Plant Physiol. (1990) 92, 37–40
0032-0889/90/92/0037/04/$01.00/0

Received for publication May 19, 1989
and in revised form August 4, 1989

Spinach Chloroplastic Carbonic Anhydrase

Nucleotide Sequence Analysis of cDNA

James N. Burnell*[1], Mark J. Gibbs, and John G. Mason
CSIRO, Division of Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia

ABSTRACT

We have determined the nucleotide sequence of a cDNA encoding spinach (*Spinacia oleracea*) chloroplastic carbonic anhydrase (CA). The open reading frame encodes a protein consisting of a transit peptide and a mature CA protein with a predicted mass of 24,116 daltons. This represents the first report of a nucleotide sequence of a plant CA.

Carbonic anhydrase (carbonate dehydratase 4.2.1.1; abbreviated CA[2]) catalyses the hydration of CO_2 to HCO_3^- (and the dehydration of bicarbonate) in a reaction that occurs at a significant rate even in the absence of a catalyst.

In plants there are different forms of CA ranging in molecular mass from 140 to 250 kD with subunit molecular masses ranging from 26 to 34 kD (7). The intra- and intercellular distribution of CA also varies among different plant types, being located in the chloroplasts of C_3 plants and in the cytosol of mesophyll cells in C_4 plants. The taxonomic diversity of plant CA has been demonstrated using anti-spinach (21) and anti-maize leaf CA antibodies (JN Burnell, unpublished results) and by polyacrylamide gel electrophoresis (1).

More is known about CA in animals, with six isozymic forms having been identified. The amino acid sequences are known for many of the animal isozymes (5, 6, 29), and the three-dimensional structure of two of them has been determined (20).

As part of a wider project to study the control of expression of CA in C_3 and C_4, and so-called C_3–C_4 intermediate plant species, we have isolated clones from a spinach leaf cDNA library in λgt11 using antibodies to spinach leaf CA and have identified and characterized these clones. In this paper we report on the isolation, identification, and sequence analysis of a cDNA for spinach CA.

MATERIALS AND METHODS

Materials

Restriction enzymes were obtained from either Boehringer-Mannheim or Pharmacia-LKB; DNA ligase from either Biolabs or Bresatec, Adelaide, South Australia; DNA polymerase I large fragment (Klenow fragment) from Bresatec; λgt11 from Promega; radioactive compounds from Bresatec or Amersham; Sephacryl S-300 from Pharmacia-LKB; DEAE-cellulose 52 from Whatman; pTZ18R, pTZ19R and the Sequenase sequencing kit from United States Biochem. Corp., Cleveland, OH. All other chemicals were from Sigma Chemical Co., Ajax Chemicals, or BDH.

Monospecific antibodies against spinach leaf CA (raised in rabbits) were generously provided by Professor S. Miyachi (University of Tokyo, Japan). (For characterization of the antibodies see ref. 21. The CA antibodies cross-reacted with only a single band on a Western blot of total spinach leaf protein and cross-reacted with only the major protein band of purified spinach leaf CA which ran with a mobility which coincided with a mol wt of 26 kd.)

Methods

Purification of Spinach Leaf Carbonic Anhydrase (for N-Terminal Sequencing)

The method of purification of spinach leaf CA was based on a previously published method (14). Spinach (*Spinacia oleracea*) leaves (350 g) were homogenized in 700 mL of buffer containing 20 mM phosphate, 0.1 M NaCl, and 1 mM EDTA, pH 6.8 (extraction buffer). Following filtration and centrifugation (10 min at 30,000*g*), the supernatant was fractionated with solid $(NH_4)_2SO_4$ and the protein precipitating between 30 and 50% saturation was redissolved in 75 mL of buffer, dialysed against extraction buffer, and run through a column (35 × 3 cm) of DEAE-cellulose 52. The active unbound fractions were pooled and concentrated by precipitating with solid $(NH_4)_2SO_4$ added to 65% saturation, and the redissolved protein was dialyzed against 20 mM Hepes-KOH (pH 7.0). The protein was applied to a column of DEAE-cellulose equilibrated with 20 mM Hepes-KOH (pH 7.0), and the CA was eluted with a 500 mL linear KCl gradient (0–0.2 M). The active fractions were pooled and the protein concentrated by $(NH_4)_2SO_4$ (65% saturation) and the redissolved concentrated protein (3 mL) was applied to a column of Sephacryl S-300 (96 × 1.5 cm) equilibrated with 20 mM phosphate, 0.1 M NaCl, and 1 mM EDTA (pH 6.8) and eluted with the same buffer. Fractions were collected and assayed for activity and protein and the peak fraction was dialysed against water for 24 h and used for N-terminal amino acid sequencing.

