American Journal of

Biochemistry and Molecular Biology

ISSN 2150-4210

Academic
Journals Inc.

www.academicjournals.com

American Journal of Biochemistry and Molecular Biology 1 (4): 319-336, 2011
ISSN 2150-4210 / DOI: 10.3923/ajbmb.2011.319.336

In silico Characterization and Homology Modeling of Cyanobacterial Phosphoenolpyruvate Carboxylase Enzymes with Computational Tools and Bioinformatics Servers

Aubrey A. Smith and Manuela C. Plazas
Department of Biology, Montgomery College, Rockville Campus, 51 Mannakee Street, Rockville, MD 20850, USA

Corresponding Author: Aubrey A. Smith, Department of Biology, Montgomery College-Rockville Campus, 51 Mannakee Street, Rockville, MD 20850, USA Tel: 240-567-5126 Fax: 240-567-7910

ABSTRACT

Phosphoenolpyruvate carboxylase (PEPC; EC4.1.1.31) catalyzes the irreversible β-carboxylation of Phosphoenolpyruvate (PEP) to yield Oxaloacetate (OAA) and inorganic Phosphate (Pi). PEPC contributes to photosynthetic and anaplerotic CO_2 fixation in higher plants and bacteria. The aim of this study was to determine the physicochemical properties of cyanobacterial PEPCs and to develop 3-dimensional models of selected enzymes. The biocomputational analyses were performed *in silico* using web-based software and servers. The alignment of cyanobacterial enzymes and secondary structure analysis revealed that there are conserved amino acid substitutions and polymorphisms between PEPCs from marine and fresh water organisms. Furthermore, some marine subgroups seem to possess unique amino acid stretches that may modulate various aspects of catalysis and regulation. Phosphoenolpyruvate carboxylase from *Synechococcus* PCC 7002, a marine organism, most closely resembles PEPC from fresh water organisms; for this reason, the enzyme was chosen for homology modeling alongside PEPCs from the fresh water strain *Anabaena variabilis* and the marine cyanobacterium *Synechococcus* RS 9917. Amino acids and domains that can distinguish between the fresh water and marine PEPCs were identified. Secondary structure analysis and homology modeling suggested that cyanobacterial PEPCs are primarily alpha helical with additional β-sheets flanking the characteristic central β-barrel. The physicochemical characteristics and the 3D models provide a framework for the purification and characterization of cyanobacterial PEPCs.

Key words: Phosphoenolpyruvate carboxylase, synechococcus, prochlorococcus, computational analysis, proteomic tools

INTRODUCTION

Photosynthetic carbon fixation is the major means of carbon dioxide assimilation in marine ecosystems. These communities consist of photosynthetic prokaryotes and eukaryotes, namely phytoplanktons, chromophytes and cyanobacteria. These organisms use the Calvin-Benson-Bassham (CBB) pathway to metabolize CO_2 (Caldwell *et al.*, 2007). The key enzyme of the CBB pathway is ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO). RubisCO-based CO_2 fixation occurs in all cyanobacteria through C_3 intermediates. A significant amount of carbon flows into C_4 acids during photosynthesis. These compounds which normally account for less than 5% of

CO_2 fixation rate in the dark, are found in significant quantities in cyanobacterial photosynthetic carbon fixation (Leegood, 2002). This indicates that a C_4 mechanism for inorganic carbon fixation is prevalent in cyanobacteria. Formation of the C_4 acid Oxaloacetate (OAA) occurs through the **β**-carboxylation of Phosphoenolpyruvate (PEP). OAA synthesis is catalyzed by three enzymes in plants and bacteria including PEP-carboxylase (PEPC) which is widespread in all higher plants and various types of bacteria (O'Leary *et al.*, 2009). Cyanobacteria are prokaryotic organisms that possess the ability to perform oxygenic photosynthesis. Many are able to fix nitrogen and they are often used as fertilizers in (Soltani *et al.*, 2007). Cyanobacteria are widely distributed and their habitats include oceans, soils, freshwater and extreme environments (Tomitani *et al.*, 2006). Extremophilic strains have been found in various ecosystems (Karthikeyan and Gopalaswamy, 2009).

Cyanobacteria in general have the unique characteristic of possessing an incomplete Tricarboxylic Acid (TCA) cycle (Tripp *et al.*, 2010). These organisms cannot derive metabolic energy from the Krebbs cycle because they lack α-ketoglutarate dehydrogenase and NADH oxidase. Because of this, PEPC has been assigned an anaplerotic role in cyanobacteria aside from its role as a key carbon fixation enzyme. Furthermore the C_4 compounds used to replenish the TCA cycle may also be required for the production of nitrogen storage molecules; such compounds may also be involved in organic stress response in plants and cyanobacteria (Khidir Ahmed, 2009). Gene expression and molecular characterization of PEPCs from C3 plants has not been extensively reported (Hammami *et al.*, 2004). Several PEPCs have been purified from plants and bacteria. The best-described bacterial PEPC is that found in *Escherichia coli*; its three-dimensional structure has been determined at 2.8 A resolution (Kai *et al.*, 1999) and its physical and chemical properties have been extensively investigated (Kwon *et al.*, 2008).

Despite their importance in net carbon dioxide fixation cyanobacterial PEPC enzymes have not been highly studied. The *pepc* gene was initially cloned in the fresh water cyanobacterial strains *Anacystis nidulans*, *Anabeana variabilis* and *Synechocystis* PCC6301 (Chen *et al.*, 2004; Izui *et al.*, 2004; Knowles and Plaxton, 2003). While The *E. coli* enzyme displays cooperativity and is regulated by Fructose-1,6-bisphosphate (FBP) and Acetyl Coenzyme A (AcCoA), PEPC from the cyanobacterium *Coccochloris peniocystis* PEPC does not possess cooperativity and is inhibited by aspartate and TCA cycle intermediates (Chen *et al.*, 2002). *C. peniocystis* PEPC resembles enzymes isolated from C_3 plants such as banana fruit PEPC. The latter is activated by hexose-monophosphates and malate, succinate, aspartate and glutamate activate it. The effect of those inhibitors varies with pH. PEPC from the cyanobacterium *Anacystis nidulans* is also a non-allosteric enzyme. The *pepc* gene from the oceanic cyanocaterium *Synechcoccus* PCC 7002 was also isolated and expressed (Smith *et al.*, 2008).

The availability of internet based tools and a server provides an excellent opportunity to characterize the physicochemical properties of cyanobacterial PEPCs as well as their primary, secondary and three-dimensional structural properties. The purpose of this study was primarily to report the *in silico* analysis and characterization of cyanobacterial PEPCs.

MATERIALS AND METHODS

PEPC protein sequences: Phosphoenolpyruvate Carboxylase (PEPC) protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) [www.ncbi.nlm.nih.gov]. The search yielded 146 bacterial sequences from which 27 cyanobacterial sequences were selected

Table 1: Phosphoenolpyruvate carboxylase sequences retrieved from the NCBI database (www.ncbi.nlm.nih.gov)

Organism	Accession number	Source	Table 2 label.
Synechococcus sp. *CC9902*	YP-377934	Marine	19
Synechococcus sp. *CC9605*	YP-380727	Marine	18
Synechococcus sp. *CC9311*	YP-729688	Marine	15
Synechococcus sp. *WH 7805*	ZP-01123374	Marine	24
Synechococcus sp. *RS9916*	ZP-01473048	Marine	14
Synechococcus sp. *RS9917*	EAQ68521	Marine	25
Synechococcus sp. *WH 8102*	CAE08562	Marine	21
Synechococcus sp. *WH 5701*	ZP-01083776	Marine	23
Prochlorococcus marinus str. MIT 9303	ABM79009	Marine	10
Prochlorococcus marinus str. MIT 9313	NP-895540	Marine	22
Prochlorococcus marinus str. MIT 9301	ABO18389	Marine	8
Prochlorococcus marinus str. AS9601	ABM71065	Marine	11
Prochlorococcus marinus str. MIT 9312	YP-398164	Marine	17
Prochlorococcus marinus str. MIT 9515	ABM72969	Marine	12
Prochlorococcus marinus subsp. *pastoris str. CCMP1986*	NP-893692	Marine	13
Prochlorococcus marinus subsp. *marinus str. CCMP1375*	NP-876121	Marine	16
Prochlorococcus marinus str. MIT 9211	ABX09626	Marine	26
Prochlorococcus marinus str. NATL1A	ABM76577	Marine	9
Prochlorococcus marinus str. NATL2A	YP-292340	Marine	20
Synechococcus sp. *PCC 7002*	AAV90837	Marine	27
Synechococcus elongatus PCC 7942	YP-401269	Fresh water	2
Anabaena variabilis ATCC 29413	YP-322649	Fresh water	3
Synechococcus elongatus PCC 6301(A. nidulans)	AAA22052	Fresh water	1
Synechocystis sp. *PCC 6803*	BAA18393	Fresh water	4
Cyanothece sp. *CCY0110*	ZP-01729388	Fresh water	6
Cyanothece sp. *PCC 7424*	ACK72608	Fresh water	5
Cyanothece sp. *PCC 8801*	ACK67434	Fresh water	7

(Table 1). *E. coli* and *Zea mays* sequence were also obtained from the NCBI database. The sequences were converted to FASTA format prior to analysis using the ReadSeQ sequence conversion server (Gilbert, 2003). The Database and sequence converters were last accessed in December 2010 from a computer terminal at Montgomery College, USA.

Sequence analysis

Sequence alignments: Multiple sequence alignments were performed with ClustalW (Larkin *et al.*, 2007). The ClustalW alignment file was used to generate a shaded output with the BoxShade server (Fig. 2). Identical and similar amino acids were shaded black and grey respectively. Phylogenetic analysis of protein sequences was generated using the alignment obtained with ClustalW (Fig. 1). The ClustalW and Boxshade servers were last accessed in December 2010 from a computer terminal at Montgomery College, USA.

