Plant Molecular Biology **16**: 487–499, 1991.
© 1991 *Kluwer Academic Publishers. Printed in Belgium.*

Partial conservation of the 5′ *ndhE-psaC-ndhD* 3′ gene arrangement of chloroplasts in the cyanobacterium *Synechocystis* sp. PCC 6803: implications for NDH-D function in cyanobacteria and chloroplasts

Shawn L. Anderson[1,2] and Lee McIntosh[1,3,*]
[1]*DOE Plant Research Laboratory;* [2]*Department of Botany and Plant Pathology;* [3]*Biochemistry Department and Genetics Program, Michigan State University, East Lansing, MI 48824, USA (*author for correspondence)*

Received 16 August 1990; accepted in revised form 9 November 1990

Key words: cyanobacteria, iron-sulfur protein, *ndh*, Photosystem I, *psaC*

Abstract

The *psaC* gene, which encodes the 8.9 kDa iron-sulfur containing subunit of Photosystem I, has been sequenced from *Synechocystis* sp. PCC 6803 and shows greater similarity to reported plant sequences than other cyanobacterial *psaC* sequences. The deduced amino acid sequence of the protein encoded by the *Synechocystis psaC* gene is identical to the tobacco PSA-C sequence. In plants *psaC* is located in the small single-copy region of the chloroplast genome between two genes (designated *ndhE* and *ndhD*) with similarity to genes encoding subunits of the mitochondrial NADH Dehydrogenase Complex I. The 5′ *ndhE-psaC-ndhD*3′ gene arrangement of higher plants is only partially conserved in *Synechocystis*. An open reading frame (ORF) upstream of the *Synechocystis psaC* gene has 85% identity to the tobacco *ndhE* gene. Downstream of *psaC* there is a 273 bp ORF with 48% identity to the 5′ portion of the tobacco *ndhD* gene (1527 bp). *psaC*, *ndhE* and the region of similarity to *ndhD* are present in a single copy in the *Synechocystis* genome. Part of the wheat *ndhD* gene was sequenced and used as a probe for the presence of the 3′ portion of the *ndhD* gene. The wheat *ndhD* probe did not hybridize to *Synechocystis* or *Anabaena* sp. PCC 7120 genomic DNA, but did hybridize to *Oenothera* chloroplast DNA. These results indicate the complete *ndhD* gene is absent in two cyanobacteria, and raises the question of what role, if any, the *ndhD* gene product plays in the facultative heterotroph *Synechocystis* sp. PCC 6803.

Introduction

Photosystem (PS) I is a multi-subunit, thylakoid membrane-bound complex which catalyzes the light-dependent transfer of electrons from plastocyanin to ferredoxin. The PS I reaction center is well conserved among higher plants, algae and cyanobacteria in terms of polypeptide organization and electron transfer components [7, 13].

The PS I core complex consists of a heterodimer of homologous 83 and 82 kDa reaction center polypeptides, encoded by the *psaA* and *psaB* genes, respectively, and is responsible for binding the electron donor P_{700} and the acceptors A_0, A_1 and F_x. The terminal PS I electron acceptors F_A and F_B, both 4 iron-4 sulfur [4Fe-4S] centers [14], are associated with a single 8.9 kDa polypeptide encoded by the *psaC*

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X53842.

gene [16, 26]. The *psaC* gene from plants shows extensive sequence conservation and is located in the small single-copy region of the chloroplast genome. The higher plant PS I core complex contains at least 8 additional polypeptides encoded by the genes *psaD* through *psaK* (see [13]). The higher plant PS I genes *psaD* through *psaH* and *psaK* are encoded in the nucleus, and *psaI* and *psaJ* are encoded in the chloroplast genome. The cyanobacterial PS I core complex contains at least 7 additional polypeptides (see [13]).

Chloroplast genes encoding proteins homologous to mammalian mitochondrial components of the respiratory chain NADH Dehydrogenase Complex I have been identified in *Chlamydomonas* [41], *Marchantia* [27], tobacco [36], rice, sugar beet and broad bean [23]. Two of these genes, *ndhE* (5′) and *ndhD* (3′), flank the highly conserved *psaC* gene in tobacco [36], *Marchantia* [27], and maize [32]. In maize, *psaC* and *ndhD* are cotranscribed, probably from a *psaC* promoter, and their expression is light-regulated, whereas *ndhE* is independently transcribed [32]. The similarity among the predicted chloroplast NDH proteins and their corresponding human mitochondrial subunits is low, ranging in tobacco, for example, from 19 to 31% identity. The similarity between corresponding predicted chloroplast NDH proteins is, in general, much higher. For example, the NDH-E and NDH-D proteins of tobacco and maize are 84 and 80% identical, respectively. The *ndhC* gene has been identified in tobacco and maize plastids and in *Synechocystis* 6803 and is cotranscribed as part of the *ndhC-psbG*-ORF157/159 operon in all three organisms [34]. The NDH-C protein of *Synechocystis* 6803 is 65 and 63% identical to the NDH-C protein of maize and tobacco, respectively. Recent work [21, 24] identified strong local similarity between PSB-G and the *Escherichia coli* NADH dehydrogenase and suggests the *psbG* gene may encode a component of a thylakoid NAD(P)H-plastoquinone oxidoreductase rather than a component of Photosystem II, as *psbG* was initially designated [37].

Transcripts for eight chloroplast genes with identity to mitochondrial subunit genes have been characterized [20, 28, 32] leading to speculation that a thylakoid NAD(P)H dehydrogenase homologous to the mitochondrial complex may be functional at some point in plastid development [19]. Chloroplast localized NAD(P)H-plastoquinone oxidoreductase activity or chlororespiration has been detected in *Chlamydomonas reinhardtii* [6, 12, 18], and more recently, chlororespiration activity involving a cyanide-sensitive component has been detected in protoplasts and open-cell preparations of tobacco and in isolated chloroplasts of pea [11].

We report here the sequence of *psaC* from *Synechocystis*, sp. PCC 6803, and the partial conservation of the 5′ *ndhE-psaC-ndhD*3′ gene arrangement in this cyanobacterium. The PSA-C protein from *Synechocystis* 6803 is more similar to higher plant PSA-C sequences than to the PSA-C proteins from three other cyanobacteria. Sequence conservation extends beyond the *psaC* coding region and includes similarity to two chlorolast genes with identity to subunits of the mitochondrial NADH Dehydrogenase Complex I, *ndhE* and *ndhD*. The *Synechocystis* 6803 *ndhD* gene is incomplete as judged by sequence and Southern analysis using the 3′ portion of the wheat *ndhD* gene as a probe.

