

Isolation and Sequence of the Phosphoenolpyruvate Carboxylase Gene of
Synechococcus PCC 7002

by

Aubrey A. Smith¹, Marguerite W. Coomes¹, and Thomas E. Smith^{1*}

1. Department of Biochemistry and Molecular Biology, Howard University College of Medicine,
Washington, DC 20059 USA

Submitted for publication in
Biomolecular Engineering

April, 2005

* To whom correspondence should be addressed at the Howard University College of Medicine Department of Biochemistry and Molecular Biology. Adams Building, Room 3430, 520 W Street, NW, Washington, DC 20059.

Telephone: 202-806-9713
Fax: 202-806-5784
Email address: tsmith@howard.edu

*Grant Number DE-FG02-97ER62453 from the Department of Energy supported this research.

Abstract

The pepc gene, which encodes phosphoenolpyruvate carboxylase (PEPC), of the marine cyanobacterium Synechococcus PCC 7002, was isolated and sequenced. PEPC is an anaplerotic enzyme, but it may also contribute to overall CO₂ fixation through β-carboxylation reactions. A consensus sequence generated by aligning the pepc genes of Anabaena variabilis, Anacystis nidulans and Synechocystis PCC 6803 was used to design two sets of primers that were used to amplify segments of Synechococcus PCC 7002 pepc. In order to isolate the gene, the sequence of the PCR product was used to search for the pepc nucleotide sequence from the publicly available genome of Synechococcus PCC 7002. At the time, the genome for this organism had not been completed although sequences of a significant number of its fragments are available in public databases. Thus, the major challenge was to find the pepc gene among those fragments and to complete gaps as necessary. Even though the search did not yield the complete gene, PCR primers were designed to amplify a DNA fragment using a high fidelity thermostable DNA polymerase. An open reading frame (ORF) consisting of 2988 base pairs coding for 995 amino acids was found in the 3066 bp PCR product. The pepc gene had a GC content of 52% and the deduced protein had a calculated molecular mass of 114,049 Da. The amino acid sequence was closely related to that of PEPC from other cyanobacteria, exhibiting 59-61% identity. The sequence differed significantly from plant and E. coli PEPC with only 30% homology. However, comparing the Synechococcus PCC 7002 sequence to the recently resolved E. coli PEPC revealed that most of the essential domains and amino acids involved in PEPC activity were shared by both proteins. The recombinant Synechococcus PCC 7002 PEPC was expressed in E. coli.

Keywords: Synechococcus PCC 7002, phosphoenolpyruvate carboxylase, pepc gene, cyanobacteria.

1. Introduction

Phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31) catalyzes the irreversible β -carboxylation of phosphoenolpyruvate (PEP) to yield oxaloacetate (OAA) and inorganic phosphate (P_i). PEPC is found in higher plants and most types of bacteria including cyanobacteria (6, 19, 5). Several PEPCs have been purified from plants and bacteria, including thermophilic bacteria (15). The enzyme contributes to photosynthetic and anaplerotic CO_2 fixation. The best-described bacterial PEPC is that found in *Escherichia coli*; its physical and chemical properties have been extensively investigated (16), and its three-dimensional structure has been determined at 2.8 Å resolution (9). *E. coli* PEPC is a tetrameric protein composed of four identical subunits of approximately 100 KDa each. The enzyme is allosterically activated by acetyl CoA, fructose-1,6-bisphosphate (F-1,6-BP), CDP, and GTP. Other activators include fatty acids and polyanions (17). Although the physiologically active *E. coli* enzyme is a tetramer, it is also catalytically active as a dimer (14). The *pepc* genes of the freshwater cyanobacteria *Anabaena variabilis*, *Anacystis nidulans*, and *Anabaena* PCC 7120 have also been cloned and sequenced (12, 6, 10).

In many marine eukaryotic microalgae, up to 80% of CO_2 assimilated may be attributed to the actions of PEPC and PEPCK. A significant amount of carbon dioxide is fixed through the same route in prokaryotic cyanobacteria. Despite the importance of PEPC in net carbon dioxide fixation, gene expression and regulation of this enzyme have not been extensively studied in cyanobacteria of marine origin. A significant portion of algal carbon metabolism is coupled to nitrogen metabolism, but little is known of the interactions between the enzymes of carbon and nitrogen metabolism. PEPC in marine cyanobacteria has not been cloned and expressed. Thus the opportunity to evaluate those interactions has not existed. As an initial step to obtain enough enzyme to begin the essential characterization, this study focuses on isolation and characterization of the *pepc* gene of *Synechococcus* PCC 7002, a halotolerant unicellular naturally transformable marine cyanobacterium (4, 18). The organism can be grown under photoautotrophic conditions in saline medium aerated with air containing CO_2 (11). *Synechococcus* PCC 7002 has a genome size of approximately 3×10^6 base-pairs (2).

2. Materials and methods

2.1 Isolation of *Synechococcus* PCC 7002 DNA

Axenic cultures of wild-type *Synechococcus* sp. PCC 7002 (formerly *Agmenellum quadruplicatum* strain PR6) were obtained from the laboratories of Drs. John Paul (The University of South Florida, St. Petersburg) and Robert Tabita (The Ohio State University, Columbus, Ohio). *Synechococcus* sp. PCC 7002 was grown in ASP-2 medium supplemented with 1 g/L $NaNO_3$. The cells were grown under constant stirring and illumination from fluorescent light bulbs at room temperature. Air supplemented with 3 to 5% (v/v) CO_2 was bubbled into the medium through gas dispersion tubes. Genomic DNA from *Synechococcus*

PCC 7002 was isolated according to a procedure by Mazur (13). An aliquot (5 g) of cyanobacterial cells was washed twice in 15 ml of buffer A (0.05 M NaCl, 0.05 M Na₂EDTA, 0.05 M Tris-Cl; pH 8.0). After centrifugation at 10,000 rpm for 5 minutes, the cells were resuspended in the same buffer. Lysozyme was added to the suspension to a final concentration of 3.5 mg/ml. This solution was placed in a water bath at 37 °C for two hours. Buffer A (15 ml) and 3% sodium dodecylsulfate (SDS) were added to the cells, and the entire suspension was frozen at -70 °C and thawed at room temperature three times before being placed at 4 °C overnight. The lysate was centrifuged at 10,000 rpm for 10 minutes, and twice the volume in ethanol was added to the supernatant. Upon precipitation, the DNA was removed from the solution with a Pasteur pipette, placed in a fresh tube and re-suspended in buffer A. The crude DNA solution was treated with RNase A (1 mg/ml) for 60 minutes and with proteinase K (1 mg/ml) for five hours. The digest was extracted twice with phenol/chloroform/isoamylalcohol (25/24/1), pH 8.0, and ethanol was added to the aqueous phase. The DNA pellet was washed in 70%, 80% and 90% ethanol, air-dried, and re-suspended in TE buffer, pH 8.0 to final concentrations of 100-500 µg/ml.

