Volume 314, number 3, 232–236 FEBS 11870 December 1992

Processing of the chloroplast transit peptide of pea carbonic anhydrase in chloroplasts and in *Escherichia coli* Identification of two cleavage sites

Inga-Maj Johansson and Cecilia Forsman

Department of Biochemistry, University of Umeå, S-901 87 Umeå, Sweden

Received 5 October 1992

The chloroplast transit peptide (cTP) of pea carbonic anhydrase was shown to be processed at two different sites, giving protein subunits of two sizes. The cleavage sites were identified and found to be localized immediately before and after a highly charged part, containing 8 acidic and 6 basic residues, of the cTP. Properties of pea carbonic anhydrase produced in *Escherichia coli* show that folding, oligomerization and catalytic activity do not depend on the presence of the acidic part or the rest of the cTP. The pattern of processing of the cTP in *E. coli* indicates that cleavage at site I is specific for a chloroplastic stromal peptidase and that cleavage at site I prevents processing at site II.

Carbonic anhydrase; Chloroplast; Transit peptide; Precursor processing; Cleavage site

1. INTRODUCTION

Carbonic anhydrase (CA; carbonate hydrolyase, EC 4.2.1.1) is a zinc-containing enzyme catalyzing the reversible hydration of CO_2. Mammalian CA's have been thoroughly investigated, while CA from higher plants is less well characterized and its physiological function is poorly understood. CA from dicotyledonous plants is believed to be an oligomer composed of six identical subunits, each of which binds one zinc ion [1]. In C_3 plants CA is found as a soluble chloroplastic protein, synthesized in the cytoplasm and transported across the envelope membranes. The majority of the proteins found in higher plant chloroplasts are nuclear encoded and targeted to the stroma by an N-terminal chloroplast transit peptide (cTP) that is subsequently removed by stromal peptidases (see [2] for a review). Common features of cTP's from higher plants are a high content of the hydroxylated residues, Ser and Thr, and very few acidic residues. In addition, the N-terminal part tends to lack charged residues, as well as Gly and Pro, and the C-terminal region seems to have a potential for forming an amphiphilic β-strand [3].

cDNA sequences coding for spinach CA [4,5] and pea CA [6,7] have been published and the derived amino acid sequences show 70% identity. A comparison between those sequences and N-terminal sequences from mature proteins shows that the cTP's are rather long, 98 and 105 residues for spinach and pea CA, respectively. The first part has the characteristics of cTP's mentioned above, but remarkably the C-terminal 30–40 amino acids contain a high number of acidic and basic residues with a net negative charge. Fawcett et al. [5] speculated that this acidic part would not be included in the cTP, meaning that there must be a stepwise processing of the precursor.

Processing intermediates have been found for the chloroplastic proteins pea Rubisco small subunit [8] and L18 ribosomal protein from *Chlamydomonas* [9]. However, the intermediate cleavage sites were not determined. In mitochondria two distinct matrix-localized protease activities, responsible for a two-step processing of some matrix targeting peptides, have been reported [10]. The reason(s) for this stepwise processing has been speculated upon. von Heijne et al. [3] proposed that the N-terminal extension is required for correct folding of the mature protein or for assembly into larger multisubunit complexes.

In this report we demonstrate that CA purified from pea leaves contains subunits of two sizes corresponding to two cleavage sites, one before and one after the acidic peptide. We have also cloned and expressed pea CA in *Escherichia coli*. By modifying the expression plasmid we have been able to produce CA carrying different lengths of the transit peptide. We have purified these proteins and used them to investigate the function of the acidic peptide and to test whether it has any importance for folding and oligomerization.

Correspondence address: C. Forsman, Department of Biochemistry, University of Umeå, S-901 87 Umeå, Sweden. Fax: (46) (90) 167 661.

Abbreviations: CA, Carbonic anhydrase; HCA II, human carbonic anhydrase II; Rubisco, ribulose-1,5-bisphosphate carboxylase; cTP, chloroplast transit peptide; IPTG, isopropyl-β-D-thiogalactopyranoside.