Structural analysis: The amino acid composition of the cyanobacterial PEPC sequences was computed using the PEPSTATS analysis tool (Rice *et al.*, 2000). The physico-chemical parameters such as the molecular weight, isoelectric point (pI), extinction coefficient, half-life, aliphatic index, amino acid property, instability index and Grand Average Hydropathy (GRAVY) were calculated

using the ProtParam tool of the Expasy proteomics server (Table 3). Secondary structure elements prediction was performed using the Network Protein Sequence Analysis (NPS@) server and the Secondary Structural Content Prediction (SSCP) server (Combet *et al.*, 2000; Eisenhaber *et al.*, 1996). The consensus secondary structure content and predicted disulfide patterns of each cyanobacterial PEPC are tabulated in Table 4. The presence of disulfide bridges was analyzed using the CYS-REC tool which predicts the most probable bonding patterns between available cysteine residues. The 3-D models of cyanobacterial PEPCs were constructed using the protein structure homology model building program SWISS-MODEL with energy minimization parameters (Arnold *et al.*, 2006). The modeled tertiary structures were built on the basis of sequence homology with the high-resolution crystal structures of the *E. coli* and *Zea mays* enzymes. The Swiss PDB viewer (Guex and Peitsch, 1997) was used to visualize and refine the models and the standalone version of PyMOL was used to generate publishable images of the PEPC models. The modeled 3D structures were evaluated and validated with the WHAT IF and RAMPAGE programs (Lovell *et al.*, 2003; Vriend, 1990). The aforementioned bioinformatics and modeling servers were accessed in between October 2009 and October 2010 from a computer terminal at Montgomery College, USA.

RESULTS AND DISCUSSION

The primary structure analysis of cyanobacterial phosphoenolpyruvate carboxylases indicates that the percentage of cysteine residues is less than 1% across species. The most abundant amino acid is leucine which makes up approximately 14% of the primary structures (Table 2). These trends

Table 2: Amino acid composition of cyanobacterial phosphoenolpyruvate carboxylases in percentage. The data was computed using the PEPSTATS analysis tool. The organisms associated with the sequence numbers are given in Table 1

	Sequence number from table 1																										
Amino acid	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
A	6.3	5.4	5.4	6.5	6.1	6.3	5.8	4.4	6.0	8.3	4.6	4.8	4.4	8.6	8.1	5.6	5.0	7.7	8.1	6.0	7.4	8.4	8.8	7.7	8.7	6.4	6.7
R	8.0	6.8	6.8	7.7	6.6	7.1	7.4	6.8	7.4	8.2	6.8	7.2	6.9	8.8	8.5	7.9	6.6	8.4	8.7	7.5	8.7	8.6	8.6	8.9	9.2	7.5	7.5
N	2.9	4.0	3.9	3.9	4.3	4.3	4.3	5.2	3.1	3.4	5.1	5.1	5.2	2.0	2.3	3.3	5.0	2.6	2.2	3.1	2.2	3.4	2.1	2.4	2.0	3.7	3.9
D	4.3	5.3	5.0	4.1	4.0	3.9	4.7	5.9	5.6	4.7	5.6	4.7	4.9	5.3	5.4	4.8	5.4	5.0	5.6	5.6	4.8	4.7	5.3	5.5	5.5	4.4	4.9
C	0.9	0.7	0.8	0.7	1.4	1.3	0.7	0.5	0.6	0.6	0.5	0.6	0.5	0.5	0.5	0.9	0.5	0.8	0.5	0.6	0.5	0.6	0.9	0.8	0.8	0.7	0.9
Q	7.5	6.7	6.5	5.6	7.1	6.5	6.5	5.3	5.9	6.7	5.2	4.9	5.2	6.3	6.2	5.9	5.3	6.8	6.6	5.8	6.9	6.7	6.6	6.5	6.7	6.2	6.8
E	8.6	7.9	8.3	9.4	8.8	9.3	8.8	7.9	7.2	7.9	8.2	8.9	8.8	7.8	7.2	9.2	8.2	7.8	7.2	7.2	7.9	7.8	7.8	7.5	7.7	8.6	7.6
G	4.8	5.3	5.2	5.2	4.8	4.5	4.6	3.5	3.9	4.3	3.5	3.4	3.5	4.5	5.1	3.8	3.5	4.8	4.7	3.9	5	4.2	4.7	4.7	4.5	4.0	4.6
H	2.3	1.8	2.0	2.1	1.8	1.6	2.1	1.4	1.9	1.9	1.5	1.6	1.6	2.4	2.2	2.1	1.4	2.4	2.2	1.9	2.5	2.0	2.1	2.1	2.1	2.1	2.2
I	4.8	5.8	5.7	5.4	6.2	5.9	5.8	5.7	4.9	3.9	5.3	5.7	5.6	3.7	4.0	5.2	5.3	3.8	3.9	4.9	4.1	3.9	4.1	3.7	3.5	5.1	6.2
L	13.5	12.8	13.1	13.1	13.2	13.1	13.1	13.4	14	14.4	13.8	13.5	13.4	14.2	14.1	14.1	13.7	14.2	14.3	14	13.7	14	14.5	14.3	14.3	13.4	12.6
K	2.5	4.3	4.3	3.4	4.7	4.9	4	6.1	4.0	2.6	6.2	6.1	6.1	1.7	1.8	3.1	6.4	2.2	2.3	4.0	1.8	2.5	1.7	1.5	1.4	3.4	3.6
M	1.5	1.4	1.5	2.0	1.5	1.3	1.8	1.6	2.1	2.1	1.6	1.9	1.8	2.0	1.9	2.3	1.6	2.2	2.0	2.1	2.2	2.3	1.8	1.5	2.1	2.0	1.6
F	4.1	4.0	3.9	3.6	4.0	4.1	3.8	4.1	3.3	3.1	4.1	3.9	4.0	3.1	3.4	3.5	4.1	3.3	3.4	3.3	3.4	3.2	3.2	3.1	3.1	4.2	3.9
P	5.4	4.6	4.4	4.4	4.5	4.1	4.2	3.8	4.7	5.3	3.7	3.8	3.8	5.3	5.9	4.1	3.7	5.4	5.4	4.7	5.9	5.3	6.2	5.2	5.3	4.5	4.5
S	6.0	6.9	7.4	6.2	5.7	6.1	7.5	9.3	9.9	8.0	9.3	8.9	9.2	9.0	8.3	9.1	8.9	7.9	7.7	9.7	8.6	7.8	6.9	8.6	8.4	9.8	7.1
T	6.1	5.4	5.0	6.3	5.2	5.4	5.0	5.0	5.5	5.2	5.1	5.1	5.1	5.2	5.7	5.4	5.2	5.2	5.6	5.6	5.1	5.2	5.2	5.7	4.8	4.5	5.2
W	1.2	1.5	1.6	1.7	1.7	1.6	1.8	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.6	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Y	3.4	3.5	3.5	3.5	3.5	3.5	3.4	2.9	2.5	2.6	2.9	2.8	2.8	2.4	2.4	2.6	3.0	2.2	2.2	2.5	2.3	2.6	2.5	2.4	2.6	2.4	3.0
V	5.8	5.8	5.9	5.2	5.0	5.3	4.9	5.7	5.6	5.5	5.7	5.7	5.7	6.0	5.6	5.6	5.9	5.5	5.5	5.6	5.5	5.5	5.7	6.3	5.7	6.0	5.3

Table 3: Parameters of cyanobacterial PEPCs calculated using the ProtParam program

Accession No.	Sequence length	MW (x10E5)	pI	EC (M^{-1} cm^{-1})	Ii	Ai	GRAVY	R-	R+
YP377934	995	1.13	5.76	113510	57.18	95.12	-0.32	128	110
YP380727	997	1.13	5.66	113510	57.89	94.13	-0.331	128	106
YP729688	1001	1.13	5.55	116070	54.16	94.84	-0.287	126	103
ZP01123374	991	1.12	5.42	116070	57.43	96.26	-0.299	129	103
ZP01473048	1003	1.13	5.51	116070	55.62	95.52	-0.293	131	105
EAQ68521	979	1.11	5.45	117350	56.67	94.59	-0.314	129	104
CAE08562	1010	1.15	5.69	114790	65.17	92.62	-0.35	128	106
ZP01083776	1002	1.13	5.36	117350	54.94	97.67	-0.259	131	103
ABM79009	1002	1.14	5.68	118630	54.17	95.43	-0.33	126	108
NP895540	1004	1.14	5.91	118630	53.37	94.56	-0.34	125	111
ABO18389	989	1.14	6.04	122470	47.29	95.4	-0.426	136	127
ABM71065	989	1.14	6.14	122470	47.03	95.11	-0.43	136	128
YP398164	989	1.14	6.28	127450	45.83	97.5	-0.405	134	128
ABM72969	989	1.14	6.66	121190	47.17	96.1	-0.41	134	131
NP893692	989	1.14	6.26	121190	49.51	95.01	-0.432	135	128
NP876121	1001	1.15	5.34	123750	54.96	97.01	-0.325	140	110
ABX09626	1005	1.15	5.63	116070	56.17	95.86	-0.297	130	109
ABM76577	994	1.13	5.88	117350	52.03	96.14	-0.329	128	114
YP292340	994	1.13	5.93	117350	51.39	96.53	-0.326	128	115
AAV90837	995	1.14	5.96	123750	51.03	95.48	-0.339	125	111
YP401269	982	1.13	5.53	125640	49.12	94.91	-0.371	130	109
YP322649	982	1.17	5.56	137120	49.78	95.58	-0.359	136	114
AAA22052	1053	1.21	5.53	120050	54.08	94.55	-0.344	136	110
BAA18393	1034	1.19	5.57	148500	50.55	93.67	-0.365	139	115
ZP01729388	1016	1.17	5.42	135025	50.71	96.09	-0.381	143	120
ACK72608	1021	1.18	5.77	142810	50.02	96.19	-0.345	131	115
ACK67434	1024	1.18	5.64	147220	54	93.43	-0.425	138	117

Mw: Molecular weight (g mol^{-1}), PI: isoelectric point, EC: Extinction coefficient (M^{-1} cm^{-1}), Ii: Instability index, Ai: Aliphatic index, GRAVY: Grand average hydropathy, -R: No. of negative residues, +R: No. of positive residues

were also observed in the *E. coli* and plant enzymes. Despite the low cysteine content, the disulfide bridge prediction tool CYS-REC computed disulfide bonds in most cyanobacterial PEPCs (Table 4). There is no periodicity among leucine residues, so the presence of specific leucine-rich motifs or domains is unlikely. Leucine residues are found as dimers in cyanobacterial PEPCs; most significantly the sequences LLRGALL and LLEVLL are found in all cyanobacterial PEPCs but they are absent in the *E. coli* and plant enzymes. Furthermore, the sequence LLKRL is found in PEPC from fresh water cyanobacteria and *Synechococcus* PCC 7002 but it is absent in the bacterial and plant enzyme. Another leucine-rich stretch, LLLAK/QE is found in cyanobacterial PEPCs exclusively.

