ELSEVIER

Biochimica et Biophysica Acta 1273 (1996) 195–197

BBA Biochimica et Biophysica Acta

Short sequence-paper

Nucleotide sequence of the *PetM* gene encoding a 4 kDa subunit of the cytochrome b_6f complex from *Chlamydomonas reinhardtii* [1]

Susan L. Ketchner *, Richard Malkin

Department of Plant Biology, 111 Koshland Hall, University of California, Berkeley, CA 94720-3102, USA

Received 6 October 1995; revised 12 December 1995; accepted 13 December 1995

Abstract

We have determined the nucleotide sequence of the *PetM* gene from the single celled alga *Chlamydomonas reinhardtii*. The gene encodes a recently characterized, small protein of the cytochrome b_6f complex, and based on this sequence, it is proposed that this protein spans the membrane by a single α-helix. Comparison of the nucleotide sequence with the deduced amino acid sequence reveals a 60-amino-acid presequence similar to a stroma-targeting peptide.

Keywords: *PetM*; Cytochrome b_6f complex; Transit peptide; (*C. reinhardtii*)

The cytochrome b_6f complex is a plastoquinol-plastocyanin oxidoreductase which transfers electrons between the Photosystem II and Photosystem I complexes in oxygenic photosynthesis. The complex is composed of four major subunits: cytochrome *f* (*petA* gene product), cytochrome b_6 (*petB* gene product), the Rieske protein (*PetC* gene product), and subunit IV (*petD* gene product). Each of these subunits is directly involved in the catalytic reaction, either by binding redox active cofactors, such as the heme groups in cytochrome b_6 and cytochrome *f*, respectively, or the Fe-S group in the Rieske subunit or by binding the quinone substrate as with subunit IV [1–3]. In contrast, small integral membrane subunits which are thought to span the membrane as a single α-helix have recently been localized to the cyt b_6f complex although these have no identified function [4,5]. The chloroplast-encoded protein petG has been identified in a number of organisms [6,7] while the nuclear-encoded PetM has been identified only in spinach and the single-celled alga, *Chlamydomonas reinhardtii* [4,5]. Elimination of the petG protein by deletion of its gene from the chloroplast of *C. reinhardtii* results in a dramatic loss of the entire cytochrome b_6f complex from thylakoid membranes [8], though the precise function of petG is unknown. In an attempt to better characterize PetM, we present its gene sequence from *C. reinhardtii*.

Degenerate oligonucleotides were designed corresponding to the 3′(GAA GAT CTA GAC AG(C/G) ACG AA(A/G) CGC AT(C/G/A) GC(C/G/A) AG) and 5′(CCT TAA GCT TAT GGG (C/T) GA GGC (G/C/T)GA GTT (C/T) A) sequences, translated from the published *PetM* protein sequence [4]. The *PetM* gene was amplified from *C. reinhardtii* genomic DNA by PCR. After amplification several products were recovered, ranging in size from approximately 140 bp to 275 bp. Sequencing revealed that the 263 bp fragment contained the expected *PetM* sequence in addition to a 173 bp intron. This fragment was used as a probe for Southern blot analysis. Hybridization with the *PetM* PCR probe suggested that the *PetM* gene was single copy (Fig. 1). *Bam*HI-*Pst*I restriction-digested total DNA was separated electrophoretically on a TBE-agarose gel and the 2.2–2.6 kb region was cut from the gel and cloned into *Bam*HI-*Pst*I-cut pUC118. The ligation was transformed into *E. coli* DH5α by standard $CaCl_2$ methods. Colonies were screened by hybridization with the *PetM* PCR product and DNA from positive colonies was used for Southern hybridization analysis. Those which yielded single fragments of 2.4 kb from a *Bam*HI-*Pst*I digestion and which hybridized positively to the *PetM* PCR product were used for sequence analysis. The sequencing strategy is seen in Fig. 2, along with the nucleotide sequence of *PetM*.

* Corresponding author. Fax: +1 510 6424995; e-mail: ketchner@mendel.berkeley.edu.

[1] The nucleotide sequence data reported in this paper have been deposited with the GenBank database under accession number U36401.

SSDI 0005-2728(95)00157-3

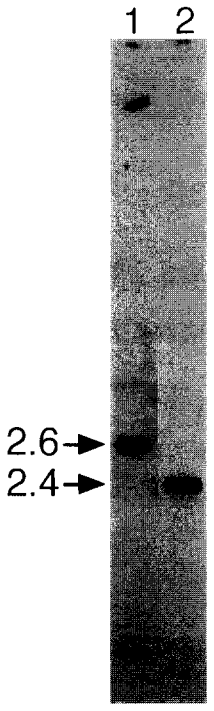

Fig. 1. Southern blot analysis of total DNA from *C. reinhardtii*. Total DNA was purified according to Herrin and Worley [9], digested with either *Pst*I (Lane 1) or *Bam*HI (data not shown) or both *Bam*HI/*Pst*I (Lane 2), separated on a 1% TAE gel and blotted onto nitrocellulose using standard methods. The filter was hybridized using the ^{32}P random-labeled *PetM* PCR. Fragments indicated are measured in kilobases (kb). The *Bam*HI cut not shown gives a band of more than 12 kb.

