Plant Molecular Biology **22**: 937–943, 1993.

Characterization of pea chloroplastic carbonic anhydrase. Expression in *Escherichia coli* and site-directed mutagenesis

Nicholas J. Provart, Nathalie Majeau and John R. Coleman*
Dept. of Botany, University of Toronto, 25 Willcocks St., Toronto, Ontario, Canada M5S 3B2
*(*author for correspondence)*

Received 10 March 1993; accepted in revised form 21 April 1993

Key words: carbonic anhydrase, *Escherichia coli*, mutagenesis, *Pisum sativum*

Abstract

A cDNA encoding the mature, chloroplast-localized carbonic anhydrase in pea has been expressed in *E. coli*. The enzyme is fully active and yields of up to 20% of the total soluble protein can be obtained from the bacteria. This expression system was used to monitor the effects of site-directed mutagenesis of seven residues found within conserved regions in the pea carbonic anhydrase amino acid sequence. The effects of these modifications are discussed with respect to the potential of various amino acids to act as sites for zinc coordination or intramolecular proton shuttles.

Introduction

The enzyme carbonic anhydrase (CA; carbonate dehydratase, EC 4.2.1.1) catalyses the reversible reaction $CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$ and is one of the more abundant proteins in the leaves of C_3 higher plants [1]. Although its role in photosynthesis is not yet fully understood, at least two functions for this enzyme have been proposed [1]. It is possible that by inducing the rapid hydration of the dissolved CO_2 as it passes across the chloroplast envelope into the alkaline stroma, CA facilitates diffusion of inorganic carbon into the chloroplast. Conversely, at the site of carboxylation, CA is thought to maintain the supply of CO_2 for the principal CO_2-fixing enzyme in C_3 plants, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco). In the absence of CA, the slow rate of uncatalysed HCO_3^- dehydration could be rate-limiting for photosynthetic carbon assimilation.

Recently, the primary structure and pattern of expression of the higher-plant chloroplast CA has been determined for a number of species [3, 6, 9, 10, 12, 13]. In pea, this nuclear-encoded enzyme is expressed as a 37 kDa precursor protein which is processed to the mature 25 kDa polypeptide after transport into the chloroplast [10, 13]. Comparison of the mature plant protein sequences with primary structures of animal CAs has revealed little if any similarity. As such, information obtained from crystal structures and active-site chemistry studies of the animal CAs cannot be used directly for an understanding of the plant enzyme [5, 16, 19, 20]. In addition, the absence of a crystal structure of the plant CA protein has also hindered the study of the catalytic mechanism. Although similar in many as-

pects of kinetics and function, the dissimilarity in primary sequences has also prompted some speculation on the evolutionary origins of plant and animal CAs. In a recent phylogenetic analysis based on amino acid sequence comparison [7], it was suggested that there are two broad categories of CAs: those with a proto-prokaryotic origin (including bacterial, cyanobacterial and chloroplastic CAs) and those with a proto-eukaryotic origin (including the periplasmic-localized algal CA and mammalian CAs).

In this paper we present our attempts to analyse the relationship between structure and function in pea CA using a cDNA coding for pea (*Pisum sativum*) chloroplastic CA cloned in a heterologous expression system in *E. coli*. The effects of 7 different amino acid mutations introduced independently by site-directed mutagenesis into the mature CA protein are described.

Materials and methods

pPCA constructs and mutants

Site-directed mutagenesis [8] was used to modify a cDNA clone coding for the unprocessed pea chloroplastic CA [10]. A novel *Nde* I site was introduced 6 bp upstream of the N-terminus Gln codon of the mature CA protein in the plasmid p1.2CA. The resultant plasmid (p1.2CA·Nde) was digested with *Eco* RI then partially digested with *Nde* I to yield a population of DNA fragments. A 950 bp fragment containing the mature protein coding region and defined at its 5′ end by the novel *Nde* I site and at its 3′ end by an *Eco* RI site was identified, purified by low-melt agarose gel electrophoresis, and force-cloned using standard techniques [14] into the *Nde* I and *Eco* RI sites of the expression vector pET5b [18]. This plasmid is designated pPCA.