Published by Elsevier Science Publishers B.V.

2. MATERIALS AND METHODS

CA was purified from homogenized light-grown pea leaves by affinity chromatography using *p*-aminomethylbenzenesulfonamide hydrochloride (Aldrich-Chemie) coupled to epoxy-activated CH-Sepharose 6B (Pharmacia, Sweden). Protein concentrations were determined by the method of Bradford [11] using bovine serum albumin as standard. N-Terminal amino acid sequences were determined in an Applied Biosystems model 477A sequencing system. Gel electrophoresis was performed on a discontinous polyacrylamide gel system [12] containing 15% acrylamide and 0.1% SDS or a gradient of 5–10% acrylamide under non-denaturing conditions. Scanning of Coomassie-stained gels was done using a LKB 2202 UltroScan Laser Densitometer.

Initial rates of CO_2 hydration were measured in a Hi-Tech stopped-flow apparatus by the changing pH–indicator method [13,14]. The buffer–indicator system was TAPS/Metacresol purple monitored at 578 nm.

Isolation of chloroplasts was done by centrifugation of homogenized pea leaves through a cushion of 40% (v/v) Percoll [15]. The chloroplasts were disrupted by osmotic shock plus sonication and centrifuged 1 h at 300,000 g. Soluble proteins were analyzed by SDS PAGE and immunoblotting using anti-pea CA IgG raised in rabbits and peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad).

Protein secondary structure was predicted using the PC/GENE software (IntelliGenetics Inc.) programs NOVOTNY based on the method of Novotny and Auffray [16], GGBMS based on the Gascuel and Golmard basic statistical method [17] and GARNIER based on the method of Garnier et al. [18].

cDNA encoding pea CA was isolated from a pea cDNA λgt11 library (Clontech Laboratories) using the polymerase chain reaction (PCR) method. The primers were designed to amplify the coding region and at the same time introduce an *Nco*I site at the initial ATG codon and a *Hin*dIII site downstream from the stop codon. PCR was run for a total of 28 cycles of the following profile: annealing 2 min at 40°C, extension 3 min at 72°C, denaturation 1 min at 94°C. The PCR products were cloned into a mutagenesis/expression vector (pACA [19]) used for production of HCAII, replacing the HCAII gene by using the *Nco*I and *Hin*dIII sites in the vector. The resulting plasmid was called pPCA. Site-directed mutagenesis was done by the method of Kunkel [20], based on host cell deficiency in the enzymes dUTPase (dut⁻) and uracil *N*-glycosylase (ung⁻). All new constructs were verified by sequencing the complete pea CA gene. For production of recombinant pea CA the *E. coli* strain BL21/DE3 [21] was transformed with pPCA and grown at 20°C in rich media to $A_{595} = 0.5$. Pea CA synthesis was induced by the addition of 0.5 mM IPTG and the cells were harvested after an additional 15 h. The enzyme was purified from lysed cells using the same kind of affinity column as described above.

