Identification, characterization and sequence analysis of the gene encoding phosphoenolpyruvate carboxylase in *Anabaena* sp. PCC 7120

IRENE LUINENBURG and JOHN R. COLEMAN*

Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, Ontario, Canada M5S 3B2

(Received 14 October 1991; revised 13 January 1992; accepted 17 January 1992)

The gene (*ppc*) encoding phosphoenolpyruvate carboxylase (PEPCase) in the cyanobacterium *Anabaena* sp. PCC 7120 has been isolated, characterized and its nucleotide sequence determined. Heterologous hybridization using the *Synechococcus* sp. PCC 7942 *ppc* gene as a probe of an *Anabaena* genomic DNA library identified an 8.2 kb *Hind*III DNA fragment that contained a 3.08 kb open reading frame encoding the cyanobacterial PEPCase. Deletion analysis of the 8.2 kb DNA fragment was used to determine sequences required for expression of enzyme activity in *Escherichia coli* cells. Primer extension data have been used to identify the cyanobacterial transcription initiation site and the position of the *ppc* start codon. Comparisons of the *Anabaena* deduced amino acid sequence with the *Synechococcus* sp. PCC 6301, *E. coli* and higher-plant *ppc* sequences have also been performed and the data are discussed with respect to conservation of specific regions of the protein.

Introduction

The cyanobacteria assimilate inorganic carbon (C_i) primarily by the C_3 photosynthetic pathway and employ ribulose biphosphate carboxylase/oxygenase (Rubisco) as the principal CO_2 -fixing enzyme. However, they can also fix large amounts of carbon in the light as C_4 acids (Coleman & Colman, 1980). These C_4 acids, primarily aspartate and malate, may represent as much as 20% of the total carbon assimilated at the onset of illumination and have been shown to be the product of phosphoenolpyruvate carboxylase (PEPCase; EC 4.1.1.31) activity (Coleman & Colman, 1981; Owttrim & Colman, 1986). This enzyme catalyses the irreversible carboxylation of PEP, utilizing HCO_3^- to form oxaloacetate and P_i . Rapid amination or reduction of the newly synthesized oxaloacetate results in the production of aspartate or malate, respectively (Maruyama *et al.*, 1966). The role of PEPCase in cyanobacterial carbon metabolism is still poorly understood. Early studies suggested that the cyanobacteria utilize primarily a C_4 mechanism for C_i

fixation (Dohler, 1974; Colman *et al.*, 1976; Lem *et al.*, 1986). High levels of PEPCase activity and typical C_4 characteristics of cyanobacterial photosynthesis, such as low CO_2 compensation points, little oxygen inhibition of photosynthetic carbon fixation and low rates of photorespiration, stimulated this speculation. It is now known that high-affinity C_i transport systems capable of generating large intracellular C_i pools in the cyanobacteria are responsible for these ' C_4 -like' characteristics of photosynthesis (Coleman, 1991). PEPCase activity is thought to play an anapleurotic role by generating carbon skeletons for biosynthesis (Owttrim & Colman, 1988). Cyanobacteria as a group are characterized by possessing an incomplete tricarboxylic acid (TCA) cycle lacking both α -ketoglutarate dehydrogenase and NADH oxidase (Smith *et al.*, 1967). As a result of these deficiencies, metabolic energy cannot be derived from this pathway and the cycle is incapable of regeneration. The C_4 acids synthesized by PEPCase activity could be used to replenish the TCA cycle intermediates, which are in turn utilized for biosynthetic reactions. The synthesis of C_4 acids may be required for the production of cyanophycin, a nitrogen storage compound accumulated by cyanobacteria (Weathers & Allen, 1978).

Methods

Cultures of *Anabaena* sp. PCC 7120 were maintained on BG-11 medium (Allen, 1968) solidified with 1.0% (w/v) Bacto-agar or grown

* Author for correspondence. Tel. (416) 978 2339; fax (416) 978 5878; E-mail coleman@vm.uts.utoronto.ca.

Abbreviations: CAM, crassulacean acid metabolism; PAR, photo-synthetically active radiation; PEP, phosphoenolpyruvate; PEPCase, phosphoenolpyruvate carboxylase.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been given the accession number M80541.

in liquid batch culture at 28 °C in an illuminated shaker (100 $\mu\text{E m}^{-2} \text{s}^{-1}$ PAR). *Escherichia coli* strains used included JM101 (plasmid host cells), LE392 and C600 *hfl* (λ host cells) and 3594 (strain deficient in PEPCase activity as a result of a point mutation in *ppc*: Glansdorff, 1965). Growth of *E. coli* strains, manipulation of plasmid pBS+ (Stratagene) and λ 1149, radioactive labelling of nucleic acids, Southern blot analysis, exonuclease digestion of cloned DNA to form nested sets of deletions, and DNA sequencing and analyses were carried out as previously described (Sambrook *et al.*, 1989; Owttrim & Coleman, 1989). Primer extension reactions were performed as described by Kjems & Garrett (1988). Cyanobacterial RNA and genomic DNA was isolated using the method described by Golden *et al.* (1987). *In vitro* PEPCase assays using clarified cell lysates (centrifuged at 13000 *g* for 10 min) and $\text{H}^{14}\text{CO}_3^-$ were performed essentially as described previously (Colman *et al.*, 1976; Luinenburg & Coleman, 1990). Total soluble protein in the clarified cell lysate was determined using the procedure of Bradford (1976).

