Plant Molecular Biology **17**: 915–917, 1991.

Update section

Sequence

Nucleotide sequence of *psbB* from *Prochlorothrix hollandica*

Karen L. Greer and Susan S. Golden*
*Texas A&M University, Department of Biology, College Station, TX 77843, USA (*author for correspondence)*

Received and accepted 20 May 1991

Prochlorothrix hollandica is an oxygen-evolving photosynthetic prokaryote that has a chlorophyll *b*-containing light-harvesting antenna rather than phycobilisomes [1, 2]. Thus *P. hollandica* has a photosynthetic membrane structure more like that of higher-plant chloroplasts than that of cyanobacteria and may be descended from the evolutionary precursor of chloroplasts. If such an evolutionary relationship exists, analysis of genes encoding conserved elements of the photosynthetic apparatus of *P. hollandica* should yield some insights into the origin of chloroplast genome structure and the constraints of coregulation of coordinately synthesized polypeptides. Towards this goal we have isolated and sequenced the *P. hollandica* genes *psbB* (presented here), *psbH*, *petB* and *petD* (manuscript in preparation).

The *psbB* gene encodes a chlorophyll *a*-binding protein associated with Photosystem II (PSII), termed CP-47 [7, 11]. The structure of the PSII reaction center appears to be conserved among all oxygenic photoautotrophs and is thought to consist of the proteins D1, D2, and cytochrome b_{559}, along with the P680 chlorophyll and other accessory pigments [4, 10]. CP-47 and CP-43, another chlorophyll *a*-binding protein, are frequently isolated with the PSII reaction center and probably contribute energy directly to the reaction center as the final light-harvesting antennae [4]. CP-47 was so named because the protein identified from some higher plants migrates at approximately 47 kDa on denaturing polyacrylamide gels [4]. *P. hollandica* PSII chlorophyll-protein complexes also release a protein that migrates at 47 kDa on polyacrylamide gels [1]. However, as is true for the higher-plant gene [11], the *psbB* sequence predicts a product of higher molecular weight. The deduced amino acid sequence from the *P. hollandica psbB* gene sequence would yield a polypeptide of 57 kDa.

The *psbB* gene was isolated from a λgt10 library of *P. hollandica* genomic DNA [9] based on hybridization with the same gene from *Anabaena* sp. strain PCC7120 [7]. The clone contained a 6.6 kb *Eco* RI fragment that included all of the gene except for the last 14 amino acids of the open reading frame. An overlapping 6.6 kb *Bam* HI fragment was cloned from a new library of *P. hollandica* genomic DNA constructed in the bacteriophage vector λL47.1 [8]. Sequencing on both strands of approximately 2.2 kb including the open reading frame was by dideoxynucleotide chain-termination reactions (Sequenase Kit, United States Biochemical Corporation). Sequence comparisons using the University of Wisconsin Genetics Computer Group programs [3] indicated that the *psbB* gene sequence is very well conserved with those of cyanobacteria and higher plant chloroplasts. It showed 83% similarity and

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X59614 P. Hollandica, psbB DNA.

73% identity at the amino acid level when compared to the gene from the maize chloroplast [11]. Compared to the *Synechocystis* 6803 *psbB* [13], the *P. hollandica psbB* has 90% similar and 82% identical amino acids.

Unlike the consensus arrangement in the chloroplast genome, *psbB* in *P. hollandica* is not found in an operon with *psbH*, *petB* and *petD*. Only *petB* and *petD* are in a cotranscribed operon in the *Nostoc* genome [6]; *psbB* and *psbH* are not closely linked in any other cyanobacteria [6, 7, 13] or in the cyanelle genome of *Cyanophora paradoxa* (D. Bryant, personal communication).

