Identification and Analysis of the Polyhydroxyalkanoate-Specific β-Ketothiolase and Acetoacetyl Coenzyme A Reductase Genes in the Cyanobacterium Synechocystis sp. Strain PCC6803

GASPAR TARONCHER-OLDENBURG, KOREN NISHINA, AND GREGORY STEPHANOPOULOS*
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received 28 April 2000/Accepted 19 July 2000

Synechocystis sp. strain PCC6803 possesses a polyhydroxyalkanoate (PHA)-specific β-ketothiolase encoded by phaA<sub>Syn</sub> and an acetoacetyl-coenzyme A (CoA) reductase encoded by phaB<sub>Syn</sub>. A similarity search of the entire Synechocystis genome sequence identified a cluster of two putative open reading frames (ORFs) for these genes, slr1993 and slr1994. Sequence analysis showed that the ORFs encode proteins having 409 and 240 amino acids, respectively. The two ORFs are colinear and most probably coexpressed, as revealed by sequence analysis of the promoter regions. Heterologous transformation of Escherichia coli with the two genes and the PHA synthase of Synechocystis resulted in accumulation of PHAs that accounted for up to 12.3% of the cell dry weight under high-glucose growth conditions. Targeted disruption of the above gene cluster in Synechocystis eliminated the accumulation of PHAs. ORFs slr1993 and slr1994 thus encode the PHA-specific β-ketothiolase and acetoacetyl-CoA reductase of Synechocystis and, together with the recently characterized PHA synthase genes in this organism (S. Hein, H. Tran, and A. Steinbüchel, Arch. Microbiol. 170:162–170, 1998), form the first complete PHA biosynthesis pathway known in cyanobacteria. Sequence alignment of all known short-chain-length PHA-specific acetoacetyl-CoA reductases also suggests an extended signature sequence, VTXGXXGIG, for this group of proteins. Phylogenetic analysis further places the origin of phaA<sub>Syn</sub> and phaB<sub>Syn</sub> in the γ subdivision of the division Proteobacteria.

Cyanobacteria produce polyhydroxyalkanoates (PHAs), a class of biodegradable polymers that are synthesized by many genera of eubacteria as well as some representatives of the archaebacteria (9). PHAs are carbon and energy storage compounds that are synthesized and deposited in the cytoplasm as insoluble inclusions. Production of PHAs by cyanobacteria for commercial purposes has attracted a great deal of attention lately, because, in contrast to other bacteria, cyanobacteria can obtain their precursors for production of PHAs from CO₂ assimilated through photosynthesis rather than more complex organic carbon sources (2).

The most well-studied type of PHA is poly-3-hydroxybutyrate (PHB) (Fig. 1). The presence of PHA inclusion bodies in cyanobacteria was first reported by Carr in 1966 following extraction of PHB from Chlorogloea fritschii (6). Since then, the occurrence of PHAs has been shown for several other species of cyanobacteria, including Gloeocapsa sp. (40), Spirulina platensis (3), Aphanothece sp. (5), Oscillatoria linosa (46), Anaebaena cylindrica (21), Synechococcus sp. (27), and Synechocystis sp. (15).

Biosynthesis of short-chain-length PHAs from the acyl-coenzyme A (CoA) precursors hydroxybutyl-CoA and hydroxvaleryl-CoA takes place in three steps, as exemplified in Fig. 1 for PHB. The first reaction consists of a Claisen type condensation of two molecules of acetyl-CoA to form acetoacetyl-CoA. This step is catalyzed by a β-ketothiolase (acetoacetyl-CoA thiolase; EC 2.3.1.9). Acetoacetyl-CoA is then reduced by an acetoacetyl-CoA reductase (EC 1.1.1.36), which yields d(-)-3-hydroxybutyryl-CoA; this is followed by a polymerization reaction catalyzed by a PHA synthase (no EC number).

The genes coding for the enzymes described above are referred to as phaA, phaB, and phaC (and phaE), respectively (47). phaE encodes the second subunit of the two-component, type III PHA synthases found in cyanobacteria, as well as other organisms (15).

The genes responsible for PHA biosynthesis were first identified, cloned, and characterized in Zooglaea ramigera (32) and Ralstonia eutropha (34, 44, 45). Since then, several more representatives of the three genes have been cloned in other organisms (13, 16, 23–25, 33, 35, 42, 43, 50, 52, 56). Recently, as a result of the availability of the full genome sequence of the cyanobacterium Synechocystis sp. strain PCC6803 (18), the first cyanobacterial PHA synthase was identified and characterized (15).

