on a 12% T and 3% C gel, according to Laemmli (1970), with only one modification. We used piperazine diacrylamide (BioRad, Hercules, California) as crosslinker instead of bisacrylamide. The separated proteins were electroblotted in a BioRad Transblott Cell onto two layers of ProBlott membranes (Applied Biosystems, Foster City, California), using 50 mM Tris-borate buffer as blotting buffer. After blotting, the membranes were stained in strongly diluted Coomassie blue until the bands became clearly visible. They were allowed to air dry.

Protein modification

Native umecyanin (50 nmol) was reduced with DTT in 50 mM phosphate buffer, pH 7.1, containing 50 mM EDTA and 6 M guanidine hydrochloride. The reaction mixture was flushed with argon and incubated for 2 h at 37 °C. Carboxymethylation as described by Crestfield et al. (1963) was performed after cooling the reaction mixture in an ice bath for 30 min. The reduced protein was then carboxymethylated with 2.5 mM iodoacetic acid and incubated at room temperature for 30 min. Desalting was performed by ultrafiltration twice through Centricon-10 tubes (Amicon, Beverly, Massachusetts) using 2 mL of Milli-Q water (Millipore).

B

Similarity (%)

	RSC	CBP	UMC	STC	CPC
RSC		27.6	23.6	19.6	16.8
CBP	27		30.2	35.5	29.9
UMC	25	32		38.3	46.7
STC	21	38	41		35.5
CPC	18	32	50	38	

Number of identities

Fig. 17. A: Dendrogram of small blue class I blue copper proteins, as reconstructed from a simple pairwise matrix. The disulphide bridge of CPC has not yet been proven, but the high homology with umecyanin and stellacyanin suggests that it does exist. **B:** Number of identical residues in the cupredoxins mentioned in Figure 16A (left) and percentage (right).

A second batch of umecyanin (20 nmol) was reduced with a very small spatula tip of DTT; the color of the solution thereby changed immediately from blue to colorless. The reduced protein was then immediately precipitated by adding trichloroacetic acid to the solution in an ice bath. After centrifugation, the precipitate was washed twice with 200 μ L Milli-Q water and redissolved in a known volume of 0.1 M Tris-HCl, pH 8.52.

Enzymatic digestions

Carboxymethylated apoprotein (23 nmol) was digested with Lys-C endoproteinase from *Achromobacter lyticus* (Wako, Osaka, Japan) in 100 mM Tris-HCl buffer, pH 8.0, for 4 h at 37 °C, using an enzyme-to-substrate ratio of 1/65 (weight/weight). Digestion with Asp-N endoproteinase from *Pseudomonas fragi* (Sigma) was performed on 10 nmol of carboxymethylated apoprotein using an enzyme-to-substrate ratio of 1/500 in 100 mM Tris-HCl buffer, pH 8.0, at 37 °C during 8 h. Arg-C endoproteinase (Boehringer, Mannheim, FRG) digestion was carried out on 7.5 nmol of carboxymethylated apoprotein at an enzyme-to-substrate ratio of 1/20, in 0.1 M Tris-HCl buffer, pH 8.0. The incubation time was 2 h at 37 °C.

The three largest Arg-C fractions were subdigested with 0.5 μ g Lys-C endoproteinase for 5 h at 37 °C in 100 mM Tris-HCl buffer, pH 8.3.

Digestion of 10 nmol of the second batch of apoprotein was performed with Lys-C endoproteinase at 37 °C for 2 h with an enzyme-to-substrate ratio of 1/60.

Peptide and protein purification

Peptides obtained after enzymatic digestions were separated by HPLC on a BioRad Bio-Sil TMS 90-3 S column of 4.6 $\times$ 100 mm. The peptides were eluted with a gradient of 100% solvent A (0.1% TFA in Milli-Q water) to 100% solvent B (0.1% TFA in 70% ACN) at a flow rate of 1 mL/min. The chromatographic equipment consisted of an 870 three-headed piston pump (DuPont, Wilmington, Delaware), an 8800 gradient controller (DuPont), a UV variable spectrophotometer (DuPont) set at 220 nm, and a Rheodyne injector equipped with a 100- μ L loop. The eluted fractions were collected manually, dried immediately in a Savant Speedvac Concentrator, and stored at -18 °C. Prior to use, the dried fractions were redissolved in 50–100 μ L of 0.1% TFA in Milli-Q water.