PEPC was slightly larger in fresh water organisms with an average molecular weight of approximately 118000 g mol^{-1}. Conversely, oceanic PEPCs had an average molecular weight of 114000 g mol^{-1}. The computed isoelectric point (pI) was below 7 indicating that the proteins will most likely precipitate in acidic buffers. The average extinction coefficient was 1.2 for fresh water enzymes and 1.0 for marine strains, the difference may be due to the higher percentage of tyrosine residues in fresh water enzymes. According to the ProtParam server, a protein whose Instability

Table 4: Predicted consensus secondary structure content and predicted disulfide patterns of cyanobacterial PEPCs

Accession No.	α-helix	β-sheet	Coil	Disulfide bridge prediction
YP377934	60.20	5.43	31.36	None
YP380727	60.08	5.22	31.90	None
YP729688	59.44	6.09	26.37	None
ZP01123374	54.99	7.27	34.51	284-443
ZP01473048	60.92	5.88	23.83	295-555
EAQ68521	61.49	4.80	30.95	106-296, 270-531
CAE08562	58.51	5.25	26.83	None
ZP01083776	63.27	4.29	25.15	287-404, 313-452
ABM79009	63.97	4.69	24.25	324-457
NP895540	64.84	4.48	23.21	23-300, 326-459
ABO18389	57.84	5.36	33.67	281-439
ABM71065	58.95	4.75	32.76	281-439
YP398164	59.86	5.36	30.74	281-439
ABM72969	62.49	4.55	29.93	57-281, 307-439
NP893692	60.06	4.65	32.05	307-439
NP876121	59.44	8.39	23.38	None
ABX09626	59.10	5.77	28.06	52-455, 55-296, 322-432
ABM76577	57.24	5.84	34.10	17-310, 284-443
YP-292340	57.14	6.14	33.60	17-310, 284-443
AAV90837	61.11	5.53	30.15	41-206, 436-618
YP401269	56.01	6.31	34.73	267-293, 413-515
YP322649	54.94	8.11	28.74	41-337, 457-559,
AAA22052	55.37	9.31	28.40	23-40, 357-518, 383-510
BAA18393	58.41	5.32	28.53	307-457, 538-936,
ZP01729388	56.86	6.86	27.84	None
ACK72608	55.24	8.42	27.82	56-768, 57-307, 361-485, 457-465
ACK67434	55.18	6.74	29.49	307-333

The data was generated from the protein sequence analysis server and CYS REC (Stultz *et al*., 1993; White *et al*., 1994)

Index (Ii) is larger than 40 may be unstable therefore, the program predicted that all cyanobacterial PEPCs would be unstable in solution. The aliphatic index which gives a measure of the relative volume occupied by alanine, valine, isoleucine and leucine, ranged from 92.62 to 97.67 for all cyanobacterial PEPCs. The relatively high Ai values indicate that cyanobacterial PEPCs maybe stable over a large range of temperatures. The Grand Average Hydropathy (GRAVY) values for PEPCs ranged from -0.287 to -0.432 indicating that the proteins will interact favorably with water (Table 3).

A consensus method was chosen in order to determine the secondary structure elements in each PEPC protein. The secondary structure consensus prediction program of the Protein Sequence Analysis server generated a secondary consensus where the most present predicted conformational state is reported for each amino acid. The methods used to generate the consensus were: The Double Prediction Method (DPM) (Deleage and Roux, 1987), Discrimination of protein Secondary structure Class (DSC) (King and Sternberg, 1996), Garnier Osguthorpe and Robson (GOR1) (Garnier *et al*., 1978), GOR3 (Biou *et al*., 1988), Hierarchical Neural Network (HNN) (Geourjon *et al*., 1999), PHDsec (Rost and Sander, 1993), the PredictProtein server (Rost *et al*.,

Table 5: Percent amino acid identity between *Synechococcus* PCC 7002 phosphoenolpyruvate carboxylase and other cyanobacterial PEPC sequences

Organism 1	Organism 2	Percent homology
A.nidulans	Syn.PCC7002	62
Nostoc	Syn.PCC7002	61
A.variabilis	Syn.PCC7002	62
Syn.PCC6803	Syn.PCC7002	62
Syn.PCC7424	Syn.PCC7002	66
Cyanothece	Syn.PCC7002	66
Syn.PCC8801	Syn.PCC7002	66
PmMIT9301	Syn.PCC7002	52
PmNATL1A	Syn.PCC7002	53
PmMIT9303	Syn.PCC7002	53
PmAS9601	Syn.PCC7002	52
PmMIT9515	Syn.PCC7002	51
PmCCMP1986	Syn.PCC7002	51
Syn.RS9916	Syn.PCC7002	54
Syn.CC9311	Syn.PCC7002	52
PmCCMP1375	Syn.PCC7002	52
PmMIT9312	Syn.PCC7002	52
Syn.CC9605	Syn.PCC7002	53
Syn.CC9902	Syn.PCC7002	53
PmNATL2A	Syn.PCC7002	53
Syn.WH8102	Syn.PCC7002	52
PmMIT9313	Syn.PCC7002	53
Syn.WH5701	Syn.PCC7002	53
Syn.WH7805	Syn.PCC7002	53
Syn.RS9917	Syn.PCC7002	54
PmMIT9211	Syn.PCC7002	53
Syn.PCC7002	*E. coli*	32
Syn.PCC7002	*Zea mays* C4	30
Syn.PCC7002	*Zea mays*	29
E. coli	*Zea mays* C4	40
E. coli	*Zea mays*	42
ZMC4	*Zea mays*	81

2004), Predator (Frishman and Argos, 1996) and the Self-Optimized Prediction Method (SOPM) (Geourjon and Deleage, 1994). The results of the consensus in Table 5 indicate that cyanobacterial PEPCs are largely alpha helical with less than 10% beta structures. This is in agreement with the X-ray crystal structure of plant and bacterial PEPCs (Matsumura *et al*., 2002).

A multiple sequence alignment of cyanobacterial, plant and *E. coli* phosphoenolpyruvate carboxylases was performed with ClustalW (Larkin *et al*., 2007). The alignment indicated that *Synechococcus* PCC 7002 PEPC shared the most identity with PEPCs from fresh water organisms and only 30% identity with *E. coli* and *Zea mays* PEPCs (Table 6). A phylogenetic tree of the alignment of cyanobacterial PEPCs was also generated; the cladogram showed that *Synechococcus* PCC 7002 PEPC is grouped with PEPC from fresh water cyanobacteria (Fig. 1). This is interesting since *Synechococcus* PCC 7002 is a marine organism. Furthermore, the alignment of cyanobacterial

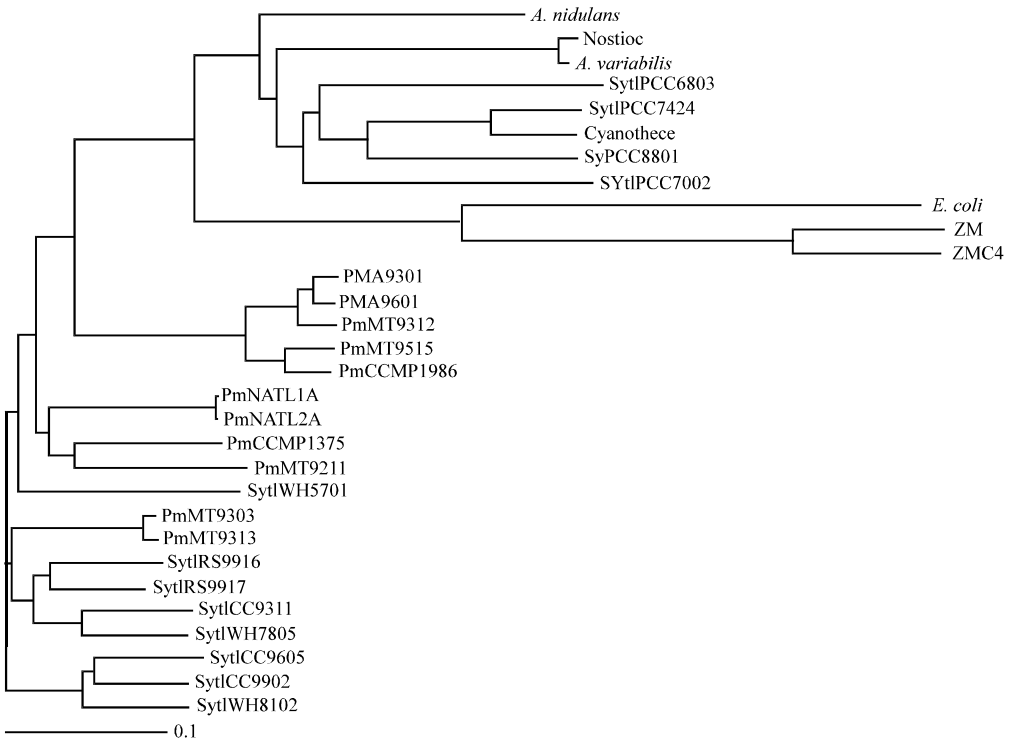

Fig. 1: Cladogram of various cyanobacterial PEPC sequences. *Synechococcus* PCC 7002 PEPC seems to share a higher degree of homology with PEPC from fresh water cyanobacteria than oceanic organisms

enzymes revealed that there are conserved amino acid substitutions between the marine and fresh water strains (Table 6). *Synechococcus* PCC 7002 PEPC shows substitution patterns associated with the enzyme from the fresh water strains. Those single amino acid substitutions appear to distinguish PEPC from marine organisms and those from fresh water organisms (Smith *et al.*, 2008). The multiple sequence alignment also revealed that there are stretches of amino acids that are exclusive to the marine and fresh water enzyme. In addition, stretches of amino acids that are unique to *Prochlorococcus marinus str. MIT 9301, Prochlorococcus marinus str. AS9601, Prochlorococcus marinus str. MIT 9312, Prochlorococcus marinus str. MIT 9515 and Prochlorococcus marinus* subsp. *pastoris str. CCMP1986* have been identified (Fig. 2). In addition to *Synechococcus* PCC 7002 PEPC sharing most sequence homology with PEPC from fresh water organisms, the enzyme's primary structure contains strictly conserved amino acids at positions unique to PEPCs of fresh water cyanobacteria. Those amino acids are listed alongside equally conserved residues found on marine PEPCs (Table 6). Aside from the listed amino acids, there are 20 other residues that share equivalent positions on *Synechococcus* PCC 7002 PEPC and PEPCs from fresh water cyanobacteria. *Synechococcus* PCC 7002 PEPC also shares homologous stretches of amino acids with fresh water enzymes rather than with the marine PEPCs (Fig. 2). The most homology between PEPCs is found at the C-terminal where there are only α-helices.