The two subunits of this two-component PHA synthase are encoded by two open reading frames (ORFs), phaE<sub>Syn</sub> (slr1829) and phaC<sub>Syn</sub> (slr1830), which are located contiguously and in the same orientation on the Synechocystis sp. strain PCC6803 chromosome. (The gene classification and nomenclature used throughout this article are in accordance with the gene classification and nomenclature of the Synechocystis sp. strain PCC6803 genome project [18; http://www.kazusa.or.jp/cyano/]; subscripts refer to the species to which the genes belong [37]).

The aims of the present study were identification and characterization of the genes coding for the enzymes responsible for the two other steps in the PHA biosynthetic pathway in Synechocystis sp. strain PCC6803, a PHA-specific β-ketothiolase and an acetoacetyl-CoA reductase. The availability of the entire genomic sequence of Synechocystis sp. strain PCC6803 facilitated the task of identifying putative ORFs for phaA<sub>Syn</sub>, the thiolase, and phaB<sub>Syn</sub>, the reductase.

**MATERIALS AND METHODS**

Organisms. Batch cultures of Synechocystis sp. strain PCC6803 were maintained at 30°C in BG<sub>11</sub> medium (39). Throughout the experiment, continuous
irradiance of ca. 250 μmol of photons m\(^{-2}\) s\(^{-1}\) was provided by cool white fluorescent bulbs. Nitrogen-starved cells (high-PHB-production conditions) were obtained by growing *Synechocystis* in BG\(_{11}\) (0) medium. Cultures were further supplemented with 10 mM acetate for increased PHB production. Cells were harvested by centrifugation at 3,200 x g for 10 min at 4°C. The pellet was deep frozen in liquid N\(_2\), freeze-thawed twice, and then resuspended in 1 ml of DNAzol (GIBCO-BRL). The suspensions were placed in an N\(_2\) cell disruption bomb, and the cells were ruptured by two consecutive cycles of nitrogen decompression at 2,000 lbm\(^{-2}\) (Parr Instrument Company, Moline, Ill.). Homogenates were processed by following the manufacturer’s protocol (DNAzol; GIBCO-BRL).

Isolation of DNA from *Synechocystis* cells. Cells (1 × 10\(^8\) to 3 × 10\(^8\) cells) in the exponential growth phase (optical density at 730 nm, 0.5 to 1) were harvested by centrifugation at 3,200 × g for 10 min at 4°C. The pellet was deep frozen in liquid N\(_2\), freeze-thawed twice, and then resuspended in 1 ml of DNAzol (GIBCO-BRL). The suspensions were placed in an N\(_2\) cell disruption bomb, and the cells were ruptured by two consecutive cycles of nitrogen decompression at 2,000 lbm\(^{-2}\) (Parr Instrument Company, Moline, Ill.). Homogenates were processed by following the manufacturer’s protocol (DNAzol; GIBCO-BRL).

**Plasmid construction and transformation of *E. coli*.** *E. coli* (Epicurian coli XL10-Gold; Stratagene, La Jolla, Calif.) was transformed with a plasmid containing ORFs slr1993, slr1829, and slr1830 (18; http://www.kazusa.or.jp/cyano/), corresponding to the *phaA* (putative), *phaB* (putative), *phaC*, and *phaSyn* genes, respectively (15). Plasmid pCRScript-PHASYN was constructed in two steps. First, the clusters containing ORFs slr1993 and slr1994 (2,545 bp) and ORFs slr1829 and slr1830 (2,746 bp) were amplified by PCR. The PCR conditions used were as follows: denaturation at 96°C for 5 min followed by 30 cycles of denaturation at 96°C for 30 s, annealing at 50°C for 1 min, and extension at 72°C for 12 min, with an additional extension step consisting of 7 min at 72°C after the last cycle. Each amplification mixture contained 10% dimethyl sulfoxide, 50 to 100 ng of genomic DNA, each primer at a concentration of 0.1 M, each of the four deoxynucleotide triphosphates at a concentration of 0.5 μM, 2.5 U of Pfu DNA polymerase (Stratagene), and the buffer provided by the manufacturer. The primers, *phaB* (F), *phaB* (R; SacI), *phaC* (F; EcoRI), and *phaC* (R) (Table 1), were designed to include a 200-bp region upstream of the first ORF of each cluster and an equivalent region downstream of the stop codon of the second ORF of each cluster in the amplification product (see Results for details). Two of the primers, *phaAB* (R; SacI) and *phaEC* (F; SacI), were designed to incorporate a SacI restriction site. Following digestion of the two PCR products with SacI, the resulting gel-extracted DNA fragments were ligated to form a single insert (5,255 bp) (T4 DNA ligase; New England Biolabs, Beverly, Mass.). This insert was then gel purified and ligated into the pCR-Script Amp SK(+) plasmid by following the instructions of the manufacturer (Stratagene). The resulting plasmid, pCRScript-PHASYN (8,216 bp) (Fig. 2), was transformed into *E. coli*. Positive colonies (ampicillin resistant) were transferred into liquid LB medium supplemented with 50 μg of ampicillin per ml and 1% glucose to enhance PHB biosynthesis.