Results and Discussion

Isolation of the *Anabaena ppc* gene

The isolation of the *Anabaena ppc* gene was accomplished by screening a λ 1149 *Hind*III library of *Anabaena* sp. PCC 7120 genomic DNA. The probe used for this screening was a 4.4 kb DNA fragment containing the *ppc* coding sequence previously isolated from the unicellular cyanobacterium *Synechococcus* PCC 7942 (Luienburg & Coleman, 1990). Genomic *Anabaena* sp. PCC 7120 DNA digested with *Hind*III and hybridized in Southern blots with the 4.4 kb *Synechococcus* DNA fragment at 65 °C resulted in a single hybridizing band of 8.2 kb (Fig. 1, lane B). An 8.2 kb fragment which corresponded to this band was isolated from the *Anabaena* genomic library (Fig. 1, lane D), ligated into the plasmid pBS+ and used to transform the PEPCase-deficient *E. coli* strain 3594. Transformants were selected for their ability to grow on minimal medium, indicating that the 8.2 kb *Anabaena* DNA fragment was able to complement the Ppc⁻ phenotype of this mutant. PEPCase assays of these transformants showed that the enzyme was active when the 8.2 kb fragment was inserted into pBS+ in either orientation, demonstrating that promoter elements located on this fragment are able to function within *E. coli*.

Sequence analysis

Restriction endonuclease mapping and PEPCase activity measurements were used to locate the *ppc* coding region on the 8.2 kb fragment, and a 4.3 kb segment spanning the coding region was sequenced in both directions using two independently isolated clones. A portion of these data, showing the entire coding region and some 5' and 3' sequence, is presented in Fig. 2. The deduced amino acid sequence is also shown. The ORF

Fig. 1. Autoradiogram of a Southern blot of *Anabaena* sp. PCC 7120 chromosomal DNA. The DNA was digested with *Eco*RI (lanes A and C) and *Hind*III (lanes B and D) and probed at 65 °C with either the *Synechococcus* sp. PCC 7942 *ppc* (lanes A and B) or the *Anabaena* sp. *ppc* (lanes C and D). The positions and molecular sizes of the hybridizing fragments are indicated.

encodes a polypeptide of 112653 Da, which is consistent with the reported monomeric molecular mass of 100500 Da as determined by SDS-PAGE (Lem *et al.*, 1986). In comparison, the deduced amino acid sequence of *ppc* from *Synechococcus* sp. generates a monomer molecular mass of 121082 Da (Katagiri *et al.*, 1985). Previous studies of the PEPCase holoenzyme, isolated from a number of sources, have shown that the active form of the enzyme is a tetramer composed of four identical subunits (O'Leary, 1982). The mass of the *Anabaena* holoenzyme would therefore be approximately 450000 Da, which is somewhat larger than that previously determined by gel filtration chromatography (Lem *et al.*, 1986) but still indicative of the active form of the enzyme existing as a tetramer.

In the absence of N-terminal sequence of the purified *Anabaena* protein, we have identified one of two methionine residues within the same ORF as the putative initiation codon for translation. The two methionine residues are located at nucleotide -19 and at nucleotide 113 (Fig. 2). Primer extension data support the hypothesis that the second ATG is the correct start site for translation in *Anabaena*. As shown in Fig. 3, the primary transcription start site is 3' of the first in-frame ATG and indicates that transcription begins at nucleotide 1. Putative -10 and -35 regions are also indicated in Fig. 2. These regions were determined on the basis of