3. RESULTS

3.1. *Processing of the pea CA chloroplast transit peptide (cTP)*

CA purified from pea leaves appears on an SDS gel as a doublet corresponding to masses of 25 kDa and 27 kDa (Fig. 1, lane 3). Gel scanning indicates that approximately 90% of the subunits are found in the 25 kDa band and 10% in the 27 kDa band. To check if this doublet is due to proteolysis during enzyme preparation or to the presence of a cytosolic isozyme, pea chloroplasts were isolated. A doublet with the same M_r values is recognized by anti-pea CA antibodies in immunoblots of the soluble fraction from broken chloroplasts separated by SDS PAGE. This result supports the idea that the two bands correspond to different subunit sizes present within the chloroplasts. In a non-denaturing gel system the purified enzyme also gives a doublet but with a mass of around 230 kDa (data not shown). These two bands were excised and run in separate lanes on the SDS-gel shown in Fig. 1. The results show that the lower band from the non-denaturing gel represents oligomers composed of small subunits only (25 kDa) (lane 1). The upper band gives a doublet corresponding to oligomers composed of equal amounts of small (25 kDa) and large (27 kDa) subunits (lane 2). No band corresponding to unprocessed precursor protein is detected by Coomassie- or immunostaining. N-Terminal sequencing of the two bands from the SDS-gel gave Gly-Lys-Gly-Tyr-Asp-Glu for the 27 kDa subunit and Thr-Thr-Ser-Ser-Ser-Asp for the 25 kDa subunit. A comparison with the published pea CA cDNA sequences [6,7] shows that the first amino acid sequence is located within the postulated cTP and the second sequence three residues after the cleavage site proposed by Roeske and Ogren. It thus seems as if the cTP can be processed at two (or three) different sites marked by arrows in Fig. 2. Cleavage at site I removes a peptide of 70 amino acid residues from the precursor protein. This peptide conforms to the characteristics of a cTP [3]. Cleavage at site II removes an additional 38 residue segment, the sequence of which is unusual for cTP's as it contains 8 acidic and 6 basic residues (cf. Fig. 2). The large subunit, thus, contains an 'extra', highly charged peptide at its N-terminus.

3.2. *Cloning and expression of pea CA cDNA in* E. coli

To test the function of this extra peptide we decided to produce the protein in *E. coli*. From a pea cDNA library we amplified and isolated two 1 kb fragments using the PCR technique. After cleavage with *Nco*I and *Hin*dIII both fragments were cloned into the vector

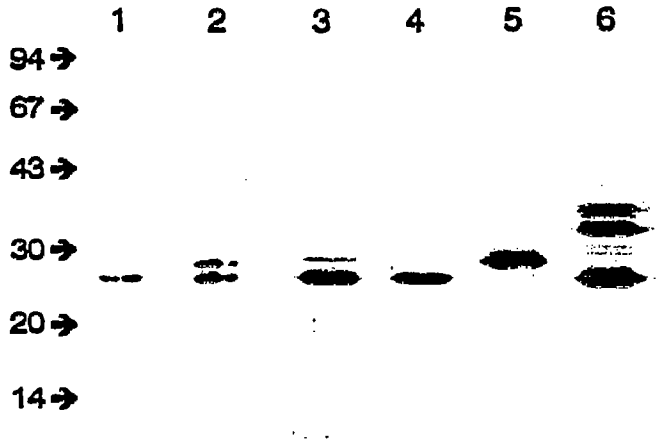

Fig. 1. SDS-PAGE analysis of pea CA. Bands excised from the doublet obtained in native PAGE were run in lane 1 (lower band) and lane 2 (upper band). Lane 3, CA purified from pea leaves; lane 4, PCAs; lane 5, PCAl; lane 6, PCAt. Molecular size markers (kDa) are shown to the left.