**Linear conversion cassette synthesis and gene disruption in *Synechocystis* sp.** Gene disruption of ORFs slr1993 and slr1994 was attained as previously described (48). The upstream and downstream sequences flanking the putative *phaA*-syn gene cluster were amplified in two separate reactions by using primers LFHphaa(F), LFHphaa(R;Kan), LFHphaab(F;Kan), and LFHphaab(R) (Table 1). In both cases, the primer for the region closest to the *phaA*-syn gene cluster consisted of a *phaAB* locus-specific 3′ half (21 bp) and a Tn903 kan gene-specific 5′ half (25 bp) [LFHphaa(R;Kan) and LFHphaab(R)] (Table 1). This generated PCR fragments with short overlapping sequences homologous to the marker module (the PCR conditions were the same as those described above). The two PCR products were subsequently used for amplification of the Tn903 kanamycin resistance gene encoding an aminoglycoside 3′-phosphotransferase. The PCR conditions were the same as those described above except that no dimethyl sulfoxide was added to the reaction mixtures, plasmid DNA was used as the template (Ncol restriction digest of pEP2::Tn903 [26, 30]), an additional 5 μl of each of the two gel-purified PCR products was added to the mixture, and the two flanking primers [LFHphaa(F) and LFHphaab(R)] were added at a concentration of 0.1 μM. Approximately 2 to 3 μg of PCR product from the second reaction was used for transformation of *Synechocystis* cells. Cells were transformed by electroporation, and transformants were selected on BG\(_{11}\) agar plates (39) containing 5 mM NaHCO\(_3\) and 50 μg of kanamycin per ml, as previously described by Chiaramonte et al. (6). The transformation efficiencies ranged from 2,000 to 5,000 colonies per 10\(^8\) *Synechocystis* cells.

**PHB analysis.** Between 30 and 50 ml of culture in the stationary growth phase was collected by centrifugation (10 min, 3,200 × g, 4°C). The resulting pellet was washed once with distilled H\(_2\)O and dried overnight at 85°C. The dry pellets were boiled in 1 ml of concentrated H\(_2\)SO\(_4\) for 60 min, diluted with 4 ml of 0.014 M H\(_2\)SO\(_4\), and filtered through a polyvinylidene difluoride filter (Acrodisc LC13 PVDF; Pall Gelman Laboratory, Ann Arbor, Mich.). Samples were then diluted

---

**FIG. 1.** PHB biosynthetic pathway. Substrates and products are underlined; boldface indicates cofactors and side products.**

![PHB biosynthetic pathway diagram](http://aem.asm.org/)

---

**TABLE 1.** Primers used in this study

<table>
<thead>
<tr>
<th>Primer(^a)</th>
<th>DNA target(^b)</th>
<th>Sequence (5′→3′)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>phaAB</em> (F)</td>
<td>slr1993</td>
<td>ACACAATCTCAAGCGGCGCGTCGAAAGAGCGTTGG</td>
</tr>
<tr>
<td><em>phaAB</em> (R; SacI)</td>
<td>slr1994</td>
<td>TTGTGAAAATCGCGAGAATCTATTTGGCCTAACGCGG</td>
</tr>
<tr>
<td><em>phaBC</em> (EcoRI)</td>
<td>slr1829</td>
<td>ACCGACGACGCCATCCTGACGAGGGCGATGAG</td>
</tr>
<tr>
<td>LFHphaa (F)</td>
<td>slr1991</td>
<td>CACGATTGTGGATGACAGTTGCTGATAGCTCCCCTT</td>
</tr>
<tr>
<td>LFHphaa (R; Kan)</td>
<td>slr1992, kan</td>
<td>CGACGATCCGGCGAAATCTCGGTAGGATGAAATGTCGACAAAGCTCA</td>
</tr>
<tr>
<td>LFHphaab (F; Kan)</td>
<td>slr1906, kan</td>
<td>TTCTTCAAGGGAGGAATCGCGAGCATTGTTGACCCCTT</td>
</tr>
<tr>
<td>LFHphaab (R)</td>
<td>slr1906</td>
<td>TAAATGTATTCAGAGCAGTCTGTCCTGTCCTCAG</td>
</tr>
<tr>
<td>TCI (F)</td>
<td>kan</td>
<td>AAGCCATTTCCACGGAAGAAACAGTTCG</td>
</tr>
<tr>
<td>TCI (R)</td>
<td>slr1993</td>
<td>CCTACATAGCGTACAAAGGTAAACGT</td>
</tr>
<tr>
<td>TCI (II)</td>
<td>kan</td>
<td>ACGACTCCCCGGGAGCTTGGTCGACTCC</td>
</tr>
<tr>
<td>TCI (IV)</td>
<td>slr1994</td>
<td>CCGAGACATTTTCCTCCGTTTACAAAGCG</td>
</tr>
</tbody>
</table>

\(^a\) F, forward primer; R, reverse primer; Kan, homology in 5′ half of the primer to *kan* (aminoglycoside 3′-phosphotransferase); TC, transformation check.