CGC CTG CCT ATT TTA CCC TTC CCT GAT GAC GAT GAA TCA ACC AAT TAA AGT TTA AAA	1766	CAG GTT GTA CCA TTA TTT GAG ACT GTC GAA GAT TTA CAA CGT TCT CGC AGC GTC ATG
ATA TAC TTC AAA AAT ATG CTC ACA TAA GTA GAT AAG AGC ATA AGT TAC CAA AAA TCA	1823	GCA CAG TTA TTT GAA TTA CCT CTA TAC CGC GCT TTG CTA GCT GGT TAC AAA AAT
-58 ATC TGT TAA CAA GTA ATT <u>GCG TTT ACA</u> AGA GAA GCA AGT ATG GGT TCT <u>GTT TTA</u> TAC	1880	ACC GAA GTA AAA GTA CCG AAT ACT GAG CTA ACA CCC CAA TCC CCA GCC CCC AGT CCC
-1 TCT TTA TCT GAA TCT GCT AAT TTA TAT CCA GCA TCG GAA TTA TTC TTG CGT CAT CGT	1937	CAG TCC GTA CTC ACC CCC GAC TTA CAA GAA GTC ATG CAG GGA TAT TCC GAC AGC AAC
96 CTC CAA ATA GTA GAA GAA CTG TGG GAG TCG GTA CTG CCG <u>CAA</u> GAA TCG GCC CAA AAT	1994	AAA GAC TCC GGC TTC TTG AGT AGC AAT TGG GAA ATT CAT AAA GCC CAA AAA TCA TTA
113 ATG GTG GAT TTG TTA CCG CAG TTG CGT GAT TTG TGT TCA CCA GAA GGG CAA GCC ACA	2051	CAG CAA ATC GCC GAA GAA TAT GGC GTA AAT TTG CGG ATT TTC CAC GGG CGC GCG GGT
Met Val Asp Leu Leu Arg Gln Leu Arg Asp Leu Cys Ser Pro Glu Gly Gln Ala Thr	2108	TCT GTA GGA CGA GGC GGC GGC CCA GCC CAC GAG GCG ATT TTG GCT CAA CCA GGA CAC
170 AAA GAC CAA GCA GTT TCT GCT GTC AAG TTA ATT GAA CAG TTG AAT ATT AAT GAA GCC	2165	ACC ATC AAC GGG CGG ATT AAA ATT ACC GAA CAA GGG GAA GTA TTA GCC TCC AAG TAC
Lys Asp Gln Ala Val Ser Ala Val Lys Leu Ile Glu Gln Leu Asn Ile Asn Glu Ala	2222	TCC TTG CTA GAT TTA GCC CTG TAC AAC TTG GAA ACC ATC ACC ACG GCG GTG ATT CAA
227 ATT CCG GCA GCT AGG GCT TTT GCT TTG TAT TTT CAG CTG ATC AAC ATC ATA GAG CAG	2279	GCC AGC CTC CTG CCA ACT GGG TTT GAT GAT ATC GAA CCT TGG AAC GAA ATT ATG GAA
Ile Arg Ala Ala Arg Ala Phe Ala Leu Tyr Phe Gln Leu Ile Asn Ile Ile Glu Gln	2336	GAA TTA CCA GCG CGA TCG CCG CAA CAT TAT CGT GGT TTA ATT TAC GAA CCG CAT GAT
284 GAA TAT GAG CAA AGG CAG CAA TTA ACT CGC TAT TCT GAC TTA GAA CCG GAA ACA GCA	2393	TTC ATT GAC TTT TTC CAC CAA GTC ACA CCC ATT GAA GAA ATC AGT CAA CTA CAA ATT
Glu Tyr Glu Gln Arg Gln Gln Leu Thr Arg Tyr Ser Asp Leu Glu Ala Glu Thr Ala	2450	AGT TCT CGT CCC GCG CGT CGT TCT GGT AAG GAA GAT TTA AGC AGT CTC GCG GCC
341 CCC CTA AAT GGT TCC GAA AAC ATT ACC TCA TCC TCT AAC CAC GAA GAT GAT GTT	2507	Att CCT TGG GTA TTT AGC TGG ACA CAA ACA CCA TTA CTA CCT TCT TGG TAC GGT
Pro Leu Asn Gly Ser Glu Asn Ile Thr Ser Ser Ser Asn His Asn Glu Asp Asp Val	2564	GTC GGT ACA GCT TTA CAA GAA TTT TTC AAC GAA GAA CCA GAA GAA CAT CTC AAA TTG
398 ATT TTC AAC CCG GGT TTA GGG ACT GAT TTC TTA GGA AAA AAT TGG ACT AAT CGT GGA	2621	ATG CGC TAC TTT TAT GTC AAG TGG CCT TTC TTC AAG ATG GTG ATT TCT AAA GTA GAA
Ile Phe Asn Arg Gly Leu Gly Thr Arg Phe Leu Gly Lys Asn Thr Thr Asn Arg Gly	2678	ATG ACC TTC GCG AAA GTA GAT ATG CAA ATG GCT GGT CAC TAC GTA CAA GAA TTG TCT
455 CAA GGG AAA CAA AAA GGT ACT TTT GCA GCG TTA TTT CCC TTA TTG TCT AAA TTG AAT	2735	GAC CCT GAA GAC AAA CCC CGG TTT GAG AAA GTA TTT GAG CAA ATC GCC AAC GAA TAT
Gln Gly Lys Phe Lys Glu Thr Phe Ala Ala Leu Phe Pro Leu Leu Ser Lys Leu Asn	2792	TAT CTC ACC AGA GAT TTA GTC TTA AAA ATC ACC GAC CAC GGT CGG CTT TTA GAT GGT
512 GTC CCA CCC CAA CAA ATT CAA CCG CTG ATT TCC CAA CTA GAT GTC CGT TTG GTA TTC	2849	GAT CCT GTA CTC CAG CGT TCT GTA CAG TTA CCG AAT GGT ACA ATC GTC CCA CTA GGG
Val Pro Pro Gln Gln Ile Gln Leu Thr Arg Tyr Ser Asp Leu Val Glu Asn Arg Ala Lys	2906	TTT ATC CAA GTT TCC CTC CTC AAG CCG CTA CCG CAG TCC AAA AAT AAT ACT GCA ACC
569 ACT GGC CAC CCA AGC GAA ATT GTC CCG CAC ACC ATT CCG GAC AAA CCG CAA GTA	2963	TCT GGT GTC ATT CAC TCT CGT TAC AGT AAG GGA GAA TTA TTG CCG GCG GCA CTG TTA
Thr Ala His Pro Thr Glu Ile Val Arg His Thr Ile Arg Asp Lys Gln Arg Gln Val	3020	ACC ATT AAT GGT ATT GCC GCA GGG ATG AGA AAT ACA GGT TGA TTT GTG AGT TCT CAG
626 GTA GAT TTG TTG CAA CAA CTC GAT GAG GTG GAA AAT CGA GCT AAG GAC GGC GGC GGC	3077	TAG TTA GTC GTC AGT GGT AAA ACA CAT ACA ACT GAC AAC TGA CAA CCG AAC AAA TGG
Val Asp Leu Leu Gln Gln Leu Asp Glu Val Glu Asn Arg Ala Lys Asp Gly Gly Gly	3134	TCA AAA ATA AAA TCT TAA TCT GTA CTG
683 TAT CCT TGG GAA GCG GCG GAA ATT CCG GAA AAA TTG CTA GAA GAA ATT CCG TTG TGG		
Tyr Pro Trp Glu Glu Gly Glu Ile Arg Lys Leu Glu Lys Leu Glu Trp		
740 TCG CGT ACA GAC GAA CTC CAC CAG TTT AAG CCC ACT GTG CTG GAT GAA GTG GAT TAT		
Trp Arg Thr Asp Glu Leu His Gln Phe Lys Pro Thr Val Leu Asp Glu Val Asp Tyr		
797 GCT TTG CAT TAC TTC CAA GAA GTT TTA TTT GAT GGG ATT CCC CAA CTG TAT AAG CGT		
Ala Leu His Tyr Phe Gln Glu Val Leu Phe Asp Gly Ile Pro Gln Leu Tyr Lys Arg		
854 TTC AAA TAC GCC CTA AAT CAA ACT TTC TCG TGG TTA GAA CCA CCA AGC AAA GAT TTC		
Phe Lys Tyr Ala Leu Asn Thr Thr Phe Ser Trp Leu Glu Pro Pro Ser Lys Asp Phe		
911 TGT TCC TTT GGT TCT TGG GTA GGT TCA GAC AGA GAT GGT AAT CCA TCG GTA ACA CCG		
Ser Ser Phe Gly Ser Trp Val Gly Ser Asp Arg Asp Gly Asn Pro Ser Val Thr Pro		
968 GAA ATT ACC TGG CAG ACA GCT TGT TAT CAG CCG AAA ATG GTG CTG GAA AGA TAT ATC		
Glu Ile Thr Trp Gln Thr Ala Cys Tyr Gln Arg Lys Met Val Leu Glu Arg Tyr Ile		
1026 AAG TCT GTC ACC CAA TGT ATT GAA TTA AGT ATC TCC ATG CAC TGG AGT GAC GTT		
Lys Ser Val Thr Gln Leu Ile Glu Leu Leu Ser Ile Ser Met His Trp Ser Asp Val		
1082 TTA CCA GAC TTA TTG GAA TCC TTG GAG TTA GAT CAG TCC CAG TTA AGT GAA GTA TAC		
Leu Pro Asp Leu Leu Glu Ser Leu Glu Leu Asp Gln Ser Gln Leu Ser Glu Val Tyr		
1139 GAT GCT CTG GCG CTG AGA TAT CCG CAA GAA CCC TAT CGT CTC AAA CTA GCT TAC GTC		
Asp Ala Leu Ala Leu Arg Tyr Arg Gln Glu Pro Tyr Arg Leu Lys Leu Ala Tyr Val		
1196 CTC AAG CCG CTG GAA AAT ACC CCG GAT CCG AAT TTA GCT TTA TAT AAA GGT GAA ACC		
Leu Lys Arg Leu Glu Asn Thr Arg Asp Arg Asn Leu Ala Leu Tyr Lys Gly Glu Thr		
1263 CCC CAA AAT GAA GAT AGC CCC ATG TAT CGT TCG GCG TCG GAA TTT TTG CCG GAA CTG		
Pro Thr Asn Glu Asp Ser Pro Met Tyr Arg Ser Gly Ser Glu Phe Leu Ala Glu Leu		
1310 CGA TTA ATT CAG CAT AAC TTG ACG GAA ACA GGT TTA AGC TGC CGA GAA TTG GAT AAT		
Arg Leu Ile Gln His Asn Leu Thr Glu Thr Gly Leu Ser Cys Arg Glu Leu Asp Asn		
1367 TTG ATC TGT CAA GTG GAA ATT TTT GAC TTT AAC CTG ACA AAA TTA GAC ATC CGT CAA		
Leu Ile Cys Gln Val Glu Ile Phe Asp Phe Asn Leu Thr Lys Asp Arg Ile Arg Gln		
1424 GAA TCC ACC AGA CAT TCG GAC GCG TTA AAC GAG ATT CTC GAC TAT CTC CAA TTA TTA		
Glu Ser Thr Arg His Ser Asp Ala Leu Asn Glu Ile Leu Asp Tyr Leu Gln Leu Thr		
1481 CCC CAG CCA TAC AAC CAG CTC TCC GAA GAA GGT GCT TGG CTA CCA ACC GAA		
Pro Gln Pro Tyr Asn Asp Leu Ser Glu Glu Gln Arg Val Ala Trp Leu Thr Thr Glu		
1538 CTG CAA ACC CGT CCG CCA TTA ATT TCG TCA GAA CTA CCA TTC TCT GAC AAA ACT AAT		
Gln Gln Thr Arg Pro Leu Ile Ser Ser Glu Leu Pro Phe Ser Asp Lys Thr Asn		
1595 GAT GTC ATT AAA ACT TTC CCG GTA GTG CGA TCG CTC CAA CAA GAA TTT GGC ATT AAT		
Asp Val Ile Lys Thr Phe Arg Val Val Arg Ser Leu Gln Gln Phe Gly Ile Asn		
1652 ATC TGC CAA ACT TAC ATC ATT AGT ATG TGT CGT CAA GTC AGC GAT GTC TTA GAA GTT		
Ile Cys Gln Thr Tyr Ile Ile Ser Met Cys Arg Gln Val Ser Asp Val Leu Glu Val		
1709 TTT CTT CTA GCC AAA GAA GCG ACA CTT TTT GAC CCA GCG ATC GCC GGT ACA ATT		
Leu Leu Leu Ala Lys Glu Ala Arg Leu Phe Asp Pro Ala Ile Ala Val Gly Thr Ile		

