

Short Sequence-Paper

Nucleotide sequences of the *psaA* and the *psaB* genes encoding the reaction center proteins of Photosystem I in *Anabaena variabilis* ATCC 29413

Karin J. Nyhus, Kintake Sonoike¹, Himadri B. Pakrasi *

Plant Biology Program, Department of Biology, Washington University, Box 1137, St. Louis, MO 63130-4899, USA

(Received 1 December 1993)

Abstract

We have determined the nucleotide sequence of the *psaAB* gene cluster from the filamentous cyanobacterium *Anabaena variabilis* ATCC 29413. These genes encode the Photosystem I reaction center proteins, which are highly homologous to similar proteins in other cyanobacteria and higher plants.

Key words: Photosystem I; Reaction center protein; *psaA*; *psaB*; Cyanobacterium, filamentous; (*Anabaena variabilis* ATCC 29413)

PsaA and PsaB, encoded by the *psaA* and *psaB* genes, respectively, are large, membrane-spanning proteins which bind most of the essential chlorophylls and cofactors in the photosynthetic reaction center I complex (PS I) (reviewed in [1]). This complex mediates electron transport from plastocyanin to ferredoxin in cyanobacteria, green algae and higher plants. In order to study the structure and function of these proteins, we chose the mesophilic, filamentous cyanobacterium *Anabaena variabilis* ATCC 29413 as our experimental organism because it can grow heterotrophically and be manipulated genetically [2,3]. Using these features, we and others have generated mutant strains of *Anabaena* 29413 by inactivation of the *psaA* and *psaB* genes [4,5].

We isolated the *psaAB* gene cluster from a λ EMBL3 library of *Anabaena* 29413 DNA [5]. We then subcloned restriction endonuclease fragments from the

genes into pUC-based plasmid vectors [6]. Determination of their nucleotide sequence using the Sanger dideoxynucleotide chain termination method revealed the presence of two large open reading frames that correspond to the *psaA* and *psaB* genes, respectively. A sequence identical to the bacterial Pribnow box lies 128 nucleotides upstream of the *psaA* coding region (TATAAT, underlined in Fig. 1). Both *psaA* and *psaB* are preceded by potential ribosome binding sites (AGGAGG and AGAGG, underlined in Fig. 1). Downstream of *psaB*, there is a potential hairpin-forming sequence similar to sequences found in similar positions in the genomes of the cyanobacteria *Synechocystis* sp. PCC 6803, *Synechococcus* sp. PCC 7002, and *Synechococcus vulcanus* [7–9]: it may serve as a transcription terminator [8].

We found no additional genes in this putative transcript. No open reading frame (ORF) is located in the intergenic spacer of 122 nucleotides, although an ORF is located in the spacer of *Synechococcus* 7002 [8]. An ORF (beginning at nucleotide 4940) lies downstream of the potential transcript-terminating hairpin sequence. The 3' end of this ORF lies past the end of the region sequenced in this study.

We predicted that the *psaA* and *psaB* genes encode proteins of 752 amino acids (83.2 kDa) and 741 amino

* Corresponding author. Fax: +1 (314) 9354432.

¹ Present address: Department of Biology, Faculty of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan.

The nucleotide sequence data reported in this paper have been submitted to the Genbank-EMBL database under the accession number L26326.

acids (83.4 kDa), respectively (Fig. 1). However, native PsaB does not retain its initiating methionine [10]. Overall, these proteins are >78% identical to the corresponding proteins of the thermophilic cyanobacte-