\(^b\) ORF numbers are from Cyanobase (http://www.kazusa.or.jp/cyano/).

\(^c\) The underlined sequences are restriction sites; boldface type indicates homology to *kan*. 

---

Downloaded from http://aem.asm.org/ on September 6, 2017 by guest
contiguous to slr1993, the putative revealed the presence of one ORF, slr1994 (720 bp), that is tase-dehydrogenase functions. A close analysis of these ORFs these latter sequences encoded enzymes with putative reduc-
phaA
phaB
phaBCvin
 sequences of several possible ORFs with high similarity to tocol. Sequence analysis. Sequence similarity searches were performed by using the Basic Local Alignment Search Tool (BLAST) (1) to screen the entire genome with the
phaA
phaBSyn
phaASyn
nome with the
phaA
phaB
phaC
 strains (Fig. 3; Table 2). PCR analysis was done to determine the correct location of the insert (primers TCI and TCII [Table 1]) and the presence or absence of the
phaA-B
cluster in the genome (primers TCIII and TCIV [Table 1]). Synechocystis cells can contain from 6 to 10 copies of the genome (20, 55), and therefore we had to be careful that total gene replacement or disruption of all
phaA-B
ORFs present in the cell occurred. The absence of a PCR product in the amplification reaction with the
phaA-B
strains confirmed that all copies of the
phaA-B
cluster were totally replaced in the transformant (data not shown). Synechocystis
RESULTS
Identification of putative
phaA
and
phaB
genes in
Synechocystis
sp.
strain PCC6803. Possible candidate ORFs for the
phaA
and
phaB
genes were identified by performing a similarity search of the
Synechocystis
sp.
strain PCC6803 genome with the
phaA and
phaB sequences of
Chromatium
vinosum.
The search for
phaA resulted in one primary candidate, slr1993 (1,227 bp), and the search for
phaB generated a list of several possible ORFs with high similarity to
phaB
. All of these latter sequences encoded enzymes with putative reductase-dehydrogenase functions. A close analysis of these ORFs revealed the presence of one ORF, slr1994 (720 bp), that is contiguous to slr1993, the putative
phaA
ORF. This cluster was thus hypothesized to encode the putative
phaA
and
phaB
genes.
Heterologous expression of the putative
phaA
and
phaB
genes in
E. coli.
E. coli
is not capable of accumulating PHB as a storage compound (11). Production of PHB in
E. coli
requires heterologous expression of three genes, the
β-ketothio-
lase, acetoacetyl-CoA reductase, and PHA synthase genes involved in PHA biosynthesis (45). We constructed a plasmid containing all three genes (four ORFs) from
Synechocystis,
phaA
, phaB
, phaE
, and phaC
. Several recombinant strains of
E. coli
were cultivated as batch cultures, and PHB accumulation was confirmed by HPLC (Fig. 3).
E. coli
was grown in LB medium with or without 1% glucose to provide sufficient carbon for PHB synthesis (15). Table 2 shows that
E. coli
was able to accumulate PHB when it was transformed with all four
Synechocystis
pha ORFs and that this accumulation was differentially regulated depending on the carbon source provided. Growth rates were not affected by the transformation (data not shown).
Targeted disruption of the putative
phaA
and
phaB
genes in
Synechocystis
sp.
The putative
phaA
and
phaB
genes were inactivated in
Synechocystis
sp.
strain PCC6803 by targeted gene disruption through long flanking homology-PCR as previously described (48, 53).
Transformation was confirmed phenotypically by HPLC determination of PHB content and genotypically by PCR analysis of the
phaA-B
strains. HPLC analysis showed the presence of PHB in wild-type
Synechocystis
cells and the absence of PHB in the
phaA-B
strains (Fig. 3; Table 2). PCR analysis was done to determine the correct location of the insert (primers TCI and TCII [Table 1]) and the presence or absence of the
phaA-B
cluster in the genome (primers TCIII and TCIV [Table 1]). Synechocystis cells can contain from 6 to 10 copies of the genome (20, 55), and therefore we had to be careful that total gene replacement or disruption of all
phaA-B
ORFs present in the cell occurred. The absence of a PCR product in the amplification reaction with the
phaA-B
strains confirmed that all copies of the
phaA-B
cluster were totally replaced in the transformant (data not shown). Synechocystis
TABLE 2. Accumulation of PHB by *Synechocystis* sp. and *E. coli* grown in different media and with or without organic carbon supplementation.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Medium</th>
<th>PHB content (% of cell dry weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Synechocystis</em> sp.</td>
<td>BG11</td>
<td>0.26 ± 0.13</td>
</tr>
<tr>
<td></td>
<td>BG11(0)</td>
<td>3.3 ± 0.95</td>
</tr>
<tr>
<td></td>
<td>BG11 + 10 mM acetate</td>
<td>0.94 ± 0.26</td>
</tr>
<tr>
<td></td>
<td>BG11 + 10 mM acetate</td>
<td>13.2 ± 1.95</td>
</tr>
<tr>
<td><em>Synechocystis</em> sp. ΔphaA-B</td>
<td>BG11 + 10 mM acetate</td>
<td>ND</td>
</tr>
<tr>
<td><em>E. coli</em> (phaA-B-E-C&lt;sub&gt;Syn&lt;/sub&gt;)</td>
<td>LB</td>
<td>2.8 ± 0.43</td>
</tr>
<tr>
<td></td>
<td>LB + 1% glucose</td>
<td>12.3 ± 2.9</td>
</tr>
</tbody>
</table>

