Plant Cell Physiol. 32(7): 969–981 (1991)

Structure of a Co-Transcribed Gene Cluster, *ndh1-frxB*[3]*-ndh6-ndh4L*, Cloned from the Filamentous Cyanobacterium *Plectonema boryanum*[4]

Yasuhiro Takahashi[1,5], Fumio Shonai[1,6], Yuichi Fujita[1], Takayuki Kohchi[2], Kanji Ohyama[2] and Hiroshi Matsubara[1]

[1] *Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka, 560 Japan*
[2] *Department of Agricultural Chemistry, Faculty of Agriculture, Kyoto University, Kyoto, 606 Japan*

The *frxB* gene found in chloroplast genomes encodes a bacterial ferredoxin-like (Fd-like) Fe-S protein, whose function has not been established. Using an *frxB* gene probe from liverwort, we have isolated and determined the nucleotide sequence of a 4.6-kb fragment of DNA that contains the *frxB* gene from the filamentous cyanobacterium *Plectonema boryanum*. The *frxB* gene from *Plectonema* encodes a protein of 194 amino acids that shows 66% sequence identity to that encoded by liverwort chloroplasts. The sequence contains two clusters of cysteine residues (C-X-X-C-X-X-C-X-X-X-C-P), typical of bacterial ferredoxins that chelate two [4Fe-4S] clusters. Such sequences are found in highly conserved regions in the seven available sequences of *frxB*-related proteins. In addition, we identified three ORFs homologous to the chloroplast *ndh1, ndh6* and *ndh4L* genes that encode the components of a putative NADH dehydrogenase. They show 60%, 50% and 65% amino acid sequence identity to the respective gene products from liverwort chloroplasts, but they show a lower degree of similarity to the mitochondrial homologues. The *ndh1, frxB, ndh6* and *ndh4L* genes, occurring as single genomic copies, are tightly linked to one another, and they are co-transcribed as a 2.8-kb mRNA which was identified by Northern blot analysis with gene-specific probes. These four genes occur in the same order in the chloroplast genomes of liverwort, tobacco and rice as they do in *Plectonema*. Therefore, the organization of *ndh1-frxB-ndh6-ndh4L* may represent a case of clustering and co-transcription of functionally related genes in both cyanobacteria and chloroplasts. These findings lend a support to the hypothesis that cyanobacteria and chloroplasts contain an enzyme complex that is closely related to the mitochondrial complex I, although such a highly organized enzyme complex has not been identified to date in an extensive survey of cyanobacterial NAD(P)H dehydrogenases.

Key words: Chloroplast DNA — Complex I — Cyanobacterial DNA — Iron-sulfur protein — NAD(P)H dehydrogenase — *Plectonema boryanum*.

The complete nucleotide sequences are known of the chloroplast genomes of liverwort (*Marchantia polymorpha*; Ohyama et al. 1986), tobacco (*Nicotiana tabacum*; Shinozaki et al. 1986), and rice (*Oryza sativa*; Hiratsuka et al. 1989). Genes encoding bacterial Fd-like proteins (*frxA* and *frxB*) have been found in these genomes (Ohyama et al. 1988, Kohchi et al. 1988). These Fe-S proteins are not related to the soluble chloroplast-type Fd which functions in photosynthetic electron transport. *frxA* encodes a 9-kDa protein of the photosystem I complex which carries two [4Fe-4S] clusters (centers A and B; Oh-oka et al. 1987, 1988), while the function of the *frxB* protein has not been established. The product of expression of the *frxB* gene has been identified at the protein level in chloroplasts of *Chlamydomonas reinhardtii* (Wu et al. 1989), tobacco and spinach (Lin and Wu 1990). The prod-

Abbreviations: Fd, ferredoxin; ORF, open reading frame.

[3] For renaming "*frxB*" two proposals have been made: "*ndhK*", by dePamphilis and Palmer (1990); and "*ndhI*", by Dupuis et al. (1991).

[4] The sequence data will appear in the DDBJ/EMBL/GenBank Nucleotide Sequence Databases under the accession number D01014.

[5] To whom correspondence should be addressed.

[6] Present address: Institute for Fundamental Research, Research Center, Suntory Ltd., Shimamoto-cho, Mishima-gun, Osaka, 618 Japan.

uct appears to be a membrane protein that is bound to the peripheral portion of thylakoids since it can be extracted by high-salt washing. This conclusion is consistent with our recent finding that the *frxB* protein can be solubilized by detergent from the thylakoid membranes and is recovered in a fraction that contains unidentified protein complexes (Shonai et al. unpublished results).

In chloroplast genomes, seven ORFs have been identified as the homologues of the mitochondrion-encoded components of complex I (NADH:ubiquinone reductase). The ORFs of liverwort chloroplasts are named *ndh1, 2, 3, 4, 4L, 5* and *6* (Ohyama et al. 1986, 1988), and the corresponding genes (*ndhA, B, C, D, E, F* and *G*) are also found in chloroplasts of tobacco and rice (Matsubayashi et al. 1987, Hiratuka et al. 1989). The *frxB* gene is located in a gene cluster that encodes four homologues of components of mitochondrial complex I: ORF392(or 393)-*ndh1*(*A*)-*frxB-ndh6*(*G*)-*ndh4L*(*E*)-*frxA*(*psaC*)-*ndh4*(*D*). It has recently been demonstrated that some of the nucleus-encoded subunits of mitochondrial complex I have also homologues encoded by ORFs in chloroplast genomes. They are the 49-, 30- and 23-kDa subunits of bovine complex I, which are related to product of liverwort ORF392, ORF169 and *frxB*, respectively (Fearnly et al. 1989, Pilkington et al. 1991a, Dupuis et al. 1991). These findings have led to the suggestion that the putative *ndh* proteins, together with the product of *frxB* are components of an NAD(P)H:plastoquinone reductase that is responsible for the respiratory electron-transport activity in *Chlamydomonas* chloroplasts (Godde 1982, Maione and Gibbs 1986). Another possibility is exemplified by the results of Wu and co-workers (1989) who showed that the salt-extracted *frxB* protein has an affinity for the origin of replication of ctDNA. However, both these piece of evidence are indirect and fail to disclose the functional role of this protein in chloroplasts.

We have studied structure-function and evolutionary relationships among various iron-sulfur proteins and we have initiated a molecular genetic study using cyanobacteria which is aimed at elucidating the functional role of the *frxB* protein. Here we report the cloning, nucleotide sequence and co-transcription of the *ndh1, frxB, ndh6* and *ndh4L* genes of a filamentous cyanobacterium, *Plectonema boryanum*. These results are discussed in terms of the putative function of the *frxB* protein.

Materials and Methods

Preparation of genomic DNA and hybridization—A filamentous, non-heterocystous cyanobacterium, *Plectonema boryanum* strain M101, was kindly provided by Dr. T. Kato (Kyoto University). Cells were grown in BG-11 medium (Rippka et al. 1979) at 30°C, in an illuminated incubator. They were harvested by centrifugation at the stationary phase of growth, and DNA was prepared as described by Fujita et al. (1991). Fractionated digests of DNA were transferred from agarose gels to nitrocellulose as described by Southern (1975). A 363-bp *Bcl* I-*Alu* I fragment, containing the truncated *frxB* gene from liverwort chloroplasts, was radioactively labeled with [α-^{32}P]dCTP by the Multiprime DNA labeling system (Amersham) and used as a hybridization probe. This probe contained nucleotides 99916 to 100278 of the liverwort ctDNA (Kohchi et al. 1988). Hybridization was carried out with the radiolabeled probe for 18 h at 45°C. The filters were washed twice in a mixture of 6× SSC and 0.5% SDS at room temperature, twice again in the same solution at 45°C, twice at 55°C, and twice more at 60°C. For other details, including hybridization conditions, see Fujita et al. (1991).