Table 6: Conserved amino acid substitutions between fresh water and oceanic cyanobacterial PEPC sequences

Amino acid from *Synechococcus* PCC 7002	Amino acid from fresh water strains	Amino acid from oceanic strains
Glycine 42	*Glycine*	*Proline*
Arginine 53	Arginine	Lysine
Leucine 77	*Leucine*	*Methionine*
Alanine 88	Alanine	Serine
Glutamine 159	Glutamine	Glutamate
Aspartate 197	Aspartate	Glutamine
Glutamate 254	Glutamate	Glutamine
Phenylalanine 275	Phenylalanine	Tyrosine
Valine 318	Valine	Methionine
Histidine 338	Histidine	Glutamine
Leucine 364	*Leucine*	*Arginine*
Glutamate 447	Glutamate	Histidine
Isoleucine 462	Isoleucine	Leucine
Phenylalanine 512	Phenylalanine	Tryptophan
Threonine 523	*Threonine*	*Valine*
Isoleucine 542	Isoleucine	Valine
Lysine 648	*Lysine*	*Isoleucine*
Lysine 658	Lysine	Isoleucine
Leucine 740	Leucine	Valine
Isoleucine 753	Isoleucine	Leucine
Tyrosine 770	Tyrosine	Histidine
Serine 818	Serine	Glutamine
Tyrosine 830	Tyrosine	Phenylalanine
Phenylalanine 854	Phenylalanine	Leucine
Lysine 857	Lysine	Arginine
Valine 864	Valine	Leucine
Glutamine 957	Glutamine	Aspartate
Lysine 973	Lysine	Arginine

The amino acid substitutions in italic were not previously reported

The homology models of phosphoenolpyruvate carboxylases from *Anabaena variabilis*, *Synechococcus* RS 9917 and *Synechococcus* PCC 7002 were modeled using the *Zea mays* enzyme as the template. Sequence homology scores of approximately 30% between the template and cyanobacterial sequences were sufficient to generate useful models. The stereochemical and energetic properties of the models were evaluated with the WHAT IF and RAMPAGE servers. All three models were deemed acceptable by the structure validation criteria within the WHAT IF server. According to the Ramachandran plots, 87 to 88% of residues were in the most favored regions, 8.1 to 8.7% of residues were in the allowed regions while 3.7 to 4.0% of residues were in generously allowed regions. These results indicate that the models were geometrically viable. The cyanobacterial PEPCs contain the characteristic **β**-barrel seen in all PEPCs (Fig. 3). Residues that are important in bicarbonate binding as well as those involved in catalysis are found near the C-terminal of the **β**-barrel (Kai *et al*., 1999). The modeled *Synechococcus* RS 9917 PEPC only contains this **β**-barrel as its lone **β** structure. On the other hand, *Anabaena variabilis* and *Synechococcus* PCC 7002 PEPCs contain one and two additional **β**-sheets, respectively (Fig. 3). One of the **β**-sheets of *Synechococcus* PCC 7002 PEPC lies just below the N-terminal side of the **β**-barrel (Fig. 4).

```
Syn_CC9605      10 EQETARMSGGGSG-AGQLLQHRLDLIEDLWKSVLRSECPPEQSERLLRLKQLSDPVSL-E
Syn_CC9902      10 QGEQPRESGVTAG-AGRLLQNRLVLVEDLWQIVLRSECPPEQSARLLRLKQLSDPVAL-E
Syn_WH8102      22 DGEQPRASGGSPG-AGRLLQHRLELVEDLWQIVLRSECPPEQSERLLRLKQLSDPVAL-E
PmMIT9303       20 DCDQPRAIGEGQQ-AGRLLQNRLELVEDLWQIVLRSECPPDQAERLLRLKQLSEPLAL-E
PmMIT9313       22 DCEQPRAIGEGQQ-AGRLLQNRLELVEDLWQIVLRSECPPDQAERLLRLKQLSEPLAL-E
Syn_RS9916      22 ESDRSGIA-EPFQAGGQLLKQRLALVEDLWETVLRSECPPDQAERLLRMKQLSDPVAS-D
Syn_RS9917       6 NGD------------DQLLQQRLELVEELWETVLRSECPPEQAERLLRMKQLSEPLAG-D
Syn_CC9311      20 ESTQPRADGNEAG-GGQLLQQRLALVEDLWRIVLRSECPPEQAEQLLRMKQLSDPVLP-G
Syn_WH7805      10 ESEQPRVAVAEAPVGGQLLQQRLALVEDLWQIVLRSECPAEQAERLLRMKQLSDPVLP-D
PmNATL1A        10 ISNHSTVCVEDQD-PGSLLQQRLELVEDLWKIVLKSECPPDQTERLLRLKQLSDPS----
PmNATL2A        10 ISNHSTVCVEDQD-PGSLLQQRLELVEDLWKIVLKSECPPDQTERLLRLKQLSDPS----
PmCCMP1375      18 DRKESSQILADHM-SGRLLQKRLELVEDLWETVVRSECPLEQVERLLRLKQLSNSSCIVG
PmMIT9211       19 SSFRNNQASTTKN-SGSLLQQRLALVEDLWETVLCSECPSDQINRVLRLKKLSNPNNFAG
Syn_WH5701      13 AQDPQGVPTAPAPRVTRLLGERLELVEDLWQIVLLSECPPDQAERLLRLKRLSDPLEAAD
PmMIT9301        9 NNNVDLIS------NNDPLDKNRLLIEDLWESVLREECPDDQAERLIQLKELSYSKQI-D
PmAS9601         9 NNNVDLIS------NNDPLDKNRLLIEDLWESVLREECPDDQAERLIQLKELSYSKQI-D
PmMIT9312        9 NNKVDLIS------TNDPLDKNRLLIEDLWESVLREECPDDQAERLIQLKELSYSKQI-E
PmMIT9515        9 NHNVDLIS------NNDPLDKNRLLIEDLWESVLREECPNDQANRLMKLKELSYCNQV-E
PmCCMP1986       9 NNNVDLIS------NNDPLDKNRALIEDLWESVLREECPDDQANRLIQLKELSYSNKV-E

Nostoc           1 ----------------------------------------MVDLLRQLRDLCSPECQ-A
A_variabilis     7 SLS---ESANLYPASELFLRHRLQIVEELWESVLRQECGQNMVDLLRQLRDLCSPECQ-A
Syn_PCC7424      7 VST---TPLQFFSTSDLFLQDRLKLVEELWEEVLQAECGQELVALLKKLRAICCEQCQ-A
Cyanothece       7 VPT---TPLQFFSTSDLFLQARLKLVEDLWEAVLKAECGQELVDLLIKLRAICSEQCQ-A
Syn_PCC8801      7 SSSSIVEEINIFSTSDLFLRQRLKLVEELWEAVLRAECGQELVDLLKQLRTWCSPECQ-L
Syn_PCC6803     11 STNWSGNGNGSNSEEESVLYQRLKMVEELWERVLQSECGQELVDLLTELRLQGTHEAI-T
Syn_PCC7002      7 PPS---AEAELLSTSQSLLRQRLTLVEDIWQAVLQKECGQKLVERLNHLRATRTADCQ-S
A_nidulans      61 VSSPLATDLDLSSPLEFFLRHRLTVVEELWEVVLRQECGQELVDILTQLRDLTSPECQ-A

Syn_CC9605      68 GRDGNSTSEAIVELIKAMDLSEAISAARAFSLYFQLINILEQRIEED----------SYL
Syn_CC9902      68 GRDGDSTSEAIVELIRAMDLSEAIAAARAFSLYFQLINILEQRIEED----------SYL
Syn_WH8102      80 GRDGESSSEAIVELIRSMDLSEAIAAARAFSLYFQLINILEQRIEED----------SYL
PmMIT9303       78 GADENSASTAIVLLIKEMDLAEAITAARAFSLYFQLVNILEQRIEED----------SYL
PmMIT9313       80 GADENSASRAIVLLIQEMDLAEAITAARAFSLYFQLVNILEQRIEED----------SYL
Syn_RS9916      80 GA--AVSSDAVAQLIREMDLVEAIAAARAFSLYFQLVNILEQRIEED----------SYL
Syn_RS9917      53 GD--SSSSDAVVQLIREMDLVEAIAAARAFSLYFQLVNILEQRIEED----------SYL
Syn_CC9311      78 EH--PAGTDALIDLIKGMDLAEAIAAARAFSLYFQLVNILEQRIEED----------TYL
Syn_WH7805      69 GN--AVSSDALVSLIRDMDLSEAIAAARAFSLYFQLVNILEQRIEED----------GYL
PmNATL1A        65 KSNQDNSSKAIVQLITKMDLAEAISAARAFSLYFQLVNILEQRIEED----------SYL
PmNATL2A        65 KSNQDNSSKAIVQLITKMDLAEAISAARAFSLYFQLVNILEQRIEED----------SYL
PmCCMP1375      77 EE-QTNQINEIVELIKEMDLAEAISAARAFSLYFQLVNILEQRIEED----------SYL
PmMIT9211       78 EEEQKNPFNEIVLLIEEMDLAESIAAARAFSLYFQLVNILEQRIEED----------TYL
Syn_WH5701      73 GT--VDITAAIVQLIREMDLSEAIAAARAFSLYFQLVNILEQRIEED----------SYL
PmMIT9301       62 GNSSKTFKNEIVDIVNSMDLAESIAAARAFSLYFQLVNILEQRVEED----------RYI
PmAS9601        62 GDSSKTFKNEIVDIVNSMDLAESIAAARAFSLYFQLVNILEQRVEED----------RYI
PmMIT9312       62 GDSSKTFKKEIVDIVNSMDLAESIAAARAFSLYFQLVNILEQRVEED----------RYI
PmMIT9515       62 EDISRTFKNEIVDIVNSMDLAESISAARAFSLYFQLVNILEQRVEED----------RYI
PmCCMP1986      62 EYSPKNFKTEIVDIVNSMDLAESISAARAFSLYFQLVNILEQRVEED----------RVI

Nostoc          19 TKDQAVSAV---KLIEQLNINEAIRAARAFALYFQLINIIEQEYEQRQQLTRYSDLEAET
A_variabilis    63 TKDQAVSAV---KLIEQLNINEAIRAARAFALYFQLINIIEQEYEQRQQLTRYSDLEAET
Syn_PCC7424     63 QKALEDSIT---QLIEQLDLVEAVRTSRAFALYFQLINIVEQHYEQRDQQL------SRR
Cyanothece      63 QKALEDSIT---QLIEKLDLNEAVRTSRAFALYFQLINIVEQHYEQRDQQL------SRR
```