Some digestion mixtures from subdigestions, as described above, were purified on the so-called "PTC C-18" column (2.1 $\times$ 220 mm) at a flow rate of 200 μ L/min, using a 140A solvent delivery system equipped with a 1000S photo diode array detector (all from Applied Biosystems). The eluents were the same as those used with the TMS column. The described chromatographic equipment was also used for the purification of the native protein.

Amino-terminal sequence and amino acid composition analysis

Automated N-terminal sequence analysis of peptides were performed in a Applied Biosystems pulsed liquid 477A and/or 475A sequenator, equipped with an on-line 120A PTH-amino acid analyzer. The retention time of PTH-Hyp was checked by apply-

ing 1.5 nmol of Hyp (Merck, Darmstadt, Germany) on the 477A sequenator. Sequence analysis of electroblotted proteins was carried out in a 476A sequencer with a specifically designed Blott cartridge (both from Applied Biosystems).

Amino acid composition analyses were performed on a 420A derivatizer (Applied Biosystems) with an on-line 130A separation system (Applied Biosystems). Gas phase hydrolysis was carried out in borosilicate glass tubes of 5 $\times$ 55 mm placed in a Pierce hydrolysis vial, using 6 N HCl as hydrolysis agent for 24 h at 106 °C.

Mass analysis

Plasma desorption mass spectra were collected on a BioIon 20K biopolymer mass analyzer (BioIon, Uppsala, Sweden) during 10^6 counts of the ^{252}Cf source at an accelerating potential of 15 kV. Samples were applied in small volumes of 0.1% TFA in Milli-Q water and spin-dried on nitrocellulose-coated targets wetted with 5 μ L 0.1% TFA in methanol.

Electrospray mass spectrometry was performed on a Bio-Q quadrupole mass spectrometer equipped with an electrospray ionization source (Fisons Instruments, Altrincham, UK). Ten microliters of sample solution in 50% ACN/1% acetic acid were injected manually in the 10- μ L loop of the Rheodyne injector and pumped to the source at a flow rate of 5 μ L/min. The solvent, 50% ACN/1% acetic acid, was delivered by the 140A solvent delivery system. Scans of 12 s over the mass range of 400–1600 a.m.u. were collected over 2 min. Calibration of the scans was performed with 50 pmol of horse heart myoglobin (Sigma).

MALDI-MS was carried out on a ToFSpec instrument using a nitrogen-laser (337 nm) and a linear time-of-flight tube of 65 cm (Fisons Instruments). Scans were accumulated over 40 laser shots, using α -cyano sinappinic acid as matrix. External calibration was realized using gramicidin S and bovine insulin (Sigma).

Carbohydrate composition analysis

Native umecyanin (1,400 pmol) was hydrolyzed in 3 M TFA for 2 h at 120 °C in Pierce Reacti-Vial. After hydrolysis, the sample was dried in a Savant Speedvac concentrator and redissolved in 50 μ L of 0.5 M PMP (Sigma) and 50 μ L of 0.3 M NaOH, according to Honda et al. (1989). After incubation for 30 min at 60 °C, 15 μ L of 1 M HCl was added to this solution and all solvents were removed in a Speedvac concentrator. The sample was then reconstituted twice in 80 μ L of chloroform and 160 μ L of Milli-Q water. After vortexing, the aqueous phase was stored for analysis. Upon analysis, the collected aqueous phase was dried in a Speedvac concentrator and redissolved in 400 μ L of 0.1% TFA in Milli-Q water. Analytical separations were carried out on the PTC C18-column (2.1 $\times$ 220 mm) at 43 °C in a home-made oven, using the 140A solvent delivery system and the 1000S photo diode array detector. Injections were done manually with a Rheodyne injector using a 100- μ L loop at a flow rate of 200 μ L/min. Detection occurred at 245 nm. The solvents used were: 0.1% TFA in Milli-Q water (solvent A) and 0.1% TFA in 90% acetonitrile (solvent B). The gradient program was: 5% B for 5 min, 5% to 35% B over 35 min and 35% to 100% B over 1 min. Calibration was made with PMP-derivatized glucosamine, mannose, fucose, and xylose. Ribonuclease A and B and horseradish peroxidase (Sigma) were used as test proteins.

