Genetic Tagging, Cloning, and DNA Sequence of the *Synechococcus* sp. Strain PCC 7942 Gene (*gnd*) Encoding 6-Phosphogluconate Dehydrogenase

SHELDON E. BROEDEL, JR., AND RICHARD E. WOLF, JR.*

Department of Biological Sciences, University of Maryland Baltimore County, Catonsville, Maryland 21228

Received 12 February 1990/Accepted 22 April 1990

A genetic approach was used for the cloning of the *Synechococcus* sp. strain PCC 7942 (*Synechococcus* strain R2) *gnd* gene which encodes 6-phosphogluconate dehydrogenase (6PGD). A restriction map of the *gnd* locus was prepared by Southern analysis using the *Escherichia coli* gene as a heterologous probe. The *Synechococcus* strain R2 gene was genetically tagged by restriction site-specific insertion of the *nptII* gene of Tn903 into a pUC19 plasmid library of *Synechococcus* strain R2 chromosomal DNA. *Synechococcus* strain R2 was transformed with this insertion mutation library, and isolates carrying the *gnd::nptII* gene were identified as mutants hypersensitive to incubation in the dark. The interrupted gene was cloned from one of the mutants. A plasmid carrying the *gnd::nptII* gene was reintroduced into *Synechococcus* strain R2, and kanamycin-resistant transformants were selected. Transformants arising by gene replacement were dark sensitive and missing 6PGD activity. Transformants arising by plasmid insertion were dark resistant and had 6PGD activity. The wild-type gene was then cloned from a transformant containing a plasmid insertion, making use of the restriction map derived from the interrupted gene. *Synechococcus* strain R2 6PGD was expressed in *E. coli* when the cloned *gnd* gene was transcribed from the *lacZ* promoter resident on the vector. The boundaries of the gene and the direction of transcription were determined from the phenotypes conferred by plasmids carrying deletions entering *gnd* from either end. The nucleotide sequence was determined. The deduced amino acid sequence of *Synechococcus* strain R2 6PGD has 56% homology to that of the *E. coli* K-12 enzyme.

Photosynthetic organisms encounter diurnal periods of darkness, an energy starvation condition. Therefore, mechanisms have evolved to provide them with energy needed for maintaining cellular functions during periods of light deprivation. Cyanobacteria store glycogen during periods of active photosynthesis and use the glycogen as a carbon and energy source during periods of light-energy limitation (17). The primary pathway for the dissimilation of this stored glycogen is the hexose monophosphate shunt (for a review, see reference 24). In nitrogen-fixing cyanobacteria, the hexose monophosphate shunt serves as a major source of reductant for nitrogen fixation (7, 11).

Despite the central role that the hexose monophosphate shunt plays in cyanobacterial metabolism, little is known about the mechanisms that regulate the level of enzyme activity. As a step toward characterizing these mechanisms, we have developed a method for genetically tagging and cloning the *Synechococcus* sp. strain PCC 7942 (formerly *Anacystis nidulans* R2 and here abbreviated as *Synechococcus* strain R2) *gnd* gene, which encodes 6-phosphogluconate dehydrogenase (6PGD), the third enzyme of the hexose monophosphate shunt. The level of this enzyme is higher in cells growing slowly in light- or CO₂-limited chemostats (15) and increases as exponentially growing cells make the transition into the stationary phase (S. E. Broedel, Jr., and R. E. Wolf, Jr., unpublished data), but it does not change during incubation in the dark (23).

Our approach took advantage of the genetic recombination properties of *Synechococcus* strain R2 (8, 9, 26) and the fact that strains missing 6PGD activity are hypersensitive to darkness, presumably because of the accumulation of 6-phosphogluconate (6). Thus, an insertion mutation library of

Synechococcus strain R2 chromosomal DNA was introduced into *Synechococcus* strain R2, and strains with *gnd* insertion mutations were identified as dark-sensitive (Drk^s) mutants. The disrupted gene was cloned and then used to identify the wild-type copy of the gene. The boundaries and nucleotide sequence of the cloned *gnd* gene were determined.

MATERIALS AND METHODS

Strains and growth conditions. *Synechococcus* sp. strain PCC 7942 was grown at 30°C with cool white fluorescent lights in liquid BG11 medium or on BG11 plates solidified with 1% agar (Difco Laboratories, Detroit, Mich.) (20). *Escherichia coli* JM109 [*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1 Δ(lac-proAB)* (F' *traD36 proAB lacI^qΔM15*); 30] was used for plasmid and M13 phage manipulations. *E. coli* RW231 [F[−] *trpR lacZ*(Am) *trpA9805*(Am) *kdgR rpsL125 recA1 Δ(edd-zwf)22 Δ(sbcB-his-gnd-rfb)*; 28] and gluconate bromthymol blue indicator plates were used for scoring phenotypic expression of the cloned cyanobacterial *gnd* gene. Gnd⁺ Edd[−] strains form green colonies on the indicator plates, while Gnd[−] Edd[−] strains form white colonies (29).

Plasmids and M13 phages. Plasmids pUC18 and pUC19 (30) were used as the vectors for cloning and deletion analysis. Plasmid pUC4K (Pharmacia, Inc., Piscataway N.J.) was the source of the *nptII* gene. M13 phages mp18 and mp19 (30) were used for subcloning the pGND9 deletions and an internal fragment of *gnd* for subsequent DNA sequencing. Phage KD14 is an M13 derivative carrying a portion of the *E. coli gnd* structural gene (19).