*BD<sub>1(0)</sub> is a modified version of BG<sub>1</sub> medium in which the concentration of NaNO<sub>3</sub> has been reduced to 1.765 mM (see Materials and Methods).

**Mean ± standard error.**

**ND, not detected.**

wild-type and ΔphaA-B strains showed similar growth rates under the same growth conditions (data not shown).

**Alignments.** ClustalW alignment of the translated putative pha<sub>ASyn</sub> ORF with eight other PHA-specific β-ketothiolases showed a high degree of homology among the sequences (data not shown). The pairwise percent identity values, percent similarity values, and GenBank accession numbers of the corresponding sequences are as follows: *Rhizobium meliloti*, 36%, 51%, and RMU17226, respectively; *Z. ramigera*, 36%, 50%, and P07097; *Paracoccus denitrificans*, 37%, 52%, and D49362; *Alcaligenes eutrophus*, 39%, 53%, and J04987; *Burkholderia sp.*, 41%, 55%, and AF153086; *Alcaligenes latus*, 40%, 53%, and ALU47026; *Allochromatium vinosum*, 40%, 55%, and P45369; *Thiocystis violacea*, 38%, 52%, and P45363; *Pseudomonas sp.* strain 61-3, 40%, 53%, and AB014757; and *Acinetobacter sp.* strain RA3849, 38%, 54%, and L37761. The lengths of the different sequences vary between 240 and 247 amino acids (data not shown).

Two maximum-likelihood trees were constructed with the alignments of the above thiolase and reductase sequences. For both enzymes, the *Synechocystis* sequence clustered together with the representatives of the γ subdivision of the division Proteobacteria (*γ-Proteobacteria*) (*Allochromatium [Chromatiurn]* vinosum, *T. violacea, Acinetobacter* sp. strain RA3849, and *Pseudomonas* sp. strain 61-3 for phaA and *A. [C.]* vinosum, *Acinetobacter* sp. strain RA3849, and *Pseudomonas* sp. strain 61-3 for phaB). Of the four representatives of the β-proteobacteria, three, *A. eutrophus, A. latus*, and *Burkholderia sp.*, clustered together in both trees, and one, *Z. ramigera*, persistently clustered with the representatives of the α-Proteobacteria, *P. denitrificans*, and *R. meliloti* (data not shown).

**DISCUSSION**

In this study we identified and characterized the PHA-specific β-ketothiolase and acetoacetyl-CoA reductase of the cyanobacterium *Synechocystis* sp. strain PCC6803. A directed similarity search of the entire genome of this organism resulted in identification of two strong candidate ORFs for the two genes in question, pha<sub>ASyn</sub> and pha<sub>B Syn</sub>. Heterologous expression of the two genes together with the PHA synthase genes of *Synechocystis, phaE<sub>Syn</sub>* and phaC<sub>Syn</sub>, resulted in PHA biosynthesis in *E. coli*. On the other hand, targeted disruption of the pha<sub>ASyn</sub> and pha<sub>B Syn</sub> genes in *Synechocystis* resulted in the loss of PHA-producing capacity in the organism. Below we discuss some of the genetic, structural, and evolutionary characteristics of the pha<sub>ASyn</sub> and pha<sub>B Syn</sub> genes.