Screening of the genomic DNA library—Approximately 2×10^4 recombinants of a *Plectonema* genomic DNA library in pBR322 (Fujita et al. 1991) were screened by the colony hybridization method of Hanahan and Meselson (1983). The probe and conditions used were as described above. Two recombinants gave a strong hybridization signal with the liverwort *frxB* probe. Both of the positive recombinants carried inserted DNA fragments whose restriction maps were indistinguishable from one another. One, named pB-45, was selected for further study.

Subcloning and DNA sequencing—A 3.1-kb *Pst* I fragment, which hybridized to the probe for *frxB*, and a 1.5-kb *Pst* I-*Hind* III fragment (see Fig. 1) were subcloned into M13mp18 and -19 vectors in both orientations. Their DNA sequences were determined by the modified dideoxy method (Sanger et al. 1977, Biggin et al. 1983) after the construction of deleted subclones by means of the ExoIII/mung bean nuclease deletion technique (Henikoff 1984). The complete sequences of the 3.1-kb and 1.5-kb fragments were determined for the entire inserts in both orientations with the aid of one internal primer. A 1.7-kb *Ssp* I-*Hind* III fragment (see Fig. 1) was also subcloned into M13mp19 which had previously been cut with *Sma* I and *Hind* III. One-sided sequencing, starting from the *Ssp* I site, gave an overlapping sequence across the *Pst* I site. Sequence comparisons were made using the GENETYX (Software Development Co., Ltd., Tokyo, Japan) programs. Predicted amino acid sequences of more than 33 amino acids in length were compared with the sequences in the databases of NBRF-PDB (National Biochemical Research Foundation) and SWISS-PROT (European Molecular Biology Laboratory).

Northern blot analysis—Total RNA was isolated from *Plectonema* cells that had been grown mixotrophically in BG-11 medium supplemented with 5 mM glucose (Fujita et al. 1991). The RNA was fractionated by electrophoresis on 1.0% agarose/formaldehyde gels and transferred to an Immobilon-N membrane (Millipore). Single-strand hybridization probes were synthesized (Sambrook et al. 1989)

