Plant Molecular Biology **14**: 81–92, 1989.
© 1989 *Kluwer Academic Publishers. Printed in Belgium.*

Characterization of a four-member *psbA* gene family from the cyanobacterium *Anabaena* PCC 7120

Jacqueline M. Vrba and Stephanie E. Curtis*
*Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614, USA (*author for correspondence)*

Received 5 July 1989; accepted in revised form 24 August 1989

Key words: *Anabaena*, cyanobacteria, D1, multigene family, *psbA* genes

Abstract

The cyanobacterium *Anabaena* PCC 7120 contains a multigene family that encodes the D1 polypeptide of Photosystem II. This family consists of four members denoted *psb*AI–IV that are each unique but highly homologous. *psb*AII, III and IV are more closely related to each other than to *psb*AI. These three copies encode identical polypeptides that differ from the *psb*AI product by 21 amino acids. The transcription initiation site for *psb*AI has been mapped to 64–65 nucleotides upstream from the coding region. Primer extension assays performed with an oligonucleotide specific for *psb*AII, III and IV transcripts suggest that one or more of these genes is also expressed. Genomic mapping and chromosome walking experiments demonstrate that none of the four *psb*A copies is within 20 kbp of another member of the gene family.

Introduction

Among photosynthetic organisms, cyanobacteria and the chloroplasts of plants share the ability to perform oxygenic photosynthesis. Conservation of this process is reflected in the high degree of similarity between the photosynthetic complexes of cyanobacteria and chloroplasts [4]. The photochemical reactions of oxygenic photosynthesis occur at two membrane protein complexes termed Photosystem I and Photosystem II (PSII), of which PSII has been most extensively studied. The core of PSII contains a heterodimer of homologous proteins, denoted D1 and D2, with which the reaction center chlorophyll (P_{680}) and electron acceptors are associated [17, 21, 34]. D1 and D2 are also structurally similar to two subunits of the reaction center complex from purple nonsulfur photosynthetic bacteria and are thought to be functionally analogous [19]. D1 binds a plastoquinone (Q_B), that is the second stable electron acceptor of PSII [34] and also contains the redox-active tyrosine residue (Y_z) that reduces the reaction center chlorophyll [9].

In plants the D1 protein is encoded by a single-copy chloroplast gene denoted *psb*A, except in *Chlamydomonas* where two identical copies of *psb*A are located within the inverted repeats of the chloroplast chromosome [25]. In contrast, all cyanobacterial species thus far examined contain a family of *psb*A genes usually comprised of three closely related but nonidentical members [6, 11, 12, 15, 20, 24]. In previous work, two *psb*A genes from the cyanobacterium *Anabaena* sp. strain PCC 7120 were cloned and their sequences determined [6]. We present here the sequence of two

additional *psbA* genes from *Anabaena*, studies on expression of the gene family, and examination of the linkage relationship of members of this family.

Materials and methods

Materials

Restriction endonucleases, DNA polymerase, T4 DNA ligase, the large fragment of DNA polymerase I, T4 polynucleotide kinase and calf intestine alkaline phosphatase were purchased from Bethesda Research Laboratories (BRL), Gaithersburg, MD; Boehringer Mannheim Biochemicals (BMB), Indianapolis, IN; Promega, Madison, WI; or International Biotechnologies (IBI), New Haven, CT. Gene Clean was purchased from BIO1O1, La Jolla, CA. Circular nitrocellulose filters were purchased from Schleicher and Schuell (S&S), Keene, NH. Magnagraph nylon membrane was purchased from Micron Separations, Westboro, MA. [γ^{32}P]ATP (7000 Ci/mmol) was purchased from ICN Radiochemicals, Irvine, CA. [α^{32}P]dATP (400 Ci/mmol) and [α^{35}S]dATP (1000 Ci/mmol) were purchased from Amersham Co., Arlington Heights, IL. M13 universal sequencing primer was purchased from Pharmacia, Piscataway, NJ. Site-specific oligonucleotides were synthesized on an automated DNA synthesizer (model 380A; Applied Biosystems, Foster City, CA). All other chemicals were reagent grade.

Identification of Anabaena *PCC 7120 psbA genes*

Anabaena genomic DNA was isolated as previously described [5]. The DNA was digested with restriction endonucleases, fractionated on 0.8% agarose gels, transferred to Magnagraph filters and hybridized with a nick-translated *psbA*I probe (5×10^6 cpm probe A, 1.1 kilobase pair (kbp) *Eco* RI-*Xba* I fragment from pAn625) [5] as described [28].

Isolation of genomic clones containing psbA genes and flanking regions

To isolate clones of *Hin*d III fragments containing *psbA* genes, *Hin*d III digested total *Anabaena* DNA was fractionated on agarose preparative gels and fragments within the correct size range (4.6, 7.6 and 8.2 kbp) were electroeluted and purified. Libraries of the size-selected *Hin*d III fragments were constructed in λCharon 28 [16]. These libraries were screened by plaque hybridization as described [28] with nick-translated probe A. The 4.6, 7.6, and 8.2 kbp *Hin*d III fragment inserts of the phage DNA's were recloned into pIBI25 to generate the plasmids pAn635, pAn622 and pAn630, respectively. To facilitate sequencing, the *Hin*d III-*Eco* RI fragments of pAn630 and pAn635 with *psbA* homology, and the *Eco* RI insert of pAn620 [6] were subcloned into M13 mp18 and mp19.

Two additional libraries, a *Sau* 3A I partial library constructed in the vector λL47.1, kindly provided by James Golden, and an *Eco* RI library constructed in λCharon4A [28], were screened to obtain regions flanking the *psbA* genes. These libraries were screened for clones that overlap the known *psbA* fragments using DNA probes isolated from the 5′ and 3′ flanking areas of *psbA*II and IV, and from the 3′ flanking DNA of *psbA*I and III. Convenient restriction sites were not available for generating a *psbA*I 5′ specific probe so an end labeled oligonucleotide (5′AGACTTGCGCCTGTC3′) specific to the *psbA*I 5′ flanking DNA was used to screen the library.