rium *Synechococcus* sp. [11] and the higher plant *Nicotiana* [12], and many variant amino acids are conservative replacements (Fig. 2). An iron-sulfur center binding domain (Cys residues 575 and 584 of PsaA; 566 and

```

1 AATATTCTGGCTATTTTGAATTATTGTTAAATGTGTCCAACAAAGTACAAGTAAAGACTTTATAATTCAACAGAAACAAATTTCCACAT
91 AGCGCTTTCATCGCTGCTCCAGCCACACTCTGGAGAAAGCGCACTAAGTATCACTTTGTGGTTTCATGATTCTCTATCACCAGGA
181 GGAGGTCGAATGACAATTAGTCTCCGAGCGAGAGGAAAAAGCAAGAGTAATCGTCGATAAAGATCCGGTTCCACCTCATTTGAA
1 M T I S P P E R E E K K A R V I V D K D P V P T S F E
271 AAATGGGCGCAGCCAGGACACTTCGACAGAAGTCTCGTAGAGGTCCAAAAACCCACCATGGATTGGAACTCTCCAGCTCTCGCCAC
28 K W A Q P G H F D R T L A R G P K T T T W I W N L H A L A H
361 GATTTTGATACACATACAAGCGATTAGAAAGACATCTCCCGCAAAATATTGCGGCGCACTTCGGCCACCTGGCAGTTGTGACAATTGG
58 D F D T H T S D L E D I S R K I F A A H F G H L A V V T I W
451 TTGAGCGGGATGATTTTCCAGCGCGAAGTTTCTAACTACGAAGCCTGGTTAAGTGACCCGTTGAACGTGAGACCCAGTGCTCAAGTC
88 L S G M I F H G A K F S N Y E A W L S D P L N V R P S A Q V
541 GTTTGGCCCATGTTTGGACAAGACATATTGAATGGTGATGTCGGTGGGATTCCACGGTATTCAGATCACCTCAGGCTTGTTCACAGTA
118 V W P I V G Q D I L N G D V G G G F H G I Q I T S G L F Q V
631 TGGCGTGGCTGGGATACCAACTCCTTCCAGCTTTACTGCACAGCTATTGGCGGCTTGGTATTAGCGGGCTTGTTCCTTTGCGCGG
148 W R G W G I T N S F Q L Y C T A I G G L V L A G L F T F A H
721 TGGTTCCAGTATACAAACCGCTCCCAACTGGAATGGTTCCAGAATGTGGAGTCGATGTTGAACCCACCTACAAAGTCTTGTCTAGG
178 W F H Y H K R A P K L E W F Q N V E S M L N H L Q V L L G
811 TGGCGTTCCCTTAGGATGGGCTGGACACTTAATTCAGTCTCCGACCAATCAACAAGCTTTTGGATGCAGGGGTGCAGTTAAAGATATC
208 C G S L G W A G H L I H V S A P I N K L L D A G V A V K D I
901 CCCTTGCTCAGAGTTTCATCTCAACAGAGCGTTTGTATGACCTGTTCCCGGCTTGTGCTGGTTTAAACCTTTCTTCCACCTTG
238 P L P H E F I L N K S V L I D L F P G F A A G L T P F F T L
991 AATTGGGCTCAGTACGCTGACTTCTTAACCTTCAAGGGCGGTTTAAACCTGTAACCGGTGGTCTGTGGCTGACTGATATTTCATC
268 N W G Q Y A D F L T F K G G L N P V T G G L W L T D I S H H
1081 CACTTGGCGATCGCTGATTTGTTTCATCATTGCAGGACCAATACCGCACTAATCGGGTATCGGTACAGCATCAAAGAAATTCATAGAG
298 H L A I A V L F I I A G H Q Y R T N W G I G H S I K E I L E
1171 AACCACAAAGGCGCTTCAACAGTGAAGGCCACAAAGGTCTCTACGAAACCTGACCACATCTTGGCAGCTCAATTGGCAACAAACCTT
328 N H K G P F T G E G H K G L Y E N L T T S W H A Q L A T N L
1261 GCCTTTTGGGTTTCCATGACCATCATCATCGCGCACATATGTACGCGATGCCTCCCTACCCATCTAGCAACTGACTACGCAACACAG
358 A F L G S L T I I I A H H M Y A M P P Y P Y L A T D Y A T Q
1351 TTGTGTATCTTACCCACCATCTCGGATCGGCGGATTCCTCATCGTGGTGGCGCTGCTCAGCAGCAATATTTATGGTGGGGACTAC
388 L C I F T H H I W I G G F L I V G G A A H A A I F M R D Y