Determination of disulfide bridges

Disulfide bridge in apo-umecyanin

After mass analysis on ESMS and MALDI-MS, fraction K'8 from the Lys-C digest on the second batch of umecyanin was subdigested with 0.08 µg Asp-N endoproteinase (Boehringer, Mannheim, Germany) in 25 mM Tris-HCl buffer, pH 8.1. After incubation at 37 °C for 2 h, the reaction mixture was chromatographed with the Pharmacia SMART system on a C18 column (3.2 × 100 mm). The solvents were: 0.05% TFA in Milli-Q water (solvent A) and 0.045% TFA in 70% acetonitrile (solvent B). The eluted peptides were collected automatically with the built-in fraction collector of the SMART system and were subjected again to both mass analyzers.

Disulfide bridge in native umecyanin

To 7 nmol of native umecyanin, dissolved in 200 µL of 50 mM phosphate buffer, pH 7.0, 100 µL of 50 mM iodoacetic acid was added and left in the dark at room temperature for 30 min. The reaction mixture was then brought to about 45 °C for 3 min in a water bath and incubated overnight at 4 °C. The mixture was further incubated at room temperature for 5 h in the dark and desalted by gel filtration over Superdex X75 3.2/30 (Pharmacia, SMART system), eluted with 10 mM ammonium bicarbonate, pH 7.4. The collected fraction was digested with 0.8 µg of Asp-N endoproteinase in 25 mM Tris-HCl, pH 8.0, at 37 °C for 3 h. After incubation, the generated fractions were separated by gel filtration on Superdex X30 3.2/30 (Pharmacia, SMART system), equilibrated and eluted with 50 mM ammonium bicarbonate pH 7.8.

After sequence analysis, a selected fraction was dried and subdigested with 3 µg Lys-C endoproteinase in 50 mM Tris-HCl, pH 8.0, for 2 h at 37 °C and separated by RP-HPLC on µRPC C2/C18 PC 3.2/3 (Pharmacia SMART system) using the elution conditions described above. After PDMS analysis, reduction with DTT was performed on the same amount of sample in 10 mM ammonium bicarbonate, pH 9, at 37 °C for 15 min.

Computer graphics

Observation of the α -carbon backbone and the secondary structure of CBP (Brookhaven file 1CBP) was performed on an Apple Quadra 950 using MacIcmdad 4.0 software (Molecular Application Group, Palo Alto, California).

Acknowledgments

This work is supported by the Belgian Concerted Research Action Program (contract 12052293 to J.V.B.) and by the UK Science and Engineering Research Council (to C.D.). We also thank Bart Devreese for his skillful help with the mass spectrometric analyses.

References

Adman ET. 1991. Copper protein structures. *Adv Protein Chem* 42:145-197.
Aikazyan VT, Nalbandyan RM. 1979. Copper-containing proteins from *Cucumis sativus*. *FEBS Lett* 104:127-130.
Ambler RP. 1974. Sequence submitted to ATLAS. Accession number A00286, reference number A00088.
Baker EN. 1988. Structure of azurin from *Alcaligenes denitrificans*. Refinement at 1.8 Å resolution and comparison of the two crystallographically independent molecules. *J Mol Biol* 203:1071-1075.
Bergman C. 1980. [thesis]. Göteborg: Chalmers University of Technology.