from the inserts of deletion clones used for sequence analysis: for *ndh1*, nucleotides 1556 to 1083 of the DNA sequence shown in Figure 2 (probe 1); for *frxB*, nucleotides 2225 to 1807 (probe 2); for *ndh6*, nucleotides 2758 to 2345 (probe 3); for *ndh4L*, nucleotides 3103 to 2905 (probe 4); for the 3' non-coding strand, nucleotides 3951 to 3624 (probe 5); for ORF105, nucleotides 1496 to 1786 (probe 6); and for ORF295, nucleotides 3526 to 4011 (probe 7). Hybridization was performed in 6× SSC, 5× Denhardt's solution, 0.5% SDS and 100 $\mu g\ ml^{-1}$ salmon sperm DNA at 65°C for 16 h. The washings were performed in 0.2× SSC and 0.1% SDS at 65°C.

Results and Discussion

Cloning and DNA sequence—The radiolabeled probe derived from the *frxB* gene from liverwort chloroplasts hybridized to specific fragments in restriction digests of DNA from *P. boryanum*, namely, the 5.6-kb *Hind* III fragment, the 4.0-kb *Eco*R I fragment, and the 2.7-kb *Hind* III-*Eco*R I fragment (data not shown). With the same gene probe, two positively hybridizing recombinants, pB-05 and pB-45, were identified in a library of *Plectonema* DNA which had been constructed with the *Hind* III digest of the genomic DNA. These two positive recombinants carried a 5.6-kb *Hind* III fragment (Fig. 1) and were indistinguishable from one another with respect to the sizes of various restriction fragments. Subsequent Southern hybridization showed that the length of the hybridizing fragment in the *Pst* I digest of the recombinants was 3.1 kb. The DNA sequence of the 3.1-kb *Pst* I fragment from pB-45 corresponds to nucleotides 1 to 3108, shown in Figure 2. As described below, the sequence did not contain the 3' portion of the *ndh4L* homologue. Therefore, the sequence was extended in the 3' direction by the use of a 1.5-kb *Pst* I-*Hind* III fragment (nucleotides 3103 to 4616) derived from pB-45. The complete sequence presented in Figure 2 is 4616 bp long. Each nucleotide was identified at least once on each strand of the DNA.

Identification of potential genes—The DNA sequence

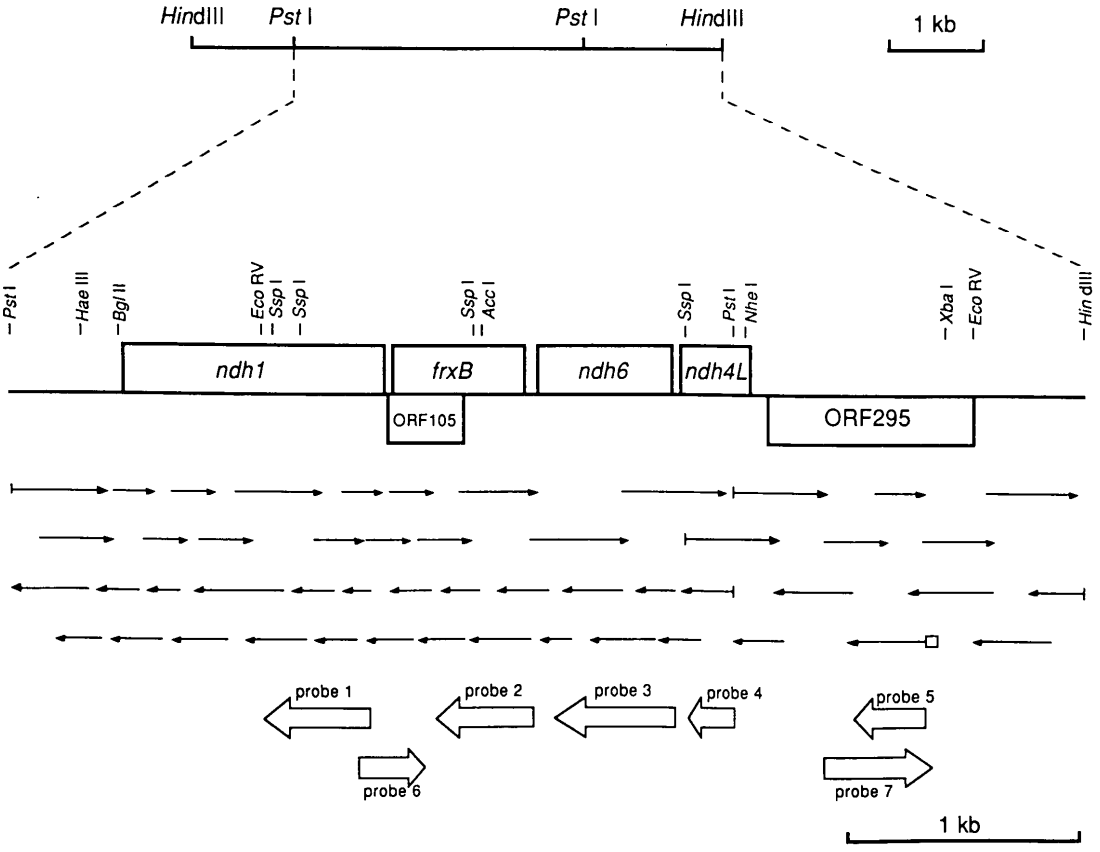
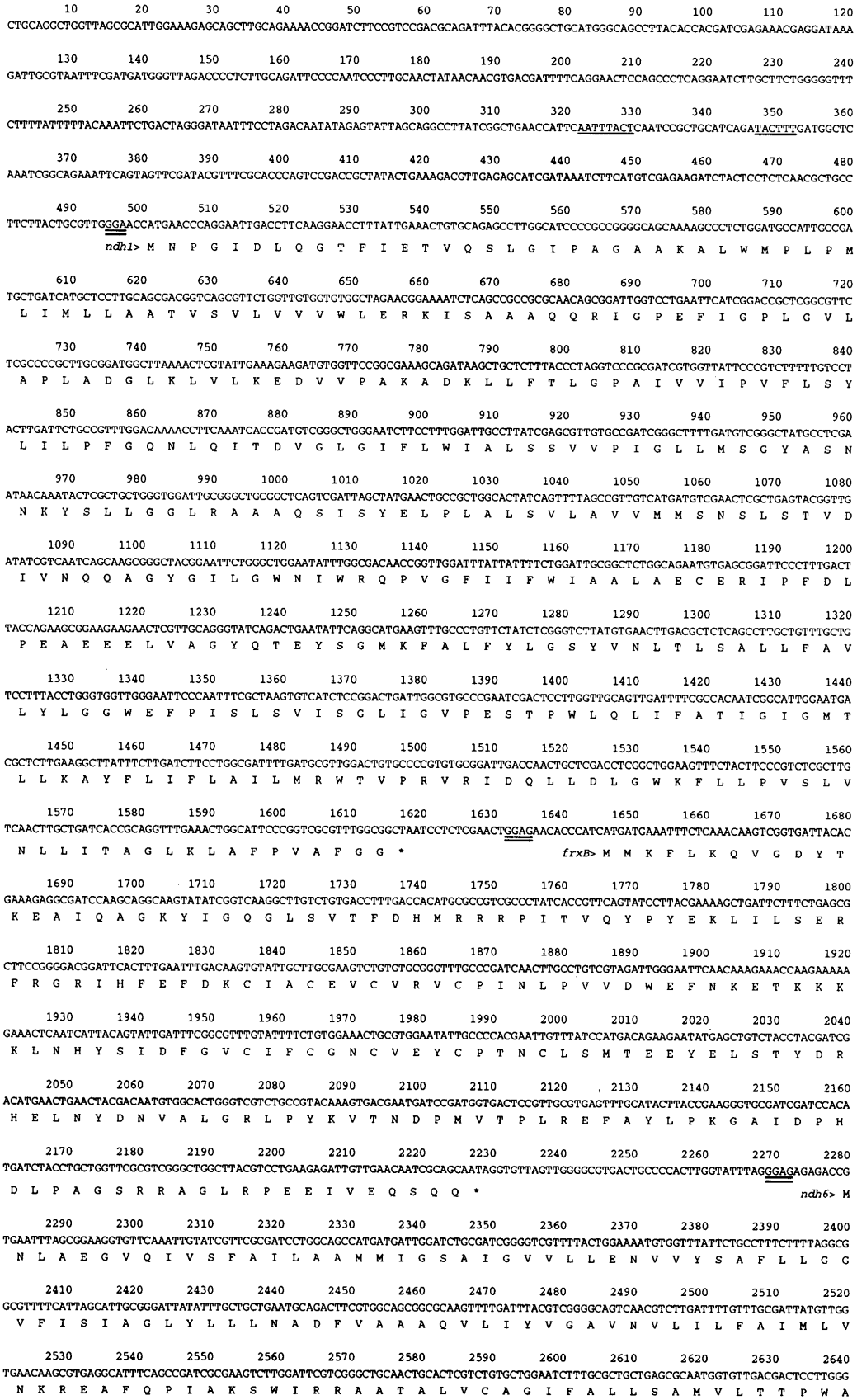