pACA and the new expression/mutagenesis plasmids pPCAa and pPCAb were obtained. The entire 1 kb inserts were sequenced in both plasmids and both of them were found to contain the whole coding region for pea CA, starting from the initiator ATG up to position 1,110, where the *Hin*dIII site was introduced in the untranslated 3′ end. One nucleotide substitution had occurred in each fragment, probably due to misincorporation by the *Taq* polymerase. In pPCAa, a T at position 168 had changed to C leading to substitution of a Thr for an Ile. In pPCAb, a C at position 132 had changed to T giving Phe instead of Ser. In both cases the substitutions are within the transit peptide before cleavage site I. The two PCR fragments differ in the untranslated 3′ end where pPCAa has a deletion of 18 bp as compaired to pPCAb. Interestingly, a comparison of the two published sequences reveals the same difference, so that the sequence of pPCAa and that of Majeau and Coleman [7] are identical in this region, while the sequence of pPCAb is identical to that of Roeske and Ogren [6] (Table I). This could mean that there are (at least) two different genes, or different alleles, encoding chloroplastic CA in pea. Different splicing of the mRNA precursor is another possibility. All our subsequent work was done on pPCAa, which was renamed pPCAt. Two other plasmids, pPCAl and pPCAs were constructed from pPCAt, allowing expression of CA containing different lengths of the cTP. pPCAt carries the entire coding region, and yields CA including the complete cTP. In pPCAl protein synthesis starts with the Met at cleavage site I, corresponding to the large subunit. In pPCAs the region coding for the entire cTP was removed and an initiator Met placed in front of the Gly at cleavage site II. This gives the small subunit as product, starting with Gly, as confirmed by N-terminal sequencing. The highest yield of CA was obtained using pPCAs, giving typically 40 mg of homogenous enzyme from 600 ml of cell culture. Expression using pPCAt gave 20% of this yield and pPCAl only 2%. Cells transformed with pPCAl grew to a lower cell density and also showed a high tendency to lose its plasmid. The three CA products were purified from *E. coli* and analyzed by SDS PAGE (Fig. 1). PCA's (lane 4) and PCAl (lane 5) are homogenous and of the same sizes as the corresponding subunits from pea leaves (lane 3). PCAt (lane 6) shows multiple bands with sizes ranging from that of an enzyme with intact cTP to that corresponding to a complete removal of the cTP. Thus the protein has been partly processed by bacterial proteases. Note, however, that PCAt does not give any band corresponding to the large subunit. Non-denaturing gel electrophoresis of the three recombinant CA's shows that oligomers of M_r = 230,000 and larger have been formed in all cases. No detectable band is found at lower molecular masses (data not shown).

Despite the different sizes and different degrees of transit peptide processing, the three recombinant CA's

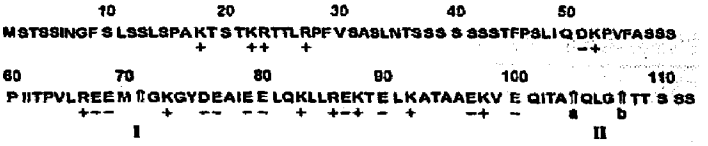

Fig. 2. Transit peptide sequence of pea CA. Charged residues are marked with + or −. Arrows indicate cleavage sites I and II. a, from [6]; b, this work.

are all fully active with kinetic parameters for CO_2 hydration equivalent to CA purified from pea leaves (Table II).

4. DISCUSSION

The cTP's of plant carbonic anhydrases, as proposed for spinach CA [4,5] and pea CA [6,7], have highly charged C-terminal regions with an excess of acidic residues (Fig. 2). In this report we demonstrate that the pea CA cTP can be cleaved at two sites, immediately before and after the acidic part, giving protein subunits of two sizes. Since site I is not recognized by *E. coli* proteases and no other cTP's have been found to contain a similar stretch of acidic amino acid residues, we propose that this site is recognized by a specific chloroplastic peptidase. The part of the cTP preceeding site I could very well be the functional transit peptide.

The second cleavage site needs some consideration. Different N-terminal sequences of the mature protein have been reported, as mentioned in section 3.1. We found that the mature protein is 3 residues shorter than reported by Roeske and Ogren [6] (Fig. 2). On the other hand, Majeau and Coleman [7] did not obtain conclusive data from N-terminal sequencing. Moreover, our cloned preenzyme PCAt is processed in *E. coli* at several sites (cf. Fig. 1). N-Terminal sequencing of the band of lowest molecular weight on the SDS-gel shows that it starts 2 amino acids before the cleavage site we found in CA isolated from pea leaves. Thus, cleavage occurs in a defined region but not necessarily at exactly the same site. It is tempting to speculate that cleavage at site II is not directed by a specific sequence, but rather by a certain secondary or tertiary structure which is prone

Table I

Differences in nuclei acid sequences between Roeske and Ogren [6], Majeau and Coleman [7] and this work

Position	Majeau and Coleman	Roeske and Ogren	This work	Amino acid difference
114	T	C	C	S → F
139–141	TTC	–	–	S → –
274	A	G	G	K → K
551	A	G	G	K → E
1054–1072	missing	present	*	

*Missing in pPCAa, present in pPCAb

to attack by proteases present in chloroplasts as well as *E. coli*.