The *PetM* gene contains two introns, one of 111 bp within the presequence and the second of 173 bp within the mature protein. The three exons containing the complete *PetM* coding sequence, including the presequence, are 106, 129 and 59 bp, respectively. For verification of the intron splice sites, a cDNA clone from a *C. reinhardtii* cDNA library was amplified by PCR with oligonucleotides corresponding to regions just 5′ and 3′ of the determined *PetM* coding sequence (indicated on the *PetM* sequence in Fig. 2). It was determined that both introns have consensus intron/exon boundaries [10]. A consensus polyadenlyation site (TGTAA) [11] is located 215 bp from the stop codon. Northern blot analysis indicates an RNA species of approximately 600 bp, which would indicate a 5′ start site approximately 50 bp from the start codon (Fig. 3).

There are two in-frame ATG codons at the beginning of the *PetM* coding sequence with the first ATG preceded by an A rich region, typical of those in higher plants (AACAAA *AUG*GC of *C. reinhardtii PetM* gene, versus consensus of AACA *AUG*GC) [13]. The placement of the initial AUG at this first position, gives an N-terminal presequence of 60 amino acids long to give total mass to the preprocessed protein at 11 762 Da.

The presequence resembles that of a stromal-targeting peptide with its short uncharged N-terminal region and a central region rich in alanine, arginine and serine that may form a positively charged amphiphilic α-helix [14]. The amino acids (A M I A) just prior to the mature protein do not match the loosely defined consensus site (V/I) X (A/C)A, found in several chloroplast transit peptides [15]. It would be possible to assign a processing site to the amino acids 25–28 (V A A A), but the resulting 'thylakoid transfer peptide' following the processing site would not contain an uninterrupted hydrophobic stretch of amino acids, typically seen in this second domain which directs nuclear-encoded proteins to the thylakoid membrane, where *PetM* is located. It is not surprising, on the other hand, that the transit peptide of this membrane bound protein would consist of only a stromal-targeting domain as several nuclear-encoded thylakoid proteins, for example, LHCII [16,17], have only stromal-targeting presequences and utilize a segment of the mature protein to direct the protein to the membrane.

The mature protein has 39 amino acid residues and a molecular mass of 4720 Da. This is in agreement with the

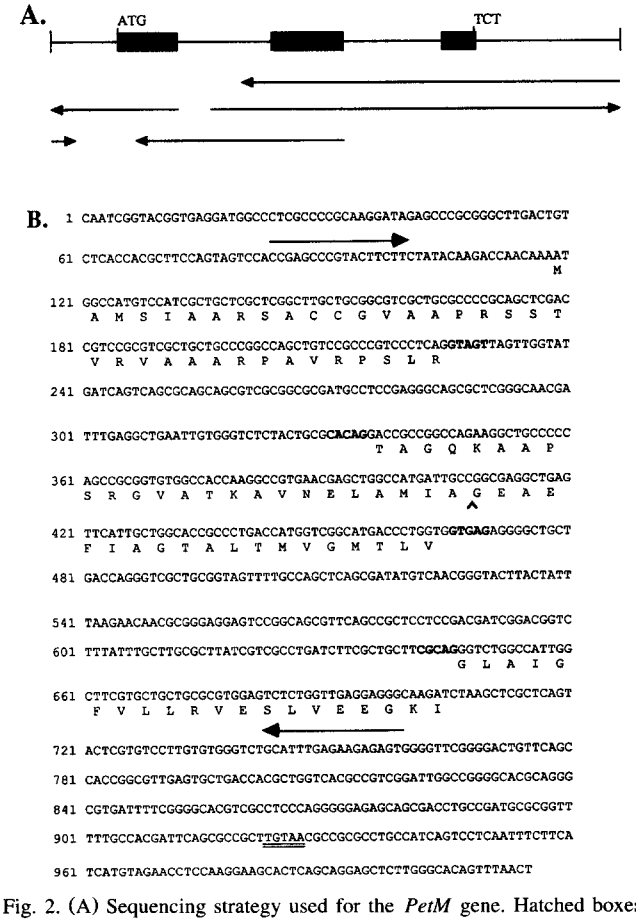

Fig. 2. (A) Sequencing strategy used for the *PetM* gene. Hatched boxes represent exon regions with the initiation ATG and termination TCT indicated. (B) Nucleotide sequence of the coding strand of the *PetM* gene. The deduced amino acid sequence is indicated beneath the nucleotide sequence with the beginning of the mature protein indicated by the arrow. Intron/exon splice sites are boldfaced; the polyadenylation signal is double-underlined. Primers used to amplify cDNA sequences are indicated by arrows above the corresponding nucleotide sequence.

apparent molecular mass of the mature chloroplast petM protein of 4.0–4.5 kDa, estimated by gel electrophoresis [4,5].

