Volume 314, number 3, 220–223 FEBS 11867 December 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

Karlheinz Mann[a], Wolfram Schäfer[a], Ulrich Thoenes[a], Albrecht Messerschmidt[a], Zara Mehrabian[b] and Robert Nalbandyan[b]

[a]*Max-Planck-Institut für Biochemie, 8033 Martinsried, Germany* and [b]*Institute of Biochemistry, Academy of Sciences of the Armenian Republic, 375044 Yerevan, Armenia*

Received 29 September 1992

We have determined the amino acid sequence of a small copper protein isolated from cucumber peelings. This cupredoxin contains 137 amino acids including a pyroglutamate as the first residue. The N-terminal 110 amino acid-long domain shows 30–37% identity to 2 other cupredoxins, stellacyanin and cucumber basic blue protein. A unique feature of this protein is a 27 amino acid-long C-terminal domain rich in 4-hydroxyproline and serine and resembling certain plant cell wall proteins. The prolines in this domain are hydroxylated to a different extent depending on the surrounding sequence.

Amino acid sequence; Copper protein; Stellacyanin; Pyroglutamate; 4-Hydroxyproline

1. INTRODUCTION

Plants and bacteria contain small copper-containing proteins (cupredoxins) involved in electron transfer reactions. These proteins bind copper at coordination sites classified as type I because of certain characteristic spectral properties, notably an intense absorption near 600 nm and an unusual EPR spectrum [1]. The three-dimensional structure of several of these cupredoxins, azurin, pseudoazurin, plastocyanin and cucumber basic blue protein, has been determined. These known cupredoxins are single-domain proteins containing one copper atom bound by two histidines, a cysteine and a methionine in a distorted tetrahedral arrangement [1].

We have determined the amino acid sequence of a cupredoxin isolated from cucumber peelings, which, in addition to a copper-containing domain, has a short C-terminal domain rich in serine and 4-hydroxyproline. Composition and sequence of this domain resemble cell wall proteins of the extensin family and related proteins.

The protein we describe has been isolated in the laboratory of one of us (Nalbandyan) under the name of stellacyanin [2] because of spectroscopic properties similar to that of laquer tree stellacyanin. This similarity probably has something to do with the nature of copper-binding residues at the coordination site. Both proteins seem to use a glutamine instead of methionine as a ligand (see below). However, because we think that the sequences of the two proteins are too different to appear under the same name and because nothing is known about the function of both proteins, we propose the name cucumber peeling cupredoxin (CPC) for the protein described here.

Correspondence address: K. Mann, Max-Planck-Institut für Biochemie, D-8033 Martinsried, Germany. Fax: (49) (89) 85782422.

2. MATERIALS AND METHODS

The cupredoxin was isolated from cucumber (*Cucumis sativus*) peelings according to Aikazyan and Nalbandyan [2]. The purified protein was reduced and carboxymethylated under denaturing conditions [3]. Reagents and buffer salts were removed by reversed-phase chromatography on a C4 column (0.46 × 25 cm, Vydac, Hesperia, California) using 0.1% trifluoroacetic acid (solvent A) and 70% acetonitrile in 0.1% trifluoroacetic acid (solvent B) with a gradient from 0–60% B in 160 min and from 60–95% B in 40 min. The flow rate was 0.25 ml/min. Enzymatic cleavages were done with 50–100 μg protein in 0.2 M ammonium hydrogen carbonate buffer at an enzyme to substrate ratio of 1:100, with trypsin (TPCK-treated; Worthington, Freehold, New Jersey) for 6 h at 30°C, with α-chymotrypsin (Worthington) for 4 h at 30°C, and with endoproteinase Asp-N (sequencing grade; Boehringer Mannheim, Germany) for 6 h at 30°C.

The resulting peptides were separated on a C18 reversed-phase column (0.46 × 25 cm, Vydac) using the same solvents and gradient as above. The blocked N-terminal tryptic peptide was further cleaved with 1 μg of thermolysin (E. Merck, Darmstadt, Germany) for 2 h at 30°C in 0.2 M ammonium hydrogen carbonate, and with 1 μg of pyroglutamate aminopeptidase (sequencing grade; Boehringer Mannheim, Germany) [4]. The C-terminal peptide resulting from endoproteinase Asp-N cleavage was further cleaved with 1 μg subtilisin (Boehringer Mannheim, Germany) in 0.2 M ammonium hydrogen carbonate. The resulting peptides were separated by C18 reversed-phase chromatography as above. The tryptic peptides of the entire protein were also separated on a Mono S cation-exchange column (HR 5/5, Pharmacia) at pH 2.5 [5]. Peptide T2 was further cleaved with CNBr following pretreatment with 5% β-mercaptoethanol in 0.2 M

Published by Elsevier Science Publishers B.V.

ammonium hydrogen carbonate for 24 h at 23°C. The dried peptide was dissolved in 75% formic acid and a small crystal of CNBr was added. The mixture was incubated for 24 h at 23°C under nitrogen in the dark and then lyophilized. The peptides were separated by HPLC as above.