[1] Present address: The Graduate School of Science and Technology, Bond University, Gold Coast, 4229, Qld., Australia.

[2] Abbreviations: CA, carbonic anhydrase; IPTG, isopropyl β-D-thiogalactopyranoside; bp, base pair.

SDS-PAGE (3) and protein determination in crude extracts and $(NH_4)_2SO_4$ fractions (2) and in purified fractions (30) were conducted as described previously.

Assay of Carbonic Anhydrase

Carbonic anhydrase was assayed at 0°C following the change in pH as described by Burnell and Hatch (4).

N-Terminal Sequence Determination

Approximately 200 pmol of protein were analyzed by Edman degradation in an Applied Biosystems 477A protein sequencer.

Isolation of Carbonic Anhydrase cDNA Clones—Screening with Antibodies

A cDNA expression library in phage λgt11 was constructed from spinach leaf poly(A^+) RNA and screened for CA antigen expression using the method of Huynh *et al.* (10). Antigens were adsorbed to IPTG-impregnated nitrocellulose filters (Amersham) for 3 h and the filters incubated with spinach CA antibodies overnight at 4°C. Bound antibodies were visualized using alkaline phosphatase bound to goat anti-rabbit antibodies (Promega).

Screening with a Nucleic Acid Probe

The λgt11 cDNA expression library was also screened using a nick-translated cDNA probe obtained from a clone initially identified using the antibody screening method. The DNA was labeled using a commercial nick translation kit (Bresatec) and [α-^{32}P]dCTP.

cDNA Nucleotide Sequence Determination

Recombinant phage λgt11 were isolated from 100 mL cultures by precipitation with polyethylene glycol 6000–8000/dextran sulfate and CsCl centrifugation as described by Miller (18). The DNA was digested with *Eco*RI, and isolated inserts were subcloned into pTZ18R/19R. Sequencing was carried out by the dideoxy method of Sanger *et al.* (25).

RESULTS

Isolation and Characterization of cDNA Coding for Carbonic Anhydrase

About 4×10^4 recombinant plaques were screened for CA-β-galactosidase fusion protein with antibodies raised against spinach leaf CA. Three confirmed positive clones were found; all three clones were subsequently shown to be the same. Pure clones were grown on Y1088 cells and DNA isolated. Digestion of the phage DNA (λCA clone LCA 48) with *Eco*RI yielded two fragments of 200 and 790 bp which were subcloned into pTZ 18R and pTZ 19R. Subclones were also generated by exploiting a *Hae*III and a *Hin*dIII sites within the 790 bp *Eco*RI fragment. Sequence of these subclones was obtained from the reverse primer site in pTZ. Sequence through the internal *Eco*RI site was obtained using a *Hin*dIII-*Kpn*I fragment from the λgt11 clone (LCA 48) containing about 800 base pairs of λ DNA, subcloned into pTZ19R. A 17 base pair oligonucleotide (ACGCAGTGTTGCACCTT) specific to a sequence in the large *Eco*RI fragment (position 390–406) was used as the primer for sequencing this subclone.

LCA 48 contains a continuous open reading frame of 809 nucleotides and 47 5′-untranslated and 170 3′-untranslated nucleotides in the noncoding region including a single putative polyadenylation site (position 920) (Fig. 1). The protein sequence predicted from the open reading frame consists of 254 amino acids with a molecular weight of 29,653.

Protein Purification and Amino Acid Analysis

The chloroplastic CA was purified from spinach leaf tissue approximately 60-fold and SDS-PAGE of the purified enzyme indicated that the CA was 98.7% pure (results not shown). The protein band indicated a molecular mass of about 26 kD. All the CA activity of a sample of purified spinach leaf CA could be titrated with spinach leaf CA antibodies. The CA antibodies cross-reacted with only the major protein band of purified leaf CA. The N-terminal amino acid sequence of the purified CA was determined using an Applied Biosystems protein sequencer. The sequence of the 20 N-terminal amino acids was determined and found to match exactly part of the deduced amino acid sequence. This also identified the proc-