Recombinant DNA procedures. Restriction enzymes were purchased from New England BioLabs, Inc. (Beverly, Mass.), Promega Biotec (Madison, Wis.), or Boehringer

* Corresponding author.

Mannheim Biochemicals (Indianapolis, Ind.) and used according to the specifications of the supplier. When specific restriction fragments were used in plasmid constructions, the fragments were gel purified on 1% low-melting-point agarose (SeaPlaque; FMC Corp., Marine Colloids Div., Rockland, Maine) by the phenol-freeze-thaw method of Silhavy et al. (22). Ligation reactions were done in 20- μ l volumes with 100 to 500 ng of DNA and 400 U of T4 DNA ligase (New England BioLabs) and incubated at 16°C for 18 h.

Isolation of *Synechococcus* strain R2 chromosomal DNA. Cultures of *Synechococcus* strain R2 were grown to 10^8 cells per ml, and chromosomal DNA was purified by the cetyltrimethylammonium bromide protocol described by Ausubel et al. (2) with the following modifications. The cells were collected by centrifugation ($10,000 \times g$ for 15 min), washed once with 0.5 volume of $1 \times$ TE (10 mM Tris chloride, pH 8.0, 1 mM EDTA), and resuspended in 0.045 volume of TE. After the sodium dodecyl sulfate (SDS)-proteinase K treatment and the cetyltrimethylammonium bromide-chloroform extraction, the aqueous phase was collected and the DNA was precipitated by adding 2 volumes of ethanol. The DNA was suspended in 400 to 1,000 μ l of $1 \times$ TE containing 10 μ g of RNase A per ml. After incubation overnight at 4°C, the DNA was extracted with phenol/chloroform-isoamyl alcohol (1:1) and ethanol precipitated.

Southern hybridization. Chromosomal DNA was digested to completion with restriction enzymes. DNA fragments were separated by electrophoresis and blotted onto Gene-Screen (New England Nuclear Corp., Boston, Mass.) by capillary action by the method of Maniatis et al. (18). For heterologous hybridization with the *E. coli gnd* probe KD14, the filter was prehybridized in 15% formamide, 1 M NaCl, 10% dextran sulfate, 0.5% SDS, and 100 μ g of salmon sperm DNA per ml at 42°C for 16 h. Probe KD14, made radioactive by primer extension (13), was added at 5×10^6 cpm per hybridization reaction. Hybridization was for 24 h at 42°C. The filter was washed twice with $2 \times$ SSC ($1 \times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-0.1% SDS at room temperature for 5 min and then with $6 \times$ SSC-0.1% SDS at 65°C for 2.5 h with two changes of buffer. The filter was then washed once in $2 \times$ SSC-0.1% SDS at room temperature for 5 min and air dried. The filter was exposed to X-ray film (XAR-5; Eastman Kodak Co., Rochester, N.Y.) at -80°C. For homologous hybridization with nick-translated (18) pUC4K as the probe, the filters were treated as described above except that they were washed in $2 \times$ SSC-0.1% SDS at 65°C.

Transformation of *E. coli* and *Synechococcus* strain R2. *E. coli* JM109 and RW231 were transformed by a low-pH method as follows. Cultures were grown in 100 ml of LB broth at 37°C to an A_{580} of 0.3. The cultures were chilled on ice for 20 min. The cells were collected by centrifugation ($4,000 \times g$ for 10 min), washed once with 50 ml of cold 10 mM NaCl, and suspended in 8 ml of cold transformation buffer (12 mM HEPES [*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid], pH 6.0, 50 mM CaCl₂, 26 mM MgCl₂, 15% glycerol). Competent cells were stored frozen at -80°C in glass vials. For transformation, the cells were thawed on ice. DNA (0.1 to 1 μ g) in a maximum volume of 20 μ l was added to 0.2 ml of competent cells, and the mixture was allowed to incubate on ice for 30 min. The cells were heat shocked at 37°C for 6 min, after which 0.8 ml of LB was added and the cells were incubated at 37°C for 60 min. The cells were plated in 0.1- or 0.2-ml portions onto the appropriate medium for selection of transformants.

Transformation of *Synechococcus* strain R2 was essentially as described previously (10), except that the cells were washed in 0.5 volume of BG11 medium and resuspended in 0.1 volume of BG11. Transformation reactions contained 0.5 ml of $10 \times$ concentrated cells and 0.5 μ g of DNA (in a volume of less than 50 μ l). The mixture was incubated at 30°C in the dark for 16 to 18 h. Cells were plated onto minimal BG11 medium, and after incubation in the light for 6 h, 1 ml of the appropriate antibiotic was added beneath the agar. The final drug concentrations in the medium were 0.5 μ g of ampicillin per ml, 5 μ g of kanamycin per ml, and 5 μ g of chloramphenicol per ml.

Construction of plasmid libraries. *Hind*III-digested *Synechococcus* strain R2 chromosomal DNA was size fractionated on a sucrose gradient (18). Fractions containing fragment sizes of 4 to 15 kilobases (kb) were pooled and ligated with *Hind*III-digested dephosphorylated pUC19. The ligated DNA was used to transform strain JM109, and Ap^r colonies were selected. The library was amplified by adding 1 ml of LB broth to each plate, scraping the colonies off the plates, and pooling them into 25 ml of LB containing 50 μ g of ampicillin per ml. This culture was grown overnight at 37°C. Half the culture was stored at -80°C; plasmid DNA was purified from the remaining half.