2.2 Amplification of pepc gene fragments by PCR and cloning of the PCR products.

PCR was used to generate homologous probes and for the isolation of the pepc gene of Synechococcus PCC 7002. A multiple sequence alignment of the pepc genes for Anabaena variabilis, Anacystis nidulans and Synechocystis sp. PCC 6803 was generated. The gene sequences were taken from the National Center for Biotechnology Information (NCBI) genebank database, and the alignment was generated with the alignment software ClustalWT™ (20). The resulting consensus sequence was used as an entry in the Primer3™ primer selection software.

The primers were set to generate PCR fragments 200-1200 bps long, and they were analyzed with NetPrimer™ for potential primer-dimer formation and secondary structures (Premier Biosoft International Inc.)

The PCR experiments were conducted in a Perkin-Elmer Cetus thermal cycler. The PCR mixtures included 0.5 µg chromosomal/target DNA and 100 pg of each primer. The MgCl₂ concentration was optimized from 1.5 mM to 3.5 mM per reaction mixture. Taq polymerase buffer (100 mM Tris-Cl, pH 8.3, 500 mM KCl, 15 mM MgCl₂, 1% Triton X-100) was used, and 1.5 U Taq DNA polymerase (Promega) was added. The annealing temperature, T_A, was also optimized from 50 °C to 60 °C. The chosen amplification sequence was: [Time delay: 2 minutes/94 °C; Thermo-cycling: (94 °C/1 minute, 50 °C/1 minute, 72 °C/3 minutes) x 30; Final extension: 72°C/5 minutes; Soak: 4 °C/∞]. The PCR products were cloned in the TA-cloning vector pGEM-T-Easy (Promega). The 10 µl ligation mixture included 2x ligase buffer (5 µl) (60 mM Tris-HCl (pH 7.8), 20 mM MgCl₂, 20 mM DTT, 2 mM ATP, 10% PEG) 50 ng of vector, 3 units of T4 DNA ligase (Promega), and deionized water. The ligation reactions were incubated at 4 °C for 16 hours. The entire reaction was used to transform E. coli strain JM109. The transformants were plated on LB-agar plates containing ampicillin, IPTG and X-Gal. Plasmid DNA was isolated from the cell cultures by alkaline lysis, and the DNA was digested with EcoRI and SacI at 10 U/µg DNA. The size of the inserts was determined by agarose gel electrophoresis. The constructs were sequenced by automatic sequencing (Veritas Inc). The recombinant plasmids were labeled pRF1 and pRF2.

2.3 In silico BLAST search of the pepc gene and amplification of the gene by PCR.

The PCR-amplified pepc probe was used to scan the incomplete genome of Synechococcus PCC 7002 with the Basic Local Alignment Search Tool (BLAST) software (1). The incomplete genome is composed of several fragments that are assumed to be uninterrupted. The probe showed significant homology to an 18.8 kb fragment (NCBI sequence ID: gnl|jmarq_320491contig051302_282). In light of the results obtained by the BLAST search, four PCR primers were designed that would amplify DNA fragments of more than 3 kb since pepc genes from other organisms are approximately 3 kb on average. PCR experiments were performed with the following reagents in a 50 μ l reaction: 5 μ l 10X reaction mixture (100 mM Tris-Cl, pH 8.3, 500 mM KCl, 15 mM MgCl₂, 1% Triton X-100), 1 μ g of Synechococcus PCC 7002 genomic DNA, 0.25 mM of each dNTP, 2.5 U Taq DNA polymerase (PGC Scientifics), 80 pmol forward primer and 80 pmol reverse primer. The reverse primer sequence was TCAACCCGTGTTCCACATTCC. ; Four forward primers were used that had the following sequences: (A), ATGCCAATCGTTAACAGTATCTC; (B), GGGTTATCCAGAACATGAAGCTATAGTCA; (C), GCAGAGGATCAAACCATCCTCACA; and (D), CCAGAACATGAAGCTATAGTCACCTATACCA. The amplification program was: [Time delay: 2 minutes/94 °C; Thermo-cycling: (94 °C/1 minute, 55 °C/1 minute, 72 °C/4 minutes) x 30; Final extension: 72 °C/10 minutes; Soak: 4 °C/∞]. Long and Accurate PCR experiments (LA-PCR) were repeated using LA-Taq DNA polymerase (Takara.Mirus.Bio). The latter is a mixture of Taq DNA polymerase with a proofreading polymerase optimized for amplification of long templates.

Additional PCR experiments were performed with primers A, D and the standard reverse primer. The 50 μ l reaction included: 5 μ l reaction mixture (200 mM Tris-Cl, pH 8.8, 100 mM KCl, 20 mM MgSO₄, 1% Triton X-100, 100 mM (NH₄)₂SO₄, 1 mg/ml nuclease-free BSA), 1 μ g of Synechococcus PCC 7002 genomic DNA, 0.25 mM of each dNTP, 2.5 U PfuTurbo DNA polymerase (Stratagene), 60 pmol reverse primer and 60 pmol forward primer. The amplification program was: [Time delay: 1 minute/95 °C; Thermo-cycling: (94 °C/30 seconds, 55 °C/30 seconds, 72 °C/4 minutes) x 30; Final extension: 72 °C/10 minutes; Soak: 4 °C/∞]. Fragments of the expected size were obtained in PCR experiments with PfuTurbo using dimethylsulfoxide (DMSO) as an additive (4% v/v) and performed using forward primers A and D. The amount of primer was reduced to 20 pmol and the amount of target DNA was reduced to 0.5 μ g. PCR products were A-tailed with Taq DNA polymerase. The 10 μ l reaction mixture contained 1 μ g of primers A or D, 1 μ l of Taq polymerase 10X buffer (100 mM Tris-Cl, pH 8.3, 500 mM KCl, 15 mM MgCl₂, 1% Triton X-100), 2 mM of deoxyadenosine triphosphate (dATP) and 5 U of Taq DNA polymerase. The reaction was placed at 70 °C for 60 minutes in a thermocycler (Perkin Elmer Cetus). The A-tailed DNA was cloned into the TA cloning vector pGEM-T-Easy vector (Promega). The ligation reaction included 50 ng pGEM-T-Easy, 1 μ l ligase buffer (300 mM Tris-HCl pH 7.8, 100 mM MgCl₂, 100 mM DTT and 10 mM ATP), 3 U T4 DNA ligase (Promega Corp.) and 6 μ l of the A-tailing reaction mixture. The ligation reaction was incubated at 4 °C for 16 hours.

Competent DH5 α cells were transformed with the ligation reactions. Positive clones were confirmed by restriction enzyme digestion and by PCR screening with the RF1 and RF2 primers. The inserts were sequenced by automated sequencing at Bioserve Biotechnologies (Laurel, MD).