1441 GATCCTGCTGTTAAACAAACACGTATTGGATCGGGTGATTCGTACAGAGATGCAATTATTTCACCTGAACTGGGTTTGTATCTTC
418 D P V V N Q N N V L D R V I R H R D A I I S H L N W V C I F
1531 CTTGGCTTCCACAGCTTTGGTTTATACATCCACACGACCAATGCGTGCCTTGGGTGCTCCCAAGATATGTTCTCCGACACCAAGT
448 L G F H S F G L Y I H N D T M R A L G R P Q D M F S D T A I
1621 CAGTTGCAGCCGGTATTGCTCAGTGGGTACAAAACCTGCATACCCTAGCTCCTGGTGGCACTGCACCAACCGCTTTAGAACCTGTTAGC
478 Q L Q P V F A Q W V Q N L H T L A P G G T A P N A L E P V S
1711 TACCCCTTGGCGGTGGTACTGGCTGTAGGTGGCAAGGTCCGAATGATGCCATATCGCATTAGGTACAGCCGACTTCTGTATCCACCAC
508 Y A F G G G V L A V G G K V A M M P I A L G T A D F L I H H
1801 ATCCACGCCCTTACCATTCACGTAAACAGTTCTGATCTCTGTTGAAAGGTGATTTGTTCCGCGCAGCTCTCGTCTGATTCAGACAAAGCG
538 I H A F T I H V T V L I L L K G V L F A R S S R L I P D K A
1891 AACTTGGGCTTCCGCTTCCCTTGGCATGGTCCCGGTGCTGGTGGTACTTGTCAAGTATCCGGTTGGGATCAGTATTCCTCGGATTATTC
568 N L G F R F P C D G P G R G G T C Q V S G W D H V F L G L F
1981 TGGATGTACAACCTCCCTATCAATTGTAATTTTCCACTTCAGCTGGAAAAATGCAGTCTGATGTCTGGGAACAGTAGACGACAGGTAAT
598 W M Y N S L S I V I F H F S W K M Q S D V W G T V D A A G N
2071 GTGTCTCAGATCACTGGTGGTAACTTTGCCCAAGTGCATCAACATCAACGGCTGGTGGTGGTCTTCTGTGGGCGCAAGCTTCTCAA
628 V S H I T G G N F A Q S A I T I N G W L R D F L W A Q A S Q
2161 GTAATCAACTCCTACGGAAGTGGCTGTCCCGCTTATGGACTAATGTTCTTGGGCGCTCACTTGTATGGGCAATTCAGCTTAATGTTCTCTG
658 V I N S Y G S A L S A Y G L M F L G A H F V W A F S L M F L
2251 TTCAGTGGTGGTACTGGCAAGAATTAATGAGTCCATTGTTTGGGCGCATAATAAAGTGAAGGTAGTCTCTGCAATCCAACCCCGC
688 F S G R G Y W Q E L I E S I V W A H N K L K V A P A I Q P R
2341 GCTCTGAGCATTAACAGGTCTGTGCTGTTGGTGTAGCTCACTACCTTGGGAGGAATTGCTACAACCTGGGATCTTCCACGACAC
718 A L S I T Q G R A V G V A H Y L L G G I A T T W A F F H A H
2431 ATACTTTCGGTAGGCTAGCAATTAGTTATAAGAGTACTGAGTATTGAGTGTGAGTACTGGGTAATAACAGGACTTAGGACTCAAACCT
748 I L S V G *
2521 CAGAACTCTGGAATCTGAAACCTGATGGCTAAAGGTGAGAGGAATATCAATGGCAACAAATTTCCAAAATTTAGCCAGGATCTAGCAC
1 M A T K F P K F S Q D L A
2611 AGGACCAACTACCCGTCGATTTGGTATGCGATGGCTATGGGCAACGACTTTGAAAGCCACGATGGCATGACCAAGAAATCTTTAC
14 Q D P T T R R I W Y A M A M G N D F E S H D G M T E E N L Y
2701 AAAAGATTTTCGCTACTCACTTCGGTCACTTGGCAATCATTTCTTATGGGCTTCTAGCCTCCTGTTCCATGTAGCCTGGCAAGGTAAC
44 K I F A T H F G H L A I I F L W A S S L L F H V A W Q G N
2791 TTGAACAGTGGATTAAAGATCCTCTACACGTCGCGCAATTTGCTACGCGATTGGGACCCCACTTCGTTAAACAGCTATTGAAGCTT
74 F T Q W I K D P L H V R P I A H A I W D P H F G K P A I E A
2881 TTACCAAGCTGGTGAACGCGCCAGTAACATTTGCTTACTCTGGTGTACCAGTGGTGTACACCATCGGTACCGGCAACAAACCGG
104 F T Q A G A N G P V N I A Y S G V Y H W W Y T I G M R T N T