Materials and methods

DNA clones and probes

Manipulations of DNA were performed according to standard protocols [30], except as noted. A bacteriophage lambda EMBL3 library containing partial *Sau*3A I digested *Synechocystis* 6803 genomic DNA inserts [17] was probed with the 303 bp *Taq*I fragment containing the wheat *psaC* gene. A 2.7 kb *Pst*I fragment isolated from a lambda clone hybridizing to the wheat *psaC* gene was subcloned into pUC119 [39] to form the plasmid pSLA2 (Fig. 1). The plasmid pSAL3B (kindly provided by J.C. Gray and T. Dyer, Cambridge University and Plant Breeding Institute, Cambridge, England) contains a 13.9 kb

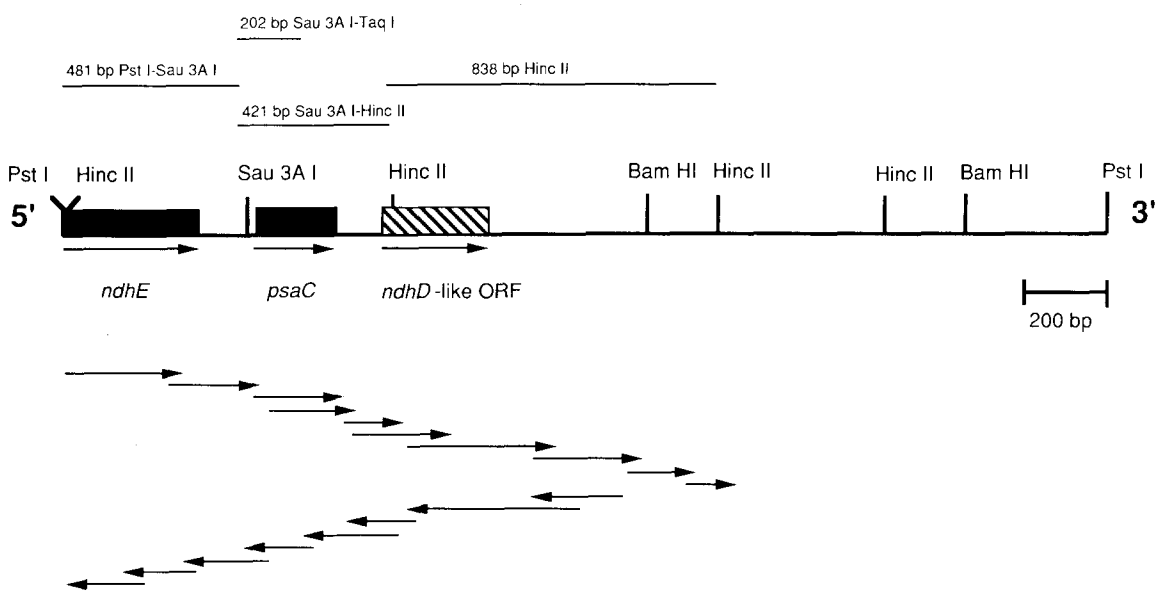

Fig. 1. The *Synechocystis* 6803 2.7 kb *Pst* I fragment of pSLA2. A partial restriction map with specific restriction fragment probes used in this study indicated above. Solid boxes represent *ndhE* and *psaC* coding regions, the hatched box represents the *ndhD*-like ORF. The direction of transcription of the coding regions is indicated below by arrows. The sequencing strategy used is indicated below.

Sal I fragment (3B) of the wheat chloroplast genome in pBR322.

Specific restriction fragments for Southern and northern hybridization were subcloned into pUC119 and then reisolated as restriction fragments by electrophoretic elution from agarose gels and purification by passage over Elutip-D columns (Schleicher and Schuell). The following restriction fragment probes were prepared by this procedure. The 306 bp *Taq* I wheat *psaC* containing fragment was subcloned as an *Eco* RI-*Hin*d III fragment from an M13 clone (kindly provided by J.C. Gray). Probes to *ndhE*, *psaC* and the region 3′ to *psaC* were subcloned from pSLA2 as 481 bp *Pst* I-*Sau*3A I, 421 bp *Sau* 3A I-*Hin*c II and 838 bp *Hin*c II fragments, respectively (see Fig. 1). A 343 bp *Bam* HI-*Pst* I fragment containing a portion of the wheat *ndhD* gene was subcloned from pSAL3B. A 550 bp *Kpn* I-*Hin*c II fragment containing the 3′ portion of the *Synechocystis* 6803 *psbA*-2 gene was isolated from the plasmid pKW1266 [17]. Probes were labeled with α-[P^{32}] dATP by the random primer method [10].

Nucleic acid isolation

Genomic DNA from *Synechocystis* 6803 was isolated from CsCl gradients [40]. *Synechocystis* 6803 RNA was isolated from light-grown, mid-log phase cultures as described [15]. *Anabaena* sp. PCC 7120 DNA was the kind gift of D. Holland and C.P. Wolk (Michigan State University, East Lansing, MI). *Oenothera hookeri* strain Johansen plastome type IV DNA was the kind gift of W.-L. Chiu and B. Sears (Michigan State University, East Lansing, MI).

Hybridization

DNA blots were routinely probed at 37 °C in hybridization buffer containing 50% (v/v) formamide, 6× SSC (1× SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 1× Denhardt's solution, 0.1% (v/v) sodium dodecyl sulfate (SDS) and 20 μg/ml denatured sonicated salmon sperm DNA. Some Southern hybridizations were performed at lower stringency, as indicated, in

hybridization buffer containing 30% (v/v) formamide with the remainder of the hybridization conditions kept the same. RNA blots were hybridized at 37 °C in buffer containing 50% formamide, 6× SSC, 1× Denhardt's solution, 0.1% (v/v) SDS, 0.03 M Tris, pH 8, 1mM EDTA and 25 μg/ml denatured sonicated salmon sperm DNA.