Amino acid compositions were determined after hydrolysis for 24 h in 6 M HCl at 110°C on a LC 5001 analyzer (Biotronik, Germany). Sequence analysis by Edman degradation was done with Applied Biosystems sequencers 470A and 473A following the manufacturer's instructions. SDS-PAGE was done according to [6]. For the electrophoretic comparison of intact and deglycosylated cupredoxin, 1 mg of lyophilized protein was further dried over P_2O_5 for 16 h and then deglycosylated with trifluoromethanesulfonic acid as in [7] with an incubation time of 3 h. Fast atom bombardment mass spectra were recorded with a mass spectrometer MAT 900 (Finnigan MAT, Bremen) equipped with a liquid secondary ion-ionization system.

3. RESULTS AND DISCUSSION

3.1. *Preparation of peptides and sequence analysis*

The purified cucumber peeling cupredoxin (CPC) migrated in SDS gel electrophoresis as a broad band with M_r ~24,000–28,000 (Fig. 1). N-terminal sequencing did not yield any PTH-amino acid derivative, indicating that the N-terminal amino acid of this protein was blocked. In order to obtain peptides for internal sequencing, 50–100 μg portions of the carboxymethylated protein were cleaved with trypsin and endoproteinase Asp-*N*. The resulting peptides were separated by reversed-phase HPLC as shown for the tryptic digest in Fig. 2. Both cleavage mixtures contained one peak which did not give a sequence and thus contained the

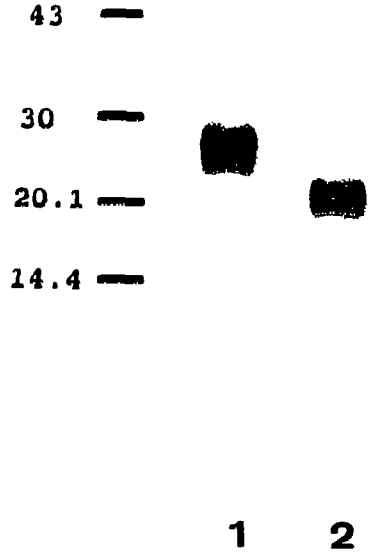

Fig. 1. SDS-PAGE of cucumber peeling cupredoxin before (lane 1) and after deglycosylation (lane 2) with trifluoromethanesulfonic acid on a 10–20% gradient gel.

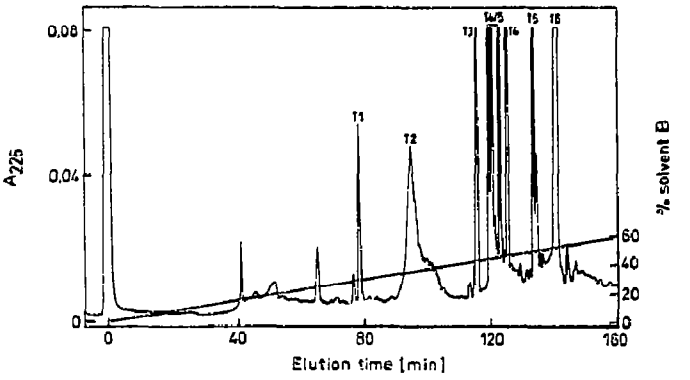

Fig. 2. Reversed-phase HPLC of tryptic peptides of cucumber peeling cupredoxin. Peptides T4 and T5 were found in several peaks for unknown reasons. The broad peak T2 contained the C-terminal peptide and peak T6 the N-terminal peptide.

blocked N terminus (EA2 and T6 in Fig. 3). The blocked peptide T6 was further cleaved with thermolysin and the separated peptides were sequenced. Together with the other peptides from these two sets of cleavage products this already established most of the sequence presented in Fig. 3. Amino acid analysis of the blocked peptide EA2 gave the composition Glu, Gly, His, Ile, Ser, Thr, Val_2. The mass of this peptide as determined by fast atom bombardment mass spectrometry was 822.5 Da, i.e. 18 Da less than expected from its composition, indicating pyroglutamate as the N-terminal residue. Subsequent treatment of peptide T6 with pyroglutamate aminopeptidase unblocked the N terminus, allowing sequence analysis to begin with the second residue (Fig. 3).