Fig. 1 Physical map of and strategy for sequencing the 4.6-kb DNA fragment that contained the *ndh1, frxB, ndh6, ndh4L* gene cluster, cloned from the filamentous cyanobacterium *P. boryanum*. The four genes and two unidentified reading frames (ORF105 and ORF295) are shown as open boxes on the map. The direction and extent of analyses of sequences of individual restriction fragments or deletion clones are indicated by the arrows below the map. An open box attached to an arrow indicates that an internal synthetic oligonucleotide primer was used in the sequence analysis. Open arrows at the bottom of the Figure correspond to single-strand DNAs used as gene-specific probes for the Northern blotting analysis in Fig. 3. For other details see Materials and Methods.



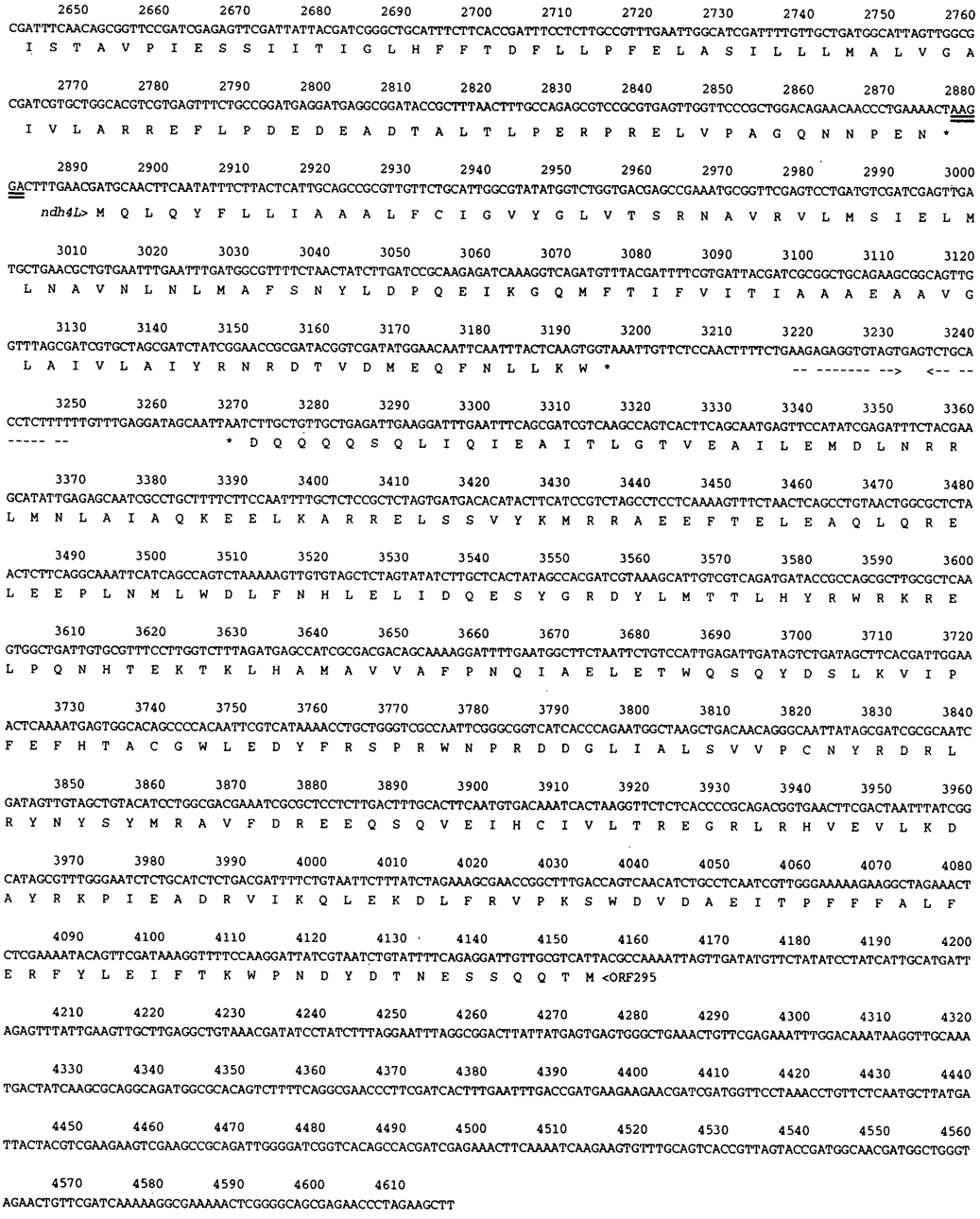
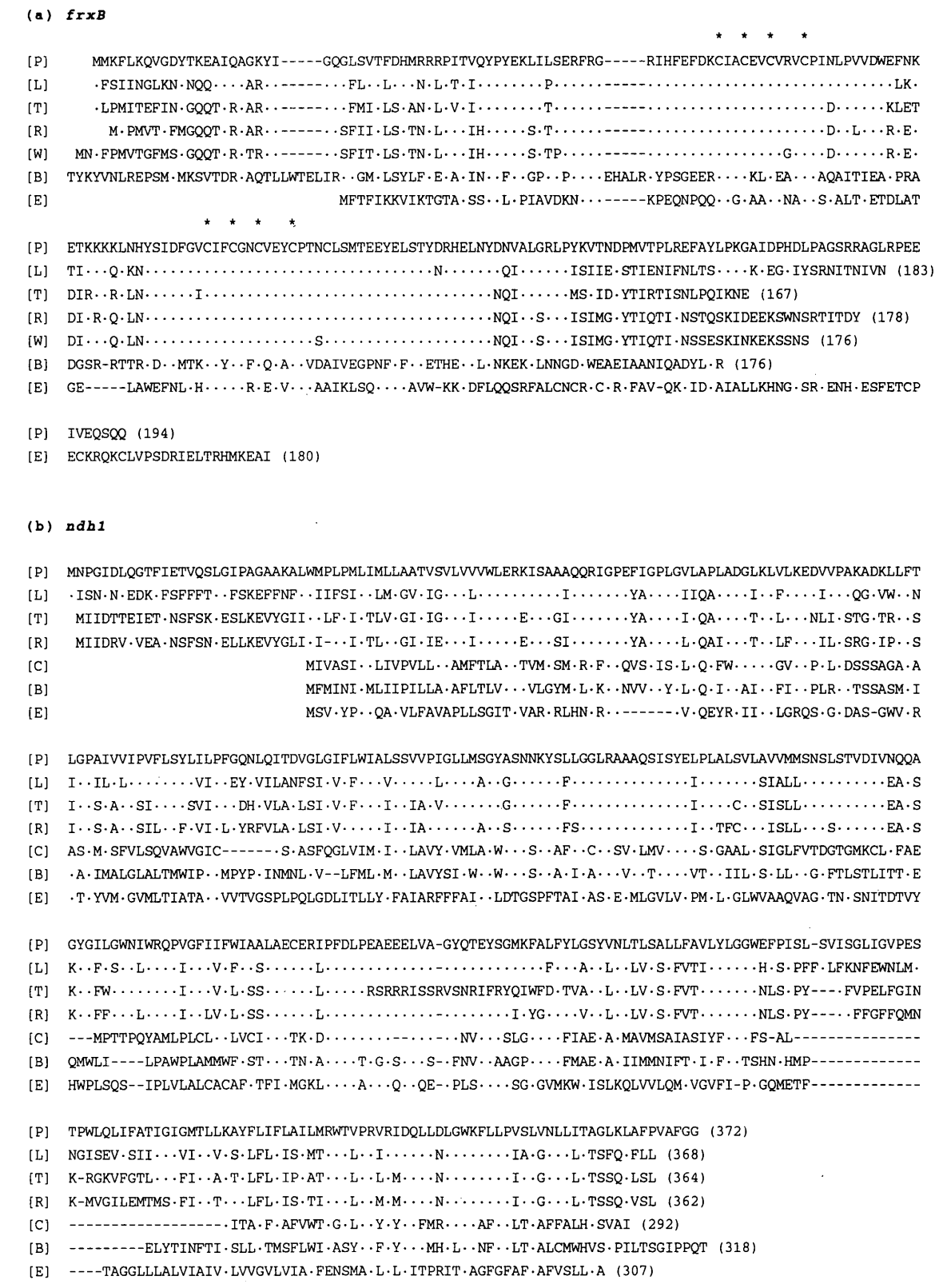


Fig. 2 Nucleotide sequence of the *ndh1, frxB, ndh6, ndh4L* gene cluster from *P. boryanum*. Nucleotide sequences are shown in the direction of transcription except in the case of ORF295. Deduced amino acid sequences are shown below the nucleotide sequence in the one-letter code. Genes with an arrowhead (> or < indicating the direction of transcription) are named at the start of their coding sequences. Putative promoter sequences (−35 and −10 sequences) and SD sequences are indicated by underlining and double underlining, respectively. Broken arrows indicate inverted repeat sequence.

was translated into amino acid sequences in all six possible reading frames and the predicted sequences were compared with known sequences contained in the databases of NBRF-PDB and SWISS-PROT. In this way, homologues of the *ndh1, frxB, ndh6* and *ndh4L* proteins of liverwort chloroplasts were identified in the same strand of DNA. In the case of *ndh6*, GTG might act as an initiation codon, as in the *frxC* gene of *P. boryanum* (Fujita et al. 1991). Possible ribosome-binding sites (Shine-Dalgarno sequences) were found preceding the initiation codons for these genes. In addition to the possible genes described above, eight ORFs encoding proteins of more than 33 amino acids in length were detected on the same and on the complementary strands, respectively. No significant ho-



mology was found between any of the predicted proteins and the sequences in the databases. The proposed arrangement of genes, including the two large ORFs (ORF105 and ORF295), is summarized in Figure 1.

The frxB gene—The *frxB* gene extends from nucleotides 1646 to 2227, as shown in Figure 2, and could encode