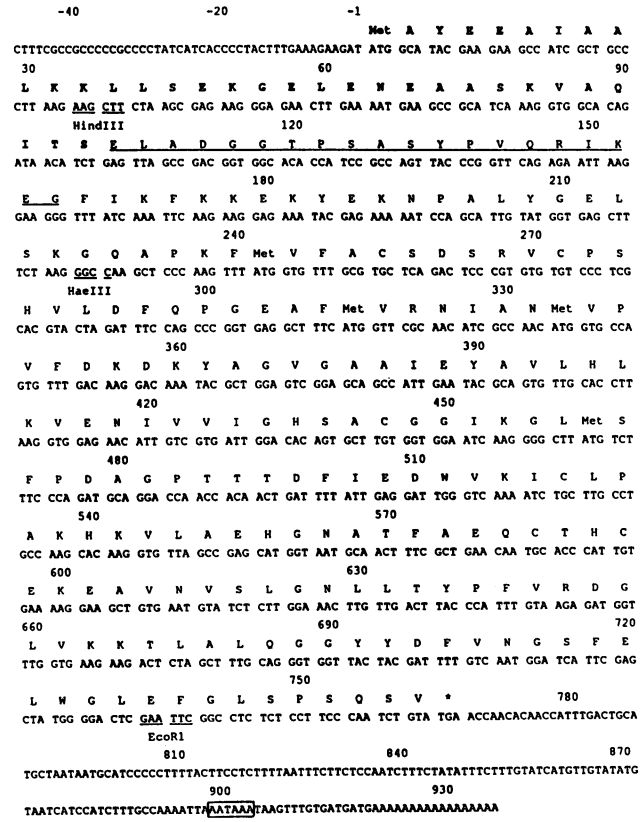

Figure 1. Nucleotide and amino acid sequence of spinach leaf carbonic anhydrase. The transit peptide sequence is shown in bold letters. The underlined sequence denotes the amino acid sequence determined by N-terminal amino acid sequencing. The boxed section denotes the polyadenylation signal.

essing site for the transit sequence and indicated that the mature protein consists of 221 amino acids.

Northern Hybridization

Northern hybridization of total spinach leaf RNA with nick-translated *Eco*RI insert (780 bp) showed a single hybridizing band approximately 1000 bases long (not shown).

DISCUSSION

We have isolated and sequenced a cDNA complementary to mRNA for spinach leaf chloroplast CA. The sequence contains an open reading frame encoding a polypeptide of 254 amino acid residues. The identity of the deduced protein with CA was verified by (a) antibodies against purified spinach leaf chloroplast CA bind to the fusion protein, (b) N-terminal amino acid analysis of purified CA, and (c) the amino acid composition of the protein deduced from the nucleotide sequence. The amino acid composition calculated for the predicted mature enzyme is in good agreement with that determined by amino acid analysis of purified spinach CA (14) (Table I).

Two CA isozymes are present in spinach leaves: a cytosolic and a chloroplastic form (1, 13, 24). The cytosolic form is the only isozyme present in etiolated leaves (24). The chloroplastic form is located in the stroma (11, 22, 23, 31). CA is encoded in the nucleus and requires a transit peptide for passage through the chloroplast membrane. The deduced polypeptide corresponds to the larger precursor and contains the transit peptide. The site of cleavage was determined by sequencing the N-terminus of the purified enzyme. The sequence of the first 20 N-terminal amino acid residues of the mature protein was identical with that deduced from the nucleotide sequence (Fig. 1) and similar to that reported for spinach leaf CA by Hewett-Emmett *et al.* (8). This result indicates that the mature enzyme and the transit peptide of CA from spinach contains 221 and 33 amino acid residues, respectively, and have mol wt of 24,116 and 3,537. In this and in previous studies (14, 28), the mol wt of the mature CA subunit was estimated to be between 26 and 28 kD as determined by SDS-PAGE. The difference between these values and the value we obtained from the sequence analysis is probably due to the inaccuracy inherent in estimations made by gel electrophoresis.

The amino acid composition of the mature enzyme as deduced from the nucleotide sequence is in good agreement with that determined by amino acid analysis of purified CA (Table 1); however, this comparison is improved when the amino acid composition is calculated assuming a mol wt of 24 kD as compared to 26 kD (see ref. 14).

In addition to the open reading frame, the nucleotide sequence also contains 47 bp of 5′-untranslated and 170 bp of 3′-untranslated sequence with a poly(A) tail. It has been recognized that poly(A^+) mRNAs possess a conserved poly(A) addition signal, AATAAA, usually found 10 to 30 nucleotides upstream from the poly(A) tail (19). This signal is present in the CA mRNA and is shown boxed in Figure 1. That the northern blot analysis showed a single hybridization band indicating that either the transcript for the cytosolic form of the enzyme is so rare that it cannot be seen or that there is little sequence homology between the chloroplastic and cytoplasmic forms. The former suggestion seems the most likely.