Fig. 2. Nucleotide and deduced amino acid sequence of the *Anabaena* sp. PCC 7120 *ppc* gene. Putative promoter regions and ribosome-binding site are underlined and double underlined, respectively. A vertical arrow indicates the site of initiation of transcription (nucleotide 1) as determined by primer extension. Putative transcription termination structures are indicated with directional arrows. Oligonucleotides used in sequencing include: 5' -¹⁶⁴CCT GAT GAC GAT GAA TC 3'; 5' -¹⁵³GAT CAA CAT CAT AGA GC 3'; 5' -¹⁰⁴⁸CGC CAA GAA CCC TAT CG 3'; 5' -¹⁶⁰²TCT AGC CAA AGA AAG CC 3'.

their similarity to characterized promoter sequences determined for *Anabaena* sp. PCC 7120 (Schneider *et al.*, 1991) and *E. coli* consensus sequences as well as their proximity to the transcript start site. The putative -35 sequence is much more similar to the bacterial consensus sequence than is the -10 region, but in general the *ppc* gene, on the basis of *E. coli* sequence similarity, would be categorized as having a weak promoter (Schneider *et al.*, 1991). A large region of dyad symmetry 3' of the translation termination codon (nucleotide 3060) would appear to be a site of transcription termination (Fig. 2).

Calculations using the primer extension and sequence analysis data suggest that the length of the transcript would be approximately 3.2 kb. This compares favourably with a maximum transcript size of approximately 3.4 kb as determined by Northern analysis in *Synechococcus* sp. PCC 7942 (data not shown).