Bergman C, Gandvik EK, Nyman PO, Strid L. 1977. The amino acid sequence of stellacyanin from the lacquer tree. *Biochem Biophys Res Commun* 77:1052-1059.
Chapman SK, Orme-Johnson WH, McGinnis J, Sinclair-Day JD, Sykes AG, Ohlsson PI, Paul KG. 1986. Kinetic, spectroscopic and structural (extended X-ray absorption fine structure) studies on the type I blue copper protein umecyanin. *J Chem Soc Dalton Trans*:2063-2068.
Clarke J, Shannon LM. 1976. The isolation and characterization of the glycopeptides from horseradish peroxidase isoenzyme C. *Biochim Biophys Acta* 427:428-42.
Crestfield AM, Moore S, Stein WH. 1963. The preparation and enzymatic hydrolysis of reduced and S-carboxymethylated proteins. *J Biol Chem* 231:622.
Dennison C, Kohzuma T, McFarlane W, Suzuki S, Sykes AG. 1994. Reversible active site protonation and electron transfer properties of *Achromobacter cycloclastes* pseudoazurin: Comparison with other type I copper proteins. *J Chem Soc Commun*:581-582.
Engeseth HR, Hermodson MA, McMillin DR. 1984. A new assignment of the disulphide linkage in stellacyanin. *FEBS Lett* 171:257-261.
Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1990. Electrospray ionisation—Principles and practice. *Mass Spectrom Rev* 9:37-70.
Fields BA, Guss JM, Freeman HC. 1991. Three-dimensional structure for stellacyanin, a "blue" copper protein. *J Mol Biol* 222:1053-1065.
Green BN, Oliver RWA. 1991. The study of intact proteins and glycoproteins by electrospray m.s. *Biochem Soc Trans* 19:29-935.
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.
Har J, Loehr T, Lu Y, Valentine JS, Averill BA, Sanders-Loehr J. 1993. Resonance Raman excitation profiles indicate multiple Cys-Cu charge transfer transitions in type I copper proteins. *J Am Chem Soc* 115:4256-4263.
Hoitink CWG, Canters GW. 1992. The importance of Asn 47 for structure and reactivity of azurin from *Alcaligenes denitrificans* as studied by site-directed mutagenesis and spectroscopy. *J Biol Chem* 267:13836-13842.
Honda S, Akao E, Suzuki S, Okuda M, Kakehi K, Okuda M, Nakamura J. 1989. High performance liquid chromatography of reducing carbohydrates as strongly ultraviolet-absorbing and electrochemically sensitive 1-phenyl-3-methyl-5-pyrazolone derivatives. *Anal Biochem* 180:351-357.
Ingledew WJ, Cobley JC. 1980. A potentiometric and kinetic study on the respiration chain of ferrous-ion-grown *Thiobacillus ferrooxidans*. *Biochim Biophys Acta* 590:141-158.
Kato S, Shiratori I, Takamiya A. 1962. Purification and some properties of spinach plastocyanin. *J Biochem* 51:32-40.
Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685.
Lu Y, Lacroix LB, Lowery MD, Solomon EI, Bender CJ, Peisach J, Roe JA, Gralla EB, Valentine JS. 1993. Construction of a blue copper site at the native zinc site of yeast copper-zinc superoxide dismutase. *J Am Chem Soc* 115:5907-5918.
Malmström BG, Reinhammar B, Vänngård T. 1970. State of copper in stellacyanin and laccase from lacquer tree *Rhus vernicifera*. *Biochim Biophys Acta* 205:48-57.
Mann K, Schäfer W, Thoenes U, Messerschmidt A, Mehrabian Z, Nalbandyan R. 1992. The amino acid sequence of a type I copper protein with an unusual serine- and hydroxyproline-rich C-terminal domain isolated from cucumber peelings. *FEBS Lett* 314:220-223.
Marchesini A, Kroneck PMH. 1979. Ascorbate oxidase from *Cucurbita pepo medullosa*. New method of purification and reinvestigation of properties. *Eur J Biochem* 101:65-76.
Markossian KA, Aikazyan VT, Nalbandyan RM. 1974. Two copper-containing proteins from cucumber (*Cucumis sativus*). *Biochim Biophys Acta* 359:47-54.
McNeil M, Darvill AG, Fry SC, Albersheim P. 1984. Structure and function of the primary cell walls of plants. *Annu Rev Biochem* 53:625-663.
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.
Pascher T, Karlsson BG, Nordling M, Malmström BG, Vänngård T. 1993. Reduction potentials and their pH dependence in site-directed mutant forms of azurin from *Pseudomonas aeruginosa*. *Eur J Biochem* 212:289-296.
Paul KG, Stigbrand T. 1970. Umecyanin, a novel intensely blue copper protein from horseradish root. *Biochim Biophys Acta* 221:255-263.
Paulson JC. 1990. Glycoproteins: What are the sugar chains for. In: Bradshaw RA, Purton M, eds. *Proteins: Form and function*. Cambridge, UK: Elsevier Trends Journal. pp 209-217.

