Plant Physiol. (1992) 99, 359–361
0032-0889/92/99/0359/03/$01.00/0

Received for publication January 22, 1992
Accepted February 10, 1992

Plant Gene Register

Molecular Cloning and Sequencing of ADP-Glucose Pyrophosphorylase from *Synechocystis* PCC 6803[1]

Genichi Kakefuda[2], Yee-yung Charng, Alberto A. Iglesias, Lee McIntosh, and Jack Preiss*

Department of Biochemistry (G.K., Y.-y.C., A.A.I., L.M., J.P.), and Michigan State University-Department of Energy Plant Research Laboratory (L.M.), Michigan State University, East Lansing, Michigan 48824-1319

ADPGlc PPase[3] is the regulatory enzyme for synthesis of starch in plants and glycogen in bacteria (8, 9). Previous work on cyanobacterial ADPGlc PPase has shown the enzyme to have intermediate characteristics to that of the higher plant and bacterial enzymes (4). ADPGlc PPase from *Synechocystis* PCC 6803 is allosterically activated by 3-phosphoglycerate and inhibited by Pi, as are the higher plant enzymes. The homotetrameric structure of *Synechocystis* ADPGlc PPase is similar to the enteric bacterial enzymes, which is in contrast with the heterotetrameric nature of all higher plant enzymes studied. Here we report the nucleotide sequence of ADPGlc PPase from *Synechocystis* PCC 6803. The *Synechocystis* clone was isolated from a *Synechocystis* PCC 6803 genomic DNA library. The probe used for screening the library was derived from PCR amplification of genomic *Synechocystis* DNA.

Amino acid sequences that were highly conserved in both higher plant and bacterial ADPGlc PPase sequences were used to design degenerate primers for PCR amplification of cyanobacterial DNA (Table I). Primer 1, which had a degeneracy of 512, was designed from the conserved amino acid sequences of the *Escherichia coli* ADPGlc PPase FBP activator binding site. The activator binding site determined for the *E. coli* enzyme is conserved in higher plants (7, 10). The conservation occurs despite the fact that FBP does not activate higher plant ADPGlc PPases. Primer 2, which had a degeneracy of 256, was designed from the conserved amino acid sequences (10) of the 8-azido-ADP-glucose affinity labeling site previously determined in the *E. coli* enzyme (5). PCR amplification of genomic *Synechocystis* DNA with these primers generated a fragment of expected size. This fragment was used to isolate a clone from a genomic library.

The nucleotide and deduced amino acid sequence of *Synechocystis* ADPGlc PPase is shown in Figure 1. The first 39 N-terminal amino acids from the deduced sequence are in identity with the sequence determined by N-terminal sequencing of the purified protein from *Synechocystis* (our unpub-

[1] Supported by grants from the National Science Foundation (DMB 86-10319), National Institutes of Health (AI22835), and U.S. Department of Agriculture/Department of Energy/National Science Foundation Plant Science Center Program (88-37271-3964).

[2] Present address: American Cyanamid Company, Plant Biotechnology, P.O. Box 400, Princeton, New Jersey 08543-0400.

[3] Abbreviations: ADPGlc PPase, ADP-glucose pyrophosphorylase; PCR, polymerase chain reaction; FBP, fructose 1,6-bisphosphate.

Table I. *Characteristics of ADP-Glucose Pyrophosphorylase Genomic DNA from Synechocystis PCC 6803*

Organism:
Synechocystis sp. strain PCC 6803.

Gene Product, Pathway:
ADP-glucose pyrophosphorylase (EC 2.7.7.27); glycogen biosynthesis.

Techniques:
PCR:
PCR primers: primer 1, 5′ GAAGCGNGCNAAXCCNGCNGT 3′
primer 2, 5′ ATCAGCNGTNCCZZGAXYCCA 3′
N = A + G + T + C, X = A + G, Y = A + T, Z = T + C.
A genomic library constructed in lambda fix II (Stratagene) was screened with a radiolabeled probe utilizing the PCR-amplified fragment as template. Other techniques included: restriction enzyme and unidirectional deletion subcloning, complete dideoxy sequencing of both strands, computer analysis, comparison, and management of sequences data (3).

Method of Identification:
Sequencing of the N-terminal 39 amino acids of ADPG1c PPase purified from *Synechocystis* PCC 6803 is identical to that deduced from the nucleotide sequence. Sequence similarity of deduced amino acids to that of ADPG1c PPases from spinach leaf 51 kD (65%), rice endosperm (63%), and *E. coli* (37%).

Features of Gene Structure:
The start codon is GTG; a Shine-Delgarno sequence located 7 bases upstream of the start codon (Fig. 1, shaded); a sequence with homology to the *E. coli* −35 and −10 box (Fig. 1) is observed.

Codon Usage:
Codons not present: TCG, TAA, TGA, CCA, CCG, ACA. (G + C) Content:
48.6% in the coding region.