```

Fig. 1. Nucleotide and deduced amino acid sequences of the *psaA* (nucleotides 190–2448) and *psaB* (nucleotides 2571–4796) genes of *Anabaena* 29413 (Genbank accession number L26326). The partial sequence of the downstream ORF is shown (nucleotide 4940–5139). Nucleotides are numbered starting at the upstream *SspI* site. The Pribnow (TATA) box and potential ribosome binding sites are underlined; a potential hairpin-forming sequence is indicated by arrows.

```

2971 AACTATATACAGGTTCACTCTTCCTGTTGTTGTCGCGTCCTTGTCTTGTGTTGCTGGTGGTGCATTTACAACCCAAGTTCGCTCCTA
134 E L Y T G S V F L L L F A S L F L F A G W L H L Q P K F R P
3061 GCCTAGCTTTGGTTAAGAGTGCTGAATCTCGCTGAAACCACTTAGCAGGTTTGTTCGGCGTTAGCTCTCTAGCTTTGGCGCGTCACT
164 S L A W F K S A E S R L N H H L A G L F G V S S L A W A G H
3151 TGATTACAGTTGCTATCCCGGAATCTCGCGGACAGCAGTAGTTGGGATAACTTCTTAACCCACAGCGCCCCACCCAGCAGGCTTACAAC
194 L I H V A I P E S R G Q H V G W D N F L T T A P H P A G L Q
3241 CCTTCTTACAGGCAACTGGGGTGTTCAGCTCAAAACCTTGATACAGCAGGTACATTTTCAGTACCTCTCAAGGTTCTGGTACAGCAA
224 P F F T G N W G V Y A Q N P D T A G H I F S T S Q G S G T A
3331 TTCTGACCTTTTGGGTGGTTTCCATCCTCAAAACAGAAATCCCTGTGGTTGACAGACATAGCTCATCACCCTTGGCGATCGCTGTACTAT
254 I L T F L G G F H P Q T E S L W L T D I A H H H L A I A V L
3421 TCATTGTTGCTGGTCAATGTACCGTACCAACTTCGGTATTGGTTCACAGCATCAAGAAATGATGAATGCCAAACTTTCTTTGGCAAAC
284 F I V A G H M Y R T N F G I G H S I K E M M N A K T F F G K
3511 CTGTTGAAGGTCCTTCAATATGCCTCACCAGGCATTTATGACACCTACAACACTCTCTGCATTTCCAATTAGGTTGGCACCTAGCTT
314 P V E G P F N M P H Q G I Y D T Y N N S L H F Q L G W H L A
3601 GTTTGGGTGTTGTACCTCTTGGGTGGCGCAGCATATGTACTCTGTGCTTCTACGCATTTCATGCTAAGGACTACACAAACAGCGAG
344 C L G V V T S W V A Q H M Y S L P S Y A F I A K D Y T T Q A
3691 CACTGTACACCCACCACCAATACATAGCCATTTTCTTAATGGTGGTGTCTTTGCTCAGCGTGCTATCTTCTTAGTCCGTGACTACGATC
374 A L Y H Q Y I A I F L M V G A F A H G A I F L V R D Y D
3781 CTGAACAAAACAAAGGTAACGTGCTTGAGCGTGTGCTACAGCACAAGAAAGCGATTATCTCCACCTCAGCTGGGTATCGCTATTCTTAG
404 P E Q N K G N V L E R V L Q H K E A I I S H L S W V S L F L