A stretch of 24 non-polar amino acids within the mature protein indicates a single membrane-spanning helix which is anchored by the presence of charged amino acids on either side of the helix, as indicated by the hydropathy plot in Fig. 4. The hydrophobic character of the protein conforms to experimental evidence of Schmidt and Malkin which suggests that the protein is membrane-bound. The more positive character of the C-terminus than the N-terminus (−1 versus −2, respectively) would orient the C-terminal end of PetM in the thylakoid lumen, according to the 'positive-inside rule' of Von Heijne [18].

It is interesting to note that Schmidt and Malkin [4] identified a similar protein in spinach, which would suggest, as with petG, that this small protein is well conserved in evolutionary distant photosynthetic organisms and would likely be an essential component for the cytochrome b_6f complex. We are now searching for the protein in cyanobacteria to determine whether this can be extended to simpler prokaryotic photosynthetic organisms.

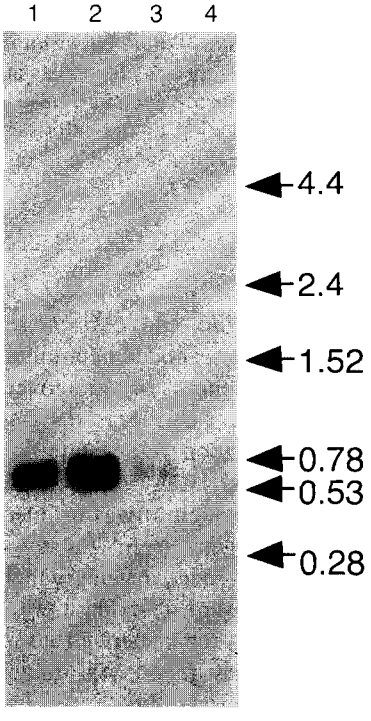

Fig. 3. Northern blot analysis of RNA from *C. reinhardtii*. RNA was isolated according to Rochaix et al. [12], separated on a Hepes-formamide gel, blotted onto nitrocellulose and hybridized with the ^{32}P random-labeled *Bam*HI-*Pst*I *PetM* genomic clone, using standard methods. Poly(A)$^+$ RNA was isolated using a Poly-T column from Bethesda Research Laboratories (Bethesda, MD). Size markers indicated are in kb. Lane (1) 0.5 μg poly(A)$^+$ RNA; (2) 1.0 μg poly(A)$^+$ RNA; (3) 5.0 μg total RNA; (4) 10.0 μg poly(A)$^-$ RNA.

Fig. 4. Hydropathy plot of the PetM mature protein sequence. The initial amino acid is the glycine residue marked by the carot on the sequence map of Fig. 2. The averaging window was 11 residues and positive values represent more hydrophobic amino acids. The plot shows evidence for the presence of a membrane spanning helix, anchored through the membrane by charged ends.

This work was supported in part by a grant from the National Institutes of Health.

References

[1] Malkin, R. (1992) Photosynth. Res. 33, 121–136.
[2] Cramer, W.A., Furbacher, P.N., Szcepaniak, A. and Tae, G.S. (1991) Curr. Top. Bioenerg. 16, 179–222.
[3] Hope, A.B. (1993) Biochim. Biophys. Acta 1143, 1–22.
[4] Schmidt, C.L. and Malkin, R. (1993) Photosyn. Res. 38, 73–81.
[5] Pierre, Y. and Popot, J.-L. (1994) C.R. Acad. Sci. (Paris) Ser. 3 316, 1404–1409.
[6] Haley, J. and Bogorad, L. (1989) Proc. Natl. Acad. Sci. USA 86, 1534–1538.
[7] Stirewalt, V.L. and Bryant, D.A. (1989) Nucleic Acids Res. 17, 10095–10095.
[8] Berthold, D.A., Schmidt, C. L. and Malkin, R. (1995) J. Biol. Chem. 270, 29293–29298
[9] Herrin, D. and Worley, T. (1990) Plant Mol. Biol. Rep. 8, 292–296.
[10] Zimmer, W.E., Schloss, J.A., Siflow, C.D., Youngblom, J. and Wetterson, D.M. (1988) J. Biol. Chem. 194, 399–411.
[11] Siflow, C.D., Chisholm, R.L., Conner, T.W., and Ranum, P.W. (1985) Mol. Cell. Biol. 5, 2389–2398.
[12] Rochaix, J.D., Mayfield, S., Goldschmidt, G.M. and Erickson, J. (1988) in Plant Molecular Biology – A Practical Approach (Shaw, C.H., ed.), pp. 253–257, IRL Press, New York.
[13] Lütcke, H.A., Chow, K.C., Michel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) EMBO J. 6, 43–48.
[14] Gavel, Y. and Von Heijne, G. (1990) FEBS Lett. 261, 455–458.
[15] Von Heijne, G., Steppuhn, J. and Hermann, R.G. (1989) Eur. J. Biochem. 180, 533–545.
[16] Smeeken, S., Weisbeek, P. and Robinson, C. (1990) Trends Biochem. Sci. 15, 73–76.
[17] Theg, S.M. and Scott, S.V. (1993) Trends Cell Biol. 3, 185–190.
[18] Gavel, Y., Steppuhn, J., Herrmann, R. and Von Heijne, G. (1991) FEBS Lett. 282, 41–48.