Peptides T2 and EA4 showed a gap in the sequence in the position corresponding to residue 109 in the complete protein (Fig. 3). This non-identified residue was followed by Thr in position 111 and could thus correspond to the consensus sequence for *N*-glycosylation, Asn-Xaa-Ser/Thr. The identity of residue 109 as a glycosylated Asn was also strongly indicated by amino acid analysis of peptide T2 which contained two Asx, one for position 104 and one for 109.

When a tryptic digest of the cupredoxin was separated on the strong cation exchanger Mono S at pH 2.5, only peptide T2 was not bound, indicating that this peptide contained no basic residue and thus represented the C terminus. Subsequently this rapid purification method was used to isolate the C-terminal peptide for further investigation. Sequence analysis of peptide T2 did not allow identification of the C-terminal residue unequivocally due to rapidly decreasing yields of PTH-amino acids at sites of Ser accumulation. However, cleavage with CNBr resulted in peptides which could be sequenced and allowed elucidation of the whole sequence of T2 (Fig. 3). This sequence was in good agreement with the amino acid composition of this peptide (not shown). Besides expected CNBr peptides, we also

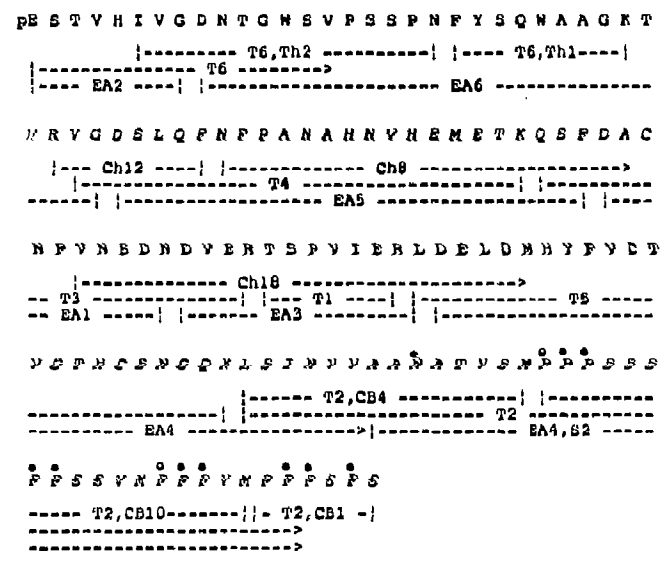

Fig. 3. Amino acid sequence of cucumber peeling cupredoxin. Broken lines indicate the extension of sequenced peptides, letters refer to the enzyme used for cleavage to obtain the indicated peptides: T, trypsin; Th, thermolysin; EA, endoprotease Asp-*N*; Ch, chymotrypsin; S, subtilisin. CB indicates cleavage by CNBr. ○, partially hydroxylated proline; ●, completely hydroxylated proline; pE, pyroglutamate; *, glycosylated asparagine.

separated and sequenced substantial amounts of longer peptides due to partial cleavage at methionines. This is usually caused by oxidation of this amino acid which prevents the reaction with CNBr from taking place. Actually, amino acid analysis of the whole protein as well as unreduced peptide T2 showed almost no genuine methionine, but only faster eluting oxidation products. Reduction with mercaptoethanol was, however, partially successful, since cleavage of the reduced peptide T2 gave rise to the peptides shown in Fig. 3. Because mass determination of T2 with fast atom bombardment mass spectrometry was not possible, we selected peptide T4 for this experiment. The result clearly indicated that a substantial part, if not all, of Met[55] was oxidized to the sulfoxide, suggesting that this may also be the case for all other methionines. Finally some very short overlaps between tryptic and endoproteinase Asp-*N*-produced peptides were extended with the aid of some chymotryptic peptides to ensure that no small peptide had been overlooked (Fig. 3).

The complete sequence contained 137 amino acids with a calculated mass of 14,748 Da for the unmodified protein. Even with one *N*-linked carbohydrate structure of probably 1,400–1,800 Da this is considerably less than the apparent molecular mass estimated from SDS gels. However, we think that the presence of carbohydrate and the unusual composition of the C-terminal sequence can account for this difference. When the carbohydrate was removed with trifluoromethanesulfonic acid, which leaves only the linkage sugar, the M_r was reduced to ~20,000 (Fig. 1).