Fig. 2: Continued

```
Syn_PCC8801      66 TDITQTPIT---AVIEQLELNESIRAARAFALYFQLINIVEQHYEQRDQQL------TRR
Syn_PCC6803      70 SEISEEVIMGITQRIEHLELNDAIRAARAFALYFQLINIVEQHYEQNEQQR------NRW
Syn_PCC7002      63 LNFSPSSIS---ELIETLDLEDAIRAARAFALYFQLINSVEQHYEQREQQQ------FRR
A_nidulans      120 PEVGGEALV---QVIETLELSDAIRAARAFALYFQLINIVEQHYEQTQYQLAYE--RSRL

Nostoc          194 EAGEIREKLLEEIRLWWRTDELHQFKPTVLDEVDYALHYFQEVLFDGIPQLYKRFKYALN
A_variabilis    238 EAGEIREKLLEEIRLWWRTDELHQFKPTVLDEVDYALHYFQEVLFDGIPQLYKRFKYALN
Syn_PCC7424     234 EAEQLKEQFKEEIRLWWRTDELHQFKPTVLDEVDYALHYFNEVLFQALPQLALRLKQTLK
Cyanothece      234 EAEEAKEQLKEEIRLWWRTDELHQFKPTVLDEVDYSLHYFNEVLFEAIPQLALRLKRTLK
Syn_PCC8801     234 EAQSAIEQLTEEIRLWWRTDELHQFKPSVLDEVDYALHYFDEVLFEVLPQLSQRLQQSLK
Syn_PCC6803     243 DAKTAIAQLTEEIRFWWRTDELHQFKPTVLDEVDYSLHYFDEVLFDAVPELSKRLGQAIK
Syn_PCC7002     213 EIENIQQQLTEEIRLWWRTDELHQFKPQVLDEVDYALHYFEEVLFDTLPELSVRLQQALK
A_nidulans      284 EAQNIRQQLTEEIRLWWRTDELHQFKPTVLDEVDYALHYFQEVLFEAIPLLYQRFRLALQ

Syn_CC9605      276 RHYPDVQVPQASFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLELYISSVQSLRQQ
Syn_CC9902      276 RHYPDVQVPQASFCTFGSWVGSDRDGNPSVTTDITWRTACYQRQLMLELYISSVQALRNQ
Syn_WH8102      288 RHYPDVQFPQASFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLELYIGSVQSLRNQ
PmMIT9303       285 ESYPDVHIPQAAFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLERYVNAVQKLRDQ
PmMIT9313       287 ESYPDVHIPQAAFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLERYVNAVQKLRDQ
Syn_RS9916      282 SSYPDVRLPPSAFCTFGSWVGSDRDGNPSVTPEITWRTSCYQRQLMLERYIAAVQALRDQ
Syn_RS9917      257 QNYPDVRLPPAAFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLDRYVSAVQDLRDQ
Syn_CC9311      280 ASYPDVRVPSSSFCTFGSWVGSDRDGNPSVTTPEITWRTACYQRQLMLDRYISAVQHLRNQ
Syn_WH7805      271 CSYPDVQLPPSSFCTFGSWVGSDRDGNPSVTTDITWRTACYQRQLMLDRYVSAVHGLRDQ
PmNATL1A        271 SSYPDVEIPNEAFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLDRYIASVQDLRDQ
PmNATL2A        271 SSYPDVEIPNEAFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLDRYIASVQDLRDQ
PmCCMP1375      282 QSYPDIDVPQEAFCTFGSWVGSDRDGNPSVTPEITWRTACYQRKLMLDRYMHSVQELRNQ
PmMIT9211       283 KSYPDVEVPRESFCTFGSWVGSDRDGNPSVTPEITWRAACYQRQLMLERYINSVQDLRDQ
Syn_WH5701      274 GSYPDVEAPRDAFCTFGSWVGSDRDGNPSVTPDITWRTACYQRQLMIERYITAVQGLRDQ
PmMIT9301       268 ENYPDVQLPPESFCNFGSWVGSDRDGNPSVTPDITWRTACYQRQLMLDRYIVATSNLRDQ
PmAS9601        268 DNYPDVQLPSESFCNFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLERYIIATSNLRDQ
PmMIT9312       268 ENYPDVQLPSQSFCNFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLERYITATSHLRDQ
PmMIT9515       268 ENYPDVQMPTESFCTFGSWVGSDRDGNPSVTPEITWRTACYQRKLMLERYISATSNLRDQ
PmCCMP1986      268 ENYPDVQMPTESFCTFGSWVGSDRDGNPSVTPEITWRTACYQRQLMLERYIKATSNLRDQ

Nostoc          254 QTFSWLEPPSKDFCSFGSWVGSDRDGNPSVTPEITWQTACYQRKMVLERYIKSVTQLIEL
A_variabilis    298 QTFSWLEPPSKDFCSFGSWVGSDRDGNPSVTPEITWQTACYQRKMVLERYIKSVKQLIEL
Syn_PCC7424     294 GAFPRLEPPKNNFCYFGSWVGGDRDGNPFVTPEVTWATACYQRNVVIEKYLQTVDELSNI
Cyanothece      294 SAFPRLRPPKNNFCYFGSWVGGDRDGNPFVTPEVTWATACYQRNVVIEKYIERVDELSNI
Syn_PCC8801     294 SSFPWLRPPKNTFCRFGSWVGGDRDGNPFVTPEVTWKTACYQRNMVLKKYLESIRDLTEI
Syn_PCC6803     303 ETFPHLRAPRANFCYFGSWVGGDRDGNPSVTPEVTWQTACYQRGLVLGKYLFSLGELVAI
Syn_PCC7002     273 ASFPTLKVPTTNFCNFGSWVGGDRDGNPSVTPDVTWKTACYQRGLVLERYIASVESLSDV
A_nidulans      344 GTFPDLQPPRYNFCQFGSWVGSDRDGNPSVTSAVTWQTACYQRSLVLDRYITAVEHLRNV

Syn_CC9605      568 LLAKEAGLVDPPNKRASLLVVPLFETVEDLQRAPEVMEGLFKTPLYRELLPVVGQQKQP-
Syn_CC9902      566 LLAKEAGLVDPAARKASLLVVPLFETVEDLQRAPAVMDELFNTPLYRDLLPMVGIQGQP-
Syn_WH8102      581 LLAKEIGLVDPQAGKASLLVVPLFETVEDLQRAPAVMDGLFQTPIYRNLLPSVGVQRQP-
PmMIT9303       577 LLAKEAGLVDPAAGIAELLVVPLFETVEDLQRAPAVMEALLSSPVYRNLLPRVSEQVQP-
PmMIT9313       579 LLAKEAGLVDPAAGIAELLVVPLFETVEDLQRAPAVMEALLSSPVYRNLLPRVSEQVQP-
Syn_RS9916      574 LLAKEAGLVDPAAEIADLLVVPLFETVEDLQRAPAVMEQLFENPLYRKLLPQVGEHVQP-
Syn_RS9917      550 LLAKEAGLVDPAAHIADLLVVPLFETVEDLQRAPEVMGQLFEHPLYRQLLPRAGEQAQP-
Syn_CC9311      572 LLAKEAGLVDPSARIADLLVVPLFETVEDLQRAPEVMEHLFQTPLYRDLLPKVGTQGLL-
Syn_WH7805      563 LLAKEAGLVEPSAGIADLLVVPLFETVEDLQRAPEVMEQLFQTPLYRNLLPRVGSQGQP-
PmNATL1A        563 LLAKESGLIDPTLGASDLLVVPLFETVEDLQHAPSVMESLLQSDVYRELLPRVGEKKQP-
PmNATL2A        563 LLAKESGLIDPTLGASDLLVVPLFETVEDLQHAPSVMESLLQTDVYRELLPRVGEKKQP-
PmCCMP1375      571 LLAKEYGLVDISSESSDLLVIPLFETVEDLQHAPSVMEELFQSEIYLKLLPRVGEKSQP-
```