Expression studies

Expression of *Anabaena* sp. PCC 7120 *ppc* in *E. coli* host cells assumes either that the *E. coli* transcriptional and

Fig. 3. Identification of the *ppc* transcription start site in *Anabaena* sp. PCC 7120. A 17-mer primer complementary to a sequence 5' of the coding region was used in primer extension assays with total *Anabaena* RNA isolated from exponentially growing cells. Products were electrophoresed concurrently with a sequence ladder generated using the same primer with double-stranded plasmid DNA containing the intact *ppc* gene. The primer used was 5' CCG CAG TAC CGA CTC CC 3'. The arrow indicates the major primer extension product and the DNA sequence in the vicinity is identified.

translational machinery is capable of utilizing cyanobacterial promoter elements or that fortuitous *E. coli* signals occur within the sequence. Examples of cyanobacterial gene expression in *E. coli* have been cited (Borthakur & Haselkorn, 1989; Kodaki *et al.*, 1985; Tandeau de Marsac & Houmard, 1987; Tumer *et al.*, 1983), and common promoter sequence elements are capable of being used in both systems (Schneider *et al.*, 1991). To study the expression of *Anabaena* sp. PCC 7120 *ppc* in *E. coli* the 8.2 kb DNA fragment containing the cyanobacterial gene was cloned into pDPL13, a vector with no additional promoters aside from that found controlling the ampicillin gene (Gendel *et al.*, 1983). Confounding promoter effects are, therefore, not expected. Exonuclease digestion of the 8.2 kb *Hind*III fragment was then used to deduce the flanking 5' and 3' DNA sequences of the gene required for expression of enzyme activity within the bacterial host cell. The constructs generated by exonuclease activity are shown in Fig. 4.

E. coli cells containing the various constructs were grown from stationary overnight cultures diluted 1:150 with fresh prewarmed LB medium supplemented with ampicillin at 50 µg ml⁻¹. The flask to medium volume ratio used was 250:15 and cultures were incubated at 37 °C and 200 r.p.m. The cells were harvested at specific time intervals over a 24 h period and at each time point the various cell lines had achieved similar OD₆₀₀ values. PEPCase activity, as measured per cell and per mg soluble protein, increased throughout the exponential growth phase of the bacterial culture and then declined slowly during stationary phase (data not shown). As shown in Table 1 (see Fig. 4 for constructs), maximal expression of cyanobacterial PEPCase activity in *E. coli*

Fig. 4. Restriction map of the *Anabaena* sp. PCC 7120 8.2 kb *Hind*III fragment encoding *ppc* (a) and diagrammatic representation of various subclones generated using restriction enzymes or exonuclease III digestion (b-f). PEPCase activities generated by these constructs when cloned in *E. coli* are shown in Table 1. The thick bars represent the *ppc* coding region and transcript size is depicted as a horizontal arrow. Vertical arrows indicate restriction enzyme sites abbreviated as: A (*Acc*I), D (*Dra*I), H (*Hpa*I), Hi (*Hind*III), P (*Pvu*II), X (*Xba*I). Restriction enzymes that do not cleave the 8.2 kb *Hind*III fragment include: *Asp*718, *Bam*HI, *Cla*I, *Eco*RI, *Nco*I, *Pst*I, *Sac*I, *Sal*I, *Sma*I, and *Sph*I.

Table 1. *In vitro* PEPCase activity of the *Anabaena* sp. PCC 7120 *ppc* constructions described in Fig. 4 expressed in cells of *E. coli* strain 3594 (*Ppc*⁻)

Values reported are means of four replicates $\pm$ standard deviation. ND, No detectable activity.

Sample	PEPCase activity [$\mu\text{mol C (mg soluble protein)}^{-1} \text{ h}^{-1}$]
3594	ND
a	2.58 ± 0.10
b	3.37 ± 0.30
c	0.06 ± 0.003
d	7.19 ± 0.29
e	ND
f	0.52 ± 0.02

(constructs *a*, *b* and *d* compared with *f*) requires sequences located 5' of the putative -35 and -10 region of the *ppc* gene. The significant reduction in expressed PEPCase activity of construct (*f*) suggests that at least one enhancer element could be located 5' of bp -119. The uniform size of the cyanobacterial *ppc* transcript when expressed in *E. coli* indicates that the same transcription start site is being used by each construct (data not shown). The much reduced activity generated by construct (*c*) would appear to be the result of deletion of the -35 region as well as an enhancer element. The elimination of PEPCase activity in construct (*e*) would appear to be the result of inability to synthesize a stable mRNA or protein. This deletion removed 12 amino acids at the carboxy-terminus of the protein as well as the termination codon. The sequence containing the putative transcription termination structure (region of dyad symmetry) has also been removed. SDS-PAGE analysis of total soluble proteins synthesized by *E. coli* containing this construct showed that the cyanobacterial PEPCase was not present (data not shown).

Sequence comparison of *ppc* genes and PEPCase from different organisms

A comparison of the deduced *ppc* amino acid sequences determined for *Anabaena* sp. PCC 7120 and *Synechococcus* sp. PCC 6301 (Katagiri *et al.*, 1985) is shown in Fig. 5. The position of the start codon for the *Synechococcus* gene is different from that determined for *Anabaena*, and was assigned by Katagiri *et al.* (1985) on the basis that it initiated the longest ORF as determined by their sequence data. The position of the initiation codon for the *Anabaena ppc* gene was determined on the basis of primer extension and sequence data as well as SDS-PAGE determination of the monomer molecular mass. There is, however, some amino acid sequence similarity between the two genes 5' of the proposed