The insertion mutation library was constructed by digesting the plasmids of the *Hind*III library described above with *Bgl*II. The digested DNA was ligated with a gel-purified 1.4-kb *Bam*HI fragment carrying the *nptII* gene of pUC4K. The ligated DNA was transformed into strain JM109, and Km^r Ap^r transformants were selected. The colonies were pooled, amplified, and stored as described above.

Identification of dark-sensitive mutants. DNA from the pooled insertion mutation library was used to transform *Synechococcus* strain R2, and Km^r colonies were selected. The transformants were scored for the dark-sensitive phenotype as follows. Colonies were patched onto BG11 minimal agar plates. One plate was incubated in the light for 5 days and served as the master plate; the second plate was wrapped with foil to exclude light, incubated for 2 days, and then returned to the light and incubated for 5 days. Patches incubated in the dark for 2 days that did not grow upon subsequent incubation in the light were scored as Drk^s. Putative Drk^s isolates were purified by streaking colonies on BG11 medium with kanamycin and rescored for the Drk^s phenotype.

Subcloning of *gnd*. Standard protocols for plasmid and phage DNA manipulations were followed (2, 18). pGND5 was constructed by deleting the 5.7-kb *Kpn*I-to-*Bst*EII fragment of pGND4. pGND7 was constructed by isolating and cloning the 4.2-kb *Hind*III fragment of pGND5 into pUC18. pGND8 was constructed by deleting the 0.6-kb *Bam*HI-to-*Bgl*II fragment of pGND7. pGND9 was constructed by subcloning the 2.9-kb *Hind*III-to-*Pst*I fragment of pGND7 into pUC19.

Unidirectional deletion mutagenesis. Unidirectional deletions were prepared with nucleases ExoIII and S1 by the method of Henikoff (12). Deletions entering *gnd* from the 5' end were made in pGND7, and deletions entering *gnd* from the 3' end were made in pGND9. With pGND9, in which *gnd* is in the same orientation as *lacZ* _{α} , some deletions extending into *gnd* from the 3' end produced in-frame *gnd-lacZ* _{α} gene fusions in strain JM109; sequencing across the fusion joints of these deletions helped define the *gnd* reading frame (3).

Nucleotide sequencing. DNA sequencing reactions were performed by the method of Sanger et al. (21) by using sequencing kits obtained from either Promega Biotec or

Pharmacia. Templates for the sequencing reactions were single-stranded M13 derivatives or double-stranded plasmids. The 6PGD amino acid sequences of *Synechococcus* strain R2 and *E. coli* were aligned by using the SEQNCE program of Delaney Software Ltd. and the Align program of BIONET.

Enzyme assays. Activities of 6PGD and glucose-6-phosphate dehydrogenase were assayed spectrophotometrically by measuring the rate of formation of NADPH. Sonic extracts were prepared from exponentially growing cells. The assays were performed as described previously (29), except that 10 mM glucose-6-phosphate was used in the reaction mixture for assays of glucose-6-phosphate dehydrogenase. The protein concentration in the extracts was determined by the method of Bradford (4).

RESULTS

Preliminary restriction map of *Synechococcus* strain R2 *gnd* locus. Restriction fragments carrying the *Synechococcus* strain R2 *gnd* gene were identified by Southern hybridization analysis of the *Synechococcus* strain R2 chromosome, by using the cloned *E. coli gnd* gene carried on phage KD14 as a heterologous hybridization probe. Positive hybridization signals were obtained for single 16.5-kb *Bam*HI and 4.2-kb *Hind*III fragments and for two *Bgl*II fragments of 4.6 and 2.7 kb (data not shown). The intensities of the hybridization signals of the two *Bgl*II fragments were approximately equal, indicating that the *Bgl*II site is in the middle of the region of homology between the *Synechococcus* strain R2 and *E. coli* sequences. Because KD14 contains only the coding sequence of the *E. coli* gene, not the flanking sequences, and because M13mp18 DNA does not hybridize with *Synechococcus* strain R2 chromosomal DNA (data not shown), we concluded that the *Bgl*II site is within the coding sequence of the *Synechococcus* strain R2 *gnd* gene and that *Synechococcus* strain R2 *gnd* could be cloned by an insertion mutagenesis scheme.

Genetic tagging of the *gnd* gene. Two classes of recombinants can be recovered when *Synechococcus* strain R2 is transformed with a nonreplicating plasmid carrying a different allele of a chromosomal gene (8, 9, 26). The predominant class of recombinants has gene replacements, which arise by either double crossovers or gene conversion (26). Addition recombinants appear at a lower frequency; they presumably arise by a single crossover that integrates the entire plasmid into the chromosome. Thus, if a target sequence were interrupted by a selectable marker, e.g., a drug resistance gene, and introduced into cells, the major class of transformants should carry insertion mutations. However, if the disrupted gene is essential, then most transformants would be addition recombinants.