DNA sequencing

DNA sequencing was done by the dideoxy method of Sanger *et al.* [31]. Subclones were prepared in pUC118/119 by insertion of suitable restriction fragments or by Exonuclease III and Mung Bean Nuclease treatment of the appropriate subcloned fragments, according to the

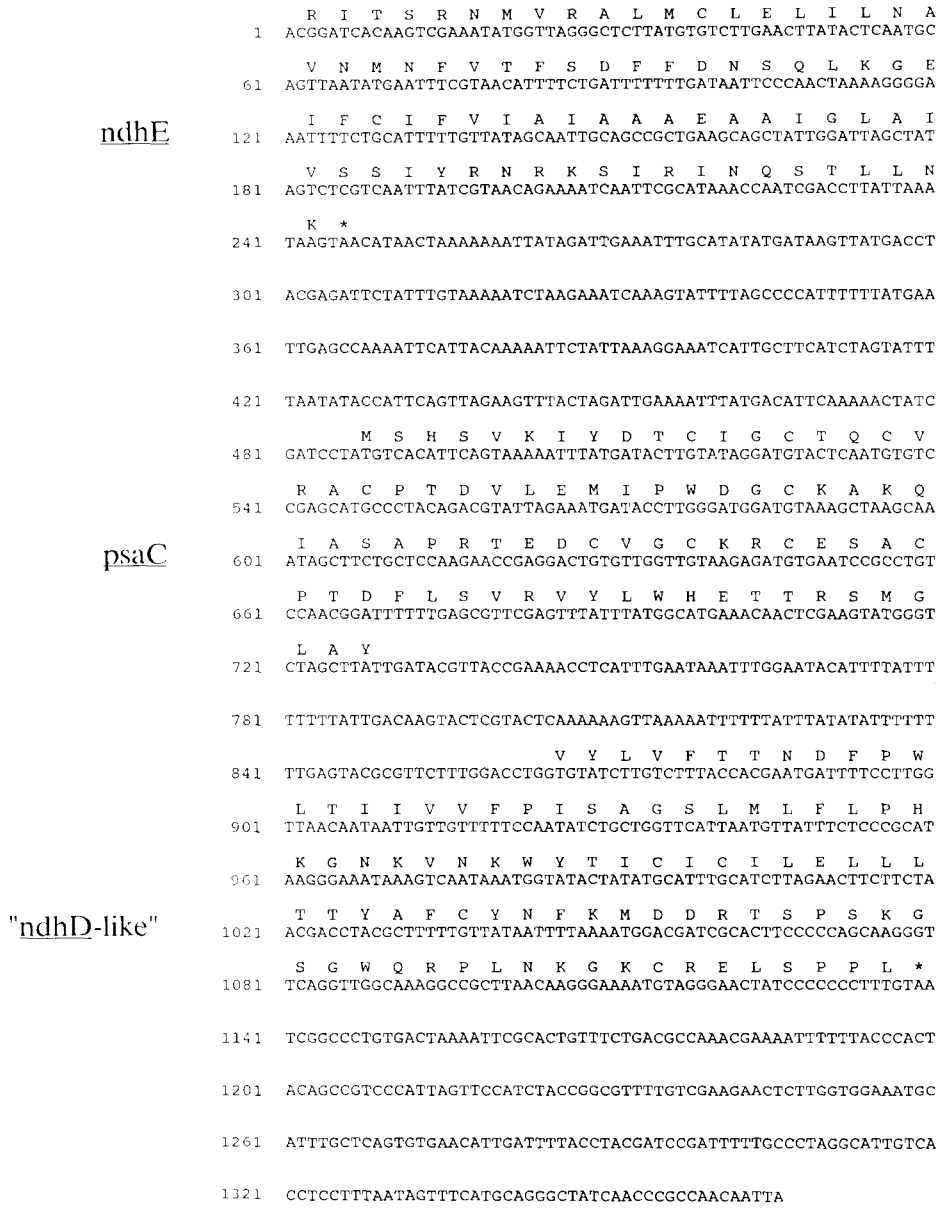

Fig. 2. Nucleotide sequence and the deduced amino acid sequence of the *Synechocystis* 6803 *ndhE* and *psaC* genes and *ndhD*-like ORF.

Bluescript Exo/Mung DNA Sequencing System Instruction Manual (Stratagene, San Diego, CA). Single-stranded templates were isolated from cells infected with the M13K07 helper phage as described [39]. Double-stranded templates were denatured for sequencing as described [43]. Sequencing of both single- and double-stranded templates was performed according to the protocol supplied with the United States Biochemical Corp. Sequenase sequencing kit. Nucleic acid and deduced protein sequences were analyzed using the Editbase Sequence Analysis Package (N. Nielson, Purdue University, West Lafayette, IN) and the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin.

Results

Cloning and sequencing of Synechocystis psaC

A lambda clone containing homology to *psaC* was identified by heterologous probing of a *Synechocystis* 6803 EMBL3 DNA library with a 306 bp *Taq* I fragment containing the wheat *psaC* gene. A partial restriction map of a 2.7 kb *Pst* I fragment subcloned from the hybridizing phage is shown in Fig. 1. The sequencing strategy used and the location of the *psaC* gene and the flanking regions containing similarity to mitochondrial NADH Dehydrogenase complex I subunit genes are indicated (Fig. 1). The DNA sequence and translation are presented in Fig. 2. The deduced *Synechocystis* 6803 *psaC*-encoded polypeptide contains 9 cysteine residues, 8 of which are arranged in the repeated CysXXCysXX-

```
                    10        20        30        40
                     .         .         .         .
Synechocystis  MSHSVKIYDTCIGCTQCVRACPTDVLEMIPWDGCKAKQIA
Nostoc           T                              A
Synechococcus                        L     V     G
S. volcanus     A T                         V     G
C. paradoxa     A T                               R N
Tobacco
Spinach
Pea                                          G
Barley
Wheat
Maize                        H
Marchantia      A A                               N


                    50        60        70        80
                     .         .         .         .
Synechocystis  SAPRTEDCVGCKRCESACPTDFLSVRVYLWHETTRSMGLAY
Nostoc          S              T         I    GA
Synechococcus   S              T         I    GA
S. volcanus     S              T         I    GA
C. paradoxa                               I              G
Tobacco
Spinach                                                  G
Pea
Barley                                         GP       A S
Wheat                                          GP       A S
Maize                                          GP       A S
Marchantia                    R                GN         S
```

Fig. 3. Comparison of the amino acid sequence of the PSA-C protein from several photosynthetic organisms. Only amino acids deviating from the *Synechocystis* 6803 sequence are indicated. The determined sequences from spinach [25] and barley [33] and the deduced sequences from *Nostoc* sp. PCC 8009, *Synechococcus* sp. PCC 7002, *Synechococcus vulcanus* [35], *Cyanophora paradoxa* [8], tobacco [16], pea, wheat [9], maize [32], and *Marchantia* [27] are included for comparison.