Figure 1 shows the nucleotide sequence of the *psbB* open reading frame, as well as upstream and downstream regions. The deduced amino acid sequence of the open reading frame is shown using the single letter code. The open reading frame is preceded by a potential ribosome binding site that matches the Shine-Delgarno consensus sequence of *Escherichia coli* [12] in distance from the ATG and in constitution. Transcripts were mapped by primer extension and S1 nuclease protection assays (data not shown). Two transcript 5′ ends were evident at −224 and −183 bases from the ATG of the open reading frame. Sequences up-

```
-350
 TGCCAGGGAT CCAGACTCGG CCTCTGGACA GTACCCCGAT CCACGTCTCA
-300
 AGGGCTGCCC ATGATGCCCC CCGATCGCCC GTGGATTTCC TTGCAATTGG
-250
 GAGGGCCACT AGTTACACTT CTTTATACCT TTTCTTTTTA GAACTTTTTG
-200
 ATGGTCAAAT TATCAGTAGA GCTTTTGTAG CAACTGCTGG GATGTGCCTG
-150
 GTTGCCCCCC GGCTTGGTGC ATCAGCGACC TGATATCAAT AGACTTGAAG
-100
 AAATGCATTT GTACTGGTTT GGTACTGGCT CCCGGCTGGA TCCCAAGCGA
-50
 GTACAGTCTT GCCAAGAGTC ATAATAATTA TTTGAGGAGG AGCGTAGTCG
1
ATGGGACTACCCTGGTATCGCGTACACACCGCCGTCATCAATGACCCCGGTCGGCTGCTA
 M G L P W Y R V H T A V I N D P G R L L
61
GCTGTACACCTAATGCACACAGCCCTGGTCTTGGGCTGGGCCGGTTCAATGGCGCTCTAT
 A V H L M H T A L V L G W A G S M A L Y
121
GAGCTAGCTGTTTTTGATCCCAGTGATGCAGTTCTCAATCCCATGTGGCGGCAAGGCATG
 E L A V F D P S D A V L N P M W R Q G M
181
TTTGTGCTGCCCTTCATGACCCGCTTGGGAGTCACCCACTCCTGGTCTGGTTGGACCGTG
 F V L P F M T R L G V T H S W S G W T V
241
ACTGGAGAGCCTTGGATTAATGAACCGGGCTTTTTAAATGCACACTTTAACTTCTGGAGC
 T G E P W I N E P G F L N A H F N F W S
301
TACGAAGGCGTAGCCCTGATGCACATTGTTCTGTCCGGTCTCTTCTTCCTGGCTGCCGTG
 Y E G V A L M H I V L S G L F F L A A V
361
TGGCACTGGGTTTACTGGGATTTGGATCTGTTTGAGGATCCCCGTACCGGCGAGCCTGCC
 W H W V Y W D L D L F E D Q R T G E P A
421
TTAGACCTACCCAAAATCTTTGGCGGTCACCTCTTCCTGCTAGGCTTCCTTTGCTTTAAC
 L D L P K I F G G H L F L L G F L C F N
481
TTCGGCACCTTCCACCTGACCGGTATCTTCGGACCCGGTATGTGGGTTTCCGATGCCTAT
 F G T F H L T G I F G P G M W V S D A Y
541
GGCTTAACGGGCCACATCGAGCATATTGCTCCTGAATGGGGTCCGGCTGGGTTTAACCCC
 G L T G H I E H I A P E W G P A G F N P
601
TTTAACCCCGGTGGTGTTGTGGCTCACCACATCGCAGCCGGTATCACCCTGATGATTGGG
 F N P G G V V A H H I A A G I T L M I G
661
GGGGTCTTCCACCTGACGGCTCGTCCCCCTGAGCGCCTTTACACCGCCCTGCGTATGGGT
 G V F H L T A R P P E R L Y T A L R M G
721
AATGTGGAAACGGCTCTGGCTAGCGCGATCGCCGCTGTCTTTGGGGCTGCTTTCGTGGTG
 N V E T A L A S A I A A V F G A A F V V
781
GCTGGAACCATGTGGTATGGCCACGTGACCACACCTATCGAACTCTTTGGACCTACCCGC
 A G T M W Y G H V T T P I E L F G P T R
841
TACCAGTGGGATCAGGGCTACTTCACCCAGGAAATTCAGCGTCGTGTGGATTCCCAACTG
 Y Q W D Q G Y F T Q E I Q R R V D S Q L
901
GCTGAAGGGGCTTCCCTGTCTGAAGCTTGGTCTTCCATCCCTGAGAAGCTAGCCTTCTAC
 A E G A S L S E A W S S I P E K L A F Y
961
GACTATGTTGGTAACAGCCCCGCTAAAGGCGGTCTGTTCCGCGTGGGTGCCATGGATAGT
 D Y V G N S P A K G G L F R V G A M D S
1021
GGTGATGGTATTGCCGAAGAGTGGCTGGGTCACCCCGTCTTCCAAGATGGTGCAGGCCGT
 G D G I A E E W L G H P V F Q D G A G R
1081
GCCCTGTCGGTGCGTCGTTTGCCCAACTTCTTTGAGAACTTCCCCGTGATCCTCACGGAT
 A L S V R R L P N F F E N F P V I L T D
1141
GGCGATGGTGTAGTTCGTGCTGATATTCCCTTCCGCCGCTCTGAGTCCCAGTACAGCTTC
 G D G V V R A D I P F R R S E S Q Y S F
1201
GAGCAAACTGGAGTAACTGTCAGCTTCTACGGCGGTGCTCTGGATGGTCAAACCTTCACC
 E Q T G V T V S F Y G G A L D G Q T F T
1261
AATCCTTCGGACGTGAAGAAGTTTGCCCGCCGTGCCCAACTGGGTGAAGCCTTCGACTTC
 N P S D V K K F A R R A Q L G E A F D F
1321
GACACCGAAACCCTCGGATCTGATGGTGTGTTCCGCACCAGCACCCGGGGTTGGTTCACC
 D T E T L G S D G V F R T S T R G W F T
1381
TTTGGCCATGCTTGCTTTGCCCTGCTCTTCTTCTTCGGCCACATCTGGCACGGTGCTCGC
 F G H A C F A L L F F F G H I W H G A R
1441
ACCCTGTTCCGTGATGTGTTCGCCGGGATCGACGCTGACCTGGGTGAGCAAATCGAATTC
 T L F R D V F A G I D A D L G E Q I E F
1501
GGTGCGTTCCAGAAGCTGGGTGACCTGACCACCCGCAAGAGCTAA
 G A F Q K L G D L T T R K S *
1546
 GACCCGAATC GCTCCCTTGG GAGTCTCTAG GCTAGGGTTT TAGAGCGCAT
1596
 CATCGGCTAA ATAATCGATC GCCTAAATAA TCGATCGCCT AAATCACCGA
1646
 TCGCATAAAC TTAAGGGGCT
```