Figure 1 shows the scheme used to construct the insertion mutation library. *Hind*III fragments (4 to 15 kb) of *Synechococcus* strain R2 chromosomal DNA were ligated with *Hind*III-digested pUC19, and the recombinant plasmids were propagated in strain JM109 as a plasmid library. The library was composed of 2,500 transformants harboring recombinant plasmids. Plasmid DNA was isolated from the library and digested with *Bgl*II. Since *Bgl*II does not cut in the vector, only plasmids with a *Bgl*II site(s) in the cloned chromosomal DNA, including the fragment carrying *gnd*, should have been linearized. These linear recombinant plasmids were then ligated with an *npt*II cartridge. Strain JM109 was transformed with the ligated DNA, and Ap^r Km^r transformants were selected. A total of 96 colonies were ob-

FIG. 1. Scheme for the construction of the insertion mutation library. The fate of a *Hind*III (H) fragment with an internal *Bgl*II (Bg) site is shown.

tained. These colonies represented a site-directed insertion mutation sublibrary, in which the *Synechococcus* strain R2 chromosomal DNA was interrupted by the *npt*II gene.

Plasmid DNA containing the library of insertion mutations was isolated and used to transform *Synechococcus* strain R2 to Km^r. Km^r transformants were obtained at a frequency of 10⁻⁵ transformants per CFU. A total of 1,000 of the Km^r transformants were screened for the Drk^s phenotype of Gnd⁻ mutants (6), as described in Materials and Methods. From the original 1,000 Km^r transformants, 3 Drk^s clones were isolated. These strains are referred to as DRK1, DRK10, and DRK28.

The specific activity of 6PGD was determined in sonic extracts of the three isolates. None of the three mutants had detectable 6PGD activity. All had glucose-6-phosphate dehydrogenase activity. Therefore, the Drk^s phenotype correlates with the absence of 6PGD activity.

The three Gnd⁻ mutants were screened for the presence of the *bla* gene carried on the pUC19 vector. DRK1 and DRK28 were Ap^s and thus were candidates for mutants that arose by gene replacement and segregation of the wild-type gene. DRK10 was Ap^r, indicating that it probably contains a copy of the vector.

The restriction map of the *gnd* locus of the chromosome of each mutant was determined by Southern hybridization. On the basis of the known sizes of the *Bgl*II fragments flanking the *npt*II insertion site (2.7 and 4.6 kb) and the size of the

FIG. 2. Scheme for the cloning of *Synechococcus* strain R2 *gnd*. Km^r Ap^r Drk^r addition recombinants were obtained after transformation of *Synechococcus* strain R2 with pGND1, as described in the text. The bracket below chromosome structure B shows the *gnd*-containing *KpnI* fragment that was cloned. The relative position of the leftward *KpnI* site in chromosome structure B was unknown at this stage of the cloning, as denoted by the dashed line. The position of *gnd* is shown by a bar below the chromosome. Restriction sites important to the cloning are shown. Bg, *BglII*; H, *HindIII*; K, *KpnI*. Open boxes indicate *Synechococcus* strain R2 chromosome; hatched shaded boxes indicate *nptII* inserted into the *BglII* site of *gnd*; closed boxes indicate pUC19.

nptII insert (1.4 kb), the predicted size of a *BglII* chromosome fragment that contains the disrupted *gnd* gene is 8.7 kb. The predicted size of the plasmid carrying the *gnd::nptII* fragment is 8.3 kb (a 4.2-kb *HindIII* fragment plus a 2.7-kb pUC19 plus a 1.4-kb *nptII* gene). Integration of this plasmid into the *Synechococcus* strain R2 chromosome at the *gnd* locus would give rise to *BglII* fragments of 11.0 or 12.5 kb, depending on which side of *nptII* the recombination event occurred. All three mutants had a *BglII* fragment of about 8.7 kb that hybridized with an *nptII* probe (data not shown). Therefore, all three mutants have a *gnd::nptII* mutation. However, other hybridization signals were obtained in the Southern analysis of the DRK1 and DRK10 chromosomes, indicating that these mutants had multiple chromosomal insertions of *nptII*. Thus, only DRK28 has a *gnd::nptII* mutation and no other copy of the drug resistance gene.

Genetic cloning of the *Synechococcus* strain R2 *gnd* gene. The disrupted *gnd* gene was cloned from strain DRK28 on the basis of its ability to confer Km^r in *E. coli*. Chromosomal DNA from DRK28 was digested with *BglII*, which cleaved outside of *gnd*, and ligated with *Bam*HI-digested pUC19. Strain JM109 was transformed with the ligated DNA, and Ap^r Km^r transformants were selected. A plasmid, pGND1, isolated from one of these transformants, carried an 8.4-kb fragment. Its *nptII* cartridge mapped 2.6 and 4.0 kb from the respective ends of the cloned DNA, closely matching the distance predicted by the Southern analysis.

To demonstrate that the disrupted *gnd* gene had been cloned, pGND1 was used to transform *Synechococcus* strain R2 to Km^r . Since pUC19 does not replicate in *Synechococcus* strain R2, Km^r transformants can arise only by recom-

bination between the disrupted *gnd* and the chromosome. Km^r transformants were obtained at a frequency of 2×10^{-5} transformants per CFU. A total of 16 Km^r colonies were purified; of the 16, 13 were Drk^s Ap^s and 3 were Drk^r Ap^r . The Drk^s Ap^s isolates had no detectable 6PGD activity and thus arose by gene replacement, while the Drk^r Ap^r isolates had wild-type levels of 6PGD activity. The Drk^r Ap^r transformants were formed by a single recombination event that integrated pGND1 into the chromosome, yielding a *gnd* duplication, with one copy being wild type and the other carrying the *nptII* insertion mutation. The ratio of gene replacements to addition recombinants was 4:1, which is less than the 100:1 ratio found by Williams and Szalay (26). Nevertheless, when pGND1 was introduced into *Synechococcus* strain R2 most Km^r transformants were Drk^s and missing 6PGD activity, thus confirming that the *gnd::nptII* gene had been cloned.