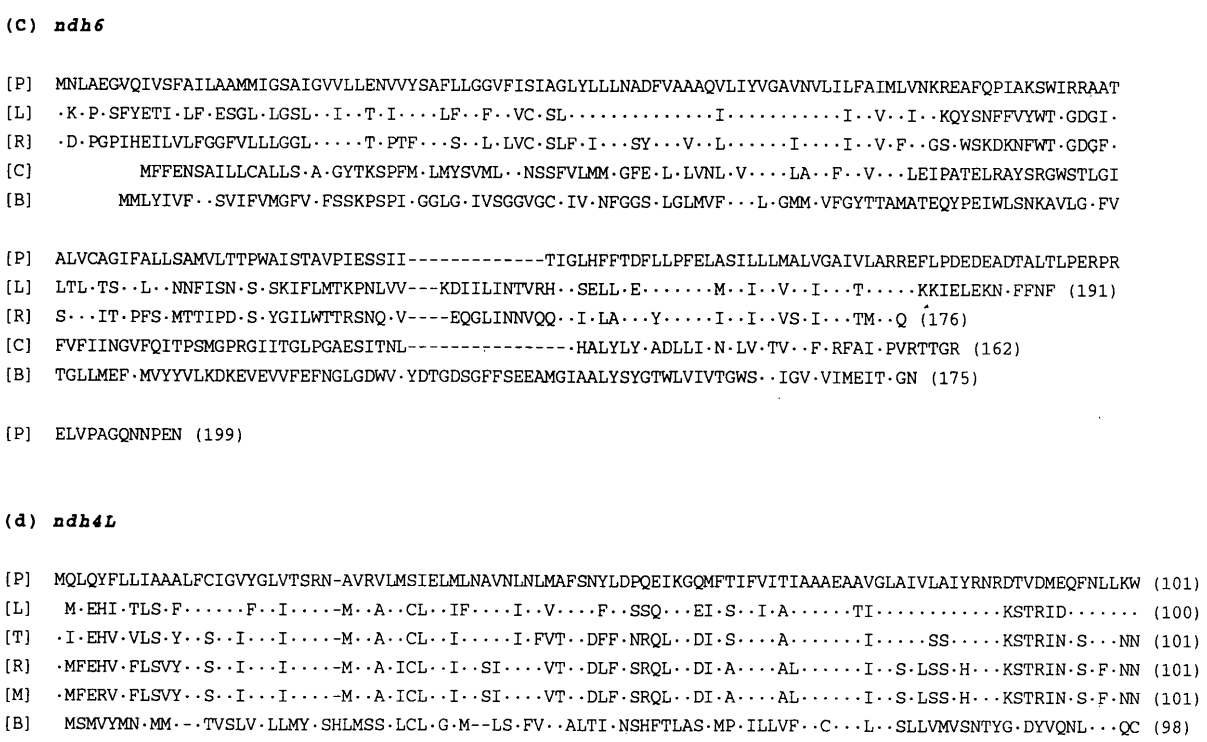


Fig. 3 Comparison of the amino acid sequences deduced from the *frxB* (a), *ndh1* (b), *ndh6* (c) and *ndh4L* (d) genes of *Plectonema boryanum* with those of related proteins. [P], *P. boryanum* (present study); [L], liverwort chloroplasts (Kohchi et al. 1988); [T], tobacco chloroplasts (Matsubayashi et al. 1987); [R], rice chloroplasts (Hiratsuka et al. 1989); [W], wheat chloroplasts (Dunn and Gray 1988); [M], maize chloroplasts (Schantz and Bogorad 1988); [B], bovine mitochondria (Anderson et al. 1982, Dupuis et al. 1991); [C], *Chlamydomonas* mitochondria (Boer and Gray 1988, 1989); [E], *E. coli* (Böhm et al. 1990). Amino acid residues identical to the corresponding residues in the proteins from *Plectonema* are denoted by dots. Bacterial Fd-like arrangements of cysteine residues found in *frxB* sequences are indicated by asterisks.

a protein of 194 amino acids with a mol wt of 22,528. It lies between the *ndh1* and *ndh6* genes; the ATG initiation codon of the *frxB* gene (at position 1646) is located 27 bases downstream from the TAA stop codon of *ndh1* gene (at position 1619), and the TAG stop codon of the *frxB* gene (at position 2230) is 50 bases upstream from the GTG initiation codon of the *ndh6* gene (at position 2280). A possible ribosome-binding site, GGAG, lies 11 bases upstream from the initiation codon ATG.

Figure 3(a) shows an alignment of the sequence of the *frxB* protein with other sequences that are currently available. The protein from *Plectonema* has 60–66% homology with each of the chloroplast proteins (Table 1). In the case of these five sequences shown, no deletions or insertions were necessary for proper alignment. The 23-kDa subunit of bovine mitochondrial complex I, probably carry-

Table 1 Percentage identities between products of *Plectonema boryanum* genes and homologous proteins from chloroplasts, mitochondria and *E. coli*

	Chloroplasts			Mitochondria		*E. coli*
	Liverwort	Tobacco	Rice	*Chlamydomonas*	Bovine	
frxB	66	64	60	—	29	25
ndh1	60	51	55	34	32	20
ndh6	50	—	43	18	12	—
ndh4L	65	57	50	—	24	—

Similarities in amino acid sequences were calculated on the basis of the sequence alignments shown in Figure 3(a–d).

ing the iron-sulfur cluster(s) known as N-2 (Dupuis et al. 1991), is also related to the cyanobacterial and chloroplast *frxB* proteins but to a lesser extent. In addition, the product of ORF6 encoded in the *hyc* operon of *Escherichia coli* (Böhm et al. 1990) was found to show a significant degree of sequence identity to the *frxB* protein. Hydropathy profiles of these proteins (not shown) also show striking similarities between them. The data suggest that these proteins are hydrophilic and contain no large hydrophobic segments that span the membrane. These profiles are consistent with the experimental observation that the *frxB* proteins from tobacco and spinach chloroplasts can be extracted from thylakoid membranes by high-salt washing (Lin and Wu 1990).

The *frxB* protein is unusual among the soluble bacterial-type Fds. A typical bacterial Fd, as exemplified by Fd from *Clostridium pasteurianum*, has two [4Fe-4S] clusters carried by a polypeptide of about 6 kDa. In these clusters, each of the eight iron atoms is chelated by cysteine residues, which are arranged in two -C-X-X-C-X-X-C-X-X-X-C-P- sequences (reviewed by Bruschi and Guerlewquin 1988). Similar arrangements of cysteine residues are also found in two highly conserved regions of the *frxB* proteins. Presumably, the conserved regions are indispensable for the operation of the Fe-S protein in electron-transfer reactions. However, the *frxB* protein is much larger than the bacterial-type Fds, with extensions mainly in the N- and C-terminal regions of about fifty residues (Kohchi et al. 1988). Obviously, these proteins must have a common ancestor and must have diverged during evolution, one to a group of proteins, such as soluble bacterial Fds, and the other to a component of a membrane-bound enzyme complex. The N- and C-terminal regions of the *frxB* protein may be involved in interactions with thylakoid membranes or with other components of the membrane-bound complex.

ndh genes—The coding region of the *ndh1* gene extends from nucleotides 501 to 1616, shown in Figure 2, and encodes a protein of 372 amino acids with a mol wt of 40,239. It terminates 27 bases upstream from the *frxB* gene. The *ndh6* gene (nucleotides 2280 to 2876; Fig. 2) follows the *frxB* gene. It encodes a protein of 199 amino acids with a mol wt of 21,267. The *ndh4L* gene extends from nucleotides 2892 to 3194. It lies 13 bases downstream from the TAA stop codon of the *ndh6* gene (at position 2879) and encodes a protein of 101 amino acids with a mol wt of 11,335. Potential binding sites for 16 *S* ribosomal RNA were located for each of the three *ndh* genes.

Alignments of the sequences of the products of the *ndh* genes with those of analogous proteins from chloroplasts, mitochondria and *E. coli* are shown in Figure 3(b–d), and their homologies are summarized in Table 1. The deduced amino acid sequences of products of the *ndh1*, *ndh6* and *ndh4L* genes from *Plectonema* show significant similarities to the corresponding gene products from chloroplasts. Although these sequences are not obviously homologous to the mitochondrial sequences, in particular in the case of the *ndh6* protein, comparisons of their hydropathy profiles reveal striking similarities (data not shown). The *ndh1*, *ndh6* and *ndh4L* proteins contain seven, three and three hydrophobic segments of more than 20 amino acids each, respectively, and these segments are consistent with the presence of transmembranous α-helical segments in these proteins.

Unidentified potential genes—ORF295 extends from nucleotides 4153 to 3269 on the opposite strand of the DNA to the *frxB* and *ndh* genes (Fig. 1 and 2). It could encode a hydrophilic protein of 295 amino acids, with a mol wt of 35,541. It terminates 74 bp downstream from the *ndh4L* gene. ORF105 extends from nucleotides 1933 to 1619, encoding a protein of 105 amino acids with a mol wt of 12,187. It lies on the complementary strand of the *frxB* coding region and the transcription of ORF105 is less probable (see below).