CysXXXCysPro pattern found in bacterial-type ferredoxins and which serve as ligands for the 2[4Fe-4S] centers F_A and F_B [42]. A comparison of the deduced and determined sequence of the PSA-C polypeptide from a number of photosynthetic organisms to the *Synechocystis* 6803 sequence is presented in Fig. 3.

psaC *flanking sequence*

Synechocystis 6803 sequences surrounding *psaC* were analyzed to determine if the 5′ *ndhE-psaC -ndhD*3′ gene arrangement found in the higher plant plastome is maintained in this cyanobacterium. A 243 bp open reading frame which extends off the 2.7 kb *Pst*I fragment is located 5′ to *psaC* (Fig. 2) and has 85 and 79% identity, respectively, to the 3′ 243 bp of the tobacco and maize *ndhE* genes (303 bp [32, 36]). Downstream of *psaC* an open reading frame of only 273 bp that has 48% identity to the 5′ end of the maize *ndhD* gene has been identified (*ndhD* -like ORF) (Fig. 2). The first 195 bp of this open reading frame is 85 and 75% identical to the tobacco and maize *ndhD* genes, respectively. A homology matrix comparison of the maize *ndhD* coding sequence to the 938 bp of *Synechocystis*

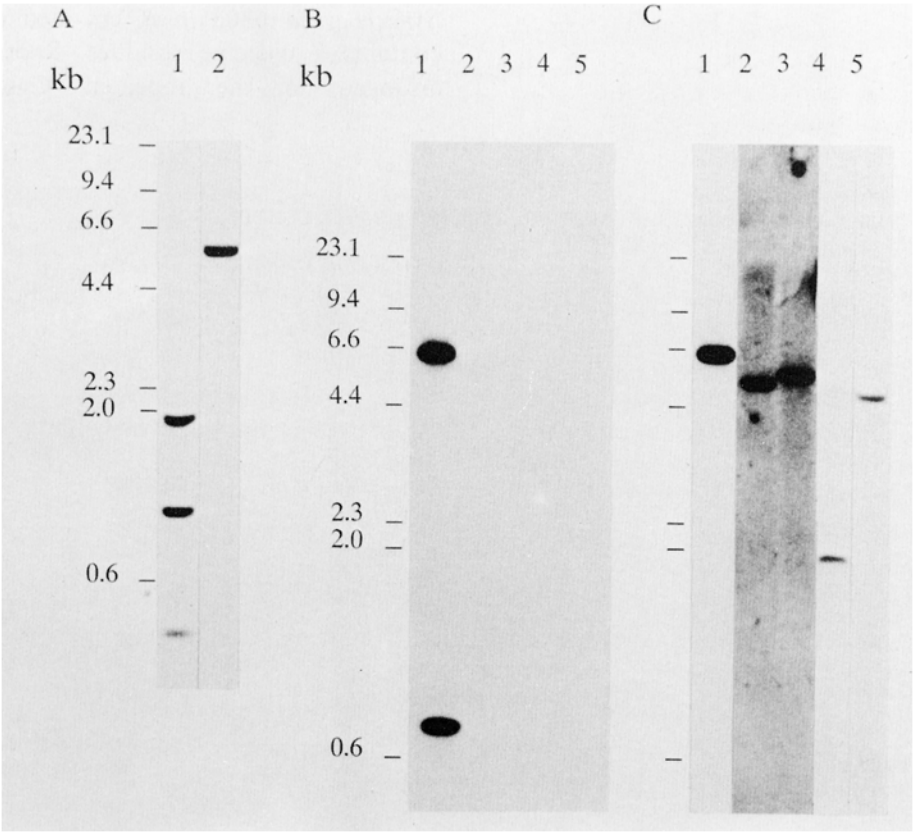

Fig. 4. Southern analysis of the *Synechocystis* 6803 genome for the presence of the *psaC*, *ndhD* and *ndhE* genes. A. *Synechocystis* 6803 DNA digested with *Hin*c II (lane 1) and *Hin*d III (lane 2) (10 μg DNA per lane) probed with the *Synechocystis* 6803 2.7 kb *Pst* I fragment at low stringency. B. Lane 1, *O. hookeri* strain Johansen plastome type IV chloroplast DNA digested with *Cla* I (1.5 μg DNA per lane). Lanes 2 and 3, *Anabaena* 7120 DNA digested with *Eco* RI and *Eco* RV, respectively (10 μg DNA per lane). Lanes 4 and 5, *Synechocystis* 6803 DNA digested with *Hin*d III and *Hin*c II, respectively (10 μg DNA per lane). Southern blot probed at low stringency with the 343 bp *Bam* HI-*Pst* I wheat *ndhD* fragment. C. Southern blot from B, stripped and reprobed at low stringency with the 421 bp *Sau* 3A I-*Hin*c II *psaC* containing fragment from *Synechocystis* 6803.

6803 sequence 3′ to the *psaC* gene starting at base 866 of the sequence did not detect sequence similarity to maize *ndhD* in the 665 bp beyond the 273 bp *ndhD*-like ORF (data not shown).

Southern analysis of the Synechocystis 6803 *genome*

Southern analysis of *Synechocystis* 6803 genomic DNA probed with the 2.7 kb *Pst* I fragment of pSLA2 detected hybridization to a single 4.9 kb *Hin*d III fragment and to 1.7, 0.9 and 0.4 kb *Hin*c II fragments (Fig. 4A). *Synechocystis* 6803 probes to *ndhE*, *psaC* and the region 3′ to *psaC* (the 481 bp *Pst* I-*Sau* 3A I, 421 bp *Sau* 3A I-*Hin*c II and 838 bp *Hin*c II fragments, respectively; see Fig. 1) all hybridized reproducibly to the 1.7 kb *Hin*c II genomic fragment (data not shown). Although sequence analysis indicates a *Hin*c II site at base 902 of the pSLA2 insert (see Fig. 1) and the *Hin*c II site has been used for subcloning, this site is apparently not digested efficiently in *Hin*c II *Synechocystis* 6803 genomic digests, resulting in the generation of a 1.7 kb *Hin*c II fragment.

To test for the presence of the remainder of the *ndhD* coding sequence elsewhere in the *Synechocystis* 6803 genome, a probe to the 3′ portion of the wheat *ndhD* gene was constructed, based on the assumption that the 5′ *ndhE-psaC-ndhD* 3′ gene arrangement is also conserved in the wheat plastome. A 343 bp *Bam* HI-*Pst* I fragment that is approximately 300 bp 3′ to the 303 bp *Taq* I fragment containing the wheat *psaC* gene was subcloned from the plasmid pSAL 3B into pUC118/119. The sequence of the 343 bp *Bam* HI-*Pst* I fragment was determined and compared to the *ndhD* sequence of maize (Fig. 5). The wheat *Bam* HI-*Pst* I fragment sequence has 86 and 95% identity, respectively, to the corresponding sequence of the tobacco and maize *ndhD* genes. The wheat *Bam* HI-*Pst* I fragment was used to probe a Southern blot of *Synechocystis* 6803 and *Anabaena* 7120 genomic DNA and *O. hookeri* chloroplast DNA at low stringency. Hybridization to two *O. hookeri* bands 6.2 and 0.7 kb in size was detected, but no hybridization to cyanobacterial DNA was detected even upon prolonged exposure (Fig. 4B). As a control, the blot was stripped and reprobed at low stringency with a 421 bp *Sau* 3A I-*Hin*c II fragment contain-

```
wheat      1 GATCCCTTAATCCAATTAAAGGAGGATTATAAATGGATAGATGTCTTCGATTTCCACTGG
maize   1694       T                          C                T

wheat     61 AGATTGGGAATCGATGGACTTTCATTAGGATCTATTTTATTGACAGGATTTATCACTACT
maize   1754                  CC

wheat    121 TTAGCTACTTTAGCAGCTTGGCCAATTACACGGAATTCGCGATTATTCTATTTCCTGATG
maize   1814                 G         GG     T           A

wheat    181 CTCGCAATGTATAGTGGTCAAATAGGATTATTTTCTTCGCGAGACCTTTTACTTTTTTTT
maize   1874    A            C                            A

wheat    241 ATCATGTGGGAGTTAGAATTAATTCCTGTTTACTTACTTTTATCCATGTGGGGGGGGAAG
maize   1934                                                          A

wheat    301 AGGCGTCTATATTCAGCTACAAAGTTTATTTTGTATACTGCAG
maize   1994          G            C
```