Fig. 2: Continued

```
PmMIT9211      575 LLAKEAGLVDISSGSADLLVVPLFETVEDLQRAPSVMEELFSSSFYLNLLPRVGEKLQP-
Syn_WH5701     572 LLAKEAGLVDPKAVQTQLLVVPLFETVEDLQRAPAVMQRLFSEPFYRTLLASSAEADKP-
PmMIT9301      559 LLAKEMGLLDQDSQISKLLVVPLFETVEDLQRAPEVMEKLFKLDFYKSLLPKVGESFKP-
PmAS9601       559 LLAKEMGLLDQNSQISKLLVVPLFETVEDLKRAPEVMEKLFKLDFYRSLLPKVGESFKP-
PmMIT9312      559 LLAKEMGLLDQNSKNSNLLVVPLFETVEDLKRAPEVMEKLFKLDFYKSLLPKVGESFKP-
PmMIT9515      559 LLAKEMGLIDQSSQISKLLVVPLFETVEDLKRAPEVMEQLFQLDFYRSLLPKVGERFKP-
PmCCMP1986     559 LLAKEMGLIDQGSQISTLLVVPLFETVEDLQRAPDVMEQLFQLDFYKSLLPKVGESFKP-

Nostoc         534 LLAKEARLFDPAIAVGTIQVVPLFETVEDLQRSRSVMRQLFELPLYRALLAGGYKNTEVK
A_variabilis   578 LLAKEARLFDPAIAVGTIQVVPLFETVEDLQRSRSVMRQLFELPLYRALLAGGYKSTEVK
Syn_PCC7424    578 LLAQEAGLYDPATSSITIRIVPLFETVEDLKRAPEIMRALFEMTLYRAALAGGYQYLAK-
Cyanothece     578 LLAQEAGLYDPATCSSTIRIVPLFETVEDLKRAPEIMRALFELTLYRATLAGGYDHLAS-
Syn_PCC8801    579 LLAKEAGLYDPATSTTTIRIVPLFETVDDLKRAPEVMEDLFKLPLYRASLAGGYDQLQPS
Syn_PCC6803    587 LLAKEAGLYDPATASNSLRIVPLFETVEDLKNAPGIMDSLFSLPFYRATLAGSYHSLKEL
Syn_PCC7002    557 LLAQEAGLYDPLTGMTTIRIAPLFETVDDLRNAPEIMQALFEIPLYRACLAGGYEPPADG
A_nidulans     631 LLAKEVGLYDPVTGKSSLQVIPLFETVEDLQNAPRVMTALFELPFYTQLNPTQSE-----

Syn_CC9605     725 ETVTTAVVQNSLVTNQLDATPSWNQLMSRLATRSRDHYRALVHDNPDLVAFFQQVTPIEE
Syn_CC9902     723 ETVTTAVVQNSLVTNQLDATPSWNQLMSRLSARSREHYRALVHDNPDLVAFFQQVTPIEE
Syn_WH8102     738 ETVTTAVVQNSLVTNQLDATPSWNQLMSRVAKRSREHYRALVHDNPDLVAFFQQVTPIEE
PmMIT9303      734 ETFTTAVLQNSLVTNQLDATPSWNQLMTRLAGRSREHYRALVHNNPDLVAFFQQVTPIEE
PmMIT9313      736 ETFTTAVLQNSLVTNQLDATPSWNQLMTRLAGRSREHYRALVHNNPDLVAFFQQVTPIEE
Syn_RS9916     731 ETVSTAVVQNSLVTNQLDATPSWNELMARLAARSRSHYRALVHDNPDLVAFFQQVTPIEE
Syn_RS9917     707 ETMSTAVVQNSLVTSQLDATPSWNDLMSRLASRSRSHYRALVHDNPDLVAFFQQVTPIEE
Syn_CC9311     729 ETVTTAVIQNSLVTNQLDATPSWNKLMSRVAKSSRRNYRALVHDNPDLVAFFQQVTPIEE
Syn_WH7805     720 ETVTTAVVQNSLVTNQLDATPSWNDLMARLARCSRREHYRALVHDNPDLVAFFEQVTPIEE
PmNATL1A       720 ETVTTAVIQNSLVTNKLDATPSWNELMTRLAARSREHYRALVHDNPDLVQFFQVVTPIEE
PmNATL2A       720 ETVTTAVIQNSLVTNKLDATPSWNELMTRLAARSREHYRALVHDNPDLVQFFQVVTPIEE
PmCCMP1375     728 ETVTTAVLQNSLVTNQWDATPSWNELMTRLAVRSRQHYRALVHDNPDLVAFFQEVTPIEE
PmMIT9211      732 ETVTTAVLQNSLVTNQLDATPSWNELMTRLAARSRRHYRSLVHDNPDLVPFFQEVTPIEE
Syn_WH5701     729 ETVTTAVLQNSLVSTNVDDTPTWNDLMGRLAARSRVHYRRLVHENPDLVAFFQQVTPIEE
PmMIT9301      716 ETVTTAVIQNSLVNSRLDATPEWNQLMSRLAETSRSHYRKLVHENPDLLNFFQEVTPIEE
PmAS9601       716 ETVTTAVIQNSLVNNRLDATPEWNQLMSRLAETSRSHYRKLVHENPDLLNFFQEVTPIEE
PmMIT9312      716 ETVTTAVIQNSLVNNRLDATPEWNQLMSRLAETSRSHYRKLVYENPDLLNFFQEVTPIEE
PmMIT9515      716 ETVTTAVIQNSLVINRLDATPEWNDLMTRLAETSRIQYRKLVHENPNLLTFFQEVTPIEE
PmCCMP1986     716 ETVTTAVIQNSLVNNRLDATPEWNDLMTRLAETSRIQYRKLVHENPNLLTFFQEVTPIEE

Nostoc         714 ETITTAVIQASLLRTGFDDIEPWNEIMEELAARSRQHYRGLIYEQPDFIDFFHQVTPIEE
A_variabilis   755 ETITTAVIQASLLRTGFDDIEPWNEIMEELAARSRQHYRGLIYEQPDFIDFFHQVTPIEE
Syn_PCC7424    754 ETIVTAVIQSSLLGCGFDDIQVWNEIMEDLAGCARKAYRSLIYEDPDFIDFFMSVTPIPE
Cyanothece     753 ETIVTAVIQSSLLGCGFDDIQVWNEIMEELASCSRKAYRALIYEEPDFVDFFMSVTPIQE
Syn_PCC8801    757 ETVSTAVIQASLLGSGFDDINPWNEIMEDLAERARKAYRGLIYEQPDFLDFFLSVTPIPE
Syn_PCC6803    766 ETLTTAVIQASLLKSSFDFIEPWNRIMEELACTARRAYRSLIYEEPDFLDFFLTVTPIPE
Syn_PCC7002    728 ESVSTAVIQSSLLASGFDDIQPWNRIMEDLSQRSRAAYRALIYEEPDFLDFFMSVTPIPE
A_nidulans     785 ETITTAVIQSSLLGSGFDDIEPWNQIMEELAARSRRHYRALVYEQPDLVDFFNQVTPIEE

NmNATL2A       839 MEMLRMLNQRWPFFRMLISKVEMTLSKVDLDVAHHYVVSLGGSDDRDAFARIFDIISSEY
PmCCMP1375     847 IELLRMLNQRWPFFRMLISKVEMTLSKVDLEVAYHYMTSLGSHENREAFNCIFEIISNEY
PmMIT9211      851 IELLKMLHQRWPFFRMLISKVEMTLSKVDLEVANHYVTSLGSRQNKEAFNKIFEVISDEY
Syn_WH5701     848 LELLRLLYQRWPFFRMLISKVEMTLSKVDLDLAHHYVQALGRPSHREAFEEIFQVIAAEF
PmMIT9301      835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVARYYVDTLGSKENKDSFDNIFEVISKEY
PmAS9601       835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVARYYVDTLGSKENKDSFDNIFEVISKEY
PmMIT9312      835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVAKYYVDTLGSEENKDSFDDIFEVISKEY
PmMIT9515      835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVAKYYVDTLGSKENAKSFGDIFDVISKEY
PmCCMP1986     835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVAKYYVDTLGSKENSKSFEEIFDVISKEY
```

Fig. 2: Continued

```
Nostoc          834 LKLMRYFYVKWPFFKMVISKVEMTLAKVDMQMAGHYVQELSDPEDKPRFEKVFEQIANEY
A_variabilis    875 LKLMRYFYVKWPFFKMVISKVEMTLAKVDMQMAGHYVQELSDPEDKSRFEKVFEQIANEY
Syn_PCC7424     874 LKLLRYFYLKWPFFRMVISKVEMTLSKVDLQIAHHYVKELSKPEDLERFERVFNQISEEY
Cyanothece      873 LKLLRYFYLKWPFFKMVISKVEMTLSKVDLQIAYHYVKELSQPEDRERFERLFEQIAQEY
Syn_PCC8801     877 LKLLRYFYLKWPFFKMVVSKVEMTLSKVDLQIAHHYVRELSKAEDKERFERVFEEISQEY
Syn_PCC6803     886 MKLLRYFYFKWPFFNMVISKVEMTLSKVDLTIASHYVQELSKPEDRERFDRLFQQIKQEY
Syn_PCC7002     848 LKLMRYFYSKWPFFRMVISKVEMTLSKVDLQMASHYVHELAEKEDIPRFEKLLEQISQEY
A_nidulans      905 LNLLRYFYEKWPFFRMVISKVEMTLAKVDLQIAHHYVHELANPEDQERFERVFSQIAAEF

PmNATL2A        839 MEMLRMLNQRWPFFRMLISKVEMTLSKVDLDVAHHYVVSLGGSDDRDAFARIFDIISSEY
PmCCMP1375      847 IELLRMLNQRWPFFRMLISKVEMTLSKVDLEVAYHYMTSLGSHENREAFNCIFEIISNEY
PmMIT9211       851 IELLKMLHQRWPFFRMLISKVEMTLSKVDLEVANHYVTSLGSRQNKEAFNKIFEVISDEY
Syn_WH5701      848 LELLRLLYQRWPFFRMLISKVEMTLSKVDLDLAHHYVQALGRPSHREAFEEIFQVIAAEF
PmMIT9301       835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVARYYVDTLGSKENKDSFDNIFEVISKEY
PmAS9601        835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVARYYVDTLGSKENKDSFDNIFEVISKEY
PmMIT9312       835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVAKYYVDTLGSEENKDSFDDIFEVISKEY
PmMIT9515       835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVAKYYVDTLGSKENAKSFGDIFDVISKEY
PmCCMP1986      835 IELLRVLHQRWPFFRMLISKVEMTLSKVDLEVAKYYVDTLGSKENSKSFEEIFDVISKEY

Nostoc          834 LKLMRYFYVKWPFFKMVISKVEMTLAKVDMQMAGHYVQELSDPEDKPRFEKVFEQIANEY
A_variabilis    875 LKLMRYFYVKWPFFKMVISKVEMTLAKVDMQMAGHYVQELSDPEDKSRFEKVFEQIANEY
Syn_PCC7424     874 LKLLRYFYLKWPFFRMVISKVEMTLSKVDLQIAHHYVKELSKPEDLERFERVFNQISEEY
Cyanothece      873 LKLLRYFYLKWPFFKMVISKVEMTLSKVDLQIAYHYVKELSQPEDRERFERLFEQIAQEY
Syn_PCC8801     877 LKLLRYFYLKWPFFKMVVSKVEMTLSKVDLQIAHHYVRELSKAEDKERFERVFEEISQEY
Syn_PCC6803     886 MKLLRYFYFKWPFFNMVISKVEMTLSKVDLTIASHYVQELSKPEDRERFDRLFQQIKQEY
Syn_PCC7002     848 LKLMRYFYSKWPFFRMVISKVEMTLSKVDLQMASHYVHELAEKEDIPRFEKLLEQISQEY
A_nidulans      905 LNLLRYFYEKWPFFRMVISKVEMTLAKVDLQIAHHYVHELANPEDQERFERVFSQIAAEF

Syn_CC9605      904 ELTRKLVLEITGQNRLLGADQGLQLSVDLRNRTIVPLGFLQVALLKRLRDQNRQPPMSET
Syn_CC9902      902 ALTRKLVLEITGQPRLLGADQGLQLSVDLRNRTIVPLGFLQVALLKRLRDQNRQPPMSES
Syn_WH8102      917 GRTLKLVLEITGQSRLLGADQNLQLSVDLRNRTIVPLGFLQVALLRRLRDQNRQPPMSES
PmMIT9303       913 SLTRRLVLNITGQPRLLGADEALQQSVDLRNRTIVPLGFLQVALLRKLRDQNRQPPMN--
PmMIT9313       915 SLTRRLVLNITGQPRLLAADEALQLSVDLRNRTIVPLGFLQVALLRKLRDQNRQPPMN--
Syn_RS9916      910 ALTLQLVLDITGQERLLDGDEGLQLSVDLRNRTIVPLGFLQVALLRRLRDQNRQPPMREA
Syn_RS9917      886 SLTRRLVLAITGQERLLDADEALQLSVDLRNRTIVPLGFLQVALLRRLRDQNRQPPMSEA
Syn_CC9311      908 ARTKELVLAITGQERLLDADEALQLSVDLRNRTIVPLGFLQVALLRRLRDQNRQPPMSES
Syn_WH7805      899 SLTHRLVLEITGQERLLDADEALQLSVDLRNRTIVPLGFLQVALLRRLRDQNRQPPMSES
PmNATL1A        899 SLTKKLILEITGKSKLLSADEALQLSVNLRNRTIVPLGFLQVALLKRLRDQNRQPPISED
PmNATL2A        899 SLTKKLILEITGKSKLLSADEALQLSVNLRNRTIVPLGFLQVALLKRLRDQNRQPPISED
PmCCMP1375      907 KLTRRLVLEITGKPKLLSADEALQLSVDLRNRTIVPLGFLQVALLCRLRDQNRQPPMSET
PmMIT9211       911 HLTKRLVLRITGKSKLLSADEALQASVELRNRTIVPLGFLQVALLRRLRDQTRQPPVSEE
Syn_WH5701      908 ELTRDLVLKISGHARLLDGDEALQLSVDLRNQTIVPLGYLQVALLRRLRDQNRQPPMSEA
PmMIT9301       895 NLTKSLILEITGKNKLLESDRDLKSSVSLRNKTIIPLGFLQVSLLRRLRDQTRQPPISEF
PmAS9601        895 NLTKSLILEITGKNKLLESDRDLKSSVSLRNKTIIPLGFLQVSLLRRLRDQTRQPPISEF
PmMIT9312       895 SLTKSLVLEITGKNKLLESDRDLKSSVSLRNKTIIPLGFLQVSLLRRLRDQTRQPPISEF
PmMIT9515       895 NLTKSLVLEITGKNKLLESDRDLRLSVNLRNKTIIPLGFLQVSLLRRLRDQTRQPPISEF
PmCCMP1986      895 NLTKSLVLEITGKNKLLESDRDLRSSVNLRNKTIIPLGFLQVSLLRRLRDQTRQPPISEF

Nostoc          894 YLTRDLVLKITDHGRLLDGDEVLQRSVQLRNGTIVPLGFIQVSLLKRLRQSKNN------
A_variabilis    935 YLTRDLVLKITDHSRLLDGDEVLQRSVQLRNGTIVPLGFIQVSLLKRLRQSKNN------
Syn_PCC7424     934 HRTCRLVLDITENERLLDGDETLQRSVQLRNGTIVPLGFLQVSLLKRLRQYNAQ------
Cyanothece      933 HRTCRLVLEITENKRLLDGDETLQRSVQLRNGTIVPLGFLQVSLIKRLRQYNAQ------
Syn_PCC8801     937 HRTRDVILNITNHQRLLDSDLSLQRSVQLRNGTIVPLGFLQVALLKRLRQYSNQ------
```