A	.....mgsvl.....y...	
S	MEAVSLRDIIGFISATVQIRPAFCVQSDVVGSGNPRMNYQNARTAMSAALQSSDAHRT	60
A	...slisesanlypaselfirhrliqveelwesvrlrqcggnmVDLLRLQLCSPEQGAT	19
S	VSSPIATDLDLSSPLEFFLRHRLTVVEELMEVVLRLQECGQELVDILTQLRLDLSPEQGAP	120
A	KQQAVSAVKLLIEQLNINEAIRAARAFALYFQLIIRIIIEQVEEQRLT.RYSDLEAETAP	77
S	EVGGEALVQVITLIELSDAIRAARAFALYFQLINIVEQHYEQYQQLAYERSRLPLPGP	180
A	LNGSEINITSNNHNEDDVIFNRGLGTFDLGKNWTRNGQKQKGTFAALFPLLSKLVNPPQ	137
S	DESPEGLHTIEIPQHOLDPFAAVIPLN.....QDPATFQTLFPRLRLQINVPQ	228
A	QIQRLISQLDVLRLVIRHPTIEIRHRTIRDRQVVDLQLQDLIVENRAKDGQYWEAGE	197
S	MIQELTDRIIRLVFTAHPTEIVRHTIRDKQRIAYLLRLQDLELET.GKNRGFRLELAQN	287
A	IREKLIIEIRLWWRITDELHQFKPTVLDEVDYALHYFQEVLDGIPOLYKRKYALNQTFE	257
S	IRQQLTEEIRLWWRITDELHQFKPTVLDEVDYALHYFQEVLDGIPOLYKRKYALNQTFE	317
A	WLEPISKDFCSFGSGWSDRDGPNSTPEITWQACYQKRMVLERVYKSVTQILIELLSIS	317
S	DIQPRFYNFCQFSGWSDRDGPNSTPEITWQACYQKRSVLDRYITAVEHLRNVLIS	407
A	MHWSDVLPDLLESLELDQSQLSEVDYDALRYRQEPYRLKLAIVLKRLENTDRNLALYK	377
S	MHWSEVLPELLSSLEQESMLFPETYEQLAVRYRQEPYRLKLSVILRLHTRDRNTRIQQ	467
A	GET.....PTNEDSPMYRSGSEFLAEIRLIQHNLITETGLSCRELDNLICQVEIFDFNL	430
S	QOEKDPITPLPEYRDGTLYQAGTAFLEDLKLQHMLKQGLSCVELEKLIQVEIFDFNL	527
A	TKLDIRQESTRHSDALNEILDYQLLPPYNDLSIEQVAVLITTEQLTRRRLISSLEPFS	490
S	VHLDIRQESSRHSDAINEICVQLILPPYNELSEARTAWLVQELKTRRLPVPARMPFS	587
A	DKTNDVTKTFRRVRSQQEFGINICOTFISRCROVSDVLEVLIAKEARLFDPDAIVGT	550
S	ESTREIETLRMVKQLQEEFGAECQTYIISMSRELSLLEVLALLAKEVGLVDPVTGKSS	647
A	IQVVFIFETVEDLQRSRVMRQLFELPLYRALLAGGYKNTVEYVNPTELTPQSPAPSPQS	610
S	LQVIFIFETVEDLQNAVRVMTALFELP.....YTQLNPTQSEF.....	686
A	VLTPLDIEVMLGYSDSNRDSGLSSNWEIHKAKSLQIAEEYGVNLRIFHGRGCSVGRG	670
S	.....IQEVMIGYSDSNKDSGLSSNWEIHKAKALGTAVDRHVKLRIFHGRGCSVGRG	741
A	GGPAHEATILAGHGSINRIKITEQGEVLASKYALPHALYNLETITTAIVQSSLLRTGF	730
S	GGPAYFAIIAQGRITTDGRKITEQGEVLASKYALPHALYNLETITTAIVQSSLLRTGF	801
A	DDIEPNEIMEELAARSQHRYGLIYEQPDFIDFHFQVTPFIEISQLQISSRPARRPSGK	790
S	DDIEPNNQIMEELAARSRRHYALVYEQPDFIDFHFQVTPFIEISKLQISSRPARRKTGK	861
A	KDLSSIRIIVFSSTQTRFLPSRYVDTLQEEFNPEEHLKLMRYFVVKWBFKRV	850
S	RDLSLRALIPVFSWTQSRFLPSRYVGTALQELQERPEQNLNLLRYFYEKWPFPRMV	921
A	ISKVENTLAKVDMQAGHYVQELSDPEDKPRFEKVFQIANEYVLTDRDLVKITDHGRLL	910
S	ISKVENTLAKVDLQIAHHYVHELANPEDQERFERVFSQIAAEFQLTCHLVLTITNHGRLL	981
A	DGDPELQRSVQLRNGTIVPLGFIQVSLIKRLKSKNNATSGVIHSRYSGKELLRGALL	970
S	DGDPELQRSVQLRNGTIVPLGFLQVALLKRLKQYRQQTETTGLMRSRYSGKELLRGALL	1041
A	INGIAAGMRNTG 982	
S	INGIAAGMRNTG 1053	

Fig. 5. Comparison of the deduced amino acid sequences for *Anabaena* sp. PCC 7120 (A) and *Synechococcus* sp. PCC 6301 (S) *ppc* as generated by an alignment program (Corpet, 1988). Numbering of the amino acid residues for *Synechococcus* is as described by Katagiri *et al.* (1985). The numbering of the *Anabaena ppc* residues is as determined on the basis of primer extension and SDS-PAGE analysis. Lower-case letters in the *Anabaena* sequence denote amino acids encoded in the ORF 5' of the putative initiating methionine. Identity at the amino acid level between the two cyanobacterial sequences is indicated by the symbol *. Amino acids conserved within the *Anabaena*, *Synechococcus*, *E. coli*, *Zea mays* and *Mesembryanthemum crystallinum ppc* sequences are indicated as open circles.

Anabaena translation start site (Fig. 5, lower-case letters) although it would seem that this region is not translated in *Anabaena*. Amino-terminal sequence analysis will be needed to determine the correct translation start site. It is possible that a mutation in the *ppc* sequence following divergence of the two genera has resulted in the