Structural Features of Protein:
Open reading frame 429 amino acids.
Calculated M_r 48,180.
Amino acid sequences similarity was found to: the *E. coli* FBP allosteric activator site (Fig. 1, amino acids 9–32) (7); the *E. coli* 8-azido-ADP-glucose affinity labeling site (Fig. 1, amino acids 95–107) (5); the spinach leaf 51 kD 3-PGA binding site (Fig. 1, amino acids 412–429) (6).

Antibodies:
Not available. Cross-reaction occurs with antibodies against either the 51 kD subunit, 54 kD subunit, or the holoenzyme of spinach leaf ADPG1c PPase (4).

GenBank Accession No: M83556.

```
   1  ATCATACGAAGCCAGGGACAGTTTACTCAGCGGCAGTTTCCGACCTTTGCCATTTCGGTT

  61  ATCCGTACCCCCACAGTGATCTGACAACTCAGCTCCGAATCCCAACGGCGATCGCCATTC
                                                            -35
 121  TTGCTTGGGGCATTAAAACCCGCTGGTTAGCCGGAATTTCCGTCCAGATTCCCTTTCCAG
                       -10
 181  ATGTCCCCCTCGGTTCTAAACTTGAGCTTCGAAGTGTGTTGTTGGCAATCGAGAGGTCTG

 241  CTTGTGAAACGTGTCTTAGCGATTATCCTGGGCGGTGGGGCCGGGACCCGCCTCTATCCT
        M  K  R  V  L  A  I  I  L  G  G  G  A  G  T  R  L  Y  P   19
 301  TTAACCAAACTCAGAGCCAAACCCGCAGTTCCCTTGGCCGGAAAGTATCGCCTCATCGAT
      L  T  K  L  R  A  K  P  A  V  P  L  A  G  K  Y  R  L  I  D   39
 361  ATTCCCGTCAGTAATTGCATCAACTCAGAAATCGTTAAAATTTACGTCCTTACCCAGTTT
      I  P  V  S  N  C  I  N  S  E  I  V  K  I  Y  V  L  T  Q  F   59
 421  AATTCCGCCTCCCTTAACCGTCACATCAGCCGGGCCTATAATTTTTCCGGCTTCCAAGAA
      N  S  A  S  L  N  R  H  I  S  R  A  Y  N  F  S  G  F  Q  E   79
 481  GGATTTGTGGAAGTCCTCGCCGCCCAACAAACCAAAGATAATCCTGATTGGTTTCAGGGC
      G  F  V  E  V  L  A  A  Q  Q  T  K  D  N  P  D  W  F  Q  G   99
 541  ACTGCTGATGCGGTACGGCAATACCTCTGGTTGTTTAGGGAATGGGACGTAGATGAATAT
      T  A  D  A  V  R  Q  Y  L  W  L  F  R  E  W  D  V  D  E  Y  119
 601  CTTATTCTGTCCGGCGACCATCTCTACCGCATGGATTACGCCCAATTTGTTAAAAGACAC
      L  I  L  S  G  D  H  L  Y  R  M  D  Y  A  Q  F  V  K  R  H  139
 661  CGGGAAACCAATGCCGACATAACCCTTTCCGTTGTGCCCGTGGATGACAGAAAGGCACCC
      R  E  T  N  A  D  I  T  L  S  V  V  P  V  D  D  R  K  A  P  159
 721  GAGCTGGGCTTAATGAAAATCGACGCCCAGGGCAGAATTACTGACTTTTCTGAAAAGCCC
      E  L  G  L  M  K  I  D  A  Q  G  R  I  T  D  F  S  E  K  P  179
 781  CAGGGGGAAGCCCTCCGCGGCATGCAGGTGGACACCAGCGTTTTGGGCCTAAGTGCGGAG
      Q  G  E  A  L  R  G  M  Q  V  D  T  S  V  L  G  L  S  A  E  199
 841  AAGGCTAAGCTTAATCCTTACATTGCCTCCATGGGCATTTACGTTTTCAAGAAGGAAGTA
      K  A  K  L  N  P  Y  I  A  S  M  G  I  Y  V  F  K  K  E  V  219
 901  TTGCACAACCTCCTGGAAAAATATGAAGGGGCAACGGACTTTGGCAAAGAAATCATTCCT
      L  H  N  L  L  E  K  Y  E  G  A  T  D  F  G  K  E  I  I  P  239
 961  GATTCAGCCAGTGATCACAATCTGCAAGCCTATCTCTTTGATGACTATTGGGAAGACATT
      D  S  A  S  D  H  N  L  Q  A  Y  L  F  D  D  Y  W  E  D  I  259
1021  GGTACCATTGAAGCCTTCTATGAGGCTAATTTAGCCCTGACCAAACAACCTAGTCCCGAC
      G  T  I  E  A  F  Y  E  A  N  L  A  L  T  K  Q  P  S  P  D  279
1081  TTTAGTTTTTATAACGAAAAAGCCCCCATCTATACCAGGGGTCGTTATCTTCCCCCCACC
      F  S  F  Y  N  E  K  A  P  I  Y  T  R  G  R  Y  L  P  P  T  299
1141  AAAATGTTGAATTCCACCGTGACGGAATCCATGATCGGGGAAGGTTGCATGATTAAGCAA
      K  M  L  N  S  T  V  T  E  S  M  I  G  E  G  C  M  I  K  Q  319
1201  TGTCGCATCCACCACTCAGTTTTAGGCATTCGCAGTCGCATTGAATCTGATTGCACCATT
      C  R  I  H  H  S  V  L  G  I  R  S  R  I  E  S  D  C  T  I  339
1261  GAGGATACTTTGGTGATGGGCAATGATTTCTACGAATCTTCATCAGAACGAGACACCCTC
      E  D  T  L  V  M  G  N  D  F  Y  E  S  S  S  E  R  D  T  L  359
1321  AAAGCCCGGGGGGAAATTGCCGCTGGCATAGGTTCCGGCACCACTATCCGCCGAGCCATC
      K  A  R  G  E  I  A  A  G  I  G  S  G  T  T  I  R  R  A  I  379
1381  ATCGACAAAAATGCCCGCATCGGCAAAAACGTCATGATTGTCAACAAGGAAAATGTCCAG
      I  D  K  N  A  R  I  G  K  N  V  M  I  V  N  K  E  N  V  Q  399
1441  GAGGCTAACCGGGAAGAGTTAGGTTTTTACATCCGCAATGGCATCGTAGTAGTGATTAAA
      E  A  N  R  E  E  L  G  F  Y  I  R  N  G  I  V  V  V  I  K  419
1501  AATGTCACGATCGCCGACGGCACGGTAATCTAGGGCCAGTTTCTTTCCTCGCACCATAGC
      N  V  T  I  A  D  G  T  V  I  *                              429
1561  CATGA
```