3.2. *Comparison to other proteins*

The sequence of cucumber peeling cupredoxin (CPC) was clearly divided into two domains, the Cu-containing domain (residues 1–110) and a shorter C-terminal tail rich in serine and 4-hydroxyproline. When compared with the FASTP program [8] to known protein sequences compiled in the MIPSX (Martinsried Institute for Protein Sequences) data collection, the N-terminal domain of CPC was found to be very similar to two other cupredoxins, stellacyanin from the lacquer-tree [9] and basic blue protein from cucumber seedlings (CBP) [10] (Fig. 4). CPC (res. 1–110) was ~36% identical to stellacyanin and ~30% identical to CBP, while the identity between CBP and stellacyanin was ~37%.

The three-dimensional structure of CBP, as determined by X-ray diffraction, has shown the single copper atom to be bound by His[39], Cys[79], His[84] and Met[89] (Fig. 4) [11]. These residues were also conserved in other small copper proteins like azurin, pseudoazurin and plastocyanin, which have been shown to have essentially the same three-dimensional structure as CBP [1] but the sequences of which are not similar enough to CPC, CBP or stellacyanin in order to be detected by the FASTP program. Sequence alignment of stellacyanin, CBP and CPC suggests that all three cupredoxins have conserved the two His and the Cys but the fourth ligand, Met, is replaced by Gln (Fig. 4) in both stellacyanin and CPC. Gln[97] has already been suggested as the

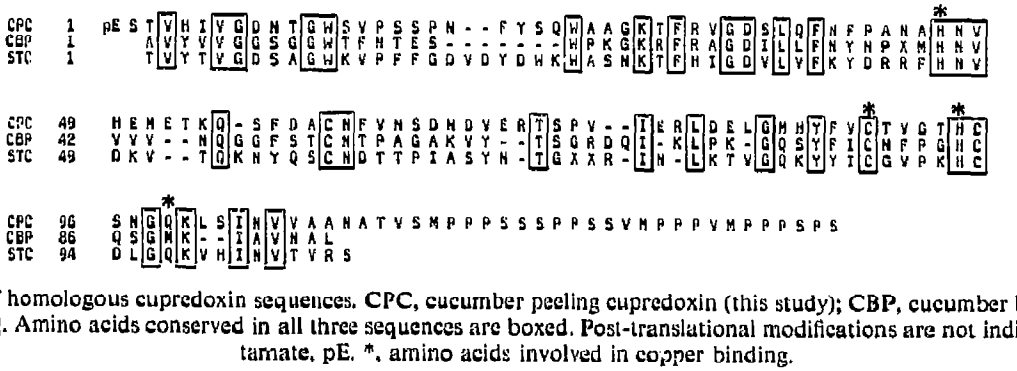

Fig. 4. Alignment of homologous cupredoxin sequences. CPC, cucumber peeling cupredoxin (this study); CBP, cucumber basic blue protein [10]; STC, stellacyanin [9]. Amino acids conserved in all three sequences are boxed. Post-translational modifications are not indicated, except pyroglutamate, pE. *, amino acids involved in copper binding.

fourth ligand in stellacyanin on the basis of model building experiments [12]. This was supported by a recent pulsed electron nuclear double resonance (ENDOR) study [13] which showed an amide to be involved in copper binding. Besides the Cys involved in copper binding, all three proteins have two more cysteines, which have been shown in stellacyanin [14] and CBP [11] to form a disulfide linkage.

The C-terminal domain of CPC is unique for a cupredoxin. It is about 27 residues long and contains blocks of hydroxylated prolines and serines resembling sequences found in plant cell-wall glycoproteins of the extensin family and related proteins [15,16], where similar repeats occur multiple times. The plant enzyme responsible for the hydroxylation of proline to 4-hydroxyproline is rather different from its animal counterpart. While the latter specifically hydroxylates prolines in the sequence Gly-Xaa-Pro-Gly [17], the plant hydroxylase seems to prefer substrates containing the sequence Ser-$(Pro)_n$ with n = 1–4 [18]. Furthermore, the latter sequence has to be present in a specific conformation [19] which is thought to be the polyproline II conformation [20], consisting of a left-handed helix of three residues per turn and a rise of 0.31 nm per residue [21]. This is different from the hydroxylation of Gly-Xaa-Pro-Gly sequences in collagen which occurs in the unfolded polypeptide chain. The C-terminal tail of CPC contains 4 short blocks of consecutive prolines and a single proline. Two of the prolines are preceded by Ser, the others by Met. Among the first prolines following Ser or Met, only those following Ser are completely hydroxylated, while the first proline following Met is hydroxylated to ~10% only (Pro^{115}, Pro^{127}) or not at all (Pro^{132}). Thus, the sequence specificity of the responsible hydroxylase is not absolute, but Ser-$(Pro)_n$ is the preferred substrate. In cell-wall proteins most of the hydroxyprolines were usually modified by the attachment of 1–4 arabinose residues [15]. One indication for the occurrence of at least partial glycosylation in CPC was the observation that the second hydroxyproline in the short proline/hydroxyproline blocks was found in rather low yields (~50%) during sequencing, when compared with neighbouring residues. This point will need further examination.