The acidic part of the cTP is present in a minor but significant fraction of the subunits within the chloroplast. High level expression of fully active pea CA from *E. coli* carrying a plasmid containing the coding region for small subunits shows that neither the acidic extension nor the rest of the cTP is necessary for correct folding, for assembly into oligomers or for full catalytic activity. Moreover, the protein seems to adopt a native conformation even in the presence of the entire cTP. Possibly the cTP forms a separate structure that is susceptible to proteolytic degradation in *E. coli*. However, when only the acidic part of the cTP is present, as in PCAl, no processing is observed. This might reflect a locally perturbed structure shielding site II from attack by proteases. If so, a step-wise processing of the precursor protein within the chloroplast seems unlikely. It seems more probable that cleavage can occur at either site I or site II, thus leading to the formation of large or small subunits. Chloroplast chaperonins have been found to bind to several chloroplast proteins in vitro [2], possibly important for folding and assembly processes in analogy with the situation in mitochondria. If this also applies to pea CA, it seems possible that the *E. coli* chaperonins bind to the cloned enzyme and fulfill this function.

Computation of the secondary structure predicts an α-helix for the 38 residue acidic peptide. Computed predictions are generally not accurate to more than about 50%, but three different methods (see section 2) all give an α-helix as the most probable structure. A helical wheel representation (Fig. 3) clearly shows that the acidic residues are all found one side of the wheel, while hydrophobic residues dominate the other side. Interestingly, not only the cTP for spinach CA contains a highly similar acidic part, but so do the cTP's for tobacco CA [24] and *Arabidopsis thaliana* CA [25]. They are all of the same length with a net charge of −2, giving similar distributions around the helical wheel. A possible interpretation is that the acidic peptide is important for interaction with some other molecule or molecules, which might be a protein component of the chloroplast envelope essential for recognition or proper translocation.

Table II

Michaelis–Menten parameters for CO_2 hydration at pH 9.0 and 25°C

	k_{cat} (ms^{-1})	K_m (mM)	K_{cat}/K_m ($\mu M^{-1} \cdot s^{-1}$)
CA from pea leaves	264	2.8	94
PCAt	368	5.6	66
PCA1	275	4.1	67
PCAs	232	3.5	66

Data were fitted to the Michaelis–Menten equation by non-linear regression analysis.

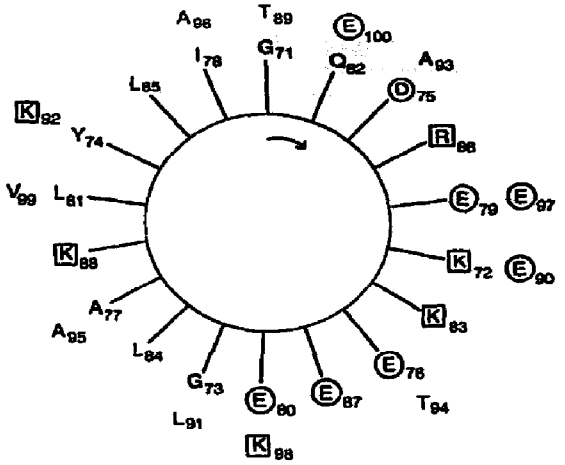

Fig. 3. Helical wheel plot of residues 71–100. Acidic amino acids are marked with a circle and basic amino acids with a box.