2.4 Expression of recombinant cyanobacterial phosphoenolpyruvate carboxylase

New PCR primers were designed based on the sequence information obtained from the sequencing of the pGEM-T-Easy recombinants. The primer sequences were (FP-S7002-PEPC-ndel) ATTCCATATGAACCAAGTCATGCATCCC and (RP-S7002-PEPC-xho1) CCGCTGAGTCAACCCGTGTTCCGCATT. The primers were synthesized with NdeI and XhoI adapters for directional cloning into the pET-15b expression vector (Novagen). The 50 μ l PCR experiments contained 20 pmol of the forward and reverse primers, 10 mM of each dNTPs, 0.5 μ g Synechococcus PCC 7002 genomic DNA, 5 μ l of PfuUltra polymerase 10X buffer (200 mM Tris-Cl, pH 8.8, 100 mM KCl, 20 mM MgSO₄, 1% Triton X-100, 100 mM (NH₄)₂SO₄, 1 mg/ml nuclease-free BSA) and 2.5 U PfuUltra High Fidelity polymerase (Stratagene). The amplification program was: [Time delay: 2 minutes/94°C; Thermo-cycling: (94 °C/1 minute, 55 °C/1 minute, 72 °C/4 minutes) x 30; Final extension: 72 °C/10 minutes; Soak: 4 °C/∞]. The 3 kb product as well as 2 μ g of pET-15b was double-digested with NdeI and XhoI. Cut pET-15b was dephosphorylated by adding 39 μ l of distilled deionized water, 10 μ l of CIAP buffer (final composition: 50 mM Tris-HCl pH 9.3, 1 mM MgCl₂, 0.1 mM ZnCl₂ and 1 mM spermidine), and 1 U calf intestinal alkaline phosphatase (CIAP) (Promega) to the 50 μ l digest. The reaction mixture was incubated at 37 °C for 60 minutes. The pepc insert was ligated to the pET-15b vector. The 20 μ l reaction included 3 U of T4 DNA ligase (GeneChoice) and 10 μ l of rapid ligation buffer; the reaction was incubated at 16 °C for 16 hours. The ligation mixture was used to transform competent JM109 cells. Recombinants were screened by restriction enzyme digestion and PCR. The pepc coding sequence was also subcloned into the pET- 30 Xa/LIC expression vector. PCR primers were designed with special adapters for ligation-independent cloning. The primer sequences were:

GGTATTGAGGGTCGCATGAACCAAGTCATGCATCCC (labeled FP-PEPC-LIC) and AGAGGAGAGTTAGAGGCAACCCGTGTTCGCGATT (labeled RP-PEPC-LIC). The 50 μ l experiments contained 20 pmoles of the forward and reverse primers, 10 mM of dNTPs, 0.5 μ g of pPCC-1, 5 μ l of PfuUltra DNA Polymerase 10X buffer (200 mM Tris-Cl, pH 8.8, 100 mM KCl, 200 mM MgSO₄, 1% Triton X-100, 100 mM (NH₄)₂SO₄, 1 mg/ml nuclease-free BSA and 2.5 U PfuUltra High fidelity polymerase (Stratagene). The amplification program was: [Time delay: 2 minutes/94 °C; Thermo-cycling: (94 °C/1 minute, 55 °C/1 minute, 72 °C/4 minutes) x 30; Final extension: 72 °C/10 minutes; Soak: 4 °C/∞]. A 5 μ l sample of the PCR mixture was removed for analysis by agarose gel electrophoresis. The PCR product was treated with T4 DNA polymerase in a 20 μ l reaction containing: 0.2 pmol of purified PCR product, 2 μ l of T4 DNA polymerase buffer (250 mM Tris-acetate, pH 7.7, 1 M potassium acetate, 100 mM magnesium acetate, 2.5 mM dGTT, 5 mM DTT, 1U T4 DNA polymerase). The reaction mixture was incubated at 22 °C for 30 minutes. The enzyme was inactivated by incubation at 75 °C for 20 minutes. The DNA polymerase-treated DNA was annealed to the pET30Xa/LIC vector (Novagen) by mixing 2 μ l containing 0.02 pmol of the insert to 1 μ l of the vector solution provided. After an incubation at 22 °C for 5 minutes, 1 μ l of 25 mM EDTA was added and the mixture was incubated for an additional 5 minutes at 22 °C. A control reaction with a Xa/LIC β -galactosidase insert was performed in parallel. An aliquot (1 μ l) of the annealing reactions was used to transform Novablue competent cells (Novagen). Positive clones were identified by agarose gel electrophoresis, PCR and sequencing. Recombinant plasmids were used to transform expression hosts BL21(DE3) pLysS and Rosetta(DE3) pLysS.

A starter culture of the pET recombinant was prepared by inoculating 3-5 ml of LB

medium containing chloramphenicol and either ampicillin or kanamycin with cells from a plate or glycerol stock. The cells were incubated at 37 °C with shaking at 250 rpm until an OD₆₀₀ of 0.4-1.0 was reached. The starter culture was either placed at 4 °C overnight until the next growth phase, or 3 ml of it was used to inoculate 100 ml of LB medium. The culture was shaken at 250 rpm and 37 °C until the OD₆₀₀ reached 0.5 to 0.7. The culture was divided into two parts. IPTG was added to one of the 50 ml cultures to a final concentration of 1 mM and no addition was made to the other part. The cells were allowed to grow for 2 to 8 hours. The OD₆₀₀ of each culture was taken, and they were chilled on ice for 5 minutes. The cells were harvested from liquid culture by centrifugation at 6000 rpm for 10 minutes using a pre-weighed centrifuge tube at 4 °C. The supernatant was removed, and the pellet was washed in cold 20 mM Tris-Cl, pH 8.0. The cells were centrifuged as above, and the weight of the pellet was determined. The pellet was frozen completely at -70 °C and then completely thawed at room temperature; the DE3 lysogens lyse upon thawing. The cell pellet was resuspended in lysis buffer (50 mM Tris-Cl, pH 7.5, 5% glycerol, 50 mM NaCl) or Bugbuster reagent (Novagen) using 7 ml of buffer per gram of cell pellet. The suspension was incubated for 20 minutes on a shaking platform at room temperature. Benzonase (Novagen) nuclease was added at 1 µl per 1 ml of lysis buffer in order to degrade all forms of nucleic acids. The extract was centrifuged at 16,000 rpm for 20 minutes. The supernatant was placed in a fresh tube. The pellet was used to prepare the insoluble cytoplasmic fraction.

The insoluble pellet was washed in 750 µl of 20 mM Tris-Cl, pH 8.0, and then it was centrifuged at 16,000 rpm for 5 minutes. The supernatant was removed, and the wash repeated. The final pellet was resuspended in 1.5 ml of 1% SDS with heating and mixing in the Minibeadbeater® (BioSpec Inc.) at 50,000 rpm for 30 seconds. SDS-PAGE samples were prepared from the soluble and insoluble fractions by combining equal amounts of lysate and 4X SDS-PAGE sample buffer (250 mM Tris-Cl, pH 6.8, 8% SDS, 10% β-mercaptoethanol, 40% glycerol and 0.02% bromophenol blue). The sample was heated immediately at 90 °C for 5 minutes and stored at -20 °C until analysis by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples (20-30 µl) were loaded on to the stacking gel. The electrophoretic buffer was Tris-glycine SDS-running buffer (25 mM Tris-Cl, pH 8.3, 98 mM glycine, 0.1% (w/v) SDS). The gel was electrophoresed at 200 V for 45 minutes.