- Ramshaw JAM, Felton A. 1982. The amino acid sequence of plastocyanin from *Cucumis sativus*. *Phytochemistry* 21:1317–1320.
- Reinhammar B. 1972. Oxidation–reduction potentials of the electronacceptors in laccases and stellacyanin. *Biochim Biophys Acta* 275:245–259.
- Romero A, Hoitink CWG, Nar H, Huber R, Messerschmidt A, Canters GW. 1993. X-ray analysis and spectroscopic characterization of M121Q azurin. A copper site model for stellacyanin. *J Mol Biol* 229:1007–1021.
- Ronk M, Shively JE, Shute EA, Blake RC II. 1991. Amino acid sequence of the blue copper protein rusticyanin from *Thiobacillus ferrooxidans*. *Biochemistry* 30:9435–9442.
- Ryden L. 1984. Structure and evolution of the small blue proteins. In: Lontie R, ed. *Copper proteins and copper enzymes, vol I*. Boca Raton, FL: CRC Press. pp 157–182.
- Ryden L, Björk I. 1976. Reinvestigation of some physico-chemical and chemical properties of human ceruloplasmin (ferroxidase). *Biochemistry* 15: 3411–3417.
- Ryden L, Hunt LT. 1993. Evolution of protein complexity: The blue copper-containing oxidases and related proteins. *J Mol Evol* 36:41–66.
- Sakurai T, Ikeda O, Suzuki S. 1990. Direct electrochemistry of the blue copper proteins pseudoazurin, plantacyanin and stellacyanin. *Inorg Chem* 29:4715–4718.
- Sørensen HH, Thomsen J, Bayne S, Højrup P, Roepstorff P. 1990. Strategies for determination of disulphide bridges in proteins using plasma desorption mass spectrometry. *Biomed Environ Mass Spectrom* 19:713–720.
- Stigbrand T. 1971. Structural properties of umecyanin, a copper protein from horseradish root. *Biochim Biophys Acta* 236:246–252.
- Stigbrand T. 1972. Oxidation–reduction potential of umecyanin. *FEBS Lett* 23:41–43.
- Stigbrand T, Malmström BG, Vänngård T. 1971. On the state of copper in the blue protein umecyanin. *FEBS Lett* 12:260–262.
- Stigbrand T, Sjöholm I. 1972. Circular dichroism studies on the copper protein umecyanin. *Biochim Biophys Acta* 263:244–257.
- Sykes AG. 1990. Active-site properties of the blue copper proteins. *Adv Inorg Chem* 36:377–408.
- Thomann H, Bernardo M, Baldwin MJ, Lowery MD, Solomon EI. 1991. Pulsed ENDOR study of the native and high pH perturbed forms of the blue copper site in stellacyanin. *J Am Chem Soc* 113:5911–5913.
- Van Beeumen J, Van Bun S, Canters GW, Lommen A, Chothia C. 1991. The structural homology of amicyanin from *Thiobacillus versutus* to plant plastocyanins. *J Biol Chem* 266:4869–4877.
- Van de Kamp M, Canters GW, Andrew CR, Sanders-Loehr J, Bender CJ, Peisach J. 1993. Effect of lysine ionization on the structure and electrochemical behaviour of the Met 44 → Lys mutant of the blue copper protein azurin from *Pseudomonas aeruginosa*. *Eur J Biochem* 218:229–238.
- Vickery LE. 1971. [thesis]. Santa Barbara, California: University of California. [Cited in: Structure and function of small blue copper proteins. In: Harrison P, Ed. *Metalloproteins*. Part 1: Metal proteins with redox roles. Weinheim, Basel: Verlag Chemie. pp 1–42.]

Note added in proof

The following paper, which was published in December 1994, indicates that the fourth copper ligand of rusticyanin is probably His 91 (following the numbering of Fig. 16A) and not His 50, as indicated in our Figure 16A.

Hunt AH, Toy-Palmer A, Assa-Munt N, Cavanaugh J, Blake RC II, Dyson JH. 1994. Nuclear magnetic resonance ^{15}N and ^1H resonance assignments and global fold of rusticyanin. Insights into the ligation and acid stability of the blue copper site. *J Mol Biol* 244:370–384.