All recombinant clones isolated with the 5′ flanking probes were rescreened with 3′ flanking probes and vice versa to determine the orientation of each clone relative to the coding region of each copy. The recombinant phage DNAs were characterized by restriction mapping and hybridization with the probes used to isolate them. The clones for each gene copy were then screened with probes to the flanking regions of the three other genes to ascertain if linkage existed within the distance spanned by the clones.

DNA sequence determinations and analysis

DNA sequences of the *psb*A genes were determined by the chain termination method [29] with M13 universal primer or *psb*A-specific oligonucleotides. The sequences of both strands were determined using either single- or double-stranded templates. Sequence analyses were performed by using the IBI DNA/Protein Sequence Analysis System computer software.

Transcript analysis

Total *Anabaena* RNA was isolated from vegetative cells grown on complete media as described [35]. Primer extension assays were performed as described [1], using the septadecanucleotide (5′CATAAATGCGGTTTTCG3′) complementary to the *psb*AI coding sequence from position 72 to 89, and the octadecanucleotide (5′GATGTATAGACGGTTGTT3′) complementary to the *psb*AII, III and IV coding sequence from position 74 to 91.

Results

Isolation and mapping of psbA genes

In previous work, *Anabaena* genomic DNA was shown to contain five *Eco* RI fragments with homology to a spinach chloroplast *psb*A probe [6]. Subsequent digests of *Anabaena* genomic DNA showed that there are four rather than five *Eco* RI fragments as well as four *Hin*d III fragments with *psb*A homology (Fig. 1). We believe that the discrepancy between the two sets of genomic mapping data is due to incomplete digestion of the genomic DNA in the first study [6]. The two lowest molecular weight fragments (An620 and An625) were previously isolated from a λCharon4A library of *Anabaena Eco* RI fragments and shown by DNA sequencing to each contain a complete *psb*A gene [6]. These genes were denoted *psb*AI (An625) and *psb*AII (An620; Figs. 1 and 2)

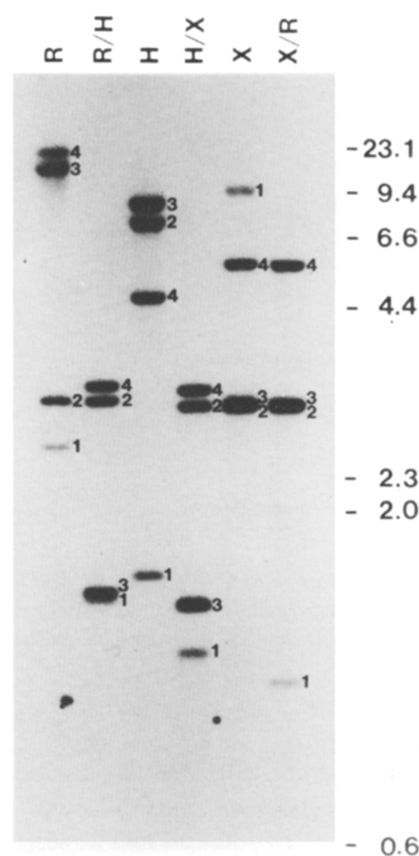

Fig. 1. Identification of *Anabaena psb*A genes. Total *Anabaena* DNA was digested with *Eco* RI (R), *Hin*d III (H) and *Xba* I (X) both singly and in pairwise combinations, electrophoresed, blotted and hybridized with a *psb*AIII coding probe. Numbers beside the bands indicated the *psb*A copy. Sizes of molecular weight standard fragments are shown at the right.

The *Eco* RI fragments of *psb*AI and II do not contain internal *Hin*d III fragments (Fig. 2) and thus the correspondence between the genomic *Hin*d III and *Eco* RI fragments was not known. Since the two largest *Eco* RI fragments identified have lengths exceeding the range of sizes contained in the λCharon 4A library, libraries of *Hin*d III fragments of *Anabaena* DNA were constructed to facilitate the isolation of additional *psb*A genes. Size-selected libraries that correspond to the 8.2, 7.6 and 4.6 kbp *Hin*d III genomic fragments with *psb*A homology were constructed in vector λCharon 28. The libraries were screened with a *psb*AI probe (Fig. 2, probe A) and three unique clones with inserts of

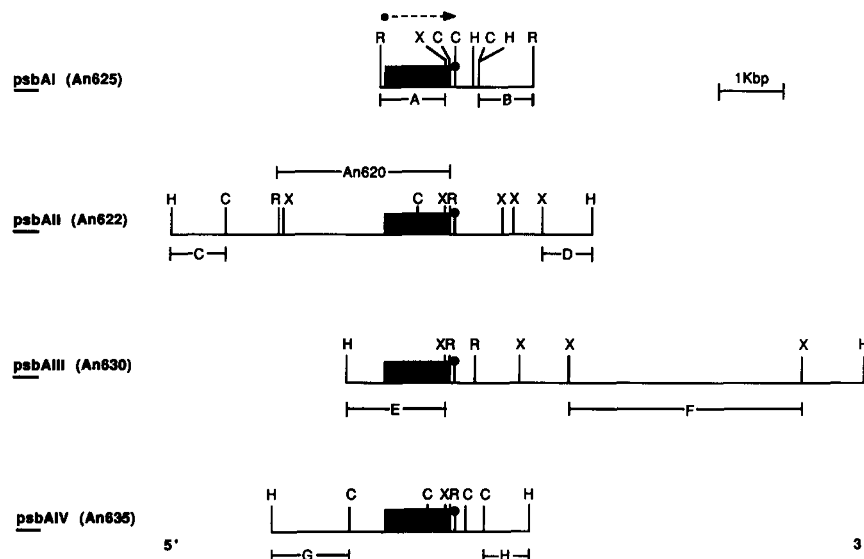

Fig. 2. Physical map of *psb*AI–IV genes. *Hin*d III (H), *Eco* RI (R), *Xba* I (X) and *Hin*c II (C) sites are indicated. Solid boxes denote *psb*A coding regions. The ball and sticks represent stem and loop structures. Fragments used as probes are indicated by bars lettered A-H beneath the maps. The dashed line above *psb*AI denotes the *psb*AI transcript. Designations of the cloned *Hin*d III and *Eco* RI fragments are indicated in parentheses.