3. Results

3.1. Probe Synthesis

In order to synthesize homologous probes for the purpose of gene cloning, an alignment of *pepc* gene from the fresh water cyanobacteria *Anabaena variabilis*, *Anacystis nidulans* and *Synechocystis* PCC 6803 was generated to design PCR primers for the amplification of segments of the phosphoenolpyruvate carboxylase gene of *Synechococcus* PCC 7002. The alignments were generated with the ClustalW software (20). Two sets of primers were chosen FP1: CTGTTACTTGGCTCTGGGT and RP1: CAGATAGGGTTACTTGGCGATA, and FP2: TATCGCCAAGTAACCCTATCGT and RP2: CACTTCCCCTGTTCGGTAAATT. The FP1/RP1 primers, derived from highly conserved regions of the genes, were expected to generate a 266 bp fragment while the FP2/RP2 primers were expected to generate a 992 bp fragment based on the numbering of the consensus sequence. The primers were analyzed with the NetPrimer software (Premiere Biosoft International) and were found to have no propensity for

secondary structures. The PCR experiments with the FP1/RP1 and FP2/RP2 primers were successful, and the sizes of the products were determined to be approximately 276 bp for the product of the FP1 and FP2 primers (RF1) and 1042 bp for the product of the FP2 and RP2 primers (RF2). These sizes are in agreement with expected values based on the alignment of cyanobacterial pepc genes. The PCR products were cloned into the pGEM-T-Easy vector (Promega Inc.). Since the RF1 and RF2 fragment represent adjacent parts of the gene, the sequences were juxtaposed and renamed Syn7002pepcΔ (or SpepcΔ). The degree of identity between SpepcΔ and corresponding regions of other cyanobacterial pepc is approximately 65%.

3.2 Isolation of pepc from Synechococcus PCC 7002 from the unfinished genome

In order to isolate the gene, the sequence of the probe (SpepcΔ) was used to search for the pepc nucleotide sequence from the publicly available unfinished shotgun genome of Synechococcus PCC 7002 (Accession number NC 003488) with the Basic Local Alignment Search Tool (BLAST) software (1). The BLAST search resulted in the identification of an 18.8 kb fragment that was renamed UF18. It was assumed that UF18 was a continuous DNA fragment, and that it may contain a number of open reading frames (ORF). Furthermore, SpepcΔ hybridized to a central position on UF18, thus, pepc was thought to be contained within that fragment. When aligned against the pepc genes of Anacystis nidulans, Anabaena variabilis and Synechocystis PCC 6803 it was found that the UF18 fragment included the 3' end of the gene that is highly conserved in all species. The UF18 sequence upstream from SpepcΔ showed a high degree of divergence from other known cyanobacterial pepcs; many PEPCs share low homology at their N-termini. An ORF search of UF18 did not find a 3 kb ORF as expected from previous knowledge of pepc genes; thus, it was evident that the sequence upstream of the probe was possibly erroneous. Nevertheless four forward primers were designed from the UF18 sequence so that the PCR experiment would yield progressively longer products slightly larger than 3 kb. The products were not expected to contain any of the UF18 sequence immediately preceding the probe sequence, but the amplification of the entire pepc coding sequence was anticipated. A reverse primer was also designed from UF18 that represented the highly conserved 3' end of the gene. This approach to isolating the gene represents an unconventional strategy to obtain the 5' end of the coding sequence that can be compared to genome walking with only partial genomic information. The products of all four reactions using the reverse primer and each of the four forward primers separately were approximately 3 kb when visualized on an agarose gel. The PCR products were cloned into the pGEM-T-Easy vector (Promega) and they were sequenced using the T7 RNA polymerase promoter forward primer and the SP6 RNA polymerase reverse primer. It seemed from the first round of sequencing that the reverse primer also acted as the forward primer in the PCR experiments. This assumption was strengthened by analyzing the sequence of the reverse primer. The 5' end of the target sequence was TCAACCCCGTTCCACTG. This differs from the reverse primer at the underlined nucleotides. When the concentration of the reverse primer was doubled and the forward primers were omitted, the PCR experiments yielded the same 3 kb product. The nucleotide sequence of the insert revealed that the open reading frame of Synechococcus PCC 7002 consisted of 2988 base pairs and encoded a polypeptide of 995 amino acid residues with a molecular mass of 114,046 Da (Figure 1). The gene was sequenced on a Beckman CEQ 800 sequencer by Bioserve Biotechnologies (Laurel, MD). The sequence was assembled with the Contig Assembly Program (CAP) (7). None of the forward primer sequences used in this work was found in the sequenced

gene. The sizes of the *pepc* gene and the predicted PEPC amino acid sequence were close to other phosphoenolpyruvate carboxylase genes and proteins isolated thus far by others. The GC content of the coding region was 52%. There were no obvious -10 and -35 sequences in the region upstream of the ATG start site. A multiple sequence alignment of *pepc* genes from *Synechococcus* PCC 7002, *Anabaena variabilis*, *Synechococcus* WH 8102, *Synechococcus* PCC 6803, *Anacystis nidulans*, *Prochlorococcus marinus* CCMP 1378, *Prochlorococcus marinus* CCMP 1375 and *Prochlorococcus marinus* MIT9313 was performed with ClustalW (20). The alignment indicated that *Synechococcus* PCC 7002 PEPC shared the most identity with *Synechococcus* PCC 6803 PEPC (63%), and only 30% identity with *E. coli* PEPC. Generally, *Synechococcus* PCC 7002 is most similar to PEPC from fresh water cyanobacteria than PEPC from marine strains (Table 1 and Figure 2). The alignment of cyanobacterial enzymes revealed that there are conserved amino acid substitutions between the marine and fresh water strains (Table 2). *Synechococcus* PCC 7002 PEPC shows substitutions associated with the enzyme from the fresh water strains. Alignment of the deduced amino acid sequence of *Synechococcus* PCC 7002 PEPC and *E. coli* PEPC was further analyzed for the presence of amino acids that have been found in PEPC from *E. coli* to be involved in catalysis and regulation. In spite of the low sequence homology between PEPC from the allosterically regulated *E. coli* PEPC and that of *Synechococcus* PCC 7002, most of the essential domains and amino acids involved in PEPC activity are present in both proteins (Figure 3). As is also the case for various cyanobacterial sequences, *Synechococcus* PCC 7002 PEPC and *E. coli* PEPC have a higher degree of homology at the C-terminal than at the N-terminal. The sequences shared 37% identity at the C-terminal and 23% identity at the N-terminal. The sequences contributing to the active site are TAHPT, DXRQES, GYSDSYKDAG, and FHGRGGXXGRGG; residues conserved between *E. coli* and cyanobacterial PEPCs are underlined (11). Replacement of the invariant arginine in the GRGG repeats was associated with a loss in catalytic activity (21). In the *E. coli* enzyme, it had been proposed that the glycine-rich loop (FHGRGGXXGRGG) helps to position substrate molecules at the active site and forms a lid that protects the reaction intermediates from attack by surrounding water molecules. Moreover, R587 of *E. coli* PEPC G-rich loop, which corresponds to R683 on the *Synechococcus* PCC 7002 PEPC, plays a role in inhibition by aspartate. Aspartate binding to this residue traps the loop and prevents it from contributing to the active site (9). Furthermore, *E. coli* and *Synechococcus* PCC 7002 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. This site is also found in PEPC from *Anacystis nidulans* even though that enzyme is not inhibited by aspartate (8).