3871 GCTTCCACACCTTAGGCTTATACGTCCACAGCAGTAGTAGTTGCTTTTCGGAACACCTGAAAAGCAAATCTTGATTGAGCCAGTATTGG
434 G F H T L G L Y V H N D V V V A F G T P E K Q I L I E P V F
3961 CTCAGTTCATTCAAGCTGCTCAGCGTAAGGTACTCTACGGCTTAGATACATTGCTGTCTAACCCCGATAGCGTTGGCTACACAGCCTATC
464 A Q F I Q A A H G K V L Y G L D T L L S N P D S V A Y T A Y
4051 CTAACACGCCAACGTTTGGCTACCAGGCTGGTTAGATGCCATTAACCTCTGGTACTAACTCCCTGTCTTAAACAATTGGCCCTGGCGACT
494 A N V W L P G W L D P N Y A I N S G T N S L F L T I G P G D
4141 TCTTGGTACACCATGCGATCGCCTTAGGTCTGCACACCAACCCCTCATCCTAGTCAAAGGTGCTTTGGATGCTCGTGGTTCCAAGCTGA
524 F L V H H A I A L G L H T T T L I L V K G A L D A R G S K L
4231 TGCCGGATAAAAGGACTTCGGCTATGCCCTTCCCTTGCAGCGTCCAGGCGGTGGCGGTACTTGCACATCTCCGCTTGGGACTCCTTCT
554 M P D K K D F G Y A F P C D G P G R G G T C D I S A W D S F
4321 ACCTATCTTTATTCTGGGCGTTAAATACAGTAGGTTGGGTAACATTCTACTGGCACTGGAACATTTAGGTATTTGGCAAGGTAACGTTG
584 Y L S L F W A L N T V G W V T F Y W H W K H L G I W Q G N V
4411 CTCAGTTCAACGAAATTCACCTACCTCATGGGCTGGTTCCGTGACTACCTCTGGGCTAACTCTGCTCAGTTGATCAACGGTTACAACC
614 A Q F N E N S T Y L M G W F R D Y L W A N S A Q L I N G Y N
4501 CCTACGGTGTGAACAACCTGTCTGTCTGGGCGTGGATGTTCCCTCTCGGACACCTAGTTTGGGCTACTGGCTTCATGTTCTCATCTCCT
644 P Y G N N L S V W A W M P L F G H L V W A T G F M F L I S
4591 GGAGAGGTTACTGGCAAGAGTTGATCGAAACCCTAGTTTGGGCGCACGAACGTACTCCTATCGCTAACCTAATTTCGCTGGAAAGACAAGC
674 W R G Y W Q E L I E T L V W A H E R T P I A N L I R W K D K
4681 CAGTGTCTCTCCATCGTTCAAGCTCGTGTAGTTGGTCTAGCTCACTCACCGTCGGCTATGCTCACCCTACCGCAGCTTCCTCATCG
704 P V T A L S I V Q A R V V G L A H F T V G Y V L T Y A F L I
4771 CCTCCACTGCTGGTAAGTTGCTGTGCTGGCTGCTAGTTTGTAGGTAATTAATAAAATCCCTGCGCAAGGTGGGGTCAATTTTTTT
734 A S T A G K F G *
-----> <-----
4861 AACTGATACACAAATTGAGTGTGCTAAAAATATACTAAACATTATCAAGTATTTCAAATATTTTGGAGTAATTTATTTGATGCCCAATTT
4951 ATCTGATAAAGAAATTAATAGAACAATTACATAGAACCTTGGGTTCATTTCTTAATTAATAGTGAACGAAAGACTGCGCAGCAATTATGAT
5041 TGATGCTGAATTGTCATTTGACTATGAATGGGATGAATACTATTACGTAATTATATTTTCAGGATTGCTTATATTCCTTCCGTTAACTT
5131 GATGTTGAC

```

Fig. 1 (continued).