Figure 1. Underlined amino acids indicate those that have been confirmed by N-terminal protein sequencing of ADPGlc pyrophosphorylase purified from *Synechocystis* PCC 6803. Double underline indicates proposed −10 and −35 box sequences. Shaded areas indicate a proposed Shine-Delgarno prokaryotic ribosome binding sequence. Asterisk indicates a stop codon.

lished results). N-terminal sequencing also confirms that the *Synechocystis* ADPGlc PPase consists of a single subunit and is not a heterotetramer with subunits of similar molecular mass. The deduced amino acid sequence of *Synechocystis* ADPGlc PPase was compared with the sequences of rice seed (1), spinach leaf (10), and *E. coli* (2). Based on the percent identity of amino acid sequences (Table I), despite being homotetrameric, the *Synechocystis* enzyme is more similar to the higher plant enzymes than to the bacterial enzymes. Furthermore, the *Synechocystis* protein is more similar to the small than to the large subunit of higher plant ADPGlc PPases.

ACKNOWLEDGMENTS

We would like to thank Mr. James Schilling and Ms. Kim Brown of the University of California, Davis Protein Structure Laboratory for the N-terminal sequence analysis of the *Synechocystis* PCC 6803 ADPGlc PPase. Brian Smith-White is gratefully acknowledged for advice on molecular techniques.

LITERATURE CITED

1. **Anderson JM, Hnilo J, Larson R, Okita TW, Morell M, Preiss J** (1989) The encoded primary sequence of a rice seed ADP-glucose pyrophosphorylase subunit and its homology to the bacterial enzyme. J Biol Chem **264:** 12238–12242
2. **Baecker PA, Furlong CE, Preiss J** (1983) Biosynthesis of bacterial glycogen: primary structure of *Escherichia coli* B ADP-glucose synthetase as deduced from the nucleotide sequence of the *glg*C gene. J Biol Chem **258:** 5084–5088
3. **Devereux J, Haeberli P, Smithies O** (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res **12:** 387–395
4. **Iglesias AA, Kakefuda G, Preiss J** (1991) Regulatory and structural properties of the cyanobacterial ADPglucose pyrophosphorylase. Plant Physiol **97:** 1187–1195
5. **Lee YM, Preiss J** (1986) Covalent modification of substrate-binding sites of *Escherichia coli* ADP-glucose synthetase. J Biol Chem **261:** 1058–1064
6. **Morell M, Bloom M, Preiss J** (1988) Affinity labeling of the allosteric activator site(s) of spinach leaf ADP-glucose pyrophosphorylase. J Biol Chem **263:** 633–637
7. **Parsons TF, Preiss J** (1978) Biosynthesis of bacterial glycogen: isolation and characterization of the pyridoxal-P allosteric activator site and ADP-glucose-protected pyridoxal-P binding site of *Escherichia coli* B ADP-glucose synthetase. J Biol Chem **253:** 7638–7645
8. **Preiss J** (1984) Bacterial glycogen synthesis and its regulation. Annu Rev Microbiol **38:** 419–458
9. **Preiss J** (1991) Biology and molecular biology of starch synthesis and its regulation. *In* B Mifflin, ed, Oxford Surveys of Plant Molecular and Cell Biology, Vol 7. Oxford University Press, Oxford, UK, pp 59–114
10. **Smith-White BJ, Preiss J** (1992) Comparison of proteins of ADP-glucose pyrophosphorylase from diverse sources. J Mol Evol (in press)