Fig. 5. DNA sequence comparison of the 343 bp *Bam* HI-*Pst* I wheat *ndhD* fragment to the corresponding region of the maize *ndhD* gene (bases 1694 to 2037, Fig. 2 [32]). Only those maize nucleotides differing from the wheat sequence are indicated.

ing the *Synechocystis* 6803 *psaC* gene. The *psaC* probe hybridized to the 6.2 kb *O. hookeri* band that also hybridized to the wheat *ndhD* probe, to 1.7 kb *Hin*c II and 4.6 kb *Hin*d III *Synechocystis* 6803 bands, and to 5.9 kb *Eco* RI and 6.2 kb *Eco* RV bands of *Anabaena* 7120 (Fig. 4C).

Gene expression

Northern analysis of *Synechocystis* 6803 RNA using the 2.7 kb *Pst* I fragment of pSLA2 as a probe detected hybridization to two transcripts of 0.9 and 0.5 kb (Fig. 6A). The 421 bp *Sau* 3A I-*Hin*c II probe (containing *psaC* and the first 12 codons of the *ndhD*-like ORF; see Fig. 1) hybridized only weakly above background to 0.9 and 0.5 kb transcripts (data not shown). Hybridization of probes specific to *psaC* (i.e., 202 bp *Sau* 3A I-*Taq* I fragment of pSLA2; see Fig. 1) or *ndhE* (481 bp *Pst* I-*Sau* 3A I fragment; see Fig. 1) to *Synechocystis* 6803 transcripts was not detected (data not shown). The 838 bp *Hin*c II fragment containing most of the *ndhD*-like ORF also hybridized to two transcripts of 0.9 and 0.5 kb (Fig. 6B). As a control, the filter used in Fig. 6A was stripped and reprobed with a 550 bp *Kpn* I-*Hin*c II fragment containing the 3′ portion of the *Synechocystis* 6803 *psbA-2* gene. This probe hybridized strongly to a 1.3 kb transcript (Fig. 6C).

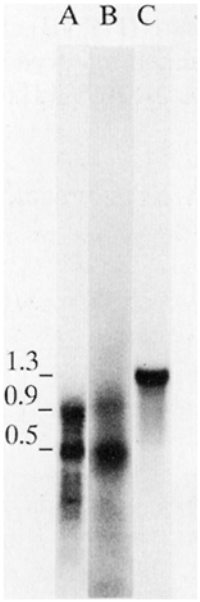

Fig. 6. Northern blot analysis. *Synechocystis* 6803 RNA (20 μg per lane) was probed with: the 2.7 kb *Pst* I fragment from *Synechocystis* 6803 (lane A); the 838 bp *Hin*c II fragment containing most of the *ndhD*-like ORF of *Synechocystis* 6803 (lane B); the filter from lane A was stripped and reprobed with the 550 bp *Kpn* I-*Hin*c II fragment of the *Synechocystis* 6803 *psbA-2* gene (lane C).

Discussion

psaC is, in general, highly conserved in a number of cyanobacteria and higher plants. The *Synechocystis* 6803 PSA-C protein, however, appears to be unique among the cyanobacteria in that it is more homolgous to PSA-C proteins from higher plants than to other cyanobacterial PSA-C proteins. The PSA-C protein of maize, spinach and tobacco are 94, 99 and 100% identical, respectively, to the protein encoded by the *Synechocystis* 6803 *psaC* gene. In contrast, the *Nostoc* sp. PCC 8009, *Synechococcus* sp. PCC 7002, *Synechococcus vulcanus* and *Marchantia* PSA-C proteins are 91, 90, 89 and 91% identical, respectively, to the *Synechocystis* 6803 PSA-C protein. The higher similarity of the *Synechocystis* 6803 PSA-C sequence to plant PSA-C sequences than to other cyanobacterial PSA-C sequences extends to the nucleotide level. The *Synechocystis* 6803 *psaC* gene is 93 and 88% identical, respectively, to the tobacco and maize *psaC* sequences, and is 72% identical on the nucleotide level to the *Synechococcus vulcanus psaC* sequence.

Analysis of the *Synechocystis*, 6803 sequence surrounding *psaC* indicates that the 5′ *ndhE -psaC-ndhD* 3′ gene arrangement of higher plants is only partially maintained in this cyanobacterium. The deduced protein encoded by the *ndhE* ORF 5′ to the *Synechocystis* 6803 *psaC* gene is 90 and 76% identical to the corresponding region of the tobacco and maize NDH-E encoded proteins, respectively. Sequence encoding the N-terminal 20 amino acid

residues of the *Synechocystis* 6803 NDH-E protein probably extends off of the 2.7 kb *Pst* I fragment and this fragment maps to the very end of the insert of the hybridizing lambda clone.