Nothing is known about the morphological location of CPC apart from the fact that it is contained in the peelings. However, the similarity of the C-terminal domain with the much longer hydroxyproline-rich cell wall proteins suggests that this protein with unknown function is also an extracellular protein which may interact with other cell wall components using the short C-terminal domain. The glycosylated Asn^{109} is located exactly between the two different domains, thus separating the globular copper-containing domain from the C-terminal domain, which is probably in the polyproline II conformation and therefore takes the form of a short rigid rod. The function of the carbohydrate could be either to protect the linking region against proteases or to keep the rigid rod away from the globular domain in order to ensure easy access to other cell wall components.

Acknowledgements: We thank W. Straßhofer for expert technical assistance and Drs. Rupert Timpl and Robert Huber for helpful comments on this manuscript.

REFERENCES

[1] Adman, E.T. (1991) Adv. Protein Chem. 42, 145–197.
[2] Aikazyan, V.T. and Nalbandyan, R.M. (1979) FEBS Lett. 104, 127–130.
[3] Glazer, A.N., Delange, R.J. and Sigman, D.S., Chemical Modification of Proteins, Elsevier, Amsterdam, 1975.
[4] Podell, D.N. and Abraham, G.N. (1978) Biochem. Biophys. Res. Commun. 81, 176–185.
[5] Mann, K., Jander, R., Korsching, E., Kühn, K. and Rauterberg, J. (1990) FEBS Lett. 273, 168–172.
[6] Laemmli, U.K. (1970) Nature 227, 680–685.
[7] Sojar, H.T. and Bahl, O.P. (1987) Arch. Biochem. Biophys. 259, 52–57.
[8] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
[9] Bergman, C., Gandvik, E.-K., Nyman, P.O. and Strid, L. (1977) Biochem. Biophys. Res. Commun. 77, 1052–1059.
[10] Murata, M., Begg, G.S., Lambrou, F., Leslie, B., Simpson, B.J., Freeman, H.C. and Morgan, F.J. (1982) Proc. Natl. Acad. Sci. USA 79, 6434–6437.
[11] Guss, M., Merritt, E.A., Phizackerley, R.P., Hedman, B., Murata, M., Hodgson, K.O. and Freeman, H.C. (1988) Science 241, 806–811.
[12] Fields, B.A., Guss, J.M. and Freeman, H.C. (1991) J. Mol. Biol. 222, 1053–1065.
[13] Thomann, H., Bernardo, M., Baldwin, M.J., Lowery, M.D. and Solomon, E.J. (1991) J. Am. Chem. Soc. 113, 55911–55913.
[14] Engseth, H.R., Hermodson, M.A. and McMillin, D.R. (1984) FEBS Lett. 171, 257–261.
[15] Showalter, A.M. and Varner, J.E., in: The Biochemistry of Plants, Vol. 15, (A. Marcus, Ed.), Academic Press, New York, 1989 pp. 485–520.
[16] Showalter, A.M. and Rumeau, D., in: Organization and Assembly of Plant and Animal Extracellular Matrix (W.S. Adair and R.P. Mecham, Eds.), pp. 247–281, Academic Press, New York.
[17] Kivirikko, K., Myllylä, R., Philajaniemi, T. (1989) FASEB J. 3, 1605–1617.
[18] Robinson, D.G., Andreae, M. and Blankenstein, P., in: Organization and Assembly of Plant and Animals Extracellular Matrix (W.S. Adair and R.P. Mecham, Eds.), Academic Press, New York, 1990, pp. 283–299.
[19] Tanaka, M., Sato, K. and Uchida, T. (1981) J. Biol. Chem. 256, 11397–11400.
[20] van Holst, G.-J. and Varner, J.E. (1984) Plant Physiol. 74, 247–251.
[21] Sasisekharan, V. (1959) Acta Chrystallogr. 12, 897–903.