3.3 Expression of recombinant *Synechococcus* PCC 7002 PEPC

The *pepc* gene of *Synechococcus* PCC 7002 was subcloned into the pET-30Xa/LIC vector (Novagen) by ligase independent cloning. The orientation of the insert was confirmed by PCR and restriction mapping. Induction of expression hosts containing plasmids carrying the *pepc* coding sequence and the β -galactosidase control (both with His tags) was carried out under various IPTG concentrations, incubation times and incubation temperatures. While induction of the β -galactosidase control produced a strong band, expression of the recombinant PEPC was weak, and the band could only be seen by Western blotting using an antibody to the His-tag adduct. Considering the possibility that the low expression was due to differences in codon usage between *Synechococcus* PCC 7002 and *E. coli*, an optimized *pepc* sequence for expression in *E.*

coli was synthesized commercially and cloned into the pET-15b vector by Geneart, Inc. Induction of this recombinant resulted in an intense band at the correct size (114 kDa). The protein, however, was insoluble (Figure 4), presumably due to formation of inclusion bodies at this very high expression level. Attempts to resolubilize the protein in a native state from these inclusion bodies have not yet been successful.

4. Conclusions

In summary, the gene encoding the phosphoenolpyruvate carboxylase of Synechococcus PCC 7002 was isolated using PCR techniques. In order to isolate the gene, a genomic library was constructed by cloning Sau3AI fragment of Synechococcus PCC 7002 genomic DNA into the BamHI-digested pGEM3Z vector.

The 1 kb RF2 riboprobe which is derived from the amplification product of a segment of the pepc gene of the cyanobacterium was generated. This probe was used in hybridization experiments with the genomic library, but this approach did not result in the isolation of the gene. In a different approach, plasmid DNA isolated from the genomic library was used to transform pepc- mutants E. coli strains MC2 (3) and EC342167 (6). Presumably the cells that carry the recombinant plasmid carrying the pepc nucleotide sequence would be able to grow in minimal medium with glucose as the only carbon source. This approach was also unsuccessful possibly due to the low specific activity of Synechococcus PCC 7002 phosphoenolpyruvate carboxylase compared to non-mutant host PEPC and to different in vivo regulatory requirements. Initial attempts at isolating the gene by PCR utilized primers based on the alignment of cyanobacterial pepc genes in the same manner as the amplification of the RF1 and RF2 amplicons were not successful. The target was the entire gene, and the consensus primers were to amplify a product of approximately 3 kb. PCR experiments with these primers failed, probably due to the lack of homology between all PEPCs at the N-terminal or the 5' end of the DNA sequences. These experiments yielded mostly bands that were less or equal to 2 kb. The successful strategy for the isolation of the phosphoenolpyruvate carboxylase gene of Synechococcus PCC 7002 was to utilize sequence information from the recently published incomplete genome of the cyanobacterium. The PCR experiments in this study effectively filled in existing gaps in the incomplete putative pepc gene from the unfinished genomic sequence while essentially using a single primer. The deduced amino acid sequence was more homologous to that of fresh water cyanobacteria than oceanic strains. For the first time, the identity of single amino acid substitutions that appear to distinguish PEPC from marine organisms and those from fresh water organisms have been revealed. The alignment of the deduced amino acid sequence of Synechococcus PCC 7002 PEPC and E. coli PEPC was further analyzed for the presence of amino acids involved in catalysis and regulation. In spite of the low sequence homology between PEPC from the two species, comparing the Synechococcus PCC 7002 sequence to the allosterically regulated E. coli PEPC revealed that the essential domains and amino acids involved in PEPC activity are shared by both proteins. The isolation and sequencing of the gene is an important first step that will permit obtaining a sufficient amount of protein for studies of the kinetics and mechanism of action and regulation of this enzyme. Furthermore, the elucidation of the sequence of this PEPC provides additional information that may allow for the design and synthesis of more appropriate probes for use in environmental studies of its involvement in CO₂ fixation.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman, DJ. *J Mol Biol* 1990;215(3):403-10.
2. Chen X, Widger WR. *J Bacteriol* 1993;175(16):5106-16.
3. Coomes, MW, Mitchell, B.K., Beezley, A., and Smith, T.E. *J Bacteriol* 1985;164:646-652.
4. Essich E, Stevens Jr. SE, Porter RD. *J Bacteriol* 1990;172(4):1916-22.
5. Gonzalez MC, Echevarria C, Vidal J, Cejudo FJ. *Plant Science* 2002;162:233-238.
6. Harrington TR, Glick BR, Lem NW. *Gene* 1986;45:113-116.
7. Huang X. *Genomics* 1992;14:18-25.
8. Ishijima S, Katagiri F, Kodaki T, Izui K, Katsuki H, Nishikawa K, Nakashima H, Ooi T. *Biochem Biophys Res Commun* 1985;133(2):436-41.
9. Kai Y, Matsumura H, Inoue T, Terada K, Nagara Y, Yoshinaga T, Kihara A, Tsumura K, Izui K. *Proc Natl Acad Sci U S A* 1999;96(3):823-8.
10. Kodaki T, Katagiri F, Asano M, Izui K, Katsuki H. *J Biochem (Tokyo)* 1985;97(2):533-9.
11. Lee CB, Hayashi H, Moon BY. *Mol Cells* 1997;7(2):296-9.
12. Luinenburg I, Coleman JR. *J Gen Microbiol* 1992;138(Pt 4):685-91.
13. Mazur BJ, Rice D, Haselkorn R. *Proc. Natl. Acad. Sci. USA* 1980;77(1):186-190.
14. McAlister LE, Evans EL, Smith TE. *J. Bacteriol* 1981;146(1):200-208.
15. Nakamura T, Yoshioka I, Takahashi M, Toh H, Izui K. *J. Biochem (Tokyo)* 1995;118(2):319-24.
16. Smith TE. *J Biol Chem* 1971;246(13):4234-41.
17. Smith TE, Balasubramanian KA, Beezley A. *J Biol Chem* 1980;255(4):1635-42.
18. Stevens SE Jr, Porter RD. *Proc. Natl. Acad. Sci. (USA)* 1980;77:6052-6056.
19. Takai K, Sako Y, Uchida A. *Microbiology* 1998;144(Pt 5):1423-34.
20. Thompson JD, Higgins DG, Gibson TJ. *Nucleic Acid Research* 1994;22:4673-80.
21. Yano M, Terada K, Umiji K, Izui K. *J. Biochem (Tokyo)* 1995;117:1196-1200.