A transit peptide of 33 amino acids in length is encoded by the nucleotide sequence assuming that translation is initiated at the methionine codon at nucleotide 1. Lutcke *et al.* (16) and Joshi (12) have surveyed a large number of plant mRNAs to define the plant AUG translation initiation consensus sequence the most frequently used consensus sequence being AACA*AUG*GC. The G and C 3′ to the AUG were preferred in 85 and 77% of the mRNAs, respectively. None of the specific nucleotides 5′ to the AUG were preferred in more than 62% of the mRNAs surveyed. Therefore, in its context the methionine codon at nucleotide 1 is a plausible AUG translation initiator.

The transit peptide consists mainly of hydrophobic and hydroxylamino acid residues like other transit peptides (26, 27); however, unlike other transit peptides it has a net negative charge due to a preponderance of glutamic acid residues. A transit peptide of 33 amino acids is about the same size as the transit peptides used to transport peptides of the light-harvesting complex of PSII and the spinach ATP synthase gamma-subunit (JG Mason, unpublished results) into the chloroplast but is smaller than those associated with the transport of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, ferredoxin, and plastocyanin (15, 26,

Table I. *Comparison of Deduced and Experimentally Determined Amino Acid Composition of Spinach Leaf Chloroplast Carbonic Anhydrase*

Amino Acid	Amino Acid Composition Derived from the Nucleotide Sequence[a]		Amino Acid Composition of Purified CA	
	Residues	mol %	Residues	mol %
Ala	19	5.60	21	8.64
Asn	8	3.79		
Asp	10	4.77	21	8.64
Arg	4	2.59	4	1.65
Cys	6	2.57	7	2.88
Gln	6	3.19		
Glu	15	8.03	22	9.05
Gly	21	4.97	24	9.88
His	6	3.41	6	2.47
Ile	9	4.22	10	4.12
Leu	18	8.45	20	8.23
Lys	18	9.57	18	7.41
Met	4	2.18	4	1.65
Phe	14	8.54	15	6.17
Pro	12	4.83	14	5.76
Ser	13	4.69	16	6.58
Thr	8	3.35	11	4.53
Trp	2	1.54	2	0.82
Tyr	8	5.41	8	3.29
Val	20	8.22	20	8.23
Total residues	221		243	
Mol wt	24,116		26,266	

[a] From ref. 14.

28), pyruvate,$P_i$dikinase (17), and spinach ATP synthase delta subunit (9). Furthermore, the transit peptide for spinach CA resembles all other transit peptides so far reported beginning with the sequence MA, but it does not conform in that it lacks the three major blocks of amino acid homology shared by the transit peptides of the light-harvesting complex of PSII or the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (see ref. 15).

There is very little sequence homology between the spinach chloroplast CA and any of the animal CA isozymes so far sequenced.

ACKNOWLEDGMENTS

We wish to thank Dr. Dennis Shaw of the John Curtin School of Medical Research, Australian National University, Canberra, Australia, for analyzing the amino acid sequence, and Dr. Adrian Gibbs of the Research School of Biological Sciences, Australian National University, Canberra, for assistance with the sequence similarity analysis.