Fig. 2: Continued

```
Syn_PCC6803    946 QLTRDFAMEITAHPHLLDGDRSLQRSVLLRNRTIVPLGLLQISLLKRLRQVTQE------
Syn_PCC7002    908 NLTKRLILEITENEALLDGDRPLQRSVQLRNGTIVPLGFLQVSLLKRLRQYTRE------
A_nidulans     965 QLTGHLVLTITNHGRLLDGDPELQRSVQLRNGTIVPLGFLQVALLKRLRQYRQQ------

Syn_CC9605     964 PGAPED--TRTYSRSELLRGALLTLNGIAAGMRNTG
Syn_CC9902     962 PGAPED--TRTYSRSELLRGALLTLNGIAAGMRNTG
Syn_WH8102     977 PGTPED--RRTYSRSELLRGALLTLNGIAAGMRNTG
PmMIT9303      971 --EAGD--GRTYSRSELLRGALLTINGIAAGMRNTG
PmMIT9313      973 --DAGD--GRTYSRSELLRGALLTINGIAAGMRNTG
Syn_RS9916     970 TSGDSD--GRTYSRSELLRGALLTINGIAAGMRNTG
Syn_RS9917     946 PNTDPD--GRTYSRSELLRGALLTINGIAAGMRNTG
Syn_CC9311     968 PSSDGD--GRTYSRSELLRGALLTINGIAAGMRNTG
Syn_WH7805     959 PS-DGD--GRTYSRSELLRGALLTINGIAAGMRNTG
PmNATL1A       959 VSIDSTQSSRTYSRSELLRGALLTINGIAAGMRNTG
PmNATL2A       959 VSIDSTQSSRTYSRSELLRGALLTINGIAAGMRNTG
PmCCMP1375     967 LLTEGD-IGRTYSRSELLRGALLTINGIAAGMRNTG
PmMIT9211      971 LFNERD-LARTYSRSELLRGALLTINGIAAGMRNTG
Syn_WH5701     968 SSGDGQ-DGRTYSRSELLRGALLTINGIAAGMRNTG
PmMIT9301      955 LIDKDE-SRRAYSRSELLRGALLTINGIAAGMRNTG
PmAS9601       955 FLDKDE-STRAYSRSELLRGALLTINGIAAGMRNTG
PmMIT9312      955 LIDKDE-SRRAYSRSELLRGALLTINGIAAGMRNTG
PmMIT9515      955 LNERNE-SERAYSRSELLRGALLTINGIAAGMRNTG
PmCCMP1986     955 IEDRIE-SKRAYSRSELLRGALLTINGIAAGMRNTG

Nostoc         948 -TATSGVIHSRYSKGELLRGALLTINGIAAGMRNTG
A_variabilis   989 -NATSGVIHSRYSKGELLRGALLTINGIAAGMRNTG
Syn_PCC7424    988 -AES-GVIHFRYSKEELLRGALLTINGIAAGMRNTG
Cyanothece     987 -AES-GVINFRYSKEELLRGALLTINGIAAGMRNTG
Syn_PCC8801    991 -AQS-GVIHFRYSKEELLRGAMLTINGIAAGMRNTG
Syn_PCC6803   1000 -AETSGVRYRRYSKEELLRGALLTINGIAAGMRNTG
Syn_PCC7002    962 -TQA-SIVHFRYSKEELLRGALLTINGIAAGMRNTG
A_nidulans    1019 -TETTGLMRSRYSKGELLRGALLTINGIAAGMRNTG
```

Fig. 2: Multiple sequence alignment of phosphoenolpyruvate carboxylases from cyanobacteria. The green block are amino acid stretches that are conserved and exclusive to fresh water cyanobacteria; the blue blocks highlight stretches that are conserved and exclusive to oceanic cyanobacteria; the yellow blocks contain sequences that are unique to a group of cyanobacteria of the prochlorococcus genus. The alignment was generated with Clustal W (Larkin *et al.*, 2007)

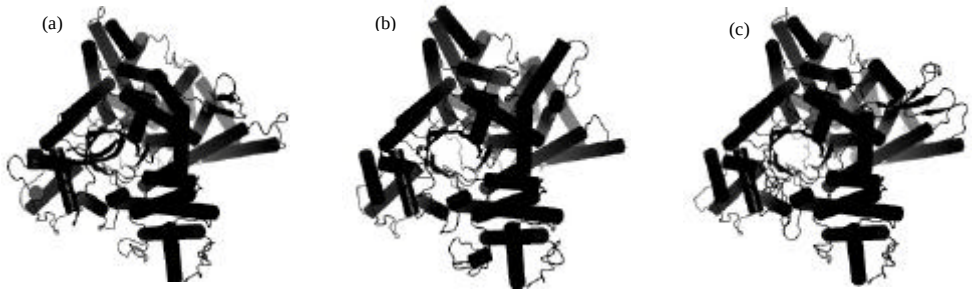

Fig. 3: Modeled structure of PEPCs from (a) anabaena variabilis, (b) synechococcus RS 9917 and (c) synechococcus PCC 7002 (c) viewed from the N-terminal end of the central β-barrel. The models were generated with SWISS MODEL and viewed with PyMOL

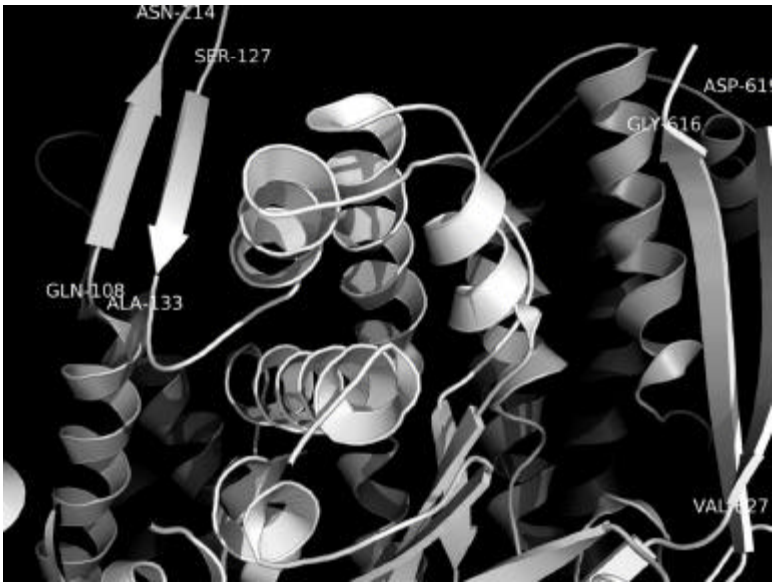

Fig. 4: Supplemental ß-strands of *Synechococcus* PCC 7002 PEPC. The model was generated with SWISS MODEL and viewed with PyMOL using the crystal structure of C4-form phosphoenolpyruvate carboxylase from maize (pdb ID: 1jqoA)

Residues that were found to be essential for catalysis and regulation in the *E. coli* and *Zea mays* enzymes are conserved in fresh water and marine cyanobacterial PEPCs. . The enzyme's active site has been found to be made of H138, R396, K546, H579, R581, R587 and R699 (*E. coli* numbering). A loop region, GRGGSIGRGG (GRGGSVGRGG in *Synechococcus* PCC 7002) is involved in catalysis and binding to aspartate, an allosteric inhibitor (Matsumura *et al*., 2002). Another loop involved in catalysis and bicarbonate binding, KRRP(G/T)GG, is found in *E. coli* and plant isoforms but it is absent in the cyanobacterial enzyme. Furthermore, E433 and R438 (*E. coli* numbering) which are involved in intersubunit contact to form the tetramer are conserved in all PEPCs.