Co-transcription of the genes in the ndh1-frxB-ndh6-ndh4L gene cluster—A prokaryotic promoter-like sequence can be found upstream from the *ndh1* gene at bases 322 to 329 (AATTTACT, −35 BOX) and 347 to 352 (TACTTT, −10 BOX). Nucleotides that can form a stem-loop structure, followed by a run of T residues, occur downstream from the *ndh4L* gene (nucleotides 3218 to 3248), which may act as a signal for ρ-independent termination of transcription. Therefore, it seems likely that the gene cluster *ndh1-frxB-ndh6-ndh4L* constitutes an operon. Further support for this proposal is provided by the presence of very short intergenic non-coding regions: 29 bp between *ndh1* and *frxB*; 52 bp between *frxB* and *ndh6*; and 15 bp between *ndh6* and *ndh4L*. In order to ascertain the significance of these transcriptional signals, Northern blot analysis was carried out. Total RNA isolated from mixotrophically grown cultures of *Plectonema* was hybridized with single-strand DNA probes for each of the coding strands, synthesized on recombinant M13 DNAs as templates. The *ndh1-*, *frxB-*, *ndh6-* and *ndh4L*-specific probes (probes 1 to 4) each recognized a transcript of about 2.8 kb, which is sufficient in terms of length to cover these genes (Fig. 4). In addition to the 2.8-kb transcript, both the *ndh1* and *frxB* probes recognized mRNA species of about 1.5 kb. Transcripts of 1.4 kb and 1.1 kb hybridized with the *ndh6* and *ndh4L* probes with different relative intensities. No signals were observed when a probe for the 3′ region (probe 5), downstream from the terminator-like sequence, was tested with the same preparation of RNA and, therefore, it appears that the clustered *ndh1-frxB-ndh6-ndh4L* genes are co-transcribed as a single mRNA. The length of the largest transcript (about 2.8 kb) is consistent with the length of the entire operon, from close to the

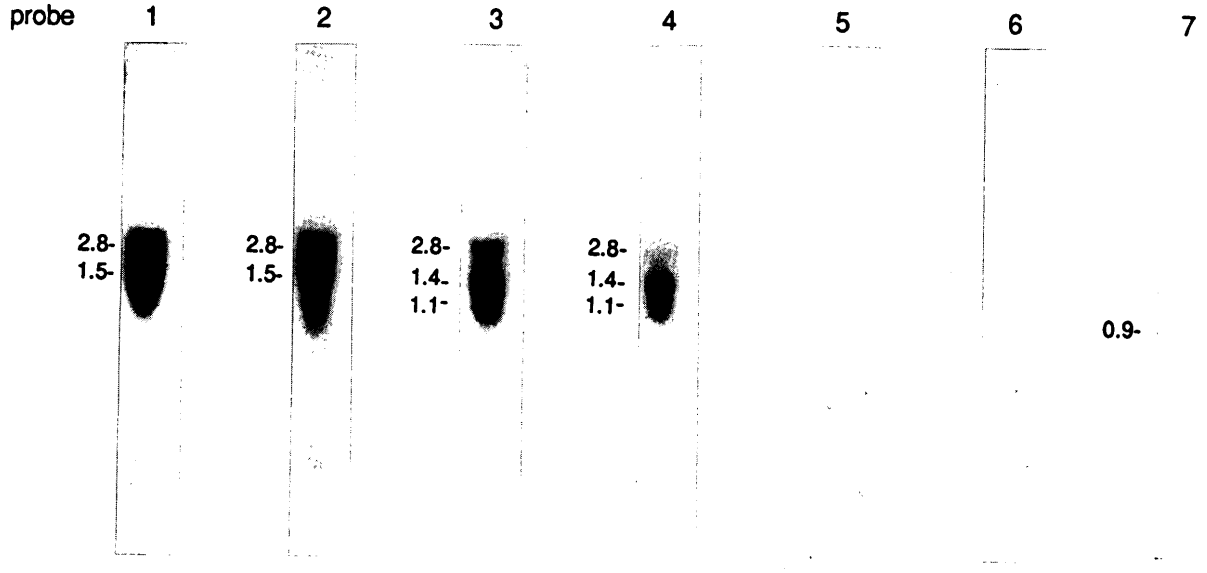


Fig. 4 Northern blot analysis of the *ndh1, frxB, ndh6, ndh4L* and ORF295 genes in *P. boryanum*. Total RNA isolated from a batch culture of mixotrophically grown *Plectonema* was tested with the gene-specific single-strand DNA probes shown in Fig. 1. The coding strand of each probe was labeled with [α-^{32}P]dCTP as described in Materials and Methods. The sizes (in kb) of hybridizing species are shown to the left of respective lanes.

"-10"-like sequence to the possible termination sequence shown in Figure 2. The presence of smaller mRNA species suggests that there may be other sites of initiation of transcription, or that the 2.8-kb transcript may be specifically processed or degraded. Essentially the same results were obtained with preparations of RNA from mixotrophically grown and photoautotrophically grown cultures of *Plectonema*, suggesting that the expression of this gene cluster is not regulated by glucose or its metabolites. Two additional hybridization probes, (probes 6 and 7; Fig. 4), were used to examine the expression of ORF105 and ORF295, respectively, both of which are found on the opposite strand of the DNA from the *frxB* and *ndh* genes. The ORF295 probe recognized a weak and diffuse band at around 0.9 kb. The mRNA is of sufficient length to cover this gene (885 bp), and it is likely to be a monocistronic transcript. Under the same conditions as described above, no transcript of ORF105 could be detected. Since we used equally labeled probes, each of almost the same radioactive intensity and specific activity, the expression of the ORF105 gene seems less likely when compared with that of other genes.

Organization of the operon—Each of the *ndh1-, frxB-, ndh6-* and *ndh4L*-specific probes (the same probes as used for the Northern blot analysis) hybridized to a single 5.6-kb *Hin*d III and a single 3.2-kb *Eco*R V fragment in digests of *Plectonema* genomic DNA (data not shown). These results indicate that these genes occur as single copies in the genome. Arrangement of the genes in *Plectonema* is shown in Figure 5 and is compared to the corresponding regions in the chloroplast genomes. The genes equivalent to *frxB* in the ctDNAs of tobacco and rice are represented by ORF167 and ORF178, respectively. In tobacco chloroplasts, a transcript of about 4 kb has been detected with *ndhA*- and *ndhE*-specific probes, and it is long enough to include genes *ndhA* to *ndhE*, if the intron is excluded (Matsubayashi et al. 1987). Since the ORF393- or *ndhD*-specific probes did not recognize the 4-kb transcript, the gene cluster of *ndhA*-ORF167(*frxB*)-*ndhG-ndhE* appears to constitute a transcription unit. The same is true for these genes in *Plectonema*. The conserved amino acid sequences, the order of genes and the co-transcription of the *ndh1, frxB, ndh6* and *ndh4L* genes in cyanobacteria and chloroplasts support an endosymbiotic origin for plastids. However, the intron contained in the plastid *ndh1* gene is not found in the *Plectonema* sequence, and the lengths and the sequences of the intergenic non-coding regions show no significant similarities. Therefore, the conserved order of genes may reflect some aspect of transcriptional or translational regulation, for example, it may be involved in the production of programmed amounts of functionally related proteins. Alternatively, the polycistronic mRNA could allow for the immediate assembly of the products of translation into an enzyme complex. The structure of the *ndhC-psbG*-ORF157/159(*ndhJ*) operon and the corresponding deduced amino acid sequences are also conserved in both the cyanobacterium *Synechocystis* PCC6803 and chloroplasts (Steinmüller et al. 1989) and, therefore, it is clear that cyanobacteria and chloroplasts contain a highly conserved enzyme complex.