The wild-type *gnd* gene was cloned from the Km^r Ap^r Drk^r transformants described above by the scheme outlined in Fig. 2. Such addition recombinants can be formed by a crossover in either of two intervals (Fig. 2). If the crossover was in region I, to the left of *nptII*, then chromosome structure A would be formed. On the other hand, if the crossover was in region II, to the right of *nptII*, then chromosome structure B would be formed. Thus, the wild-type *gnd* gene could be cloned onto the adjacent pUC19 vector by digesting chromosomal DNA with the appropriate restriction enzyme and carrying out the ligation under conditions favoring intramolecular ligation. For example, if a given addition recombinant has chromosome structure B, then the intact gene could be cloned from a *KpnI* digest.

FIG. 3. Restriction maps of *gnd* plasmids. The maps are drawn to scale. The *Bgl*II site in *gnd* that was the target for *npt*II insertion is in boldface type. The position of *gnd* is shown by a bar under the maps. Arrows beneath *gnd*, *bla*, and *lacZp* show the direction of transcription. (A) Restriction maps of pGND4, pGND5, pGND7, pGND8, and pGND9. The size of the plasmid is listed below the plasmid name. The Gnd phenotype conferred by the plasmids is listed to the right of the plasmid name. pGND4 is the plasmid isolated by the method outlined in Fig. 2. (B) Detailed restriction map of *gnd* extending from the leftward *Hind*III site to the $\Delta 47$ fusion joint (see text for fusion construction). The 5' and 3' deletion endpoints used to define the boundaries of *gnd* are shown. The Gnd phenotype of each deletion is indicated below the map. The 3' deletions shown are all *lacZ*_g gene fusions either to *gnd* ($\Delta 44$, $\Delta 50$, and $\Delta 30$) or to the downstream ORF ($\Delta 27$ and $\Delta 47$). The closed bar shows the position of the *gnd* coding sequence. The hatched bar is the ORF coding sequence discussed in the text.

Note that the *gnd::npt*II gene could be cloned by *Kpn*I digestion of strains with structure A.

Chromosomal DNA from three *Km*^r *Ap*^r *Drk*^r addition recombinants was digested with *Kpn*I, ligation was carried out at a low DNA concentration, and *Ap*^r transformants of strain JM109 were selected. Two of the addition recombinants yielded *Km*^s transformants, whereas the transformants from the third were *Km*^r. A plasmid harbored by the *Ap*^r *Km*^s transformants should carry the wild-type *gnd* gene. Plasmid pGND4, isolated from one of these transformants, was mapped with restriction enzymes *Hind*III, *Bgl*II, *Kpn*I, and *Sal*I. The mapping showed that pGND4 contained *Synechococcus* strain R2 chromosomal DNA sequences with the restriction sites predicted by the restriction map of pGND1. Additional restriction mapping of pGND4 allowed

subcloning of smaller fragments carrying portions of the *gnd* locus (Fig. 3A).

Expression of *Synechococcus* strain R2 *gnd* in *E. coli* and determination of the boundaries of the gene. To confirm that pGND4 indeed carried an intact *gnd* gene and to define better the boundaries and orientation of the gene, complementation tests with pGND4 and subclones pGND5, pGND7, pGND8, and pGND9 were done (Fig. 3). The plasmids were introduced into *E. coli* RW231, and the Gnd phenotype was assessed on gluconate-bromthymol blue indicator plates as described in Materials and Methods. Strains carrying pGND7 and pGND9 were Gnd⁺ and had 6PGD activity in sonic extracts. Strains with plasmids pGND4, pGND5, and pGND8 were Gnd⁻ and lacked 6PGD activity. Thus, expression of *Synechococcus* strain R2 6PGD is

dependent on the orientation of *gnd* in the respective plasmid. Evidently, the *Synechococcus* strain R2 *gnd* promoter, if present, does not function in *E. coli*, so transcription depends on the *lac* promoter. Accordingly, the 6PGD open reading frame (ORF) is oriented left to right, as the maps of Fig. 3 are drawn.

The boundaries of the coding sequence were further defined by the properties of two sets of unidirectional deletions entering the gene from its respective ends. The 5' deletions extended from the leftward *Hind*III site of pGND7 toward the N-terminal coding sequence. The 3' deletions extended from the *Pst*I site of pGND9 and removed the C-terminal coding sequences. These deletions were identified by their *LacZ*_α-complementing activities in strain JM109, which resulted from the in-frame fusion of *gnd* and *lacZ*_α (3), as described in Materials and Methods. The two sets of nested deletions facilitated the DNA sequencing of the gene (see below). The deletion plasmids were isolated from JM109 and introduced into strain RW231. The Gnd phenotype conferred by the respective deletion plasmids is shown in Fig. 3B. Deletion Δ5 was the largest 5' deletion that was Gnd⁺. The next smaller deletion, Δ8, was Gnd⁻. Therefore, the 5' end of the *gnd*-coding sequence must be between the ends of these two deletions. Deletion Δ27 was the largest 3' deletion that was Gnd⁺. The next largest 3' deletion was Δ30, which was Gnd⁻. Therefore, the 3' end of the coding sequence must be between the endpoints of deletions Δ30 and Δ27. By locating the deletion endpoints on the restriction map, we could deduce that *Synechococcus* strain R2 *gnd* must be smaller than 1,620 base pairs (bp) (the distance between the Δ5 and Δ27 ends) and larger than 1,220 bp (the distance between the Δ8 and Δ30 ends), a size sufficient to encode 6PGD, which in *E. coli* has a molecular weight of 51,500 (19).