Single-stranded p1.2CA · Nde DNA (generated by transfecting plasmid containing *E. coli* CJ236 cells with the helper phage R408) was used as a substrate for subsequent site-directed mutagenesis reactions. Custom oligonucleotides (HSC Biotech Centre, Toronto) were used to generate the 7 mutations in the CA coding sequence of p1.2CA.Nde, following the method of Kunkel *et al.* [8]. Sequence analysis of each mutant was used to confirm the specific bp changes. Sequencing was performed with 2 μg of double-stranded plasmid DNA using [^{35}S]dATPαS and the dideoxy chain termination method of Sanger *et al.* [15] as described in the Sequenase kit (U.S. Biochem. Inc.). The sequence-modified fragments, defined at their 5′ ends by the novel *Nde* I site and at their 3′ ends by the *Eco* RI site, were cloned independently into pET5b as described above to generate the specific mutant pea CA-expressing constructs.

Expression in E. coli *and enzyme assays*

Pea CA sequences cloned in the expression vector pET5b were transformed into BL21(DE3) cells (*E. coli* B strain [F^- *ompT* $r_B^- m_B^-$] lysogenized with a lambda derivative, DE3 which carries the immunity region of phage 21, the *lacI* gene, the *lac*UV5 promoter, the start of the *lacZ* gene and the gene for T7 RNA polymerase inserted into the *int* region of the phage). The *lac*UV5 promoter directs the transcription of T7 RNA polymerase upon the addition of isopropyl-*β-D*-thiogalactopyranoside (IPTG) [18].

Expression was optimized by assessing the CO_2 hydration activity in cell lysates obtained from *E. coli* cell cultures of varying starting cell densities (OD_{600} 0.5, 1.0 and 1.2) and measured at 100 minute intervals after induction by 0.5 mM IPTG and growth at 37 °C. Cultures were also assessed at each time point for cell density and soluble protein content of cell lysates. For activity and protein measurements, cells from 10 ml culture samples were harvested by centrifugation (10000 × *g*, 10 min), resuspended in a half volume of Veronal buffer (20 mM, pH 8.3) containing 1 mM DTT and crushed in a French pressure vessel at 110 MPa. All manipulations were carried out at 4 °C. Cell lysates were clarified by centrifugation (10000 × *g*, 10 min) and used without further processing. Soluble protein content was determined as described previously [17] and

CO_2 hydration activity was determined electrometrically using the method of Wilbur and Anderson [21].

For the experiments comparing the effects of different buffers on CO_2 hydration activity, a similar protocol was used however the cell lysate buffer was varied to match the assay buffer. Buffer compositions were as follows: Veronal: 20 mM Veronal buffer pH 8.3; HEPES/KOH: 20 mM HEPES (N-[2-hydroxyethyl]piperazine-N′-[2-ethanesulphonic acid]) buffer pH 8.3; imidazole: 20 mM imidazole buffer pH 8.3. In all cases, CO_2 hydration activity was normalized on the basis of the amount of total soluble CA protein, as determined by scanning densitometry of Coomassie-stained protein gels on a Hoefer scanning densitometer (Hoefer Scientific Instruments, CA) at 580 nm. Peak areas were calculated using the Dynamax MacIntegrator (Rainin Instrument Company) on an Apple MacIntosh computer. SDS gel electrophoresis and immunoblotting techniques of plant and *E. coli* proteins were as described previously [10].