expected size were isolated. The lambda clone inserts were subcloned into plasmid vectors, mapped with restriction endonucleases, and the regions of *psb*A homology defined by hybridization with the *psb*AI probe (Fig. 2). The *psb*AII *Eco* RI fragment was shown to be within the restriction map of the 7.6 kbp *Hin*d III fragment (An622). Maps of the 8.2 kbp (An630) and 4.6 kbp (An635) fragments indicated that these contained additional *psb*A genes that were designated *psb*AIII and IV, respectively (Fig. 2).

*Sequence of psb*AIII *and psb*AIV *coding regions*

The *Hin*d III-*Eco* RI fragments of pAn630 and pAn635 with *psb*A homology were subcloned into M13 vectors and their sequences determined using oligonucleotide primers specific to regions conserved between *psb*AI and II. Each clone was shown to contain a complete *psb*A gene (Fig. 3). The coding sequences of *psb*AIII and IV were compared with those of *psb*AI and II [6] (Fig. 3). All four members of the gene family are highly homologous but each copy has a unique sequence. The *psb*AII, III and IV genes are much more closely related to each other than to *psb*AI, with each pair of genes in this group displaying 98–99% homology. The *psb*AIII and IV genes are the most similar with differences in only five nucleotides; *psb*AII differs from *psb*AIII by seventeen nucleotides and *psb*AII differs from *psb*AIV by fifteen nucleotides. In contrast the *psb*AI gene is 88–89% homologous to each of the other copies.

The *psb*AII, III and IV genes can encode identical polypeptides since the nucleotide substitutions among these copies are silent changes (Fig. 3). The products of *psb*AII–IV differ from the *psb*AI translation product by 21 amino acids (94% amino acid identity). All but three of the changes are located in the amino terminal half of the protein and thirteen of the amino acid substitutions are conservative replacements [8].

*Comparison of psb*AI-IV *5′ flanking DNA sequence*

To extend the 5′ flanking sequences for *psb*AI and II, additional cloning was undertaken. A

lambda clone that overlaps the 5′ end of the *psb*AI gene was isolated and a portion of it was subcloned into M13 vectors, as was the insert of pAn620 (*psb*AII). Regions 5′ to each of the *psb*A genes were sequenced with an oligonucleotide complementary to all four noncoding strands just within the coding region. Oligonucleotides complementary to the coding strand in the 5′ flanking region were synthesized based on sequence of the first strand and subsequently used to confirm the sequence.

Comparisons of the 5′ flanking DNA of *psb*AI–IV show that homology observed in the gene coding regions extends into flanking DNA (Fig. 4). In all four copies, similarity is maintained for at least 80 base pairs (bp) upstream from the coding sequence and two regions of near perfect identity are apparent. Immediately preceding the methionine start codon [9] is a conserved block of twelve nucleotides in which the *psb*AI sequence differs from that of the other copies by one nucleotide (Fig. 4). The 5′ flanking sequences of *psb*AII and IV are identical for an additional 9 bp. From 54 to 80 bp upstream from the *psb*AI–IV coding sequences are additional regions of homology 8 bp in length. These regions conform to the sequence of TGTAA^{A}/T^{G}/T^{T} (Fig. 4). The 5′ flanking DNA of *psb*AII contains two identical copies of this sequence separated by 2 bp.

*Comparison of psb*AI-IV *3′ flanking DNA sequence*

To obtain 3′ flanking sequence for *psb*AII–IV, pAn622, pAn630 and pAn635 were sequenced as double stranded templates using oligonucleotides complementary to the noncoding strand just within the 3′ end of the coding region. The derived sequences were compared to previously determined *psb*AI 3′ flanking sequence [6] (Fig. 5). As with coding region comparisons, *psb*AII, III and IV are more homologous to each other than to *psb*AI. The 3′ flanking regions of *psb*AII, III and IV are identical for 55 base pairs following the translation termination codon with the exception of a single-nucleotide change in *psb*AIV (Fig. 5). In *psb*AIII and IV identity continues for four additional base pairs and diverges thereafter (Fig. 5).

The 3′ flanking sequences of all four genes contain large inverted repeats at a similar distance from the coding region. These inverted repeats have the potential to form stable stem and loop structures in the transcription product (Fig. 5). The single nucleotide change in *psb*AIV relative to *psb*AII and III occurs at a mismatch within the stem of the structure (Fig. 5).

*psb*A *transcript analysis*

In previous work, a *psb*AI probe was shown to hybridize with a single RNA transcript of 1.25 kb [6]. This result suggested that either a single *psb*A gene is expressed, or that all *psb*A transcripts have a similar length. In addition, a primer extension assay was performed using a *psb*AI fragment that would recognize all *psb*AI transcripts. In these experiments, several transcript endpoints were mapped to 55–60 nucleotides upstream from the *psb*AI coding region [6].

In the present study, primer extension experiments were performed with oligonucleotide primers to determine whether the multiple *psb*A genes are expressed. An oligonucleotide specific to the *psb*AI coding region was used in a primer extension assay with total cellular RNA. Primer extended products were mapped to 64 and 65 nucleotides upstream from the *psb*AI coding region (Figs. 4, 6A). This region is just downstream from the eight bp block of sequence conserved among all four *psb*A copies (Fig. 4).