**Identification of putative phaA and phaB genes in *Synechocystis* sp. strain PCC6803.** Recently, the type III, two-component PHA synthase of *Synechocystis* sp. strain PCC6803 was identified (15). The two ORFs encoding the two subunits of the synthase, pha<sub>E Syn</sub> and pha<sub>C Syn</sub> (slr1829 and slr1830, respectively), are located on the genome and are believed to constitute a single operon. The PHA synthase activity of these numbers of the corresponding sequences are as follows: *R. meliloti*, 40%, 53%, and RMU17226, respectively; *Z. ramigera*, 41%, 56%, and P23238; *P. denitrificans*, 42%, 56%, and P50204; *A. eutrophus*, 38%, 52%, and P14697; *Burkholderia sp.*, 38%, 51%, and AAF23366; *A. latus*, 38%, 51%, and AAD10276; *A. vinosum*, 38%, 53%, and A27012; *Pseudomonas* sp. strain 61-3, 40%, 53%, and BAA36196; and *Acinetobacter* sp. strain RA3849, 38%, 54%, and L37761. The lengths of the different sequences vary between 240 and 247 amino acids (data not shown).

FIG. 4. Alignment of the N-terminal sequences of several β-ketothiolases and the *Synechocystis* sp. strain PCC6803 translated pha<sub>4</sub> (ORF slr1993) product. The first 13 amino acids indicate that there may be posttranslational modification of the protein. The numbers in parentheses are the positions corrected for the first 13 amino acids.
two ORFs has been demonstrated by heterologous expression in E. coli (15), phenotypic complementation of a PHA-negative mutant of A. eutrophus (15), and targeted deletion in Synechocystis sp. (48). PHA synthase subunits have previously been shown to be cotranscribed in other bacteria, such as A. (C.) vinosum (24).

Up- and downstream analysis of neighboring ORFs revealed the absence of other genes related to PHA metabolism in the

FIG. 5. Nucleotide sequence of the Synechocystis sp. strain PCC6803 phaA-phaB gene cluster (ORFs slr1993 and slr1994; positions 1437310nt to 439522nt in the Synechocystis sp. strain PCC6803 genome [http://www.kazusa.or.jp/cyano/]). The underlined sequences are putative promoter recognition sites; boldface type indicates a putative SD sequence; the numbers in parentheses are the amino acid positions corrected for the first 13 amino acids (see Discussion); the residues enclosed in boxes are conserved catalytic or structural residues (see Discussion); and the cross-hatched bar identifies a newly proposed cofactor binding site consensus sequence in PHA-specific acetoacetyl-CoA reductases.
immediate vicinity of the *Synechocystis* PHA synthase genes (15, 37). Only three other type III PHA synthases have been characterized to date (those of *A. C. vinosum* [24], *T. violacea* [23], and *Thiocapsa pfennigii* [22]). In all three cases the other two biosynthetic genes, *phaA* and *phaB*, are clustered with the *phaE* and *phaC* genes. The primary annotation of the *Synechocystis* genome (18; http://www.kazusa.or.jp/cyano/) fails to identify any candidates for *phaA* or *phaB*. Based on this information, our approach consisted of first identifying possible candidate ORFs for the *phaASyn* and *phaBSyn* genes by a similarity search of the *Synechocystis* genome with the *phaA* and *phaB* sequences of *C. vinosum*. The search for *phaA* resulted in one primary candidate, slr1993, and the search for *phaB* generated a list of several possible ORFs with high levels of similarity to *phaBCvin*. All of these latter sequences encoded enzymes with putative reductase-dehydrogenase functions. A close analysis of these ORFs revealed the presence of one ORF, slr1994, that is contiguous to slr1993, the putative *phaASyn* ORF. The short intergenic region (101 nucleotides) between the two ORFs, the absence of any sequences showing significant similarity to the enterobacterial 70-like 235/210 promoter consensus sequence in the intergenic region upstream of slr1994, and the presence of *E. coli* 70-like 235 (TTGcCA) and 210 (cATAAT) promoter consensus elements (the lowercase c indicates a discrepancy compared with the 70 consensus region) and a Shine-Dalgarno (SD) sequence (AGGCGG) upstream of *phaA* provided evidence that slr1993 and slr1994 might constitute a single operon (Fig. 5). It was also intriguing to notice the identity of the three regulatory elements, the 35 and 10 promoter consensus sequences and the SD element, with the equivalent regions upstream of *phaESyn* (slr1829) (Fig. 6). This indicated that there could be parallel regulation of the three PHA synthesis genes in *Synechocystis*.