Results

The cloning strategy employed has produced an expression system for the synthesis of what is essentially the mature form of the pea chloroplast CA. The addition of the *Nde* I site into the pea CA cDNA sequence results in the generation of a Met residue as well as the inclusion of two transit-peptide residues (T104 and A105, as described in [10]) to the N-terminal Gln of the mature protein. pPCA, when used to transform *E. coli* BL21(DE3) and in the presence of IPTG results in the production of large quantities of recombinant pea CA with values up to 20% of the total soluble protein achieved in a short period of time. Maximum CA activity was obtained 300 min after the addition of 0.5 mM IPTG to an *E. coli* culture in log phase growth at 37 °C and having an initial cell density of 1.2 when measured at 600 nm (data not shown). As described in Fig. 1, the amount of CA expressed after 5 h represents approximately 16.5% of the total soluble protein of an *E. coli* lysate and by SDS-PAGE analysis

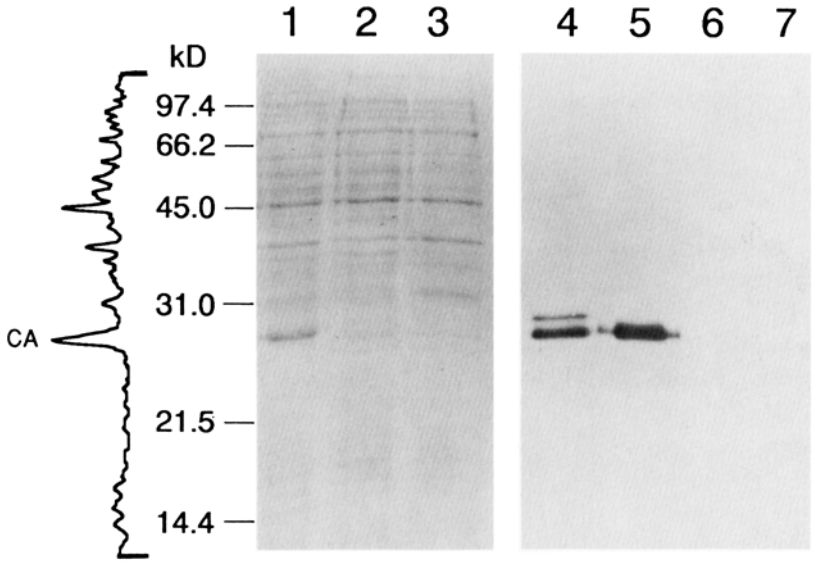

Fig. 1. Characterization of expression of pea CA in *E. coli*. SDS-PAGE analysis of soluble protein extracts from a BL21(DE3) culture in the presence or absence of pPCA and 5 mM IPTG. Cells were grown to an initial OD_{600} of 1.2 followed by the addition of 0.5 mM IPTG where required and another 5 hours growth at 37 °C. Lanes 1–3 have been stained with Coomassie Brilliant Blue and lanes 4–7 have been electroblotted and probed with pea CA antisera. Lanes 1 and 5: BL21(DE3) with pPCA plus IPTG. Lanes 2 and 6: BL21(DE3) with pPCA, uninduced. Lanes 3 and 7: BL21(DE3) with pET5b plus IPTG. Lane 4: soluble proteins in a pea leaf extract. The trace to the left of the stained gel represents the output of the scanning densitometer used to quantify the protein levels in lane 1. The CA-labelled peak represents 16.5% of the total protein loaded.

has the same monomeric molecular mass (25.5 kDa) as the major form of the enzyme found in pea leaf tissue. The minor higher-molecular-mass monomer (27.5 kDa) observed in the immunoblot of pea leaf tissue (Fig. 1, lane 4) represents a partially processed form of the CA protein which contains some transit peptide residues. No expression in *E. coli* of the plant CA protein is observed in the absence of the inducer IPTG or pPCA (Fig. 1, lanes 6 and 7, respectively). The protein is readily visible on a protein gel after Coomassie Brilliant Blue staining and can be easily quantified using a scanning densitometer. With respect to activity, no discernible differences were observed when maximal rates of CA-catalysed CO_2 hydration were determined for enzyme extracted from pea leaves and compared to that exhibited by the recombinant CA protein (data not shown).