introduction of the new initiation codon in *Anabaena*. The alignment of the deduced amino acid sequences of the two cyanobacteria proteins reveals considerable homology, with approximately 65% identity over the coding region. In contrast, the DNA comparison between these genes suggests that there has been considerable divergence of the two sequences, with a significant reduction in identity over the coding region. Fig. 5 also shows those residues which have been conserved in the deduced amino acid sequences of the PEPCase proteins isolated from the C₄ plant *Zea mays* (Izui *et al.*, 1986), the CAM induced and non-induced (C₃) plant, *Mesembryanthemum crystallinum* (Cushman *et al.*, 1989; Rickers *et al.*, 1989), and *E. coli* (Fujita *et al.*, 1984). Overall, the amino acid sequence comparisons made between the *Anabaena* PEPCase and the other proteins reveal approximately 34%, 29%, 31% and 32% identity with the *Zea*, C₃ and CAM *Mesembryanthemum* and *E. coli* enzymes, respectively. The amino acid sequence of the carboxy-terminal half of the monomer is somewhat more conserved than the amino-terminus. These data are consistent with the hypothesis that the carboxy-terminal region contains the catalytic domain, whereas those domains involved in enzyme-specific allosteric regulation of activity by metabolites are found within the more variable amino-terminal region (Ishijima *et al.*, 1985; Jiao *et al.*, 1990). As has been observed by others (Izui *et al.*, 1986; Rickers *et al.*, 1989), and as illustrated in Fig. 5, much of the amino acid sequence identity found in the various PEPCase proteins appears as discrete blocks suggestive of functional domains. For example, the sequence ⁵⁷⁸FHGRGGSIGRGGAP⁵⁹¹ (*E. coli* coordinates, Fujita *et al.*, 1984) has been proposed as a PEP-binding region as it is a unique, flexible and basic charged motif within the PEPCase sequence. As has been found in all other PEPCase proteins, it is highly conserved in the *Anabaena* sequence (⁶⁶⁰FHGRGGSvGRGGp⁶⁷³, Fig. 5). Most recently, site-specific mutagenesis of the *E. coli* enzyme has shown that although not obligatory for catalysis, H⁵⁷⁹ has a role in both catalytic and regulatory functions of the bacterial protein (Terada *et al.*, 1991). The *Anabaena* PEPCase also contains a highly conserved sequence ⁶¹⁷QEVMLGYSDSNKDSG⁶³¹ (Fig. 5) which contains a number of invariant residues including K⁶²⁸. Inactivation of this lysine, by the addition of pyridoxal 5'-phosphate and reduction by sodium borohydride, was correlated with the loss of activity of the maize enzyme and it has been suggested that this lysyl residue is a component of the active site (Jiao *et al.*, 1990).

Conservation of sequence, however, may not always indicate the existence of a functional domain involved in the binding of substrates or allosteric effectors. In the *E. coli* PEPCase, the substitution of threonine¹⁸⁸ to

isoleucine results in the elimination of fructose-1,6-bisphosphate activation of PEPCase activity (Sutton *et al.*, 1986). It is interesting to note that this residue (T²³²) as well as some flanking sequence is also present in both the *Anabaena* and the *Synechococcus* sp. sequence; however, it has been shown that cyanobacterial PEPCase is not activated by fructose bisphosphate (Owtttrim & Colman, 1986). These data would imply that the threonine residue itself is not necessarily involved in activator binding but that the replacement of the threonine residue by isoleucine has resulted in a conformational change in a regulatory site of the *E. coli* enzyme. In the future, the identification and characterization of regulatory sites will be assisted by the recent success in crystallization of PEPCase (Inoue *et al.*, 1989). A comparison of the X-ray crystallographic structure of the essentially non-allosteric cyanobacterial enzyme with the highly regulated bacterial or higher-plant enzymes would certainly be very useful.

The authors gratefully acknowledge Ms K. Leonhardt for the gift of the *Anabaena* sp. PCC 7120 genomic DNA library and the excellent technical assistance provided by Ms Thanh Nguyen. This research was funded by a grant to J. R. C. from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

- ALLEN, M. M. (1968). Simple conditions for growth of unicellular blue-green algae on plates. *Journal of Phycology* **4**, 1–4.
- BORTHAKUR, D. & HASELKORN, R. (1989). Nucleotide sequence of the gene encoding the 33 kDa water oxidizing polypeptide in *Anabaena* sp. strain PCC 7120 and its expression in *Escherichia coli*. *Plant Molecular Biology* **13**, 427–439.
- BRADFORD, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248–254.
- COLEMAN, J. R. (1991). The molecular and biochemical analyses of CO₂-concentrating mechanisms in cyanobacteria and microalgae. *Plant, Cell and Environment* **14**, 861–867.
- COLEMAN, J. R. & COLMAN, B. (1980). Demonstration of C₃ photosynthesis in the blue-green alga, *Coccochloris penicostis*. *Planta* **149**, 318–320.
- COLEMAN, J. R. & COLMAN, B. (1981). Photosynthetic carbon assimilation in the blue-green alga, *Coccochloris penicostis*. *Plant, Cell and Environment* **4**, 285–290.
- COLMAN, B., CHENG, K. H. & INGLE, R. K. (1976). The relative activities of PEP carboxylase and RuDP carboxylase in blue-green algae. *Plant Science Letters* **6**, 123–127.
- CORPET, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research* **16**, 10881–10890.
- CUSHMAN, J. C., MEYER, G., MICHALOWSKI, C. B., SCHMITT, J. M. & BOHNERT, H. J. (1989). Salt stress leads to differential expression of two isogenes of phosphoenolpyruvate carboxylase during crassulacean acid metabolism induction in the common ice plant. *The Plant Cell* **1**, 715–725.
- DOHLER, G. (1974). C-Weg der Photosynthese in der Blaualga *Anacystis nidulans*. *Planta* **118**, 259–269.
- FUJITA, N., MIWA, T., ISHIJIMA, S., IZUI, K. & KATSUKI, H. (1984). The primary structure of phosphoenolpyruvate carboxylase of *Escherichia coli*. Nucleotide sequence of the *ppc* gene and deduced amino acid sequence. *Journal of Biochemistry* **95**, 909–916.