Nucleotide sequence of *gnd*. The *gnd* gene was sequenced by the chain termination method (21). Plasmid and M13 phage templates derived from subcloned *gnd* restriction fragments and deletion constructs were used. A detailed restriction map is shown in Fig. 3B. Figure 4 shows the sequence of *Synechococcus* strain R2 *gnd*. The nucleotide sequence extends from the leftward *Hind*III site to the Δ47 endpoint and spans 2,053 bp. More than half the sequence was determined for both strands of the DNA. The deletion plasmids, pGND9Δ44, -Δ50, -Δ30, -Δ26, and -Δ47, were isolated from *LacZ*⁺ transformants of strain JM109 as described in Materials and Methods. Thus, these plasmids carried gene fusions to *lacZ*_α. Hence, deletions Δ44, Δ50, and Δ30 should be *gnd-lacZ*_α gene fusions that define the *gnd* reading frame. The start of the structural gene was identified from the facts that there is only one AUG codon between the Gnd⁺ Δ5 and the Gnd⁻ Δ8 deletions and that an AGGA Shine-Dalgarno sequence is located 8 nucleotides upstream. The ORF starting at this AUG is 467 codons (1,401 bp), one codon shorter than the coding sequence for the *E. coli* gene (19). As expected, the fusion joints of Gnd⁻ *LacZ*⁺ fusions Δ44, Δ50, and Δ30 are in this reading frame. This ORF ends with a UAG stop codon that lies between the Δ30 and Δ27 deletion ends. Thus, the size of the ORF and the correlation between the position of sequence punctuations and genetically defined gene boundaries suggest that this ORF encodes the *Synechococcus* strain R2 6PGD polypeptide.

Properties of the *Synechococcus* strain R2 *gnd* gene. The polypeptide deduced from the DNA sequence has a molecular mass of 50,108 daltons, which is nearly identical to that of the *E. coli* K-12 enzyme (19). The deduced amino acid sequence of *Synechococcus* strain R2 6PGD was compared

with the sequence of the *E. coli* K-12 enzyme. A total of 56% of the residues matched exactly (Fig. 5), and overall, 74% of the sequences had chemically conserved amino acids. The segment important to 6-phosphogluconate binding (1) is present at residues 256 to 268 in the *Synechococcus* strain R2 polypeptide and at residues 253 to 264 in *E. coli* 6PGD. The homology to *E. coli* 6PGD and the presence of amino acid sequences for substrate binding strengthen our conclusion that the ORF shown in Fig. 4 is that of *Synechococcus* strain R2 6PGD. The degree of homology between the *Synechococcus* strain R2 and *E. coli* K-12 6PGD polypeptide sequences is consistent with the evolutionary distance between *E. coli* and *Synechococcus* sp. (27).

Tandeau de Marsac and Houmard (25) analyzed the codon usage pattern for six *Synechococcus* sp. strain PCC 6301 genes and found that codons with G and C residues in the third position are preferred. The *Synechococcus* strain R2 *gnd* gene follows the same pattern. Specifically, *Synechococcus* strain R2 *gnd* uses certain codons for a given amino acid more frequently than others, and generally, these codons have G or C in the third position. For example, the Leu codons CTG, TTG, and CTC are used 20, 11, and 12 times, respectively, while codons CTA, TTA, and CTT are used 4, 2, and 1 times, respectively. Moreover, a comparison of codon usage patterns for the *Synechococcus* strain R2 and *E. coli* K-12 *gnd* genes showed that several codons used often in the *E. coli* gene were rarely used in the *Synechococcus* strain R2 gene.

Two plasmids with 3' unidirectional deletions, Δ27 and Δ47, conferred *LacZ*_α-complementing activity and were Gnd⁺ (Fig. 3). Since these deletion endpoints are downstream of the *gnd* stop codon, they probably lie in a gene that is downstream of *gnd*. Indeed, nucleotide sequencing showed that an ORF lies downstream of *gnd* and that the Δ27 and Δ47 fusion joints are between the ORF and *lacZ*_α in the respective plasmids. This reading frame is different than that of the *gnd*-coding sequence. The ORF starts with an AUG that is preceded by a GGAG Shine-Dalgarno sequence spaced 6 nucleotides upstream. Therefore, another gene starts about 80 nucleotides downstream of *gnd*. A search of the bacterial nucleotide sequences in the GenBank data base (as of 4 September 1989) did not reveal any significant homology between this *Synechococcus* strain R2 ORF and any other identified bacterial gene.