Organism	Length (aa)	Accession Number	Percent Identity (%)
<u>Anabaena variabilis</u>	1026	P28594	57
<u>Anacystis nidulans</u>	1053	P06516	59
<u>Synechococcus</u> PCC 6803	1034	P74279	63
<u>Prochlorococcus marinus</u> CCMP 1375	1001	Q7V9U4	50
<u>Prochlorococcus marinus</u> CCMP1986	989	Q7U2TO	50
<u>Prochlorococcus marinus</u> MIT9313	1004	Q7V561	51
<u>Synechococcus</u> WH 8102	1010	NP898138	50

Table 1. Percent amino acid identity to Synechococcus PCC 7002 phosphoenolpyruvate carboxylase of other cyanobacterial PEPC sequences.

Amino acid from <i>Synechococcus</i> PCC 7002	Amino acid from fresh water strains	Amino acid from oceanic strains
Arginine 53	Arginine	Lysine
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
Glutamate 447	Glutamate	Histidine
Isoleucine 462	Isoleucine	Leucine
Phenylalanine 512	Phenylalanine	Tryptophan
Isoleucine 542	Isoleucine	Valine
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

Table 2. Conserved amino acid substitutions between fresh water and oceanic cyanobacterial PEPC sequences. The PEPC sequences used for the comparisons were from *Anabaena variabilis*, *Synechococcus* WH 8102, *Synechococcus* PCC 6803, *Anacystis nidulans*, *Prochlorococcus marinus* CCMP 1378, *Prochlorococcus marinus* CCMP 1375 and *Prochlorococcus marinus* MIT9313.

Figure Captions

Figure 1: Nucleotide sequence of Synechococcus PCC 7002 pepc and deduced primary structure. The gene consists of 2988 bp and encodes a polypeptide of 955 amino acids residues with a calculated molecular mass of 114,049 Da. The GC content of the Synechococcus PCC 7002 pepc coding region is 52%. (Genebank ID number AY830136).

Figure 2: 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.

Figure 3: Alignment of Synechococcus PCC 7002 PEPC and E. coli PEPC. The sequences were aligned with ClustalW and shaded with Boxshade. The deduced amino acid sequence of Synechococcus PCC 7002 phosphoenolpyruvate carboxylase and E. coli PEPC share 30% homology compared to 51-63% for cyanobacterial PEPCs. The amino acids that are involved in catalysis are highlighted in yellow boxes, and those that are involved in aspartate binding are in red boxes.

Figure 4: (A) SDS-PAGE experiment of induced BL21(DE3)pLysS cells containing the recombinant plasmid pSSPC1. The cells containing the E. coli-optimized pepc gene were induced with IPTG for three hours. Soluble fraction after treatment with 0.01M IPTG and 0.1M IPTG, respectively, is in lanes 1 and 2. Insoluble fraction after the same IPTG treatment is in lanes 3 and 4. The expected PEPC band at 114 kDa is present in the insoluble fraction.
(B) SDS-PAGE gel of induction the insoluble fraction of Synechococcus PCC 7002 PEPC in BL21 cells. Lane 1: + pET15-b, - IPTG; Lane 2: + pET15-b, + IPTG; Lane 3: + pSSPC1, - IPTG; Lane 4: + pSSPC1 + IPTG; Lane 5: + pSSPC1, - IPTG; Lane 6: + pSSPC1, + IPTG.
(C) Western blot analysis of expression of PEPC of Synechococcus PCC 7002. Lane 1: - plasmid, - IPTG; Lane 2: - plasmid, + IPTG; Lane 3: p15G1, - IPTG; Lane 4: p15G1, + IPTG; Lane 5: pCp30A1, - IPTG; Lane 6: pCp30A1, + IPTG; Lane C contains standard markers.
(D) Western blot analysis of BL21 cells containing pSSPC3 after removal of the His tag. The top bar and numbers represent increasing concentrations of IPTG.

ggatcctcaaccccggttccactgcactacccgatcgaacctaacccttacaat 56
 catcaagatcttcttcacagcagacagccatttcctagccccaaacgtctaagtacc 116
 atgaaccaagtcatgcaccccccaagtggcaagctgaactttgtccactccaaatct 176
 M N Q V M H P P S A E A E L L S T S Q S 20
 ctccctccggcaacgcctcacgctggtaagacatttggcaggcggtgtcagaaagaa 236
 L L R Q R I T L V E D I W Q A V L Q K E 40
 tgggtcaaaaactcgtaacgcgtcaatcatctccggcaacccgaaccggcgtggc 296
 C G Q K L V E R L N H L R A T R T A D G 60
 caaaggctaaactctccaagtagtattctgagctgatcgaacccctagatctagaa 356
 Q S L N F S P S S I S E L I E T L D L E 80
 gatgcgatccggcagcacgggctttgcctctattccagctgatcaatagcgtcag 416
 D A I R A A R A F A L Y F Q L I N S V E 100
 caacattacaacagcgcgaacaacaattcggcggaaacctcgccctcgccatgcc 476
 Q H Y E Q R E Q Q F R R N L A S A N A 120
 agcgaagccaacggtaacagcgccatacagaaattggcccccacccaagccggcac 536
 S E A N G N S V H T E I A P T Q A G T F 140
 gattggttttcccatctcaagcacaaaaatatgcgcggccagacgatccagcgtctt 596
 D W L F P H L K H Q N M P P Q T I Q R L 160
 ctcaatcagttggatattcgactggatttacggccacccacggagattgtccgcccc 656
 L N Q L D I R L V F T A H P T E I V R H 180
 acgattcgaataagcagcggcgattgcgggatccatcggcaactggatcaaaccgaa 716
 T I R N K Q R R I A G I L R Q L D Q T E 200
 gaagggctaaaaagctggggacaagcgattttggaaattgaaaatattcagcagcaa 776
 E G L K S L G T S D S W E I E N I Q Q Q 220
 ctaccgaagaaattcggctctgtggcgccacccatggagctgcaccaattaaaccc 836
 L T E E I R L W W R T D E L H Q F K P Q 240
 gtgttggacaagtggactatgcgttgcattattttgaggaagttctgtttgacacc 896
 V L D E V D Y A L H Y F E E V L F D T L 260
 cccgaattatcggtacggctccaaacaggccctaaaggcatcttcccaaccctcaagg 956
 P E L S V R L Q Q A L K A S F P T L K V 280
 cccaccaccaattttgtatatttggctctgggtcgccccgatcgcgtggcaaccct 1016
 P T T N F C N F G S W V G G D R D G N P 300
 tctgtcactcctgacgttacgttgcggatctggctctgggtcgccccgatcgcgtgg 1076
 S V T P D V T W K T A C Y Q R G L V L E 320
 cgctacatcgccctcggtggaaatccctctcgatgtgtgagcttacgcgtgcactgg 1136
 R Y I A S V E S L S D V L S L S L H W S 340
 aatgtgctggcgatctgtgcattcatcgatcggatcggatcggatctggatcggat 1196
 N V L P D L L H S S Y Q D Q N I F P D V 360
 tatcaaagcctggccagtcgtcaccgtcaagaaccctatcgatcgtcaagtgcgtac 1256
 Y Q S L A S R H R Q E P Y R L K S A Y I 380
 aagcggcgcttggaaaatacctggagcgcacccgtctgtggatcgatcggatcggat 1316
 K R R L E N T L E R N R R L A N M P A W 400
 gaaaataaaagtagaagcggccgatgacaaggttatatttgcggccaggagttctcg 1376
 E N K V E A A D D K V Y I C G Q E F L A 420
 gacctaaaactgatccgcggaaaggctggatcagacggaaatcaactgtgcggccctgg 1436
 D L K L I R E S L V Q T E I N C A A L D 440
 aaactcattgccaagtcgaaatttttagcttgcgtcactcgccctgtatcgccaa 1496
 K L I C Q V E I F S F V L T R L D F R Q 460
 gaatcaacgcgcactccgatgcgtccggaaattttgtgtggactacgtggatcgatc 1556
 E S T R H S D A I A E I V D Y L S I A Q 480
 atcctacaacgacctaagcgacgcagaaaaaccacccgtggatcgatcgtggatcgatc 1616
 I L Q R P K R R K N H L G L V Q E L K 500
 acccgccgtcccttgcggaaatgcactttcgaaagaaccgtcgagacgatc 1676
 T R R P L I P K E M H F S E R T V E T I 520