The lack of similarity to *ndhD* in the 665 bp of sequence beyond the 273 bp *ndhD*-like ORF prompted us to examine whether this represented the presence a cyanobacterial insertion sequence in the *ndhD* gene, an unusual case of a split gene in a cyanobacterium or a truncated region of sequence similarity to *ndhD*. Insertion sequences have been demonstrated in the filamentous cyanobacteria *Calothrix* sp. PCC 7601 [22] and *Anabaena* sp. strain M-131 [5]. Sequences with similarity to the insertion sequence of *Anabaena* sp. strain M-131 (IS891) were also detected in *Anabaena* 7120 [5]. If the sequence beyond the 273 bp *ndhD*-like ORF of *Synechocystis* 6803 represents a cyanobacterial insertion sequence, multiple chromosomal locations might be expected. Southern analysis of *Synechocystis* 6803 genomic DNA probed with the 2.7 kb *0st* I fragment indicates *psaC*, *ndhE*, the *ndhD*-like ORF and the sequence downstream of the *ndhD*-like ORF are present as a single copy in the genome. Sequence with similarity to the 3′ portion of the higher plant *ndhD* gene would be expected elsewhere in the cyanobacterial genome if an insertion sequence or an intron were present in the *Synechocystis* 6803 *ndhD* gene. The absence of hybridization of the wheat *Bam* HI/*Pst* I probe to the *Synechocystis* 6803 genomic DNA is consistent with the conclusion that extensive sequence identity to the 3′ portion of the higher plant *ndhD* gene is not present in the genome of *Synechocystis* 6803. The wheat *Bam* HI/*Pst* I probe also did not hybridize to *Anabaena* 7120 genomic DNA indicating *Synechocystis* 6803 is not unique among the cyanobacteria in its lack of a complete *ndhD* gene. No sequence similarity to *ndhE* or *ndhD* can be identified in the limited reported *Synechococcus vulcanus psaC* flaking sequence [35].

Analysis of the expression of *psaC*, *ndhE* and the *ndhD*-like ORF in *Synechocystis* 6803 was complicated by the small size of the gene-specific probes required and the difficulty in detecting specific transcript pools in this prokaryote. Strong hybridization of the *psbA-2* probe to a 1.3 kb transcript indicates the RNA is intact and messages larger than 1 kb can be readily detected. We were unable to detect transcripts hybridizing to *psaC*-specific probes, but did detect weak hybridization of the 421 bp probe containing the *psaC*-coding region and the first 12 codons of the *ndhD*-like ORF to 0.9 and 0.5 kb transcripts which also hybridized to the 838 bp *Hinc* II and 2.7 kb *Pst* I fragments of pSLA2. Neither transcript is large enough to code for a complete NDH-D protein (515 and 509 codons in maize and tobacco, respectively [32,36]), but the larger transcript is long enough to code for both a PSA-C and a truncated NDH-D protein. The 0.5 kb transcript may represent a processing product encoding PSA-C or a truncated NDH-D protein. In maize chloroplasts, a *ndhD*-specific probe hybridized to 2.1 and 1.7 kb transcripts and a *psaC*-specific probe hybridized to 2.1, 0.5 and 0.4 kb transcripts [32]. In the study presented here, hybridization of an *ndhE* specific probe to a specific *Synechocystis* 6803 mRNA was not detected. Schantz and Bogorad [32] observed a 0.5 kb transcript hybridizing to a *ndhE*-specific probe, but the *ndhE* transcript pool in maize was small and difficult to detect (see Fig. 4 [32]), preventing a developmental study of the expression of *ndhE* in maize chloroplasts. In contrast, all of the tobacco *ndh* genes are actively transcribed in leaves in both the light and dark, and the transcripts are estimated to be as abundant as some of the highly transcribed ribosomal protein genes [20]. A probe to the tobacco *ndhD* gene hybridized to 2.5 and 2.0 kb major transcripts. The tobacco *ndhE* gene is apparently cotranscribed with *ndhA* and a *ndhE*-specific probe hybridized to three transcripts of 4.0 2.5 and 0.9 kb.

The apparent incomplete *ndhD* gene in *Synechocystis* 6803 and the inability to detect hybridization of a maize *ndhD* probe in *Euglena* or *Chlamydomonas* [32] prompts examination of the requirement of the *ndhD* encoded polypeptide in this facultative heterotroph [4] and perhaps in chloroplasts, as well. The *Synechocystis* 6803 *ndhD*-like ORF consists of 91 codons and is

predicted to encoded a protein of 10 kDa. The *Synechocystis* 6803 NDH-D-like protein is predicted to use GTG as an initiation codon, similarily, the NDH-C protein of *Synechocystis* 6803 is also predicted to use GTG as an initiation codon [38]. Alignment of the predicted amino acid sequence of the *Synechocystis* 6803 *ndhD*-like ORF with the predicted NDH-D sequences of maize and tobacco and the predicted protein sequence of the human mitochondrial homologue of NDH-D, ND4, reveals some variability in the length of the encoded proteins manifested primarily at the N-terminus (Fig. 7). The maize NDH-D protein contains 6 additional codons relative to tobacco NDH)D, 9 codons relative to the *Synechocystis* 6803 NDH-D-like protein and 33 additional codons relative to the human ND4 protein. The first 20% of the chloroplast NDH-D protein sequences is the least conserved with respect to the human mitochondrial ND4 protein relative to the remainder of the chloroplast NDH-D protein sequence. Interestingly, it is only the 5′ approximately 20% of the *ndhD* gene which is maintained in *Synechocystis* 6803. The ND proteins of mitochondria are believed to be part of the hydrophobic or HP fragment of the NADH Dehydrogenase Complex I, but their function is still unknown [29]. Our ability to asses the significance of an incomplete *ndhD* gene in *Synechocystis* 6803 is limited by our lack of knowledge of the function of the intact *ndhD* gene product in plastids, however, several possible interpretations exist. The *ndhD*-like ORF may represent the disfunctional evolutionary remnant of a once functional gene. In this case one would expect accumulation of mutations in this region of se-

```
S  .........VYLV TTND      I  F  S    ML        VNK    C  I  L  T      N
M  MGFSGQNVSCLYYFIMSYFPWLTILVVLPIFAGSLIFFLPHKGNKIVRWYTIAICLLEFLLMTYAFCYHF
T  ......M QVYLV TTN          F                RVI     C  I  L  T
H  ................................M KLIV TIMLLPLT LSKKHMIWINTTTHSLIISII


S  KMD RTSPS GSGWQRPLNKGKC ELSP  ....................................
M  QLEDPLIQLKEDSKWIDVFDFHWRLGIDPLSLGSILLTGFITTLATLAAWPVTRNSQLFYFLMLAMYSGQ
T   SD      V  Y   NF          G  I  P                    D  R  H
H  P LFFNQINNNLFSCSPT SSDPLTTPLLMLTTWL PLTIMASQRH SSE LS .KK YLSMLISLQISL


M  IGLFSSRDLLLFFIMWELELIPVYLLLSMWG.GKRRLYSATKFILYTAGGSIFFLIGVLGMGLYGSNEPG
T    S                        C    .  K                  V  L  M    LA       T
H  IMT TATE IM YIFF TT   TLAIITR  NQPE  NAG Y LF  LV  LPL  ALIYTHNTLGSLNI


M  LDLERLINQSYPTTLEILLYFGFLIAYAVKLPIIPLHTWLPYTHGEAHYSTCMLLAGILLKMGAYGLIRV
T   NF TSV      VV   I  I  F  F   S          D
H   L TLTAQELSNSWANN MWLAYTM FM  M LYG  L   KA V  PIAGS V  AV   L G  MM L


M  NMELLPHAHYLFSPWLVIIGAVQIIYAASTSLGQRNFKKRIAYSSVSHMGFIIIGIGSITNIGLNGAILQ
T          SI     M   TI      L          L                          S L DT     L
H  TLI N LTKHMAY F .LSLWGMI TSS CL Q DLK LI     I   ALVVTA LIQ PWSFT  VIL


M  ILSHGFIGATLFFLAGTACDRMRLVYLEELGGISIPMPKIFTMFSSFSMASLALPGMSGFVAELVVFFGL
T   I       A       TY  I     D M   A     M                                I    I
H  M A  LTSSL  C  NSNYE THSRIMILSQ LQTLL LMAFWWLLA L N    PTINLLG  S LVTT


M  ITSPKFMLMPKMLITFVMAIGMILTPIYLLSMLRQMFYGYKLFHVPNKNFVDSGPRELFLLICIFLPLIG
T    GQ YL I  I                S  S          NA KDS F          S S    V
H  FSWSNIT LLTG NML T LYSLYMFTITQWGSLTHHINNMKPSFTRE TLMFMHLSPI  LSLNPDI T


M  IGIYPDLVLSLSVDRVEVLLSNYTK
T        F·   A  K   I   FFYR
H  GFSS
```