DISCUSSION

In this paper, we describe a method for the cloning of cyanobacterial genes by genetic techniques and its application to the cloning and DNA sequencing of the *Synechococcus* strain R2 *gnd* gene. The approach takes advantage of several properties of unicellular cyanobacteria, which permit the genetic manipulation of the chromosome. Many unicellular strains of cyanobacteria are naturally competent for DNA uptake and thus can be readily transformed with appropriate vectors. *Synechococcus* strain R2 is one such strain, and for this reason it has become a widely used model organism for studies of cyanobacterial genetics. After transformation of unicellular cyanobacteria, circular DNA with homology to the chromosome can be inherited by double homologous recombination or at a lower frequency by a single crossover integration (26). Thus, by using the appropriate vector and markers, two types of recombinants can be distinguished genetically, gene replacements and addition recombinants (8, 9).

Our technique involves the construction of a site-directed

```

-200      -190      -180      -170      Δ5      -150
A AGCTTGCCCC CGTCATCAGG GCCTGACCGG TCGGATCGCG TCAAGAAGAA TCCACCCAG
-140      -130      -120      -110      -100      -90      -80
GCATTCGAATT TTTCTCTGAT CTGTAAATAA ATCCAGCGAT CGCGCGTGAT CCCCGCAAAG AAAGTCTAAA
-70      -60      -50      -40      -30      -20      -10
GAGGTTGCCA GAGAACCCCA AATTGGCAGG CTAACCTGAG GTCATTCTGT GTCCTAGCAG GAATTTCTCT
15      30      45      60
ATG GCA CTG CAA CAG TTC GGC CTC ATC GGG CTG GCA GTG ATG GGT GAG AAT CTC GCC CTG
75      90      105      120
AAC ATT GAG CGC AAC GGT TTT TCC TTG ACG GTC TAT AAC CGC ACG GCC GAA AAG ACT GAA
135      150      165      180
GCG TTC ATG GCC GAT CGC GCC CAA GGC AAG AAC ATT GTG CCG GCT TAC AGT CTG GAA GAC
195      210      225      240
TTT GTT GCC AGC TTG GAA CGT CCG CGC CGC ATT TTG GTG ATG GTT AAA GCG GGC GGC CCG
255      270      285      300
GTG GAT GCC GTG GTC GAG CAG CTC AAA CCC CTG CTA GAT CCC GGT GAC TTG ATC ATC GAT
315      330      345      360
GGG GGC AAC TCA CTG TTT ACC GAC ACC GAG CGC CGC GTC AAA GAT CTA GAA GCG CTA GGT
375      390      405      420
CTG GGT TTC ATG GGC ATG GGC GTC AGT GGT GGC GAA GAG GGG GCT TTG AAT GGC CCC AGT
444      450      465      480
TTG ATG CCG GGG GGC ACC CAA GCG GCC TAC GAA GCC GTG AGC CGA TCT GTG CCG ACG ATT
495      510      525      540
GCT GCC CAA GTC GAC GAT GGC CCC TGC GTT ACC TAC ATC GGT CCT GGT GGC TCT GGC CAC
555      570      585      600
TAC GTC AAG ATG GTG CAC AAC GGC ATT GAA TAT GGC GAT ATG CAG CTG ATC GCC GAA GCC
615      630      645      660
TAT GAC CTC CTC AAA TCG GTG GCA GGC CTC AAT GCC AGT GAA CTG CAT GAT GTC TTC GCC
675      690      705      720
GCC TGG AAC AAA ACG CCG GAA CTC GAC TCG TTC TTG ATT GAA ATC ACG GCG GAT ATT TTC
735      750      765      780
ACC AAA GTC GAT GAC TTG GGT ACT GGT CAG CCC TTG GTC GAG CTG ATT TTA GAT GCC GCT
795      810      825      840
GGT CAA AAA GGA ACC GGT CGC TGG ACG GTG GAA ACG GCA CTG GAA ATT GGC GTT GCG ATT
855      870      885      900
CCA ACC ATC ATT GCT GTC AAC GCC CGC ATC CTG TCC TCG ATC AAA GCC GAG CGT CAG
915      930      945      960
GCA GCT TCT GAA ATT CTT TCA GGA CCG ATT ACC GAG CCC TTC AGT GGC GAT CGC CAA GCC
975      990      1005      1020
TTT ATC GAC AGT GTG CGC GAT GCG CTC TAC TGC TCG AAA ATT TGC TCC TAT GCC CAA GGC
1035      1050      1065      1080
ATG GCG CTG CTA GCC AAA GCC TCC CAG GTC TAC AAC TAC GGT CTG AAT TTA GGT GAA CTG
1095      1110      1125      1140
GCG CGG ATT TGG AAA GGC GGC TGC ATC ATT CGG GCG GGT TTC CTC AAC AAG ATT AAG CAG
1155      1170      1185      1200
GCC TAT GAT GCT GAC CCA ACG CTG GCG AAT CTG TTG CTG GCA CCC GAA TTC CGG CAG ACG
1215      1230      1245      1260
ATT CTC GAT CGC CAG TTG GCT TGG CGG CGC GTG ATT GCG ATC GCG GCA GAA CGA GGA ATT
1275      1290      1305      1320
CCG GTT CCG GCT TTC AGT GCT TCC CTG GAT TAC TTC GAC AGC TAC CGC GCA TCG CCT GCC
1335      1350      1365      1380
CAA AAC CTC ACC CAG GCT CAG CGC ACT ACT TGC GCT CAT ACC TAC AAG GCA CCG ATC GCT
1395      1410      1420      1430
CTG CTC TTC CAT GCC ATG TTC TAG GCGACA AACTCCGGTC TCGGCTGGTC AGACGGACCA
1450      1460      1470      1480      1484      1488      1492
ACTCCTTTGG CAGAATCAGG ACTGGCAAGC GATCGGAGAG CGTG ATG GAT GTG TTG GTG
1514      1529      1544      1559
GTG CGC GGC ATT CGC GCT ACG GCT ACA CGG GGT TCT TTG ATG CCG AGC AAG AAC TGG GCC
1574      1589      1604      1619
AGT GGT TTG AGG TGG ACT TGC GCG TCT GGG TCG ATC TTC AGC CTG CTG GGG ACA GCG ATC
1634      1649      1664      1679
AGC TCG GCG ACA CCC TCG ACT ACA GCC AAG CGG TAT AGG TCG TGC AAA CGG GGA TTC GCC
1694      1709      1724      1739
AGA GTC GCT TTC GCA CGA TCG AGC GAT TGC GCT GAT CTC TCG ATC GCC TGT TTC AGC AGT
1754      1769      1784      1799
TTC CAG CCC TAC CGA ATT GAA CTG ACG TTG CGG AAA CTG GCC CCT CCC ATT CCT GAC TTT
1814      1829      1844      1850
AGT GTG ATG TCA GCT TGG TTC TGC GCG ATC GCG AAG TTA ACG GGT CGGATC CG