ccaacgctccagggtgctgcggcgtctcagcaggaatttggatcggcatttgcagacc 1736
 Q T L Q V L R R L Q Q E F G I G I C Q T 540
 tacatcatcagcatgaccaacgaagtcaagtatgtcttagaggtgctgcttggccag 1796
 Y I I S M T N E V S D V L E V L L L A Q 560
 gaagcggggtttagatgacccctcacagggatgaccaccatccgcattggccctttt 1856
 E A G L Y D P L T G M T T I R I A P L F 580
 gagacgggtggacacctgcgttaacgccccgaaatcatgcaggcggtttgaaattccc 1916
 E T V D D L R N A P E I M Q A L F E I P 600
 ctctatcgcgctttagctggctatgaaccccccgcgtggccgtgtatgaa 1976
 L Y R A C L A G G Y E P P A D G R C D E 620
 acctttggcgatgcgcgtgtccgaatctccagggaaattatgctcgcttccgacagc 2036
 T F G D R L V P N L Q E I M L G Y S D S 640
 aacaaagattccggcttcctcagtagtaattggggaaatccacaaagccagaaaaatctc 2096
 N K D S G F L S S N W E I H K A Q K N L 660
 caacaagtggccgatccctacggcattggcattttccatggcgggtggat 2156
 Q Q V A D P Y G I D L R I F H G R G G S 680
 gttggcggggcgccctgcctacggcgtatggcccaaccggccaaaccatc 2216
 V G R G G P A Y A A I L A Q P P N T I 700
 aatggccggattaaatcaccgaacagggggaaagtgtggcgtctaaatattccctccca 2276
 N G R I K I T E Q G E V L A S K Y S L P 720
 gacttggcgcctatcaccctcgaaagcggttctacggcgggtgattcaatcgagttgtta 2336
 D L A L Y H L E S V S T A V I Q S S L L 740
 gccagtgccgtttagatattcagccctggaaaccggattatggaaagacctctccagcga 2396
 A S G F D D I Q P W N R I M E D L S Q R 760
 tcgcggcggcttaccggcggttaattacgaagaacccgacttcctcgacttttcatg 2456
 S R A A Y R A L I Y E E P D F L D F F M 780
 tccgtgacggcattccggaaattagccaactgcaaattcagttccgtccggccgtcgc 2516
 S V T P I P E I S Q L Q I S S R P A R R 800
 aaaaaaggcaataaagatttgagtagtctccggcgattccctgggtcttagctggacc 2576
 K K G N K D L S S L R A I P W V F S W T 820
 caaagccgcttcctcgccgtggatgggtatgggtgggtaccggcgtccaggcgttttt 2636
 Q S R F L V P A W Y G V G T A L Q G F F 840
 gaagaggatcccgtcgagaacctgaagttatgcgcattttctacagtaatggccctt 2696
 E E D P V E N L K L M R Y F Y S K W P F 860
 ttccggatggatctcaaaggtagagatgacccttctaaagggtggacttacaaatggc 2756
 F R M V I S K V E M T L S K V D L Q M A 880
 agccactatgtccacgaactggcgaaaaagaagatattcccggttggaaaagcttcta 2816
 S H Y V H E L A E K E D I P R F E K L L 900
 gagcagatttcccgagaatacacaccaactcacaagggctaattccgtcaattaccgaaat 2876
 E Q I S Q E Y N L T K R L I L E I T E N 920
 gaagcccttcgtatggcgatggccctccagcgttctgtccaaactgagaaatggcacc 2936
 E A L L D G D R P L Q R S V Q L R N G T 940
 attgtccccctcggttctgcaagtctcattgctcaaacggctacgtcaatacaccagg 2996
 I V P L G F L Q V S L L K R L R Q Y T R 960
 gaaacccaagccagtatcggtcacttcgctacagcaaagaagaactctgcggggcgcc 3056
 E T Q A S I V H F R Y S K E E L L R G A 980
 cttttaaccatcaatggcatgcggccggaaatgcggaaacacgggttga 3104
 L L T I N G I A A G M R N T G - 995

Figure 1.

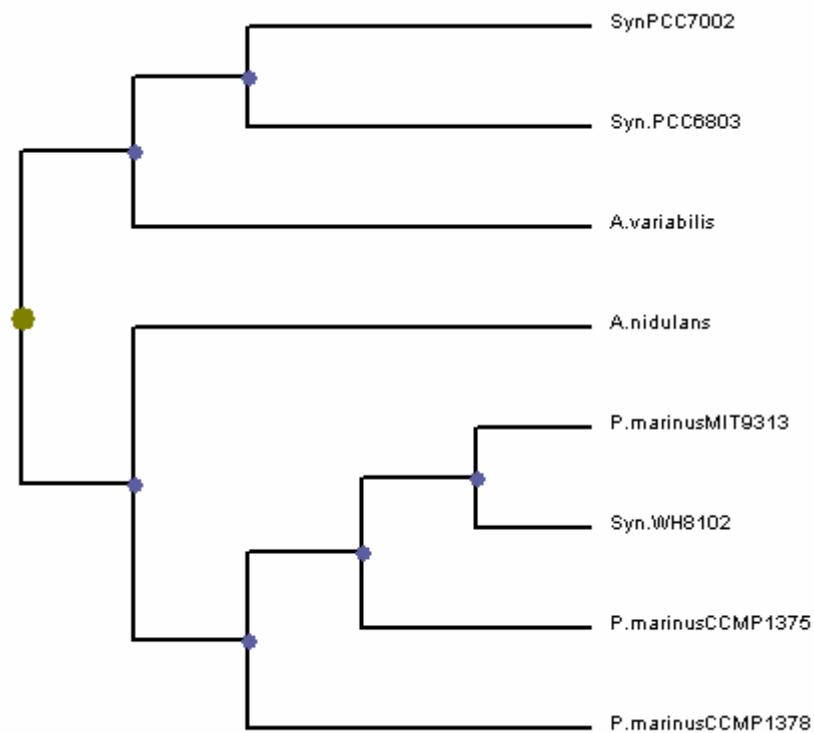


Figure 2.