Seven different amino acids of the mature CA-coding sequence were mutated independently and the positions of these modifications are shown in Fig. 2. The effects of each of these mutations on the solubility of the enzyme as expressed in *E. coli*, as well as the maximum rate of CO_2 hydration, are shown in Table 1. Activities are expressed as a percentage of wild-type pPCA activity in the appropriate buffer and were normalized on the basis of recombinant CA protein as determined by SDS-PAGE and densitometry. In the case of H169N, there was little change in the hydration activity. Four of the mutants were completely inactive, C160S, E204A, H220N and C223S. C160S and H220N were found to be mostly insoluble, although a small fraction of the total CA produced was soluble and this fraction could be increased somewhat by decreasing the growth temperature from 37 °C to 30 °C during induction. Even when accounting for the reduced expression of soluble protein, no discernible rate of catalysed CO_2 hydration was achieved by these mutants. E204A and C223S both produced wild-type levels of soluble protein at 37 °C however neither of these proteins has any CO_2 hydration activity (Table 1).

The remaining two mutations, H209N and E276A are soluble but are somewhat defective in catalytic activity, to varying degrees in different buffers. Both show near wild-type levels of activity in Veronal buffer. However, H209N has only 6% of wild-type activity in Hepes/KOH buffer,

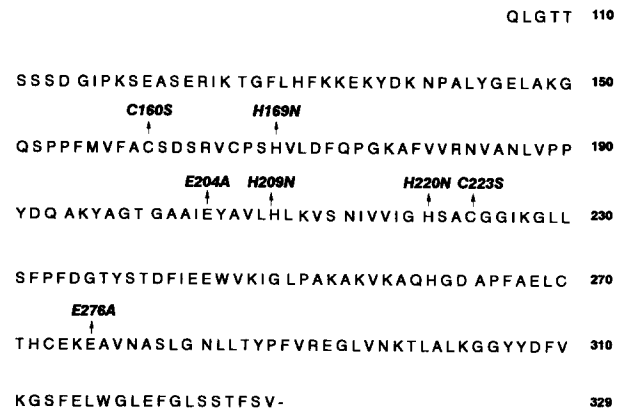

Fig. 2. Pea CA protein sequence showing mutation sites. The amino acid sequence displayed represents the mature form of the pea CA as determined by N-terminal sequence analysis of the chloroplast localized protein [13]. The residue numbering system used starts with the methionine at the N-terminus of the transit peptide as described [10].

Table 1. Characteristics of site-directed mutant pea CA proteins.

Mutation	Protein[1]	Veronal[1]	HEPES/KOH[2]	Imidazole[2]
		percentage of wild-type activity		
C160S	insoluble[3]	0	0	0
H169N	soluble	83	84	99
E204A	soluble	0	0	0
H209N	soluble	82	6	41
H220N	insoluble[3]	0	0	0
C223S	soluble	0	0	0
E267A	soluble	77	44	78

[1] Refers to the form of the majority of the CA protein found in the *E. coli* lysate.

[2] Buffers used in assays of activity. Values expressed are a percentage of activity generated by an equal amount of unmodifed recombinant CA protein when assayed in the same buffer system. Values are averages of replicate assays from one experiment and replicates vary by less than 5%. Replicate experiments show similar relationships.

[3] Although most of the protein expressed by these mutants is insoluble, the small percentage of soluble protein generated was used in these assays. Insoluble protein is also inactive.

while E276A has 40% activity. In the imidazole buffer, the activity is restored to 44% of wild type and 77.5% of wild type, respectively (Table 1).