Phylogenetic implications—Chloroplast genomes con-

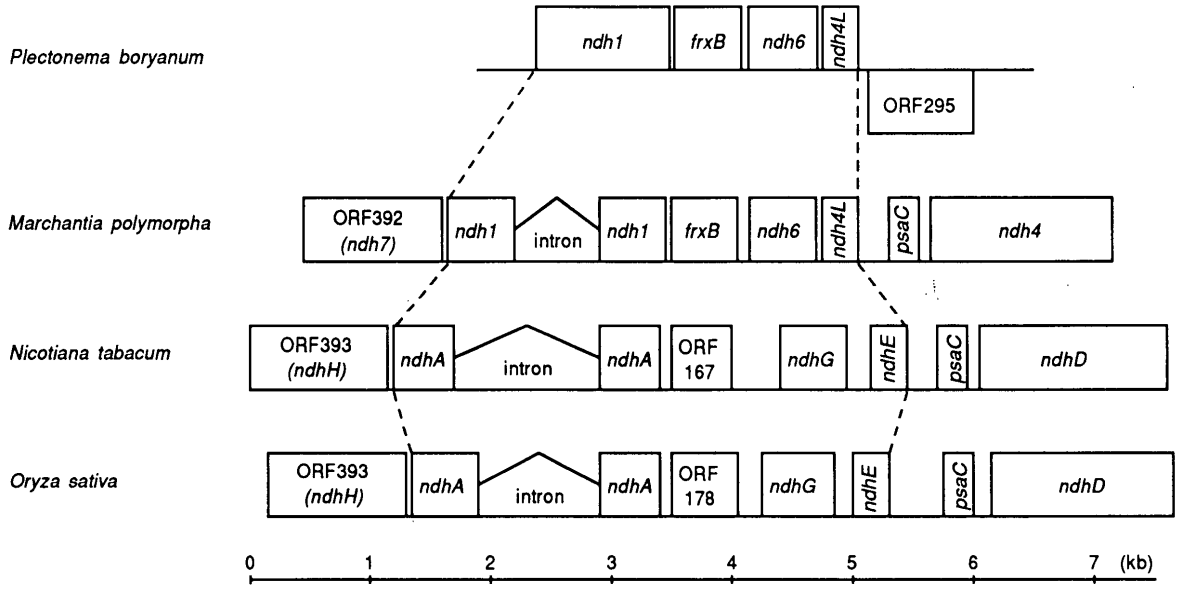


Fig. 5 Arrangement of cyanobacterial and chloroplast genes in the vicinity of *frxB*. The diagram is based on the sequence data published by Kohch et al. (1988), Shinozaki et al. (1986) and Hiratsuka et al. (1989). *frxB* homologues in the ctDNAs of tobacco and rice are denoted by ORF167 and ORF178, respectively. ORF392(393) is homologous to the nuclear-encoded 49-kDa subunit of bovine mitochondrial complex I (Fearnley et al. 1989). *psaC* (formerly *frxA*) encodes a component of photosystem I which carries two [4Fe-4S] clusters. The arrangement of genes that is conserved among the four species is enclosed by dotted lines.

tain ten genes that encode homologues to the components of mitochondrial complex I. They are *ndh1* (*A*), *2* (*B*), *3* (*C*), *4* (*D*), *4L* (*E*), *5* (*F*), *6* (*G*), *7* (*H*), *9* (*J*), and *frxB* (Ohyama et al. 1988, Matsubayashi et al. 1987, Hiratsuka et al. 1989, Fearnly et al. 1989, Pilkington et al. 1991a, Dupuis et al. 1991). In addition, it has been suggested, from indirect evidence, that the *psbG* gene is a candidate for an *ndh* gene but its product is not considered to be a component of photosystem II (Nixon et al. 1989, Mayes et al. 1990). In cyanobacterial genomes, eight related genes have been identified; *ndh2* (*B*), *3* (*C*), *9* (*J*) and *psbG* from *Synechocystis* (Steinmüller et al. 1989, Mayes et al. 1990, Ogawa 1991), and *ndh1* (*A*), *4L* (*E*), *6* (*G*) and *frxB* from *Plectonema* (this study). These findings lend support to the hypothesis that chloroplasts and cyanobacteria contain an enzyme that is related to mitochondrial complex I. Consequently, an NADPH:plastoquinone oxidoreductase has been proposed as the most probable candidate (Fearnly et al. 1989, Pilkington et al. 1991a, Dupuis et al. 1991). However, it seems premature to draw a firm conclusion because of the lack of biochemical evidence. Furthermore, the following sets of observations are in disagreement with the putative assignation of NADPH:plastoquinone oxidoreductase in both cyanobacteria and chloroplasts.

First, the amino acid sequences of the mitochondrial components are poorly conserved in the cyanobacterial and chloroplast homologues. As shown in Table 1, the extent of sequence identities between the components of *Plectonema* and bovine mitochondria is about 30% or less: not sufficiently high to allow us to postulate an identical function in these organisms. On the other hand, sequence similarities have been reported between components of formate hydrogenlyase in *E. coli* and the products of the *ndh* genes of chloroplasts and mitochondria (Böhm et al. 1990). The enzyme from *E. coli*, releasing hydrogen via the reduction of protons that is coupled to the oxidation of formate, is used in the non-energy-conserving system to eliminate excess formate. Four of the eight ORFs identified in a *hyc* operon of *E. coli*, namely, ORF3, 4, 5 and 7, exhibit homologies with products of the *ndh4* (*D*), *ndh1* (*A*), *ndh7* (*H*) and *psbG* genes, respectively (Böhm et al. 1990). In addition, we found sequence similarities between the product of ORF6 from *E. coli* and the products of *frxB* from *Plectonema* and chloroplasts (Fig. 3, Table 1). The similarity to ORF6 is not restricted to the Fd-like arrangement of cysteine residues but also extends to the other parts of the amino acid sequence. These observations highlight the difficulty in assigning a functional role from the limited homologies.

Second, not all of the components of mitochondrial complex I have been correlated with the products of chloroplast or cyanobacterial genes. Mitochondrial complex I is an enzyme composed of at least 25 subunits (Hatefi 1985). With the limited sequence data available to date, at least six components of bovine mitochondrial complex I appears to have no homologues encoded in ctDNAs (Pilkington et al.

1991b, Masui et al. 1991). One of the subuhnits is the functionally well-defined 51-kDa subunit which binds the substrate NAD(H) and contains an FMN (Pilkington et al. 1991b). We can speculate that some of these genes have been transferred to the plant nucleus after endosymbiosis, although none of the additional genes has been identified so far, even in the *ndh* operons of cyanobacteria. Weiss and co-workers (1991) have proposed the hypothesis that the mitochondrial complex I was constructed originally by the combination of at least two preexisting enzyme complexes. If so, it is not surprising that the chloroplast and cyanobacterial enzymes contain some but not all of the mitochondrial components.

Third, no enzyme resembling mitochondrial complex I has been identified so far in an extensive survey of cyanobacterial NAD(P)H dehydrogenases. Because the thylakoid membranes of cyanobacteria are known to participate in respiration (Scherer et al. 1988a, Scherer 1990), there have been several attempts to identify the cyanobacterial NAD(P)H dehydrogenases that provide reducing equivalents for the respiratory transport of electrons (Sandmann and Malkin 1983, Scherer et al. 1988b, Alpes et al. 1989). Such an NADH dehydrogenase has recently been purified from *Anabaena variabilis* and it consists of 17-kDa polypeptide(s). Compared to mitochondrial complex I, the enzyme from *Anabaena* contains FAD but not FMN and is only weakly inhibited by rotenone (Alpes et al. 1989, Scherer 1990). As for the cyanobacterial NADPH dehydrogenase, evidence has been obtained to suggest that the Fd-$NADP^+$ oxidoreductase is responsible for the respiratory oxidation of NADPH in *Anabaena* cells (Scherer et al. 1988b).

The limited data available today suggest the presence of functionally divergent enzyme complexes in cyanobacteria (chloroplasts) and mitochondria and, therefore, the possibility can not be ignored that the products of the *ndh* genes, together with the *frxB* protein, may not be involved in the respiratory electron-transport systems. It has recently been demonstrated that a mutant of *Synechocystis*, lacking the *ndhB* gene, is defective in the transport of CO_2 (Ogawa 1991). Further genetic studies of cyanobacteria will assist in the biochemical characterization of the enzyme complex.

We are grateful to Drs. S. Wakabayashi and R. Masui for their critical comments and valuable discussion. This work was supported in part by Grants-in-Aid for Encouragement of Young Scientists (Nos. 63740376 and 01740396) from the Ministry of Education, Science, and Culture of Japan.

References

Alpes, I., Siegfried, S. and Böger, P. (1989) The respiratory NADH dehydrogenase of the cyanobacterium *Anabaena variabilis*: purification and characterization. *Biochim. Biophys. Acta* 973: 41–46.