If so, the receptor must be specific for CA, since similar acidic parts have not been found in other sequenced cTP's. In contrast, competition experiments using synthetic cTP's and purified precursor proteins expressed in *E. coli* [22,23] indicate the presence of one common type of receptor. However, only a limited number of proteins have been tested so far. The possibility that the acidic peptide has a function within the chloroplast must also be considered. We know that some subunits within the chloroplast retain this peptide. Although the peptide is not necessary for catalytic activity, it might be important for the interaction with some other protein within the chloroplast in order to fulfill the physiological function of plant CA. There is not much evidence for a role for CA in photosynthesis, but the obvious hypothesis has been that chloroplastic CA catalyzes the dehydration of HCO_3^-, thus generating CO_2 needed for carboxylation by Rubisco [1]. Although an association between CA and Rubisco has not yet been demonstrated, this remains an interesting possibility. Studies of the interaction between CA and Rubisco should be facilitated by the availability of cloned plant CA with and without the acidic peptide.

Acknowledgements: Financial support from Stiftelsen Lars Hiertas Minne and O E och Edla Johanssons vetenskapliga stiftelse are greatly acknowledged. We thank Eleonore Skärfstad for skilful technical assistance and Prof. Sven Lindskog and Dr. Bengt-Harald Jonsson for valuable discussions of this manuscript.

REFERENCES

[1] Reed, M.L. and Graham, D. (1981) in: Progress in Phytochemistry, vol. 7 (Reinhold, L., Hornborne, J.B. and Swain, T., eds.) pp. 47–94, Pergamon Press, Oxford.

[2] de Boer, A.D. and Weisbeek, P.J. (1991) Biochim. Biophys. Acta 1071, 221–253.

[3] von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–545.
[4] Burnell, J.N., Gibbs, M.J. and Mason, J.G. (1990) Plant Physiol. 92, 37–40.
[5] Fawcett, T.W., Browse, J.A., Volokita, M. and Bartlett, S.G. (1990) J. Biol. Chem. 265, 5414–5417.
[6] Roeske, C.A. and Ogren, W.L. (1990) Nucleic Acids Res. 18, 3413.
[7] Majeau, N. and Coleman, J.R. (1991) Plant Physiol. 95, 264–268.
[8] Robinson, C. and Ellis, R.J. (1984) Eur. J. Biochem. 142, 343–346.
[9] Schmidt, R.J., Gillham, N.W. and Boynton, J.E. (1985) Mol. Cell. Biol. 5, 1093–1099.
[10] Kalousek, F., Hendrick, J.P. and Rosenberg, L.E. (1988) Proc. Natl. Acad. Sci, USA, 85, 7536–7540.
[11] Bradford, M.M. (1976) Anal. Biochem. 72, 248–254.
[12] Jergil, B. and Ohlsson, R. (1974) Eur. J. Biochem. 46, 13–25.
[13] Khalifah, R.G. (1971) J. Biol. Chem. 246, 2561–2573.
[14] Steiner, H., Jonsson, B.-H. and Lindskog, S. (1975) Eur. J. Biochem. 59, 253–259.
[15] Joy, K.W. and Mills, W.R., Methods Enzymol. 148, 179–188.
[16] Novotny, J. and Auffray, C. (1984) Nucleic Acids Res. 12, 243–255.
[17] Gascuel, O. and Golmard, J.L. (1988) CABIOS 4, 357–365.
[18] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[19] Fierke, C.A., Krebs, J.F. and Venters, R.A. (1991) in: Carbonic Anhydrase. From Biochemistry and Genetics to Physiology and Clinical Medicine (Botre', F., Gros, G. and Storey, B.T., eds.) pp. 22–36, Verlag Chemie, Weinheim.
[20] Kunkel, P.A. (1985) Proc. Natl. Acad. Sci. USA 82, 488–492.
[21] Studier, F.W. and Moffatt, B.A. (1986) J. Mol. Biol. 189, 113–130.
[22] Perry, S.E., Buvinger, W.E., Bennett, J. and Keegstra, K. (1991) J. Biol. Chem. 266, 11882–11889.
[23] Schnell, D.J., Blobel, G. and Pain, D. (1991) J. Biol. Chem. 266, 3335–3342.
[24] Majeau, N. and Coleman, J.R., EMBL accession no. M94135, unpublished.
[25] Raines, C., Horsnell, P.R., Holder, C. and Lloyd, J.C., EMBL accession no. X65541, unpublished.