Discussion

In spite of the phylogenetic distance (and obvious amino acid sequence divergence) between pea CA and human CA, it might prove useful to use the human enzyme as a model for understanding the catalytic mechanism of the plant enzyme. In human CAII, three histidine residues, H94, H96 and H119, are responsible for coordinating the zinc atom required for activity [5]. Histidines 94 and 96 and histidine 119 are on adjacent β-sheets in the native enzyme and they are situated such that the zinc they coordinate is at the apex of the cone-shaped active site that penetrates into the heart of the enzyme [5]. In addition, a fourth histidine, H64, acts as a proton shuttle during enzyme regeneration [20]. As the plant enzyme is known to contain Zn [1], we initially chose to mutate histidine residues 169, 209 and 220 in pea CA based on the requirement of the enzyme for Zn binding. The mutant H169N has very nearly a wild-type phenotype in terms of CO_2 hydration activity, and consequently we assume that it plays no major role in the functioning of the enzyme in terms of active site chemistry. The mutant H220N is completely inactive and is also largely insoluble, indicating a major disruption in the protein structure. One hypothesis to explain these observations could be that when the protein cannot fully coordinate the zinc atom it assumes an aberrant folding pattern that results primarily in the generation of protein inclusion bodies in *E. coli*. These protein aggregates, as well as the small percentage of soluble CA, exhibit no catalytic activity. There are data in the literature however, that suggest that the metal co-factor is not required for correct folding. Studies using denatured bovine CA II apoprotein have shown that a functional enzyme is generated even when the metal ion is added well after transfer to renaturating conditions [2]. In an attempt to generate a functional enzyme with an alteration at the H220 site, another zinc binding amino acid was substituted for this histidine. However, the mutant H220C exhibits no activity and is also mostly insoluble (data not shown). A similar phenotype of inactivity and insolubility was found with the mutant C160S. Cysteine 160 was a candidate for mutagenesis as this amino acid can coordinate zinc and does so in some zinc-finger proteins [4]. A homology search of pea, spinach, *E. coli* and *Synechococcus* PCC7942 CAs has also identified this residue as a conserved amino acid [7].

The three residues E204, C223 and E276 were also candidates for mutagenesis primarily on the basis of sequence conservation among the plant and prokaryote CAs, and the capacity of both glutamate and cysteine residues to coordinate Zn atoms [4]. Both of the mutants E204A and C223S produced large amounts of soluble CA protein but were completely inactive. The soluble nature of these two inactive, modified CA proteins suggests that correct folding has occurred but it is not known if the Zn atom has been integrated into the polypeptide.

The two remaining mutants H209N and E276A are perhaps the most interesting as they are both soluble, but defective in catalytic activity to varying degrees in different buffers. For example H209N exhibits near wild-type levels of activity in Veronal buffer however, it displays only six percent of wild-type activity when assayed in HEPES buffer. Interestingly, a mutant with a similar phenotype has been generated in human CAII. When the human enzyme contains the lesion H64A, there is a nine-fold decrease in k_{cat} when catalytic activity is assayed in MOPS (3-[*N*-morpholino] propane sulphonic acid) buffer as compared to imidazole buffer [20]. It has been postulated that H64 is part of an intramolecular proton transfer network necessary for the regeneration of the enzyme, as seen in the following equation 2 of the simplified model for the catalytic mechanism [20]:

$$EZnOH^- + CO_2 \rightleftharpoons EZn(OH^-)CO_2 \rightleftharpoons EZnHCO_3^- \rightleftharpoons EZnH_2O + HCO_3^- \quad (1)$$

$$EZnH_2O + B \rightleftharpoons EZnOH^- + BH^+ \quad (2)$$

where E is the enzyme active site and B is a buffer molecule.

In the absence of H64, proton transfer to a buffer can still occur but the efficiency of transfer apparently depends on the accessibility of the active site to the buffer molecule [20]. MOPS and HEPES both have bulky side groups which could hinder their circulation in the active site, thereby limiting proton transfer to water alone, which is a relatively poor buffer. Veronal and imidazole, on the other hand, are relatively compact buffer molecules which either lack or have short side chains and can presumably reach the active site with little hindrance.

Since E276A exhibits a similar phenotype with respect to the different buffers, it too could be involved in proton transfer. Alternately, it could be involved in stabilizing, via hydrogen bonds, the positions of the reactants within the active site, as is the case with E106 in human CAII [11]. Interestingly, the CA protein containing the modification E276A is far less stable than the wild-type enzyme and the other two mutants which express activity. It has a half-life of roughly two hours in the crude extract, whereas the other three forms have half-lives several times that (data not shown). It is possible that modification of intramolecular distances within the active site results in an unstable Zn atom placement or disruption of stabilizing hydrogen bonds. Alternatively, modification of the protein structure could have increased the susceptibility of catalytically important residues to proteolysis.