Due to the extensive sequence conservation between *psb*AII, III and IV in both coding and 5′ proximal sequences, it was not possible to synthesize oligonucleotides that would discriminate among copy-specific transcripts under conditions used for the primer extension assays. To determine collectively if copies other than *psb*AI are expressed, an oligonucleotide was synthesized that would anneal to *psb*AII, II and IV transcripts (*psb*AII–IV general primer). When primer extension was performed with the general primer

```
1             thr                     ser                         arg         thr
I             A                     T  GT       T                  G          ACA       C
II    ATG ACC GCA ACC TTA CAA CAG CGC AAA AGC GCC AAC GTA TGG GAA CAG TTC TGC GAG TGG ATT   63
III                             A               T
IV                              A
2-4   met thr ala thr leu gln gln arg lys ser ala asn val trp glu gln phe cys glu trp ile   21

1                 glu         ile     val
I                 G A       C A T   T G T                   G T G       T   T     C
II    ACC AGC ACC AAC AAC CGT CTA TAC ATC GGT TGG TTC GGC GTA CTA ATG ATC CCC ACC TTG CTA  126
III
IV
2-4   thr ser thr asn asn arg leu tyr ile gly trp phe gly val leu met ile pro thr leu leu   42

1                 val                         val
I       C   T     GT                    T     G T           T               T   T   T
II    GCT GCA ACC ACC TGC TTC ATC ATC GCC TTC ATC GCT GCA CCC CCA GTA GAC ATC GAC GGC ATC  189
III
IV
2-4   ala ala thr thr cys phe ile ile ala phe ile ala ala pro pro val asp ile asp gly ile   63

1
I                                   G                               T                   T
II    CGT GAA CCA GTA GCA GGT TCC TTA ATC TAC GGA AAC AAC ATC ATC TCC GGT GCA GTT GTT CCC  252
III                               C                                                     A
IV                                C                                                     A
2-4   arg glu pro val ala gly ser leu ile tyr gly asn asn ile ile ser gly ala val val pro   84

1
I                       T   T                   C   C               T           T
II    TCC TCC AAC GCT ATC GGC TTG CAC TTC TAC CCA ATT TGG GAA GCA GCA TCC TTA GAC GAG TGG  315
III                                                                           C
IV                                                                            C
2-4   ser ser asn ala ile gly leu his phe tyr pro ile trp glu ala ala ser leu asp glu trp  105

1                                                                 ile     cys ala
I                                                                  TC   A TGC GCT
II    TTG TAC AAC GGT GGC CCT TAC CAA TTG GTA ATT TTC CAC TTC TTG ACA GGC GTA TTC TGC TAC  378
III                                                           C A   T
IV                                                            C A   T
2-4   leu tyr asn gly gly pro tyr gln leu val ile phe his phe leu thr gly val phe cys tyr  126

1                 gln                                                         val     tyr
I     C C   T     C G         T G   T       T       T                         G     T  A
II    TTA GGA CGT GAA TGG GAA CTA TCC TAC CGC TTA GGA ATG CGT CCT TGG ATC TGC CTA GCA TTC  441
III
IV
2-4   leu gly arg glu trp glu leu ser tyr arg leu gly met arg pro trp ile cys leu ala phe  147

1                 leu     ser
I               T T G     T T   T   A                           C
II    TCT GCA CCA GTA GCA GCA GCA ACC GCA GTA TTC TTG ATC TAC CCA ATC GGA CAA GGT TCC TTC  504
III
IV
2-4   ser ala pro val ala ala ala thr ala val phe leu ile tyr pro ile gly gln gly ser phe  168

1
I       T
II    TCA GAT GGT ATG CCC TTG GGT ATC TCC GGC ACC TTC AAC TTC ATG ATC GTG TTC CAA GCA GAA  567
III                                             A
IV                                              A
2-4   ser asp gly met pro leu gly ile ser gly thr phe asn phe met ile val phe gln ala glu  189

1
I               C                       C
II    CAC AAC ATT CTG ATG CAC CCC TTC CAT ATG TTG GGT GTA GCT GGT GTA TTC GGT GGT TCC TTG  630
III             C T
IV              C T
2-4   his asn ile leu met his pro phe his met leu gly val ala gly val phe gly gly ser leu  210

1                                                                                 ile
I                               T           T   C       G                          T
II    TTC TCT GCA ATG CAC GGT TCC TTG GTA ACC TCT TCC TTA GTT CGT GAA ACA ACC GAA AAC GAA  693
III
IV
2-4   phe ser ala met his gly ser leu val thr ser ser leu val arg glu thr thr glu asn glu  231
```

```
1
I         T   G                           T                                       A   G
II      TCA CAA AAC TAC GGT TAC AAA TTC GGA CAA GAA GAA GAA ACC TAC AAC ATC GTT GCT GCA CAC  756
III
IV
2-4     ser gln asn tyr gly tyr lys phe gly gln glu glu glu thr tyr asn ile val ala ala his  252

1
I         C               T                       T       T               C
II      GGT TAC TTC GGT CGC TTG ATC TTC CAA TAC GCA TCC TTC AAC AAC AGC CGT CAA CTG CAC TTC  819
III                           A   T                           T                 T
IV                            A   T
2-4     gly tyr phe gly arg leu ile phe gln tyr ala ser phe asn asn ser arg gln leu his phe  273

1
I                                             C               G   G   C
II      TTC CTA GCT GCA TGG CCT GTA ATC GGT ATT TGG TTT ACC GCT TTA GGT GTA AGC ACA ATG GCG  882
III
IV                                            C               G   G
2-4     phe leu ala ala trp pro val ile gly ile trp phe thr ala leu gly val ser thr met ala  294

1
I                             T                                               C
II      TTC AAC TTG AAC GGC TTC AAC TTC AAC CAA TCC ATC ATC GAC TCA CAA GGT CGT GTA ATC AAT  945
III
IV
2-4     phe asn leu asn gly phe asn phe asn gln ser ile ile asp ser gln gly arg val ile asn  315

1
I
II      ACC TGG GCT GAC ATC ATC AAC CGC GCT AAC TTG GGT ATG GAA GTA ATG CAC GAG CGT AAC GCT 1008
III
IV
2-4     thr trp ala asp ile ile asn arg ala asn leu gly met glu val met his glu arg asn ala  336

1                                                                       leu thr
I                             C   A                                   G T    C
II      CAC AAC TTC CCT CTA GAT TTG GCT GCT GGT GAA GTT GCT CCT GTT GCA ATA AGC GCT CCT GCT 1071
III                                                                   T   C
IV
2-4     his asn phe pro leu asp leu ala ala gly glu val ala pro val ala ile ser ala pro ala  357

1
I
II      ATC AAC GGT TAA 1083
III
IV
2-4     ile asn gly  *   360
```