Fig. 7. Comparison of the deduced amino acid sequence of the *Synechocystis* 6803 NDH-D-like protein (S) with that of the maize NDH-D protein (M) [32], the tobacco [20] NDH-D protein (T) and the human mitochondrial homologue of NDH-D, ND4 (H) [3]. Only those amino acids deviating from the deduced maize NDH-D sequence are indicated.

quence at a high rate. However, the *ndhD*-like ORF is relatively well conserved with 48% identity to the tobacco *ndhD* gene over 243 bp and 85% identity over the first 195 bp of the *ndhD*-like ORF. The relatively high sequence conservation of the *ndhD*-like ORF may have resulted because the *ndhD*-like ORF encodes a functional component of an NADH dehydrogenase complex or some other protein complex in the cyanobacterium. Indeed, probes to the *ndhD*-like ORF hybridize to specific *Synechocystis* 6803 transcripts, although, a specific protein product or its function in *Synechocystis* 6803 has yet to be identified. Finally, the sequence downstream of *psaC* may be conserved because it functions in *psaC* message stability, however, no strong candidates for stem loop structures or inverted repeats were identified in this region of sequence.

In *Anabaena variabilis* ATCC 29413 NADPH oxidation is performed by ferredoxin-$NADP^+$ oxidoreductase, whereas NADH oxidation is performed by a separate NADH dehydrogenase [2]. The thylakoid-bound respiratory NADH dehydrogenase isolated from *Anabaena* 29413 is more similar to bacterial than to mitochondrial NADH dehydrogenases on the basis of subunit number, sensitivity to inhibitors, and prosthetic group [1]. Although the high sequence conservation of the *Synechocystis* 6803 *ndhC*, *ndhD*-like and *ndhE* sequences imply a functional role, the presence of an *Anabaena* 29413 type NADH dehydrogenase in *Synechocystis* 6803, as yet unconfirmed, may obviate the need for a complete *ndhD* gene in this cyanobacterium.

It has been suggested that the chloroplast *ndh* genes may have arose by transposition events from the mitochondrial to the chloroplast genome [36]. The presence of the conserved *ndh* sequences in *Synechocystis* 6803, however, are evidence that the *ndh* genes of chloroplasts are probably derived from an ancestor common to both higher-plant chloroplasts and the cyanobacterium *Synechocystis* 6803.

Acknowledgements

We would like to thank J.C. Gray and T. Dyer for the generous gift of two wheat chloroplast clones, and D. Holland, C.P. Wolk, W.-L. Chiu and B. Sears for the generous gifts of genomic DNAs. The technical advice of J. Fitchen and the help of L. Smart with the DNA sequence analysis programs are gratefully acknowledges. This research was supported in part by contract DEACO2-76ERO)1338 (to L.M.) and M.S.U.-R.E.F. (to L.M.).

References

1. Alpes I, Scherer S, Boger P: The respiratory NADH dehydrogenase of the cyanobacterium *Anabaena variabilis*: purification and characterization. Biochim Biophys Acta 973: 41–46 (1989).
2. Alpes I, Schrautemeier B, Scherer S, Boger P: Different enzymes involved in NADH- and NADPH-dependent respiration in the cyanobacterium *Anabaena variabilis*. FEMS Microbiol Lett 26. 147–151 (1985).
3. Anderson S, Bankier AT, Barell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young GI: Sequence and organization of the human mitochondrial genome. Nature 290: 457–465 (1981).
4. Astier C, Elmorjani K, Meyer I, Joset F, Herdman M: Photosynthetic mutants of the cyanobacterium *Synechocystis* sp. strains PCC 6714 and PCC 6803: sodium p-hydroxymercuribenzoate as a selective agent. J Bact 158: 659-664 (1984).
5. Bancroft I, Wolk CP: Characterization of an insertion sequence (IS891) of novel structure from the cyanobacterium *Anabaena* sp. strain M-131. J Bact 171: 5949–5954 (1989).
6. Bennoun P: Evidence for a respiratory chain in the chloroplast. Proc Natl Acad Sci USA 79: 4352–4356 (1982).
7. Bryant DA: The cyanobacterial photosynthetic apparatus: comparison of those of higher plants and photosynthetic bacteria: In: Platt T, Li WKW (eds) Photosynthetic Picoplankton, Canadian Bulletin of Fisheries and Aquatic Sciences, vol. 214, pp. 423–500. Dept. of Fisheries and Oceans, Ottawa (1987).
8. Bryant DA, Rhiel E, De Lorimier R, Zhou J, Stirewalt VL, Gasparich GE, Dubbs JM, Snyder W: Analysis of phycobilisome and photosystem I complexes of cyanobacteria. In: Baltscheffsky M (ed) Current Research in Photosynthesis, vol. II, Proceedings of the VIIIth International Conference on Photosynthesis,