In conclusion, substitution of residues C160, E204, H220 or C223 appears to have a severe impact on catalytic activity. It is unknown if the elimination of catalytic activity is the result of modified or absent zinc coordination, or simply the result of aberrant folding. We are currently attempting to analyse these mutant proteins for their zinc content using atomic absorption spectroscopy. Our data suggest that H209 could be involved in the intramolecular proton-shuttling network, acting in a fashion similar to H64 in the human isoform CAII. Residue E276 may also be involved in the active site mechanism or structural stability in a way as yet to be determined. The anticipated elucidation of the plant CA crystal structure will be used to interpret the effect of these specific amino acid substitutions on catalytic activity.

Acknowledgements

This research was sponsored in part by a grant to J.R.C. from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. Atkins CA, Patterson BD, Graham D: Plant carbonic anhydrases I. Distribution of types among species. Plant Physiol 50: 214–217 (1972).
2. Bergenhem N, Carlsson U: The role of the metal ion in the refolding of denatured bovine Co(II)-carbonic anhydrase II. Biochim Biophys Acta 998: 277–285 (1989).
3. Burnell JN, Gibbs MJ, Mason JG: Spinach chloroplastic carbonic anhydrase: Nucleotide sequence analysis. Plant Physiol 92: 37–40 (1990).
4. Christianson DW: Structural biology of zinc. Adv Protein Chem 42: 281–355 (1991).
5. Eriksson E, Jones TA, Liljas A: Refined structure of human carbonic anhydrase II at 2 A resolution. Proteins: Structure, Function Genet 4: 274–282 (1988).
6. Fawcett TW, Browse JA, Volokita M, Bartlett SG: Spinach carbonic anhydrase primary structure deduced from the sequence of a cDNA clone. J Biol Chem 265: 5414–5418 (1990).
7. Fukuzawa H, Suzuki E, Komukai Y, Miyachi S: A gene homologous to chloroplast CA (*icfA*) is essential to photosynthetic carbon fixation by *Synechococcus*. Proc Natl Acad Sci 89: 4437–4441 (1992).
8. Kunkel TA, Roberts JD, Zakour RA: Rapid and efficient site-specific mutagenesis without phenotypic selection. Meth Enzymol 154: 367 (1987).
9. Majeau N, Coleman JR: Nucleotide sequence of a complementary DNA encoding tobacco chloroplastic carbonic anhydrase. Plant Physiol 100: 1077–1078 (1992).
10. Majeau N, Coleman JR: Isolation and characterization of a cDNA clone coding for pea chloroplastic carbonic anhydrase. Plant Physiol 95: 264–268 (1991).
11. Merz KM: Insights into the function of the zinc-hydroxide-Thr199-Glu106 hydrogen bonding network in carbonic anhydrases. J Mol Biol 214: 799–802 (1990).
12. Raines CA, Horsnell PR, Holder C, Lloyd JC: *Arabidopsis thaliana* carbonic anhydrase: cDNA sequence and effect of CO_2 on mRNA levels. Plant Mol Biol 20: 1143–1148 (1992).

13. Roeske CA, Ogren WL: Nucleotide sequence of pea cDNA encoding chloroplast carbonic anhydrase. Nucl Acids Res 18: 3413 (1990).
14. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, NY (1989).
15. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74: 5463–5467 (1977).
16. Sheridan RP, Allen LC: The active site electrostatic potential of human carbonic anhydrase. J Am Chem Soc 103: 1544–1550 (1981).
17. Stoscheck CM: Quantitation of protein. Meth Enzymol. 182: 50–68 (1990).
18. Studier W, Rosenburg AH, Dunn JJ, Dubendorff JW: The use of T7 polymerase to direct the expression of cloned genes. Meth Enzymol 185: 60–89 (1990).
19. Tashian RE: The carbonic anhydrases: Widening perspectives on their evolution, expression and function. Bioessays 10: 186–192 (1989).
20. Tu C, Silverman DN, Forsman C, Jonsson B-H, Lindskog S: The role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site specific mutant. Biochemistry 28: 7913–7918 (1989).
21. Wilbur KM, Anderson NG: Electrometric and colorimetric determination of carbonic anhydrase. J Biol Chem 176: 147–154 (1948).