Fig. 3. Nucleotide sequence of the *psb*AI–IV genes and their translation products. The complete sequence of *psb*AII gene is shown, with changes in the *psb*AI gene given above and the *psb*AIII and IV genes given below. *psb*AII, III and IV encode identical translation products shown in the lower line. Amino acid changes for the *psb*AI product are shown on the upper line. Sequences of *psb*AI and *psb*AII are from Curtis and Haselkorn [6]. The sequence of the *psb*AI-specific oligonucleotide and the sequence of the *psb*AII, III and IV specific oligonucleotide used for primer extension analysis are underlined under the *psb*AI and II sequences, respectively.

and run in parallel with a *psb*AIII ladder generated with the same primer, one major and four minor products were observed (Fig. 6B). The major product maps to a position 48 nucleotides upstream from the *psb*A coding regions. The minor products map to positions 55, 51, 35 and 31 nucleotides upstream from the *psb*AIII coding region.

Chromosome walking experiments

To examine the genomic linkage relationships among the four *psb*A genes, chromosome walking experiments were performed by isolating additional lambda clones that overlap the 5′ and 3′ *psb*AII–IV flanking regions and the *psb*AI 3′ flanking region. Copy-specific probes made to the 3′ end of all four genes (probes B, D, F and H, Fig. 2), and the 5′ end of *psb*AII and IV (probes

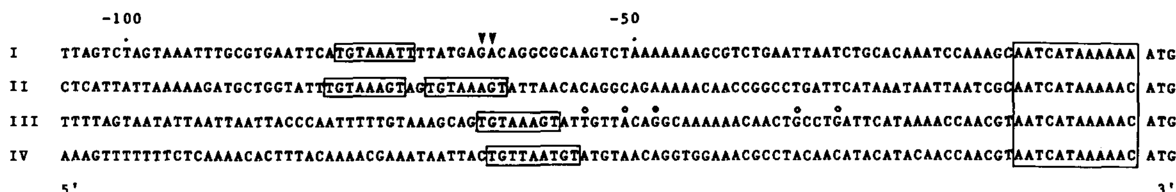

Fig. 4. Nucleotide sequence of *psb*AI–IV 5′ proximal regions. The triplets at the right indicate the translation initiation codons for the *psb*A genes. The boxed region immediately upstream from the translational start is conserved among all the *psb*A genes except for a single nucleotide change in the *psb*AI sequence. Sequences upstream that conform to an *E. coli* promoter −10 region are also boxed. The arrowheads above the *psb*AI sequence indicate the positions of transcript endpoints derived from primer extension using a *psb*AI specific oligonucleotide. The closed and open circles above the *psb*AIII sequence represent the major and four minor endpoints respectively, derived from primer extension analysis using a *psb*AII, III and IV specific oligonucleotide.

C and G, Fig. 2) were used to screen a *Sau* 3A I partial λL47.1 and a λCharon 4A *Eco* RI library. Overlapping clones were obtained for all but the 5′ flanking region of *psb*AIII. The overlapping clones isolated were mapped with restriction enzymes and the flanking probes from each gene copy to determine whether any of the *psb*A genes are closely linked. None of the restriction maps obtained for *psb*AI–IV were found to overlap (data not shown). Genomic mapping data (Fig. 1) indicate that *psb*AIII and *psb*AIV are located on separate *Eco* RI fragments greater than 20 kbp.