- GENDEL, S., STRAUS, N., PULLEYBLANK, D. & WILLIAMS, J. (1983). Shuttle cloning vectors for the cyanobacterium, *Anacystis nidulans*. *Journal of Bacteriology* **156**, 148–154.
- GLANDSORFF, N. (1965). Topography of cotransducible arginine mutations in *Escherichia coli* K-12. *Genetics* **51**, 167–179.
- GOLDEN, S. S., BRUSSLAN, J. & HASELKORN, R. (1987). Genetic engineering of the cyanobacterial chromosome. *Methods in Enzymology* **153**, 215–231.
- INOUE, M., HAYASHI, M., SUGIMOTO, M., HARADA, S., KAI, Y., KASAI, N., TERADA, K. & IZUI, K. (1989). First crystallization of a phosphoenolpyruvate carboxylase from *Escherichia coli*. *Journal of Molecular Biology* **208**, 509–510.
- ISHIJIMA, S., KATAGIRI, F., KODAKI, T., IZUI, K., KATSUKI, H., NISHIKAWA, K., NAKASHIMA, H. & OOI, T. (1985). Comparison of amino acid sequences between phosphoenolpyruvate carboxylases from *Escherichia coli* (allosteric) and *Anacystis nidulans* (non-allosteric): identification of conserved and variable regions. *Biochemical and Biophysical Research Communications* **133**, 436–441.
- IZUI, K., ISHIJIMA, S., YAMAGUCHI, Y., KATAGIRI, F., MURATA, T., SHIGESADA, K., SUGIYAMA, T. & HATSUKI, H. (1986). Cloning and sequence analysis of cDNA encoding active phosphoenolpyruvate carboxylase of the C₄-pathway from maize. *Nucleic Acids Research* **14**, 1615–1628.
- JIAO, J., PODESTA, F. E., CHOLLET, R., O'LEARY, M. H. & ANDREO, C. S. (1990). Isolation and sequence of an active-site peptide from maize leaf phosphoenolpyruvate carboxylase inactivated by pyridoxal 5'-phosphate. *Biochimica et Biophysica Acta* **1041**, 291–295.
- KATAGIRI, F., KODAKI, T., FUJITA, N., IZUI, K. & KATSUKI, H. (1985). Nucleotide sequence of the phosphoenolpyruvate carboxylase gene of the cyanobacterium *Anacystis nidulans*. *Gene* **38**, 265–269.
- KJEMS, J. & GARRETT, R. A. (1988). Novel splicing mechanism for the ribosomal RNA intron in the archaeobacterium *Desulfurococcus mobilis*. *Cell* **54**, 693–703.
- KODAKI, T., KATAGIRI, F., ASANO, M., IZUI, K. & KATSUKI, H. (1985). Cloning of phosphoenolpyruvate carboxylase gene from a cyanobacterium, *Anacystis nidulans*, in *Escherichia coli*. *Journal of Biochemistry* **97**, 533–539.
- LEM, N. W., PENROSE, D. M. & GLICK, B. (1986). Partial purification and characterization of phosphoenolpyruvate carboxylase from the cyanobacterium *Anabaena variabilis*. *Biochemistry and Cell Biology* **64**, 427–433.
- LUINENBURG, I. & COLEMAN, J. R. (1990). A requirement for phosphoenolpyruvate carboxylase in the cyanobacterium *Synechococcus* PCC 7942. *Archives of Microbiology* **154**, 471–474.
- MARUYAMA, H., EASTERDAY, R. L., CHANG, H. C. & LANE, M. D. (1966). The enzymatic carboxylation of phosphoenolpyruvate carboxylase. I. Purification and properties of phosphoenolpyruvate carboxylase. *Journal of Biological Chemistry* **241**, 2405–2412.
- O'LEARY, M. H. (1982). Phosphoenolpyruvate carboxylase: an enzymologist's view. *Annual Review of Plant Physiology* **33**, 297–315.
- OWTTRIM, G. W. & COLEMAN, J. R. (1989). Regulation of expression and nucleotide sequence of the *Anabaena variabilis* recA gene. *Journal of Bacteriology* **171**, 5713–5719.
- OWTTRIM, G. W. & COLMAN, B. (1986). Purification and characterization of phosphoenolpyruvate carboxylase from a cyanobacterium. *Journal of Bacteriology* **168**, 207–212.
- OWTTRIM, G. W. & COLMAN, B. (1988). Phosphoenolpyruvate carboxylase mediated carbon flow in a cyanobacterium. *Biochemistry and Cell Biology* **66**, 93–99.
- RICKERS, J., CUSHMAN, J. C., MICHALOWSKI, C. B., SCHMITT, J. M. & BOHNERT, H. J. (1989). Expression of the CAM-form of phosphoenolpyruvate carboxylase and nucleotide sequence of a full length cDNA from *Mesembryanthemum crystallinum*. *Molecular and General Genetics* **215**, 447–454.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SCHNEIDER, G. J., LANG, J. D. & HASELKORN, R. (1991). Promoter recognition by the RNA polymerase from the vegetative cells of the cyanobacterium *Anabaena* 7120. *Gene* **105**, 51–60.
- SMITH, A. J., LONDON, J. & STANIER, R. Y. (1967). Biochemical basis of obligate autotrophy in blue-green algae and thiobacilli. *Journal of Bacteriology* **94**, 972–983.
- SUTTON, F., BUTLER, E. T., III & SMITH, T. E. (1986). Isolation of the structural gene encoding a mutant form of *Escherichia coli* phosphoenolpyruvate carboxylase deficient in regulation by fructose 1,6-bisphosphate. Identification of an amino acid substitution in the mutant. *Journal of Biological Chemistry* **261**, 16078–16081.
- TANDEAU DE MARSAC, N. & HOUMARD, J. (1987). Advances in cyanobacterial molecular genetics. In *The Cyanobacteria*, pp. 251–302. Edited by P. Fay & C. van Baalen. Amsterdam: Elsevier.
- TERADA, K., MURATA, T. & IZUI, K. (1991). Site directed mutagenesis of phosphoenolpyruvate carboxylase from *E. coli*: The role of His⁵⁷⁹ in the catalytic and regulatory functions. *Journal of Biochemistry* **109**, 49–54.
- TUMER, N. E., ROBINSON, S. J. & HASELKORN, R. (1983). Different promoters for the *Anabaena* glutamine synthetase gene during growth using molecular or fixed nitrogen. *Nature, London* **306**, 337–342.
- WEATHERS, P. J. & ALLEN, M. M. (1978). Variations in short-term products of inorganic carbon fixation in exponential and stationary phase cultures of *Aphanocapsa* 6308. *Archives of Microbiology* **116**, 231–234.