**Heterologous expression of *phaA-Bsyn* and *phaE-Csyn* in *E. coli*.** As mentioned previously, the PHA synthase activity of slr1829 and slr1830 was demonstrated through heterologous expression in *E. coli* of these two ORFs together with the *A. eutrophus phaAAe* and *phaBAe* genes coding for a ketothiolase and a reductase, respectively (15, 34). The *phaE-Csyn* cluster in that experiment was under exclusive control of its native promoter. The identity between the *E. coli* 70-like 235 and 210 promoter consensus elements and the SD sequence upstream of the *phaA-BSyn* cluster and the *phaE-CSyn* operon led us to believe that combined transformation of all four *Synechocystis* *pha* genes with their native promoter regions into *E. coli* could result in expression of the genes. The results corroborated this hypothesis (Fig. 3 and 6; Table 2).

**PHB accumulation in the transformed *E. coli* strains was analyzed under two different growth conditions (Table 2).** The PHB accumulation dynamics reported here are similar to those described previously when there was no addition or addition of an organic carbon source, glucose, to *E. coli* transformed with PHA biosynthesis genes (15, 44).

**Disruption of the slr1993-slr1994 gene cluster in *Synechocystis* sp.** Construction of ΔphaA-B strains resulted in a total loss of PHB accumulation capacity in *Synechocystis* (Fig. 3; Table 2). Growth of wild-type *Synechocystis* under different physiological conditions conducive to different levels of PHB accumulation results in a PHB content that accounts for up to...
approximately 13% of the cell dry weight (Table 2). Several transformed strains showed different degrees of gene leakage (data not shown). PCR analysis (see Materials and Methods) of these strains showed the presence of both the phaA-B cluster and kan in the genomic DNA (data not shown). These strains represent intermediate stages of full replacement of the phaA-B clusters in all copies of the *Synechocystis* genome. Addition of kanamycin and further subculturing ultimately resulted in a total loss of PHB accumulation capacity.

**β-Ketothiolase.** The slr1993 ORF shows a very high similarity to those of PHA-specific β-ketothiolases from other PHA-accumulating bacteria. The alignment of the N-terminal region of the *Synechocystis* sequence with those of nine other PHA-specific ketothiolases shows the presence of an “extra” 13 residues at the beginning of the derived *Synechocystis* thiolase amino acid sequence that do not have counterparts in any of the other thiolases (Fig. 4). This feature indicates very strongly that there may be N-terminal posttranslational modification of the protein (41). Close analysis of the translated sequence further reveals the presence of all catalytic amino acid residues conserved in this family of thiolases, as recently determined from the crystal structure of the *Z. ramigera* PHA-specific β-ketothiolase (29) (the numbers in parentheses are the positions corrected for the extra N-terminal 13-amino-acid sequence); the catalytic residues His365(352) (activation of Cys101(88)), Cys101(88) (covalent acyl-CoA intermediate formation), and Cys395(382) (substrate activation); the substrate binding residue Ser264(251); and the hydrogen-bonding network residues Gly21, Gly15, Gly19, and Gly21 (54); and the acetoacetyl-CoA reductase-specific binding fold motif GlyXXXGlyXGly (Gly15, Gly19, and Gly21) (54); and the acetoacetyl-CoA reductase-specific NADP binding motif GlyXXXGlyXGly (Thr14, Gly15, Gly16, and Gly19) (36). The last two segments, located at the turn between the first β-sheet strand and the α-helix in the N-terminal region of the protein, include two additional conserved residues in all the PHA-specific reductases aligned in this study: Val14 and Ile20 (the numbers are the residue positions in the *Synechocystis* sequence). This consensus motif can thus be redefined as ValThrGlyXXXGlyIleGly for the PHA-specific acetoacetyl-CoA reductases (Fig. 5).

**Acetoacetyl-CoA reductase.** The alignment of the deduced ORF slr1994 amino acid sequence with the sequences of 10 other short-chain-length PHA-specific acetoacetyl-CoA reductases, including the sequences of *Pseudomonas aeruginosa* RhlG, an NADPH-dependent β-ketoacyl reductase involved in rhamnolipid synthesis (4), and FabG, another *P. aeruginosa* NADPH-dependent β-ketoacyl reductase that provides precursors for medium-chain-length PHA biosynthesis when it is expressed in *E. coli* (38), shows a high degree of similarity among representatives of this subgroup of proteins, which are members of the large superfamilly of short-chain dehydrogenase/reductases (SDRs). SDRs are characterized by several conserved residues and motifs that are related to their function and structure (17). Some of these residues in the *Synechocystis* reductase have the following identities: catalytic residues Lys151, Tyr147, and Ser134 (14); the SDR-specific coenzyme binding motif GlyXXXGlyXGly (Gly15, Gly19, and Gly21) (54); and the acetoacetyl-CoA reductase-specific NADP binding moiety sequence ThrGlyXXXGly (Thr14, Gly15, Gly16, and Gly19) (36). The last two segments, located at the turn between the first β-sheet strand and the α-helix in the N-terminal region of the protein, include two additional conserved residues in all the PHA-specific reductases aligned in this study: Val14 and Ile20 (the numbers are the residue positions in the *Synechocystis* sequence). This consensus motif can thus be redefined as ValThrGlyXXXGlyIleGly for the PHA-specific acetoacetyl-CoA reductases (Fig. 5).