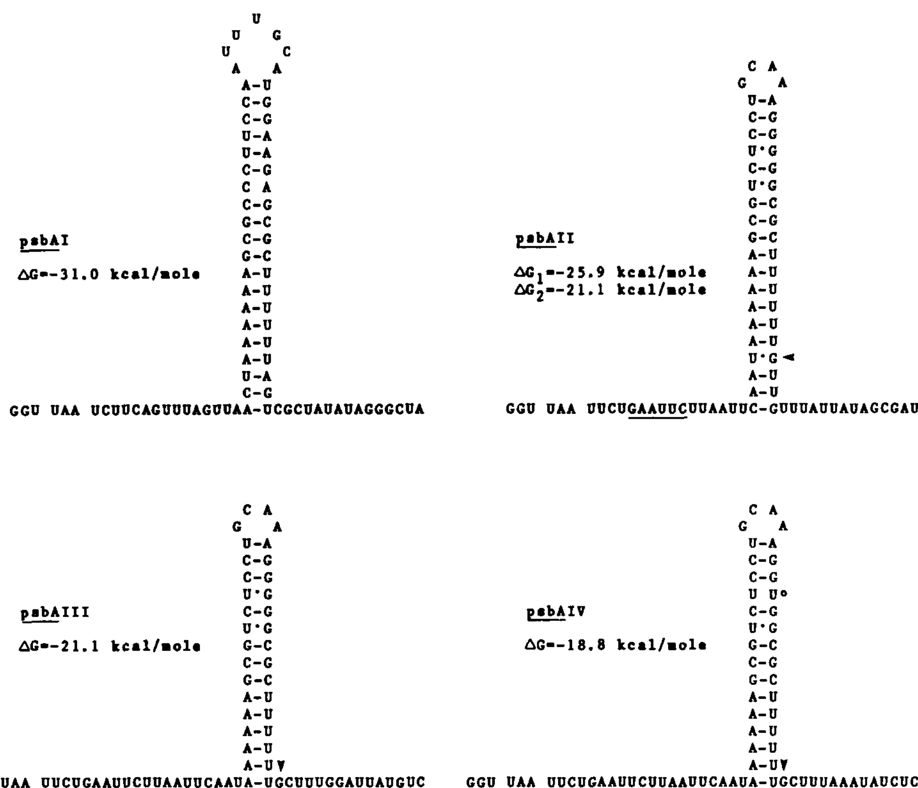

Fig. 5. Nucleotide sequences of regions 3′ to the *psb*AI–IV genes capable of forming stable stem and loop structures in mRNA. The last codon and stop triplet of the genes are shown at the left. Regions that correspond to the *Eco* RI sites in *psb*AII–IV are underlined. Arrowheads indicate where the *psb*AII, III and IV sequences diverge and the open circle indicates where the *psb*AIV sequence differs from that of *psb*AII and III in this region. The potential stem and loop structures are shown in their most stable base pairing scheme. Note that the *psb*AII sequence can also form an alternate stem and loop structure identical to that of *psb*AIII. The calculated free energy is given for each stem and loop structure [33].

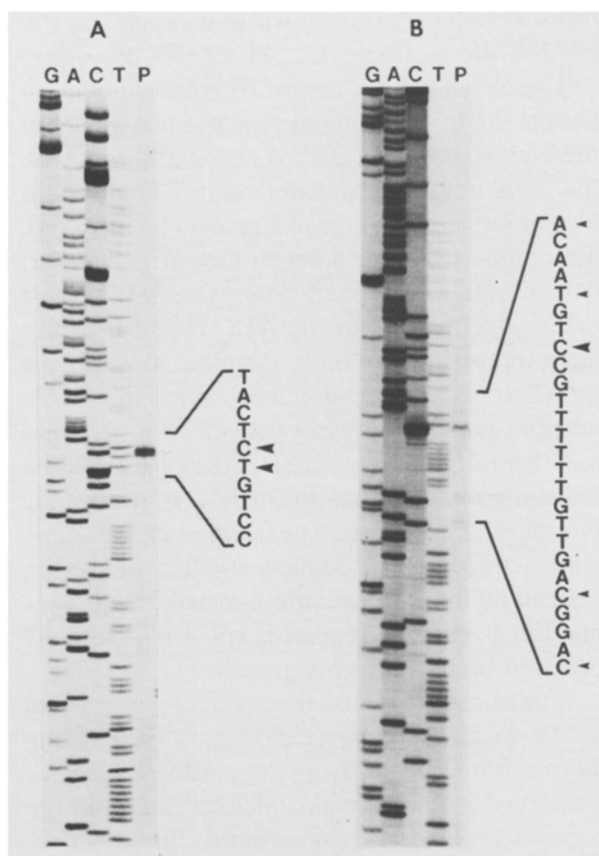

Fig. 6. Identification of *psb*A transcription initiation sites. Primers specific to the *psb*AI coding region (A) or the *psb*AII, III and IV coding regions (B) were used in primer extension assays with total *Anabaena* RNA. The extended products (lane P), were electrophoresed in parallel with a sequence ladder (G,A,C,T) generated by using the same primer on the noncoding strand of the *psb*AI DNA (A) or *psb*AIII DNA (B).

These data together with the restriction maps derived from chromosome walking demonstrate that no member of the gene family is within 20 kbp of another *psb*A gene.

The *psb*AI gene is closely linked to a sequence with structural similarity to prokaryotic insertion elements (J. Alam, J. Vrba and S. Curtis, in preparation). This putative insertion element is repeated several times in the *Anabaena* genome, but no additional members of this second repeated family are closely linked to the other *psb*A copies.

Discussion

We report here that the filamentous cyanobacterium *Anabaena* PCC 7120 contains a family of *psb*A genes with four nonidentical members. Three of the four genes, *psb*AII, III and IV, are more closely related to each other than to *psb*AI, and would encode identical polypeptides that differ from the *psb*AI product by 21 amino acids. Thus, the *psb*AII, III and IV genes may have resulted from relatively recent duplication events, or selection constraints may have prevented divergence. Genomic restriction mapping, in conjunction with chromosomal walking experiments indicate that none of the four *psb*A genes from *Anabaena* PCC 7120 is within 20 kbp of another on the chromosome. This result is in agreement with chromosome mapping of *Anabaena variabilis* ATCC 29413 in which *psb*A genes were shown to map to three different cosmid linkage groups [14]. The total number of *psb*A genes in *Anabaena variabilis* is not known. In *Anabaena* PCC 7120, *psb*AI was found to be closely linked to one member of a family of putative insertion elements. No other member of this second family maps within 10 kbp of the remaining *psb*A genes.

We have shown that the *Anabaena psb*AI gene is expressed and provided evidence that one or more of the other three *psb*A genes may also be expressed. The 5′ ends of the *psb*AI transcripts were mapped to 64–65 bp upstream from the coding region by primer extension analysis using a *psb*AI-specific oligonucleotide (Fig. 6). Preceding the *psb*AI initiation site is an eight bp region that conforms both in sequence and position to the −10 consensus sequence for *Escherichia coli* promoters [27]. A similar sequence is found at approximately the same position in *psb*AIII and IV, and twice in *psb*AII. When a primer extension assay was performed with an oligonucleotide specific for *psb*AII, III and IV transcripts, one major and several minor products were observed suggesting that one or more of the *psb*AII–IV genes is transcribed. The major product and two of the minor products differ by a few bases in length and map downstream from the region where the *E coli* −10 promoter sequences

are found in *psb*AII, III and IV. A sequence resembling the – 35 consensus sequence of *E. coli* promoters is not found at the appropriate distance 5′ to the *psb*AI transcript endpoint, nor upstream of the other *psb*A genes. A similar situation is observed with the *Anabaena rbc* L [23], *pet* F1 [1], *atp* 1 [7] and *atp* 2 [18] operons in which the transcription initiation sites are preceded by sequences that conform to the – 10 region of the *E. coli* promoter however corresponding – 35 sequences are absent. The sequences required for the promotion of transcription in *Anabaena* have not been identified, but in vitro transcription studies using RNA polymerase purified from vegetative cells suggest that the *Anabaena* enzyme prefers promoters close to the *E. coli* consensus [30].