All PEPCs contain the aspartate binding site homology which is composed of three domains: EM(T/V)(L/F)(S/A)K, LRN(G/I)(T/Y) and MRNTG. There may be other residues and domains involved in regulation by aspartate as *Anabaena variabilis* is not subject to control by aspartate (Izui *et al*., 2004). The following sequences were found in all cyanobacterial PEPCs but they were absent in *E. coli* and plant PEPCs: ^{150}LNVPPX(Q/K)X(E/Q)XL^{160}, ^{245}VDYALHYFQEVLF^{257}, ^{308}WXTACYQR^{315}, ^{336}S(M/L)(H/Q)W(S/C)XVXXXLLESLE^{351} (*Synechococcus* PCC 7002 numbering). In spite of the low sequence homology between cyanobacterial PEPCs and the allosterically regulated *E. coli* and Maize enzymes, the previously reported domains and residues involved in catalysis are shared by all PEPC proteins analyzed thus far. Unique structural features of PEPC may afford cyanobacteria with a CO_2 concentrating mechanism that is similar to what is found in plants (Sikolia *et al*., 2009).

CONCLUSION

In this study cyanobacterial phosphoenolpyruvate carboxylases were selected and characterized from a physicochemical perspective. For these enzymes, molecular weight, theoretical isoelectric

point, molar extinction coefficient, aliphatic index, instability index, number of negative residues, Grand Average Hydropathy (GRAVY) and disulfide bond proclivity were computed. Physicochemical parameters provided useful data for the purification of cyanobacterial PEPCs. The primary structure of cyanobacterial PEPCs was further analyzed and amino acids as well as domains were identified that can distinguish between the fresh water and marine enzyme. Secondary structure analysis revealed that the proteins are largely alpha helical; this was supported by the homology modeling of PEPCs from *Synechococcus* RS 9917, *Anabaena variabilis* and *Synechococcus* PCC 7002. The latter is a marine organism but its PEPC is most similar to PEPCs from fresh water cyanobacteria in all aspects. The 3D model of *Synechococcus* PCC 7002 PEPC suggests that the molecule contain two **β**-sheets in addition to the central **β**-barrel. Those **β**-sheets may play a role in regulation and catalysis. While mutagenesis will be crucial in determining the function of the various residues and structures identified in this study, the 3D models can be used for functional analysis until crystal structures become available for cyanobacterial PEPCs.

REFERENCES

Arnold, K., L. Bordoli, J. Kopp and T. Schwede, 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics, 22: 195-201.

Biou, V., J.F. Gibrat, J.M. Levin, B. Robson and J. Garnier, 1988. Secondary structure prediction: Combination of three different methods. Protein Eng., 2: 185-191.

Caldwell, P.E., M.R. MacLean and P.R. Norris, 2007. Ribulose bisphosphate carboxylase activity and a Calvin cycle gene cluster in *Sulfobacillus* species. Microbiology, 153: 2231-2240.

Chen, L.M., K.Z. Li, T. Miwa and K. Izui, 2004. Overexpression of a cyanobacterial phosphoenolpyruvate carboxylase with diminished sensitivity to feedback inhibition in *Arabidopsis* changes amino acid metabolism. Planta, 219: 440-449.

Chen, L.M., T. Omiya, S. Hata and K. Izui, 2002. Molecular characterization of a phosphoenolpyruvate carboxylase from a thermophilic cyanobacterium, *Synechococcus vulcanus* with unusual allosteric properties. Plant Cell. Physiol., 43: 159-169.

Combet, C., C. Blanchet, C. Geourjon and G. Deleage, 2000. NPS@: Network protein sequence analysis. Trends Biochem. Sci., 25: 147-150.

Deleage G. and B. Roux, 1987. An algorithm for protein secondary structure prediction based on class prediction. Protein Eng., 1: 289-294.

Eisenhaber, F., F. Imperiale, P. Argos and C. Froemmel, 1996. Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods. Proteins Struct. Funct. Design, 25: 157-168.

Frishman, D. and P. Argos, 1996. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. Protein Eng., 9: 133-142.

Garnier, J., D.J. Osguthorpe and B. Robson, 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol., 120: 97-120.

Geourjon, C. and G. Deleage, 1994. SOPM: A self-optimized method for protein secondary structure prediction. Protein Eng., 7: 157-164.

Geourjon, C., P. Gallinari and G. Deleage, 1999. Improved performance in protein secondary structure prediction by inhomogeneous score combination. Bioinformatics, 15: 413-421.

Gilbert, D., 2003. Sequence file format conversion with command-line readseq. Curr. Protoc Bioinformatics.

Guex, N. and M.C. Peitsch, 1997. SWISS-MODEL and the Swiss-PDBViewer: An environment for comparative protein modeling. Electrophoresis, 18: 2714-2723.

Hammami, S.S., R. Chaffai and E. El Ferjani, 2004. Effect of cadmium on sunflower growth, leaf pigment and photosynthetic enzymes. Pak. J. Biol. Sci., 7: 1419-1426.

Izui, K., H. Matsumura, T. Furumoto and Y. Kai, 2004. Phosphoenolpyruvate craboxylase: A new era of structural biology. Annu. Rev. Plant Biol., 55: 69-84.

Kai, Y., H. Matsumura, T. Inoue, K. Terada and Y. Nagara *et al.*, 1999. Three-dimensional structure of phosphoenolpyruvate carboxylase: A proposed mechanism for allosteric inhibition. Proc. Natl. Acad Sci. USA., 96: 823-828.

Karthikeyan, C.V. and G. Gopalaswamy, 2009. Studies on acid stress tolerant proteins of cyanobacterium. Int. J. Biol. Chem., 3: 1-10.

Khidir Ahmed, O., 2009. Sugar regulation of plastid reversion in citrus epicarp is mediated through organic acid metabolism. Pak. J. Biol. Sci., 12: 246-251.

King, R.D. and M.J.E. Sternberg, 1996. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Sci., 5: 2298-2310.

Knowles, V.L. and W.C. Plaxton, 2003. From genome to enzyme: Analysis of key glycolytic and oxidative pentose-phosphate pathway enzymes in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Cell Physiol., 44: 758-763.

Kwon, Y.D., S.Y. Lee and P. Kim, 2008. A physiology study of *Escherichia coli* overexpressing phosphoenolpyruvate carboxykinase. Biosci. Biotechnol. Biochem., 72: 1138-1141.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna and P.A. McGettigan *et al.*, 2007. Clustal W and Clustal X version 2.0. Bioinformatics, 23: 2947-2948.

Leegood, R.C., 2002. C_4 photosynthesis: Principles of CO_2 concentration and Prospects for its introduction into C_3 plants. J. Exp. Bot., 53: 581-590.

Lovell, S.C., I.W. Davis, W.B. Arendall, P.I. de Bakker and J.M. Word *et al.*, 2003. Structure validation by Cα geometry: CΨ and Cβ deviation. Proteins, 50: 437-450.

Matsumura, H., Y. Xie, S. Shirakata, T. Inoue and T. Yoshinaga *et al.*, 2002. Crystal structures of C_4 form maize and quaternary complex of *E. coli* phosphoenolpyruvate carboxylases. Structure, 10: 1721-1730.

O'Leary, B., S.K. Rao, J. Kim and W.C. Plaxton, 2009. Bacterial-type Phosphoenolpyruvate Carboxylase (PEPC) functions as a catalytic and regulatory subunit of the novel class-2 PEPC complex of vascular plants. J. Biol. Chem., 284: 24797-24805.

Rice, P., I. Longden and A. Bleasby, 2000. Emboss: The european molecular biology open software suite. Trends Genet., 16: 276-277.

Rost, B. and C. Sander, 1993. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. Proc. Natl. Acad. Sci., 90: 7558-7562.

Rost, B., G. Yachdav and J. Liu, 2004. The predictprotein server. Nucleic Acids Res., 32: W321-W326.

Sikolia, S., J.C. Onyango, E. Beck and J.I. Kinyamario, 2009. The distribution of C3 and C4 *Photosynthetic* species of the centrospermeae along an altitudinal gradient in Western Kenya. Int. J. Bot., 5: 47-57.

Smith, A.A., M.W. Coomes and T.E. Smith, 2008. Isolation and sequence of the phosphoenolpyruvate carboxylase gene of the marine cyanobacterium synechococcus PCC 7002. J. Boil. Sci., 8: 1261-1270.

Soltani, N., G. Zarrini, Y. Ghasemi, S. Shokravi and L. Baftehchi, 2007. Characterization of a soil cyanobacterium *Fischerella* sp. FS 18 under NaCl stress. J. Boil. Sci., 7: 931-936.

Stultz, C.M., J.V. White and T.F. Smith, 1993. Structural analysis based on state-space modeling. Protein Sci., 2: 305-314.

Tomitani, A., A.H. Knoll, C.M. Cavanaug and T. Ohno, 2006. The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives. Proc. Natl. Acad. Sci., 103: 5442-5447.

Tripp, J.H., S.R. Bench, K.A. Turk, R.A. Foster and B.A. Desany *et al.*, 2010. Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. Nature, 464: 90-94.

Vriend, G., 1990. What If: A molecular modeling and drug design program. J. Mol. Graphics, 8: 52-56.

White, J.V., C.M. Stultz and T.F. Smith, 1994. Protein classification by stochastic modeling and optimal filtering of amino-acid sequences. Mathematical Biosci., 119: 35-75.