**Organization of PHA biosynthesis genes in *Synechocystis*.** In many of the bacteria analyzed to date, phaA, phaB, and phaC(E) form a single cluster in the genome (37). Only in some organisms, including *Z. ramigera*, *Aeromonas caviae*, *Methyllobacterium extorquens*, *Nocardia corallina*, *R. mellitii*, *Rhodococcus ruber*, *P. denitrificans*, *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*, and *Rhodobacter capsulatus* (most of
which are α-proteobacteria; the exceptions are Z. ramigera, a β-proteobacterium, A. caviae, a γ-proteobacterium, and N. corallina, a firmicute), do the genes not colocalize (37). As shown here, the corresponding genes in Synechocystis are also localized in different sections of the genome. Our analysis, together with previous work on pha-E (15), shows that the Synechocystis pha genes are most closely related to those of the γ-proteobacteria (see below). It is thus interesting to notice that the organization of the pha genes is similar to that encountered in α-proteobacteria and not in γ-proteobacteria, where the genes are always clustered. It is further intriguing that Z. ramigera also has its pha genes structured in a way similar to the way in which the genes of α-proteobacteria are structured. As shown in our phylogenetic analysis, the Z. ramigera phaA and phaB genes persistently cluster within the α-proteobacterial clade and not the β-proteobacterial group.

Possible coevolution of the phaA-B and phaE-C clusters? The sequence data and analysis presented here show that the Synechocystis PHA-specific thiolute and reductase are most similar to those found in representatives of the γ-Proteobacteria. It was previously shown that the Synechocystis type III PHA synthase also has the highest similarity to the heterodimers of three γ-proteobacteria, A. vinosum, T. violacea, and T. pfeunigii, and that type III PHA synthases occur only in γ-proteobacteria (15, 37). In these three organisms the PHA biosynthetic pathway genes are clustered together (37). In the case of Synechocystis the genes form two separate clusters.

It has been postulated that the PHA biosynthesis genes might have been acquired by Synechocystis through lateral gene transfer (15). Several examples of lateral gene transfer involving Synechocystis have been described, and the mechanism has been analyzed (7, 12, 57). The fact that ORFs slr1993, slr1994, and slr1830 are not clustered in Synechocystis indicates that the genes might have been acquired in separate events. Alternatively, two possible explanations can be invoked for this phenomenon. One is that separation of genes belonging to a cluster occurred during acquisition of foreign DNA. It was recently suggested, based on a comparative analysis of complete bacterial genomes, that Synechocystis has a tendency to shuffle gene locations following transfer, which results in a lower degree of operon-cluster structure than that found in other bacteria (51). The other possibility is the occurrence of internal transposition through insertion, a phenomenon that has also been described in Synechocystis (31). The present analysis thus suggests some intriguing questions regarding the evolution and plasticity of the PHA biosynthetic pathway and, on the basis of similarity, of other gene clusters coding for complete or partial biosynthetic pathways in cyanobacteria.

In this study we identified and characterized two of the genes involved in PHA biosynthesis in Synechocystis sp. strain PCC6803, phaASyn and phaBSyn, which code for a PHA-specific β-ketothiolase and acetoyethyl-CoA reductase, respectively. These two genes, together with the recently characterized PHA synthase genes of Synechocystis (15), form the first complete PHA biosynthesis pathway known in cyanobacteria. Detailed analysis of the sequences has revealed several peculiarities of the Synechocystis pha genes: their organization in two separate clusters on the genome, the almost identical −35, −10, and SD promoter sequences in the phaASyn−phaBSyn and phaESyn−phaCSyn clusters, and the possible γ-proteobacterial origin of these genes in Synechocystis. This is the first step in understanding PHA biosynthesis in cyanobacteria. With the imminent availability of more complete cyanobacterial genome sequences we will be able to gain better insight into the significance and evolution of this pathway in this group of organisms.


54. Yura, I., H. Toh, and M. Go. 1999. Putative mechanism of natural transformation as deduced from genome data. DNA Res. 6:75–82.