At a similar distance downstream from all four *psb*A genes are long inverted repeats potentially capable of forming stable stem and loop structures in the transcription product (Fig. 5). The stems of the structures are composed predominantly of G/C pairs and are followed by a string of U residues. These features are characteristic of *E. coli* rho-independent transcription terminator structures [26]. Although transcription termination sequences have not been identified in *Anabaena*, the inverted repeats may serve an analogous function to those in *E. coli.* The structures in Fig. 5 show the greatest degree of base pairing possible within the inverted repeats, with the 3′ string of U residues base-paired with a string of A residues at the 5′ end of all four inverted repeats. If the inverted repeats represent transcription termination sequences, they could function in a bidirectional manner since transcripts made from the complementary strand would contain a stem and loop structure followed by a string of U residues. Immediately downstream from *psb*AI is an element that would be transcribed convergently with *psb*AI. The stem and loop structure could function to terminate both transcription units and prevent readthrough of the downstream sequences. Another potential function of the inverted repeat sequences is a role in transcript stability. Stem and loop structures at the 3′ end of some prokaryotic and chloroplast transcripts have been shown to increase the stability of the message [13, 22].

The distance from the *psb*AI transcription initiation site to the inverted repeat downstream of the coding region (1.2 kbp) correlates well with the *psb*A transcript size determined by northern blot analysis using a *psb*AI probe (1.25 kb) [6]. This probe would also detect transcripts derived from the three other *psb*A genes. Our data suggest that one or more of the *psb*AII, III and IV genes have transcript endpoints mapping close to the *psb*AI transcription initiation site. If the 3′ inverted repeats represent transcription terminators, transcripts from *psb*AII, III or IV would be indistinguishable from the *psb*AI transcript by Northern blot analysis. The fact that all four *psb*A copies are similar in sequence within the region defined by the 5′ *E. coli* promoter-like sequences and the 3′ inverted repeats is consistent with expression of all four *psb*A genes.

Immediately 5′ to the coding regions of *psb*AI–IV is a twelve bp region conserved among all four *psb*A genes. By analogy with *E. coli*, this conserved block of nucleotides is located in the region where one would expect to find a Shine-Dalgarno ribosome binding sequence (GGAGG) [32], but no such sequence is observed. Several other *Anabaena* genes also lack Shine-Dalgarno sequences at the appropriate position [5, 18, 23]. The 3′ end of the 16S rRNA gene from *Anabaena* has recently been sequenced (J. Martin and S. Curtis, unpublished). The portion of the *Anabaena* 16S rRNA corresponding to the mRNA-binding region of the *E. coli* 16S rRNA [2] was found to be identical. Thus, the lack of Shine-Dalgarno sequences preceding many *Anabaena* genes is unexpected.

All cyanobacteria thus far studied with the exception of *Anabaena* PCC 7120 have a family of closely related *psb*A genes comprised of three members. Although the *Anabaena psb*A family is larger than that observed in other cyanobacteria, the relationship of members in the family is analogous to that observed in *Synechococcus* PCC 7942 (*Synechococcus*), the other cyanobacterial species where all *psb*A genes have been characterized [12]. In *Synechococcus*, two of the three

*psb*A genes (*psb*AII and III) encode the same product (Form II) that differs from the *psb*AI product (Form I) by 25 amino acids [12]. Thus in both *Anabaena* and *Synechococcus*, two forms of the D1 proteins can be derived from a family of three or four *psb*A genes.

All three members of the *Synechococcus psb*A family are transcribed with the steady state level of *psb*AI mRNA much greater than that of *psb*AII and III under normal growth conditions [3, 12]. Our results suggest that all members of the *Anabaena psb*A family may also be expressed. Schaefer and Golden have recently shown that both Form I and II of D1 can be isolated from photosynthetic membranes of *Synechococcus* [31]. The ratio of the D1 forms varies in strains cultured at different light intensities and reflects the relative abundances of *psb*A transcripts, with the amount of Form I highest in low light, and the amount of Form II highest in high light [31]. Their results suggest that Form II may play a role in high light intensity adaptation. If Form I and II do differ in their roles in photosynthesis, the question remains as to why cyanobacteria have multiple genes that encode the Form II of D1.

Acknowledgements

This work was supported by Public Health Service grant GM32766 from the National Institutes of Health and grant DMB8614424 from the National Science Foundation. J.M.V. was supported in part by a graduate fellowship provided by the North Carolina Consortium for Research and Education in Plant Molecular Biology. Paper No 12265 of the Journal Series of the North Carolina Agricultural Research Service.