575 of PsbB) is especially conserved among all the organisms. This domain is also conserved in the photosynthetic reaction center proteins of the anaerobic bacteria *Chlorobium* and *Helio bacterium* [reviewed in 13], even though overall homology between PsbA/PsbB and these proteins is low.

Based on hydropathy analysis, we found that PsbA and PsbB of *Anabaena* 29413 each contain eleven hydrophobic α -helices (data not shown). The eleven predicted membrane helices of PsbA and PsbB contain 32 and 27 histidine residues, respectively, which can serve as ligands to chlorophyll. These are insufficient to ligate the ~ 100 chlorophylls associated with Photosystem I [10]. However, the eleven helices contain asparagine and glutamine residues which may also serve as ligands to chlorophyll. The inclusion of these residues yields a total of 85 possible chlorophyll ligands.

Small regions of PsbA and PsbB differ significantly among organisms. First, the *Anabaena* 29413 PsbA protein contains two histidine residues (745 and 747) in a potential membrane-spanning helix not found in other PsbA proteins. Second, all sequences from thermophilic cyanobacteria, represented here by *Synechococcus* sp., contain a variable region including an insertion of four amino acids around residues 256–259 of the *Anabaena* 29413 PsbA protein (Fig. 2). Third, *Anabaena* 29413 shares two domains of PsbB with thermophilic cyanobacteria that are absent from other mesophilic cyanobacteria and higher plants (residues 308–324 and 493–498; Fig. 2).

This work was supported by a grant (GM41841) from the National Institutes of Health to H.B.P. K.S. was partially supported by a short term fellowship from the Human Frontier Science Program.