Stockholm, pp. 1–9. Kluwer Academic Publishers, Dordrecht (1990).
9. Dunn PPJ, Gray JC: Localization and nucleotide sequence of the gene for the 8 kDa subunit of photosystem I in pea and wheat chloroplast DNA. Plant Mol Biol 11: 311–319 (1988).
10. Feinberg AP, Vogelsten BA: A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13 (1983).
11. Garab G, Lajko F, Mustardy L, Marton L: Respiratory control over photosynthetic electron transport in chloroplasts of higher-plant cells: Evidence for chlororespiration. Planta 179: 349–358 (1989).
12. Godde D, Trebst A: NADH as electron donor for the photosynthetic membrane of *Chlamydomonas reinhardii*. Arch Microbiol 127: 245–252 (1980).
13. Golbeck JH, Bryant DA: Photosystem I. Current Topics in Bioenergetics (1990). (In press).
14. Golbeck JH, McDermott AE, Jones WK, Kurtz DM: Evidence for the existense of [2Fe-2S] as well as [4Fe-4S] clusters among F_A, F_B and F_X. Implications for the structure of the photosystem-I reaction center. Biochim Biophys Acta 891: 94–98 (1987).
15. Golden SS, Brusslan J, Haselkorn R: Genetic engineering of the cyanobacterial chromosome. Meth Enzymol 153: 215–231 (1987).
16. Hayashida N, Matsubayashi T, Shinozaki K, Sugiura M, Inoue K, Hiyama T: The gene for the 9 kd polypeptide, a possible apoprotein for the iron-sulfur centers A and B of the photosystem I complex, in tobacco chloroplast DNA. Curr Genet 12: 247–250 (1987).
17. Jansson C, Debus RJ, Osiewacz HD, Gurevitz M, McIntosh L: Construction of an obligate photoheterotrophic mutant of the cyanobacterium *Synechocystis* 6803. Plant Physiol 85: 1021–1025 (1987).
18. Maione T, Gibbs M: Association of the chloroplastic respiratory and photosynthetic electron transport chains of *Chlamydomonas reinhardii* with photoreduction and the oxyhydrogen reaction. Plant Physiol 80: 364–368 (1986).
19. Marder JB, Barber J: The molecular anatomy and function of thylakoid proteins: Plant Cell Environ 12: 595–614 (1989).
20. Matsubayashi T, Wakasugi T, Shinozaki K, Yamaguchi-Shinozaki K, Zaita N, Hidaka T, Meng BY, Ohto C, Tanaka M, Kato A, Maruyama T, Sugiura M: Six chloroplast genes (*ndhA-F*) homologous to human mitochondrial genes encoding components of the respiratory chain NADH dehydrogenase are actively expressed: determination of the splice sites in *ndhA* and *ndhB* pre-mRNAs: Mol Gen Genet 210: 385–393 (1987).
21. Mayes SR, Cook KM, Barber J: Nucleotide sequence of the second *psbG* gene in *Synechocystis* 6803: Possible implications for *psbG* function as a NAD(P)H dehydrogenase subunit gene: FEBS Lett 262. 49–54 (1990).
22. Mazel D, Castets A-M, Houmard J, Tandeau de Marsac N: Cyanobacterial insertion elements: characterization and potential. VI Int Symp Photosynthetic Prokaryotes, p. 227 (abstract) (1988).
23. Meng BY, Matsubayashi T, Wakasugi T, Shinozaki K, Sugiura M, Hirai A, Mikami T, Kishima Y, Kinoshita T: Ubiquity of the genes for components of a NADH Dehydrogenase in higher plant chloroplast genomes. Plant Sci 47: 181–184 (1986).
24. Nixon PJ, Gounaris K, Coomber SA, Hunter CN, Dyer TA, Barber J: *psbG* is not a photosystem two gene but may be an *ndh* gene. J Biol Chem 264: 14129–14135 (1989).
25. Oh-oka H, Takahashi Y, Kuriyama K, Saeki K, Matsubara H: The protein responsible for center A/B in spinach photosystem I: Isolation with iron-sulfur cluster(s) and complete sequence analysis. J Biochem 103: 962–968 (1988).
26. Oh-oka H, Takahashi Y, Wada K, Matsubara H, Ohyama K, Ozeki H: The 8 kDa polypeptide in photosystem I is a probable candidate of an iron-sulfur center protein coded by the chloroplast gene *frxA*. FEBS Lett 218: 52–54 (1987).
27. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y,, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H: Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA: Nature 322: 572–574 (1986).
28. Ohyama K, Kohchi T, Sano T, Yamada Y: Newly identified groups of genes in chloroplasts: TIBS 13: 19–22 (1988).
29. Ragan CI: Structure of NADH-ubiquinone reductase (Complex I). In: Lee CP (ed) Current Topics in Bioenergetics, pp. 1–36: Academic Press, San Diego (1987).
30. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
31. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors: Proc Natl Acad Sci USA 74: 5463–5467 (1977).
32. Schantz R, Bogorad L: Maize chloroplast genes *ndhD*, *ndhE*, and *psaC*. Sequences, transcripts and transcript pools. Plant Mol Biol 11: 239–247 (1988).
33. Scheller HV, Svendsen I, Moller BL: Amino acid sequence of the 9-kDa iron-sulfur protein of photosystem I in barley. Carlsberg Res Commun 54: 11–15 (1989).
34. Scheller HV, Svendsen I, Moller BL: Subunit composition of photosystem I and identification of center X as a [4Fe-4S] iron-sulfur cluster. J Biol Chem 264: 6929–6934 (1989).
35. Shimizu T, Hiyama T, Koike H, Inoue Y: Nucleotide sequence of the *psaC* gene of the cyanobacterium *Synechococcus vulcanus*. Nucleic Acids Res 18: 3644 (1990).
36. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T,

Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5: 2043–2049 (1986).

37. Steinmetz AA, Castroviejo M, Sayre RT, Bogorad L: Protein PSII-G: An additional component of Photosystem II identified through its plastid gene in maize. J Biol Chem 261: 2485–2488 (1986).
38. Steinmueller K, Ley AC, Steinmetz AA, Sayre RT, Bogorad L: Characterization of the *ndhC-psbG--ORF157/159* operon of maize plastid DNA and of the cyanobacterium *Synechocystis* sp. PCC 6803: Mol Gen Genet 216: 60–69 (1989).
39. Vieira J, Messing J: Production of single-stranded plasmid DNA. Meth. Enzymol 153: 3–11 (1987).
40. Williams JGK: Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. Meth Enzymol 167: 766–778 (1988).
41. Wu M, Nie ZQ, Yang J: The 18-kD protein that binds to the chloroplast DNA replicative origin is an iron-sulfur protein related to a subunit of NADH dehydrogenase. Plant Cell 1: 551–557 (1989).
42. Yasunobu KT, Tanaka M: The isolation and primary structures of various types of ferredoxin. Meth Enzymol 69: 228–239 (1980).
43. Zhang H, Scholl R, Browse J, Somerville C: Double stranded DNA sequencing as a choice for DNA sequencing. Nucleic Acids Res 16: 1220 (1988).