Fig. 1. Nucleotide sequence and deduced amino acid sequence of *psbB* from *P. hollandica*. Transcript 5′ ends are indicated by double underlines; single underlines identify potential promoter regions upstream from the first transcript 5′ end. A potential Shine-Delgarno ribosome binding site preceding the open reading frame is designated by italics.

stream from the −183 transcription start site do not resemble those of consensus *E. coli* promoters at −10 and −35 [5]. However, the second transcription start site (−224) is preceded by potential promoter elements that match, in sequence and spacing, the most conserved bases of *E. coli* consensus promoters [5]. Mapping of *psbB* transcripts in *Anabaena* [7] and *Synechococcus* sp. strain PCC7942 (R. Kulkarni and S. Golden, unpublished data) also indicates two transcription start sites. Multiple transcripts arise from the *psbB-psbH-petB-petD* operon in the chloroplast genomes of higher plants, presumably by processing from a single long transcript that contains all four of the genes [11].

Acknowledgements

We thank Jean Lang and Robert Haselkorn for the *Anabaena psbB* probe. This research was supported by grants from the National Science Foundation (DMB 8958089) and the Texas Advanced Research Program.

References

1. Bullerjahn GS, Matthijs HCP, Mur LR, Sherman LA: Chlorophyll-protein composition of the thylakoid membrane from *Prochlorothrix hollandica*, a prokaryote containing chlorophyll *b*. Eur J Biochem 168: 295–300 (1987).
2. Burger-Wiersma T, Veenhuis M, Korthals HJ, Van de Wiel CCM, Mur LR: A new prokaryote containing chlorophylls *a* and *b*. Nature 320: 262–264 (1986).
3. Devereux J, Haberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 385–398 (1984).
4. Ghanotakis DF, Yocum CF: Photosystem II and the oxygen-evolving complex. In: Briggs WR, Jones RL, Walbot V (eds) Annual Reviews of Plant Physiology and Plant Molecular Biology, vol 41, pp. 255–276. Annual Reviews, Inc., Palo Alto, CA (1990).
5. Hawley DK, McClure WR: Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucl Acids Res 11: 2237–2255 (1983).
6. Kallas T, Spiller S, Malkin R: Characterization of two operons encoding the cytochrome b_6-f complex of the cyanobacterium *Nostoc* PCC 7906, J Biol Chem 263: 14334–14342 (1988).
7. Lang JD, Haselkorn R: Isolation, sequence and transcription of the gene encoding the photosystem II chlorophyll-binding protein, CP-47, in the cyanobacterium, *Anabaena* 7120. Plant Mol Biol 13: 441–456 (1989).
8. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
9. Morden CW, Golden SS: *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts. Nature 337: 382–385 (1989).
10. Nanba O, Satoh K: Isolation of a photosystem II reaction center consisting of D-1 and D-2 polypeptides and cytochrome b-559. Proc Natl Acad Sci USA 84: 109–112 (1987).
11. Rock CD, Barkan A, Taylor WC: The maize plastid *psbB-psbF-petB-petD* gene cluster: spliced and unspliced *petB* and *petD* RNAs encode alternative products. Curr Genet 12: 69–77 (1987).
12. Stormo GD: Translation initiation. In: Reznikoff W, Gold L (eds) Maximizing Gene Expression, pp. 195–224. Butterworths, Boston (1986).
13. Vermaas WFJ, Williams JGK, Arntzen CJ: Sequencing and modification of *psbB*, the gene encoding the CP-47 protein of Photosystem II, in the cyanobacterium *Synechocystis* 6803. Plant Mol Biol 8: 317–326 (1987).