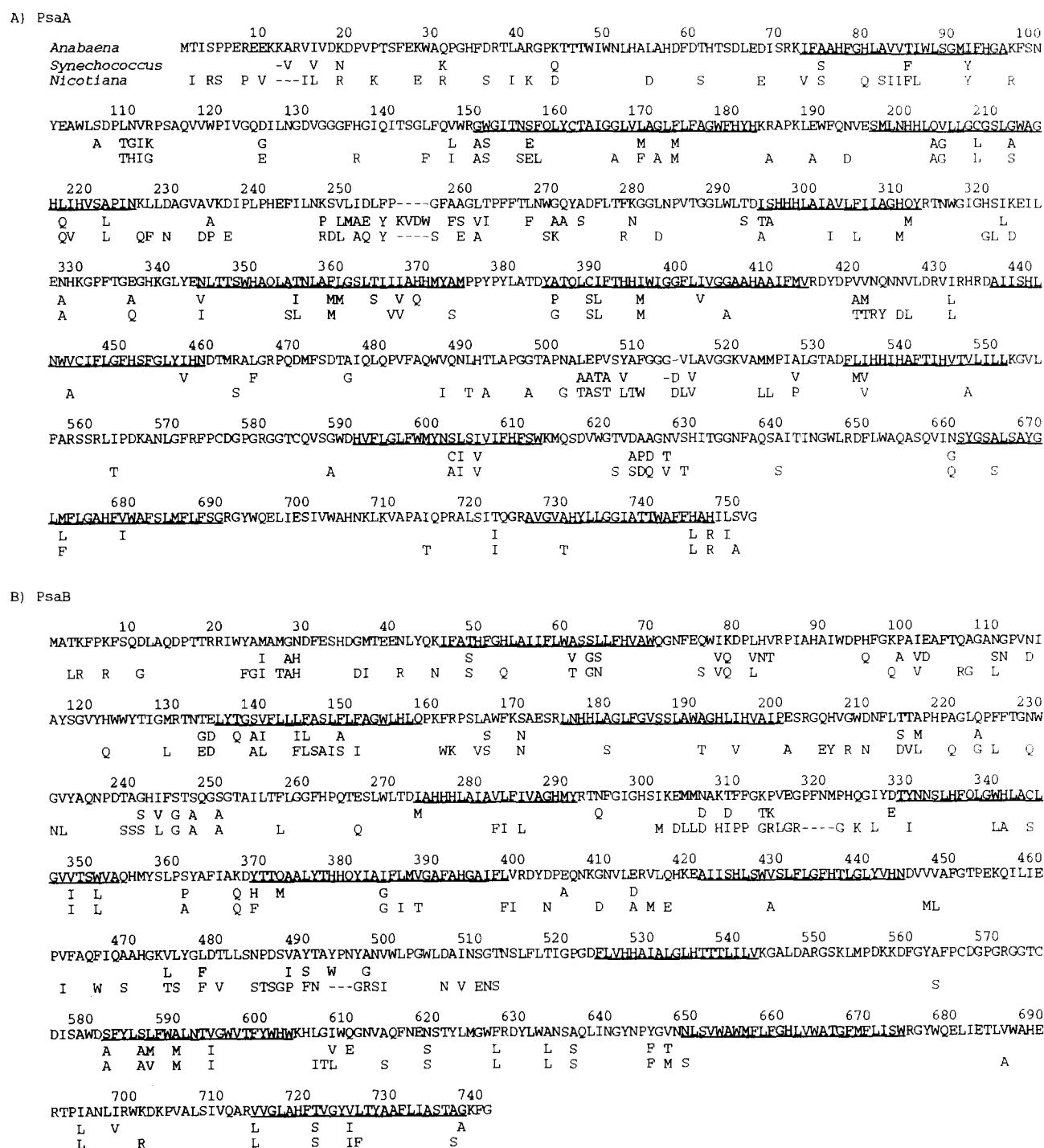


Fig. 2. Comparison of the deduced amino acid sequences of the PsaA and PsaB proteins from *Anabaena* 29413 with those from the thermophilic cyanobacterium *Synechococcus* sp. [11] and the higher plant *Nicotiana* [12]. For the latter two sequences, only residues differing from the *Anabaena* 29413 sequence are indicated. Potential membrane-spanning α -helices are underlined.

References

- [1] Bryant, D.A. (1992) in Current Topics in Photosynthesis (Barber, J., ed.), Vol. 11, pp. 501–549, Elsevier, Amsterdam.
- [2] Wolk, C.P. and Shaffer, P.W. (1976) Arch. Microbiol. 110, 145–147.
- [3] Maldener, I., Lockau, W., Cai, Y. and Wolk, C.P. (1991) Mol. Gen. Genet. 225, 113–120.
- [4] Toelge, M., Ziegler, K., Maldener, I. and Lockau, W. (1991) Biochim. Biophys. Acta 1060, 233–236.
- [5] Nyhus, K.J., Thiel T., Pakrasi, H.B. (1993) Mol. Microbiol. 9, 979–988.
- [6] Vieira, J. and Messing, J. (1987) Methods Enzymol. 153, 3–11.
- [7] Smart, L.B. and McIntosh, L. (1991) Plant Mol. Biol. 17, 959–971.
- [8] Cantrell, A. and Bryant, D.A. (1987) Plant Mol. Biol. 9, 453–468.

- [9] Shimizu, T., Hiyama, T., Ikeuchi, M. and Inoue, Y. (1992) *Plant Mol. Biol.* 18, 785–791.
- [10] Nyhus, K.J., Ikeuchi, M., Inoue, Y., Whitmarsh, J. and Pakrasi, H.B. (1992) *J. Biol. Chem.* 267, 12489–12495.
- [11] Mühlenhoff, U., Haehnel, W., Witt, H. and Herrmann, R.G. (1993) *Gene* 127, 71–78.
- [12] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, M., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K. Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) *EMBO J.* 5, 2043–2049.
- [13] Lockau, W. and Nitschke, W. (1993) *Physiol. Plant.* 88, 372–381.