References

1. Alam J, Whitaker RA, Krogmann DW, Curtis SE: Isolation and sequence of the gene for ferredoxin I from the cyanobacterium *Anabaena* sp. strain PCC 7120. J Bact 168: 1265–1271 (1986).
2. Brosius J, Dull TJ, Sleeter DD, Noller HF: Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. J Mol Biol 148: 107–127 (1981).
3. Brusslan J, Haselkorn R: Resistance to the photosystem II herbicide diuron is dominant to sensitivity in the cyanobacterium *Synechococcus* sp. PCC 7942. EMBO J 8: 1237–1245 (1989).
4. Bryant DA: The cyanobacterial photosynthetic apparatus: comparison of those of higher plants and photosynthetic bacteria. In: Platt T, Li WKW (eds) Photosynthetic Picoplankton, pp. 423–500, Canadian Bulletin of Fisheries and Aquatic Science, vol 214. Department of Fisheries and Oceans, Ottawa (1986).
5. Curtis SE, Haselkorn R: Isolation and sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from the cyanobacterium *Anabaena* 7120. Proc Natl Acad Sci USA 80: 1835–1839 (1983).
6. Curtis SE, Haselkorn R: Isolation, sequence and expression of two members of the 32kd thylakoid membrane protein gene family from the cyanobacterium *Anabaena* 7120. Plant Mol Biol 3: 249–258 (1984).
7. Curtis SE: Genes encoding the beta and epsilon subunits of the proton-translocating ATPase from *Anabaena* sp. strain PCC 7120. J Bact 169: 80–86 (1987).
8. Dayhoff MO, Schwartz RM, Orcutt BC: A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of Protein Sequence and Structure, vol 5, pp. 345–352. National Biomedical Research Foundation, Washington, D.C. (1978).
9. Debus RJ, Barry BA, Sithole I, Babcock GT, McIntosh L: Directed mutagenesis indicates that the donor to P_{680} in photosystem II is tyrosine-161 of the D1 polypeptide. Biochemistry 27: 9071–9074 (1988).
10. Eyal Y, Goloubinoff P, Edelman M: The amino terminal region delimited by Met_1 and Met_{37} is an integral part of the 32 kDa herbicide binding protein. Plant Mol Biol 8: 337–343 (1987).
11. Gingrich JC, Buzby JS, Stirewalt VL, Bryant DA: Genetic analysis of two new mutations resulting in herbicide resistance in the cyanobacterium *Synechococcus* sp. PCC 7002. Photosynth Res 16: 83–99 (1988).
12. Golden SS, Brusslan JA, Haselkorn R: Expression of a family of *psb*A genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5: 2789–2798 (1986).
13. Gruissem W, Barkan A, Deng X-W, Stern, D: Transcriptional and postranscriptional control of plastid mRNA levels in plants. Trends Genet 4: 258–263 (1988).
14. Herrero A, Wolk P: Genetic mapping of the chromosome of the cyanobacterium *Anabaena variabilis*. J Biol Chem 261: 7748–7754 (1986).
15. Jannson CRJ, Debus J, Osiewacz HD, Gurevitz M, McIntosh L: Construction of an obligate photoheterotrophic mutant of the cyanobacterium *Synechocystis* 6803. Plant Physiol 85: 1021–1025 (1987).
16. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual, pp. 270–281. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
17. Marder JB, Chapman DJ, Telfer A, Nixon PJ, Barber J: Identification of *psb*A and *psb*D gene products, D1 and

D2, as reaction centre proteins of photosystem 2. Plant Mol Biol 9: 206–210 (1987).

18. McCarn DF, Whitaker RA, Alam JA, Vrba JM, Curtis SE: Genes encoding the alpha, gamma, delta, and four F_o subunits of ATP synthase constitute an operon in the cyanobacterium *Anabaena* sp. PCC 7120. J Bact 170: 3448–3458 (1988).
19. Michel H, Deisenhofer J: Relevance of the photosynthetic reaction center from purple bacteria to the structure of photosystem II. Biochemistry 27: 1–7 (1988).
20. Mulligan B, Schulters N, Chen L, Bogoard L: Nucleotide sequence of a multiple-copy gene for the B protein of photosystem II of a cyanobacterium. Proc Natl Acad Sci USA 81: 2693–2697 (1984).
21. Nanba O, Satoh O: 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).
22. Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF: Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell 48: 297–310 (1987).
23. Nierzwicki-Bauer SA, Curtis SE, Haselkorn R: Cotranscription of genes encoding the small and large subunits of ribulose-1,5-bisphosphate carboxylase in the cyanobacterium *Anabaena* 7120. Proc Natl Acad Sci USA 81: 5961–5965 (1984).
24. Nierwicki-Bauer SA, Haselkorn R: Differences in mRNA levels in *Anabaena* living freely or in symbiotic association with *Azolla*. EMBO J 5: 29–35 (1986).
25. Palmer JD: Comparative organization of chloroplast genomes. Ann Rev Genet 9: 325–354 (1985).
26. Platt T: Transcription termination and the regulation of gene expression. Ann Rev Biochem 55: 339–372 (1986).
27. Reznikoff WS, Siegele DA, Cowing DA, Gross CA: The regulation of transcription initiation in bacteria. Ann Rev Genet 9: 355–388 (1985).
28. Rice D, Mazur BJ, Haselkorn R: Isolation and physical mapping of nitrogen fixation genes from the cyanobacterium *Anabaena* 7120. J Biol Chem 257: 13157–13167 (1982).
29. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chaintermination inhibitors. Proc Natl Acad Sci USA 74: 5463–5467 (1977).
30. Schneider, GJ, Tumer NE, Richaud C, Borbely G, Haselkorn R: Purification and characterization of RNA polymerase from the cyanobacterium *Anabaena* 7120. J Biol Chem 262: 14633–14639 (1987).
31. Shaefer MR, Golden S: Light availability influences the ratio of two forms of D1 in cyanobacterial thylakoids. J Biol Chem 264: 7412–7417 (1989).
32. Shine J, Dalgarno L: The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71: 1342–1346 (1974).
33. Tinoco I Jr, Borer PN, Dengler B, Levine M, Uhlenbeck OC, Crothers DM, Gralla J: Improved estimation of secondary structure in ribonucleic acids. Nature 246: 40–41 (1973).
34. Trebst A: The topology of the plastoquinone and herbicide binding peptides of photosystem II in the thylakoid membrane. Z Naturforsch 41c: 240–245 (1986).
35. Tumer NE, Robinson SJ, Haselkorn R: Different promoters for the *Anabaena* glutamine synthetase gene during growth using molecular or fixed nitrogen. Nature 306: 337–341 (1983).