Anderson, S., De Bruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F. and Young, I. G. (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* 156: 683–717.

Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* 80: 3963–3965.

Boer, P. H. and Gray, M. W. (1988) Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *EMBO J.* 7: 3501–3508.

Boer, P. H. and Gray, M. W. (1989) Nucleotide sequence of a region encoding subunit 6 of NADH dehydrogenase (ND6) and $tRNA^{Trp}$ in *Chlamydomonas reinhardtii* mitochondrial DNA. *Nucleic Acids Res.* 17: 3993.

Böhm, R., Sauter, M. and Böck, A. (1990) Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenlyase components. *Mol. Microbiol.* 4: 231–243.

Bruschi, M. and Guerlesquin, F. (1988) Structure, function and evolution of bacterial ferredoxins. *FEMS Microbiol. Rev.* 54: 155–176.

dePamphilis, C. W. and Palmer, J. D. (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* 348: 337–339.

Dunn, P. P. J. and Gray, J. C. (1988) Nucleotide sequence of the *frxB* gene in wheat chloroplast DNA. *Nucleic Acid Res.* 16: 349.

Dupuis, A., Skehel, J. M. and Walker, J. E. (1991) A homologue of a nuclear-coded iron-sulfur protein subunit of bovine mitochondrial complex I is encoded in chloroplast genome. *Biochemistry* 30: 2954–2960.

Fearnly, I. M., Runswick, M. J. and Walker, J. E. (1989) A homologue of the nuclear coded 49-kd subunit of bovine mitochondrial NADH-ubiquinone reductase is coded in chloroplast DNA. *EMBO J.* 8: 665–672.

Fujita, Y., Takahashi, Y., Shonai, F., Ogura, Y. and Matsubara, H. (1991) Cloning, nucleotide sequence and differential expression of the *nifH* and *nifH*-like (*frxC*) genes from the filamentous nitrogen-fixing cyanobacterium, *Plectonema boryanum*. *Plant Cell Physiol.* 32: 1093–1106.

Godde, D. (1982) Evidence for a membrane-bound NADPH-plastoquinone-oxidoreductase in *Chlamydomonas reinhardtii* CW-15. *Arch. Microbiol.* 131: 197–202.

Hanahan, D. and Meselson, M. (1983) Plasmid screening at high colony density. *Methods Enzymol.* 100: 333–342.

Hatefi, Y. (1985) The mitochondrial electron transport and oxidative phosphorylation system. *Annu. Rev. Biochem.* 54: 1015–1069.

Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoint for DNA sequencing. *Gene* 28:

351–359.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C., Meng, B., Li, Y., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217: 185–194.

Kohchi, T., Shirai, H., Fukuzawa, H., Sano, T., Komano, T., Umesono, K., Inokuchi, H., Ozeki, H. and Ohyama, K. (1988) Structure and organization of *Marchantia polymorpha* chloroplast genome. IV. Inverted repeat and small single-copy regions. *J. Mol. Biol.* 203: 353–372.

Lin, C-H. and Wu, M. (1990) A ferredoxin-type iron-sulfur protein gene, *frxB*, is expressed in the chloroplasts of tobacco and spinach. *Plant Mol. Biol.* 15: 449–455.

Maione, T. and Gibbs, M. (1986) Association of the chloroplastic respiratory and photosynthetic electron transport chains of *Chlamydomonas reinhardtii* with photoreduction and the oxyhydrogen reaction. *Plant Physiol.* 80: 364–368.

Masui, R., Wakabayashi, S., Matsubara, H. and Hatefi, Y. (1991) The amino acid sequences of two 13-kDa polypeptides and partial amino acid sequence of 30-kDa polypeptide of complex I from bovine heart mitochondria. *J. Biochem.* 109: 534–543.

Matsubayashi, T., Wakasugi, T., Shinozaki, K., Yamaguchi-Shinozaki, K., Zaita, N., Hidaka, T., Meng, B. Y., Ohto, C., Tanaka, M., Kato, A., Maruyama, T. and Sugiura, M. (1987) 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.

Mayes, S. R., Cook, K. M. and Barber, J. (1990) Nucleotide sequence of the second *psbG* gene in *Synechocystis* 6803. Possible implication for *psbG* function as an NAD(P)H dehydrogenase subunit gene. *FEBS Lett.* 262: 49–54.

Nixon, P. J., Gounaris, K., Coomber, S. A., Hunter, C. N., Dyer, T. A. and Barber, J. (1989) *psbG* is not a photosystem two gene but may be an *ndh* gene. *J. Biol. Chem.* 264: 14129–14135.

Ogawa, T. (1991) A gene homologous to the subunit-2 gene of NADH dehydrogenase is essential to inorganic carbon transport of Synechocystis PCC6803. *Proc. Natl. Acad. Sci. USA* 88: 4275–4279.

Oh-oka, H., Takahashi, Y., Kuriyama, K., Saeki, K. and Matsubara, H. (1988) 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.

Oh-oka, H., Takahashi, Y., Wada, K., Matsubara, H., Ohyama, K. and Ozeki, H. (1987) The 8-kDa polypeptide in photosystem I is a probable candidate of an iron-sulfur protein coded by the chloroplast gene *frxA*. *FEBS Lett.* 218: 52–54.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572–574.

Ohyama, K., Kohchi, T., Sano, T. and Yamada, Y. (1988) Newly identified groups of genes in chloroplasts. *Trends Biochem. Sci.* 13: 19–22.

Pilkington, S. J., Skehel, J. M. and Walker, J. E. (1991a) The 30-kilodalton subunit of bovine mitochondrial complex I is homologous to a protein coded in chloroplast DNA. *Biochemistry* 30: 1901–1908.

Pilkington, S. J., Skehel, J. M., Gennis, R. B. and Walker, J. E. (1991b) Relationship between mitochondrial NADH-ubiquinone reductase and a bacterial NAD-reducing hydrogenase. *Biochemistry* 30: 2166–2175.

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. and Stanier, R. Y. (1979) Genetic assignment, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* 111: 1–61.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual. Second edition.* Cold Spring Harbor Laboratory Press.

Sandmann, G. and Malkin, R. (1983) NADH and NADPH dehydrogenases from the blue-green alga, *Aphanocapsa. Arch. Microbiol.* 136: 49–53.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463–5467.

Schantz, R. and Bogorad, L. (1988) Maize chloroplast genes *ndhD, ndh4L*, and *psaC*. Sequences, transcripts and transcript pools. *Plant Mol. Biol.* 11: 239–247.

Scherer, S. (1990) Do photosynthetic and respiratory electron transport chains share redox proteins? *Trends Biochem. Sci.* 15: 458–462.

Scherer, S., Almon, H. and Böger, P. (1988a) Interaction of photosynthesis, respiration and nitrogen fixation in cyanobacteria. *Photosynth. Res.* 15: 95–114.

Scherer, S., Alpes, I., Sadowski, H. and Böger, P. (1988b) Ferredoxin-$NADP^+$ reductase is the respiratory NADPH dehydrogenase of the cyanobacterium *Anabaena variabilis. Arch. Biochem. Biophys.* 267: 228–235.

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

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503–517.

Steinmüller, K., Ley, A. C., Steinmetz, A. A., Sayre, R. T. and Bogorad, L. (1989) Characterization of the *ndhC-psbG-ORF157/159* operon of maize plastid DNA and of the cyanobac-

terium *Synechocystis* sp. PCC6803. *Mol. Gen. Genet.* 216: 60–69.

Weiss, H., Friedrich, T., Hofhaus, G. and Preis, D. (1991) The respiratory-chain NADH dehydrogenase (complex I) of mitochondria. *Eur. J. Biochem.* 197: 563–576.

Wu, M., Nie, Z. Q. and Yang, J. (1989) 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.

(Received April 27, 1991; Accepted July 8, 1991)