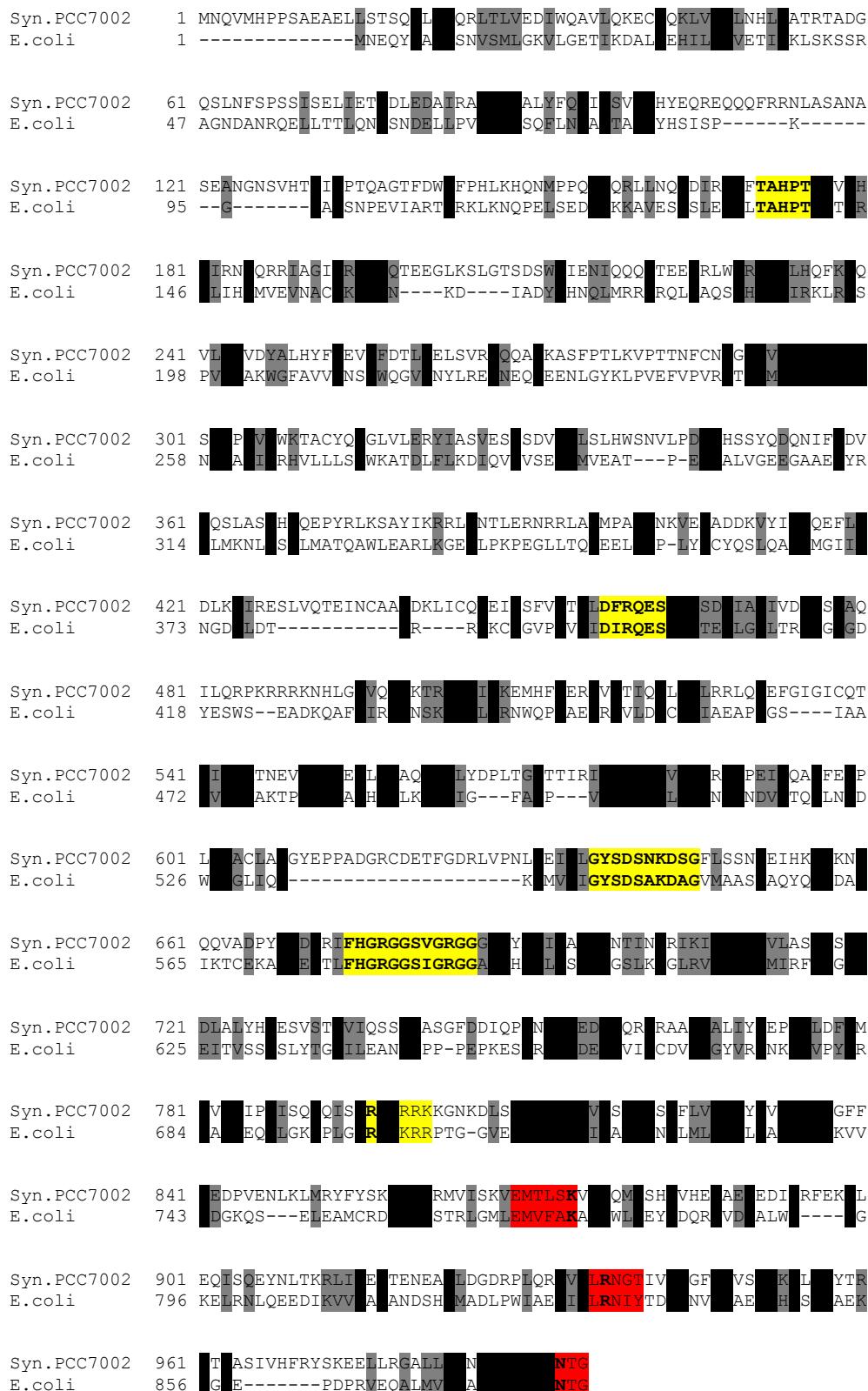
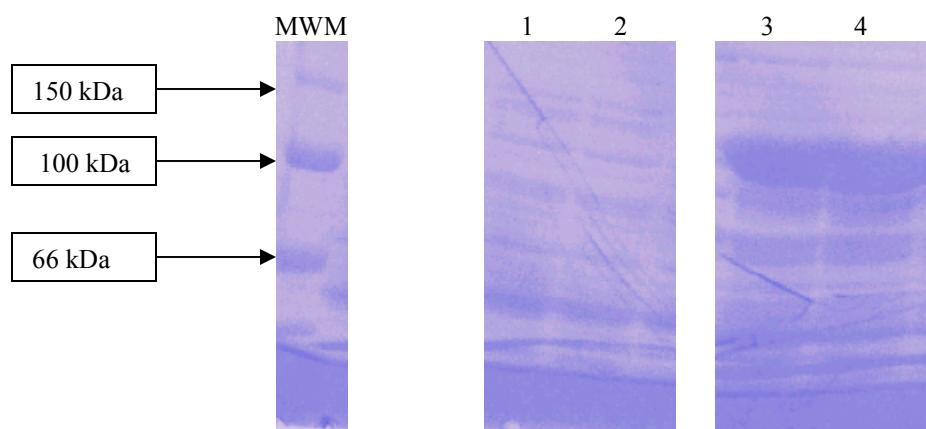
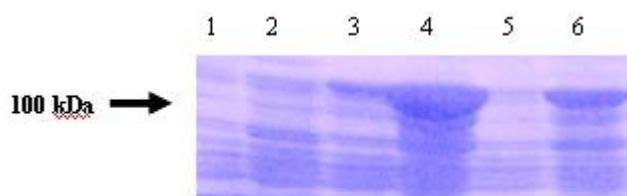


Figure 3.

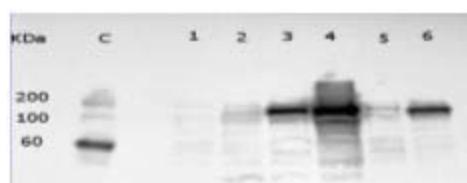
A.



B.



C.



D.

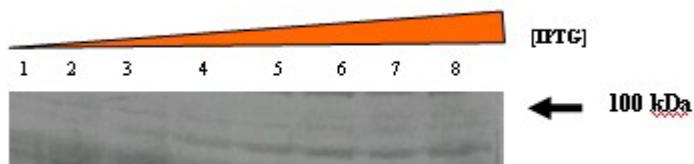


Figure 4.