```

FIG. 4. DNA sequence of *Synechococcus* strain R2 *gnd*. The DNA sequence of 2,053 bp of the *gnd* region is shown. The sequence begins at the leftward *Hind*III site and extends to the Δ47 fusion joint (Fig. 3). The coding sequence begins at +1 and extends to +1401. Deletion endpoints that are shown in Fig. 3 are labeled above the sequence at their respective locations. The beginning of the downstream ORF discussed in the text is at +1485. The putative Shine-Dalgarno sequences and AUG translation start sites for 6PGD and the ORF are underlined.

insertion mutation library which contains recombinant chromosomal DNA interrupted by a drug resistance cassette. The insertion mutation library is used to identify transformants of *Synechococcus* strain R2 harboring the insertionally inactivated gene of interest by first selecting for the presence of a drug resistance marker and then scoring for specific mutant phenotypes. Since most transformants arise by gene replacements, insertionally inactivated genes can be readily obtained. The method is limited to nonessential genes that have an assayable phenotype associated with the null mutation. Nonetheless, a large number of genes can be genetically tagged by this technique. By the appropriate choice of a restriction enzyme, the wild-type gene can be

readily cloned onto the adjacent *E. coli* vector from the chromosome of an addition recombinant that had been isolated by integration of the cloned disrupted gene into the wild-type chromosome.

A different insertion mutation scheme for genetically tagging *Synechocystis* sp. strain PCC 6803 genes has recently been reported (5). This method involves the ligation of restriction fragments of total chromosomal DNA with a drug resistance marker followed by selection for transformants which harbor the recombinant products integrated into the chromosome. Mutants defective in photosystem II were identified among the transformants. One limitation of this method is that the selectable marker must be ligated to two

- 115:369–376.
16. Labarre, J., F. Chauvat, and P. Thuriaux. 1989. Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. *J. Bacteriol.* 171:3449–3457.
 17. Lehmann, M., and G. Wober. 1976. Accumulation, mobilization and turn-over of glycogen in the blue-green bacterium *Anacystis nidulans*. *Arch. Microbiol.* 111:93–97.
 18. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 19. Nasoff, M. S., H. V. Baker II, and R. E. Wolf, Jr. 1984. DNA sequence of the *Escherichia coli* gene, *gnd*, for 6-phosphogluconate dehydrogenase. *Gene* 27:253–264.
 20. Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* 111:1–61.
 21. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463–5467.
 22. Silhavy, T. J., M. L. Berman, and L. W. Enquist. 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 23. Singer, R. A., and W. F. Doolittle. 1975. Control of gene expression in blue-green algae. *Nature (London)* 253:650–651.
 24. Smith, A. J. 1982. Mode of cyanobacterial carbon metabolism, p. 47–86. In N. G. Carr and B. A. Whitton (ed.), *The biology of cyanobacteria*. University of California Press, Los Angeles.
 25. Tandeau de Marsac, N., and J. Houmard. 1987. Advances in cyanobacterial molecular genetics, p. 251–302. In P. Fay and C. Van Baalen (ed.), *The cyanobacteria*. Elsevier Science Publishers, B.V., New York.
 26. Williams, J. G. K., and A. A. Szalay. 1983. Stable integration of foreign DNA into the chromosome of the cyanobacterium *Synechococcus* R2. *Gene* 24:37–51.
 27. Woese, C. R. 1987. Bacterial evolution. *Microbiol. Rev.* 51: 221–271.
 28. Wolf, R. E., Jr. 1980. Integration of specialized transducing bacteriophage λ c1857 *St68 h80 dgnd his* by an unusual pathway promotes formation of deletions and generates a new translocatable element. *J. Bacteriol.* 142:588–602.
 29. Wolf, R. E., Jr., and F. M. Shea. 1979. Combined use of strain construction and affinity chromatography in the rapid, high-yield purification of 6-phosphogluconate dehydrogenase from *Escherichia coli*. *J. Bacteriol.* 138:171–175.
 30. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33:103–119.