LITERATURE CITED

1. **Atkins CA, Patterson BD, Graham D** (1972) Plant carbonic anhydrases I. Distribution of types among species. Plant Physiol **50:** 214–217
2. **Bradford MM** (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:** 248–253
3. **Burnell JN, Hatch MD** (1984) Regulation of photosynthesis: catalytic phosphorylation as a prerequisite for ADP-mediated inactivation of pyruvate, P_i dikinase. Biochem Biophys Res Commun **118:** 65–72
4. **Burnell JN, Hatch MD** (1988) Low bundle sheath carbonic anhydrase is apparently essential for effective C_4 pathway operation. Plant Physiol **86:** 1252–1256
5. **Deutsch HF** (1987) Carbonic anhydrases. Int J Biochem **19:** 101–113
6. **Fernley RT** (1988) Non-cytoloplasmic carbonic anhydrases. Trends Biochem Sci **13:** 356–359
7. **Graham D, Reed ML, Patterson BD, Hockley DG** (1984) Chemical properties, distribution and physiology of plant and animal carbonic anhydrases. Ann NY Acad Sci **429:** 222–237
8. **Hewett-Emmett D, Hopkins PJ, Tashian RE** (1984) Origins and molecular evolution of carbonic anhydrase isozymes. Ann NY Acad Sci **429:** 338–358
9. **Hermans J, Rother CH, Bichler J, Steppuhn J, Herrmann RG** (1988) Nucleotide sequence of cDNA clones encoding the complete precursor for subunit delta of thylakoid-located ATP synthase from spinach. Plant Mol Biol **10:** 323–330
10. **Huynh TV, Young RA, Davis RW** (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. *In* DM Glover, ed, DNA Cloning, Vol 1. IRL Press, Oxford, pp 49–78
11. **Jacobson BS, Fong F, Heath RL** (1975) Carbonic anhydrase of spinach: studies on its location, inhibition and physiological function. Plant Physiol **55:** 468–474
12. **Joshi CP** (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res **15:** 6643–6652
13. **Kachru RB, Anderson LE** (1974) Chloroplast and cytoplasmic enzymes. V. Pea-leaf carbonic anhydrase. Planta **118:** 235–240
14. **Kandel M, Gornall AG, Cybulsky DL, Kandel SI** (1978) Carbonic anhydrase from spinach leaves. Isolation and some chemical properties. J Biol Chem **253:** 679–685
15. **Karlin-Neumann GA, Tobin EM** (1986) Transit peptides of nuclear-encoded chloroplast proteins share a common amino acid framework. EMBO J **5:** 9–13
16. **Lutcke HE, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA** (1987) Selection of AUG initiation codons differ in plants and animals. EMBO J **6:** 43–48
17. **Matsuoka M, Ozeki Y, Yamamoto N, Hirano H, Kano-Murakami Y, Tanaka Y** (1988) Primary structure of maize pyruvate, orthophosphate dikinase as deduced from cDNA sequence. J Biol Chem **263:** 11080–11083
18. **Miller H** (1987) Practical aspects of preparing phage and plasmid DNA: Growth, maintenance, and storage of bacteria and bacteriophage. Methods Enzymol **152:** 145–170
19. **Nevins JR** (1983) The pathway of eukaryotic mRNA formation. Annu Rev Biochem **52:** 441–466
20. **Notstrand B, Vaara I, Kannan KK** (1975) Structural relationship of human erythrocyte carbonic anhydrase isozymes B and C. *In* CL Markert, ed, Enzymes: Molecular Structure, Vol 1. Academic Press, New York, pp 575–599
21. **Okabe K, Yang S-Y, Tsuzuki M, Miyachi S** (1984) Carbonic anhydrase: its content in spinach leaves and its taxonomic diversity studied with anti-spinach leaf carbonic anhydrase antibody. Plant Sci Lett **33:** 145–153
22. **Poincelot RP** (1972) The distribution of carbonic anhydrase and ribulose diphosphate carboxylase in maize leaves. Plant Physiol **50:** 336–340
23. **Poincelot RP** (1972) Intracellular distribution of carbonic anhydrase in spinach leaves. Biochim Biophys Acta **258:** 637–642
24. **Reed ML** (1979) Intracellular location of carbonate dehydratase (carbonic anhydrase) in leaf tissue. Plant Physiol **63:** 216–217
25. **Sanger F, Nicklen S, Coulson AR** (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA **74:** 5463–5467
26. **Schmidt GW, Mishkind ML** (1986) The transport of proteins into chloroplasts. Annu Rev Biochem **55:** 879–912
27. **Smeekens S, Weisbeck P** (1988) Protein transport towards the thylakoid lumen; post-translational translocation in tandem. Photosynth Res **16:** 177–186
28. **Tobin AJ** (1970) Carbonic anhydrase from parsley leaves. J Biol Chem **245:** 2656–2666
29. **Venta PJ, Montgomery JC, Tashian RE** (1987) Enzymes. *In* MC Rattazzi, JG Scandalios, GJ Whitt, eds, Current Topics in Biological and Medical Research, Vol 14. Alan R. Liss, New York, pp 59–72
30. **Warburg O, Christian W** (1941) Isolierung und kristallization des garungsferments enolase. Biochem Z **310:** 384–421
31. **Werdan K, Heldt HW, Geller G** (1972) Accumulation of bicarbonate in intact chloroplasts following a pH gradient. Biochim Biophys Acta **283:** 430–441