ELSEVIER

Biochimica et Biophysica Acta 1188 (1994) 443–446

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

Nucleotide sequence of the *petB* gene encoding cytochrome b_6 from the mesophilic cyanobacterium *Synechocystis* PCC 6803: implications for evolution and function

Jochen Kruip [a], Peter J. Nixon [b], Heinz D. Osiewacz [c], Matthias Rögner [a,*]

[a] *Institute of Botany, University of Münster, Schlossgarten 3, D-48149 Münster, Germany*
[b] *Department of Biochemistry, Imperial College of Science, Technology and Medicine, London SW7 2AY, UK*
[c] *FSP Angewandte Tumorvirologie, Abteilung: Molekularbiologie der Alterungsprozesse, Im Neuenheimer Feld 242, D-69120 Heidelberg, Germany*

Received 1 September 1994

Abstract

The gene encoding the cytochrome b_6 subunit (*petB*) of the cytochrome b_6f complex has been isolated, cloned and sequenced by nonradioactive methods from genomic DNA of the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. The coding region consists of 666 nucleotides, coding for a polypeptide with a molecular mass of 25.02 kDa. In contrast to higher plant *petB* sequences an aminoterminal extension of seven amino acids occurs. Aminoterminal sequencing of the isolated protein excludes – different from higher plants – the existence of an intron after the first amino acids but indicate the posttranslational removal of three amino acids from the amino terminus. The aminoterminal extension – found only in non-nitrogen-fixing, unicellular cyanobacteria – shows a high degree of homology between different species.

Keywords: Cyanobacterium; Cytochrome b_6f complex; Evolution; Nucleotide sequence; *petB*; Photosynthesis

The cytochrome b_6f complex is a plastoquinol : plastocyanin oxidoreductase and serves as a link between Photosystem II and Photosystem I in oxygenic photosynthesis [1]. A better understanding of the relationship between its structure and function is of primary interest since this step is the rate limiting step of the photosynthetic electron transport chain [2]. Moreover, the cytochrome b_6f complex seems to be the crosspoint between photosynthesis and respiration in cyanobacteria [3]. This complex is multisubunit and transmembrane, consisting of at least four proteins: cytochrome *f* (*petA* gene product), the Rieske protein (*petC* gene product), subunit IV (*petD* gene product) and cytochrome b_6 (*petB* gene product) [4]. In all organisms studied so far *petB* and *petD* are closely linked and are cotranscribed [5–16].

In this work the *petB* gene of the mesophilic cyanobacterium *Synechocystis* sp. PCC 6803 was amplified from genomic DNA by PCR using two synthetic oligonucleotides, specific for the 3′ (TTCCGTAAATCGGGATC) and 5′ (CCACCCTTGGTCAAACC) flanking regions of the *petB* gene ([17] and Osiewacz, unpublished results). The resulting 948 bp PCR fragment was cloned into the SrfI site of the pCR-Script SK(+) vector (Stratagene, Germany). The nucleotide sequence of three clones was determined using the dideoxynucleotide chain termination method in combination with direct blotting (Hoefer, USA) and chemiluminescent detection (Tropix, USA). In addition, a genomic copy of the gene was partially sequenced by conventional radioactive methods. The DNA sequence contains two open reading frames (ORFs) which code for cytochrome b_6 and parts of subunit IV, respectively (Fig. 1). The first ORF is homologous (70–80% identity) with reported *petB* genes and predicts a highly hydrophobic 25 019.9 Da protein. The second ORF, with only the 5′ part being present, corresponds to the *petD* gene according to [17], thus showing that the

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ nucleotide sequence databases under the accession number Z31580.

* Corresponding author. E-mail: rogner@dmswwu1a.uni-muenster.de. Fax: +49 251 833823.

SSDI 0005-2728(94)00163-4

```
1   ccaccctggtcaaaccgtcatagtacgctgattatatgcgattttacggctccggacgtcaacccctcc

1                                  Met Phe Ser Lys Glu Val Thr Glu Ser Lys
72  atccctgatccttaagagaagtctagcttca atg ttt tca aaa gaa gtc acc gaa tca aaa

11  Val Phe Gln Trp Phe Asn Asp Arg Leu Glu Val Gln Ala Ile Ser Asp Asp Ile
133 gtt ttt caa tgg ttc aat gat cgt ctg gaa gtg caa gcc atc tcc gat gac att

29  Ala Ser Lys Tyr Val Pro Pro His Val Asn Ile Phe Tyr Cys Leu Gly Gly Leu
187 gcc agc aaa tac gtt cct ccc cac gtc aat att ttt tac tgt ctg ggg ggc ctg

47  Thr Leu Thr Cys Phe Leu Ile Gln Phe Ala Thr Gly Phe Ala Met Thr Phe Tyr
241 acc ctg acc tgc ttc ctg atc cag ttt gcc act ggg ttc gcc atg acc ttc tac

65  Thy Lys Pro Thr Val Thr Glu Ala Phe Ala Ser Val Gln Tyr Ile Met Asn Glu
295 tac aaa ccc acg gtc acg gaa gcc ttc gcc tct gtc caa tac atc atg aat gaa

83  Val Asn Phe Gly Trp Leu Ile Arg Ser Ile His Arg Trp Ser Ala Ser Met Met
349 gtc aac ttt ggt tgg ctg att cgc tcc att cac cgt tgg tcc gcc agc atg atg

101 Val Leu Met Met Ile Leu His Val Phe Arg Val Tyr Leu Thr Gly Gly Phe Lys
403 gtg ctg atg atg att ctc cac gtt ttt cgg gtt tac ctc acc ggt ggt ttc aag

119 Lys Pro Arg Glu Leu Thr Trp Val Val Gly Val Met Leu Ala Val Thr Thr Val
457 aag ccc cgg gaa ttg acc tgg gtt gtg ggt gta atg cta gcc gtc acc act gtc

137 Thr Phe Gly Val Thr Gly Tyr Ser Leu Pro Trp Asp Gln Val Gly Tyr Trp Ala
511 acc ttt ggt gta acc ggt tac tcc ctt ccc tgg gac cag gtg ggt tac tgg gcg

155 Val Lys Ile Val Ser Gly Val Pro Ala Ala Ile Pro Val Val Gly Asp Gln Leu
565 gta aaa atc gtt tcc ggt gta ccg gcc gct att ccc gtg gtt ggg gat caa ttg

173 Val Thr Leu Met Arg Gly Ser Glu Ser Val Gly Gln Ala Thr Leu Thr Arg Phe
619 gtg acc ctc atg cga ggt agt gaa agc gtt ggt cag gct acc ttg acc cgc ttc

191 Tyr Ser Leu His Thr Phe Val Leu Pro Trp Ala Ile Ala Val Leu Leu Leu Leu
673 tac agc ctc cac acc ttc gtg ctt ccc tgg gcg atc gcc gtg ttg ctg ttg ttg

209 His Phe Leu Met Ile Arg Lys Gln Gly Ile Ser Gly Pro Leu ***
727 cac ttc ctg atg atc cgc aaa caa ggc att tcc gga cct ttg taa ttcctgtcaac

783 ccaccgatctgagtttgatcgggcaatttcgtttaaaccgttgccaaattgctaaagattggaacagatta

1                                  Met Ser Ile Ile Lys
854 gaaaatctagaaaacttctaaaaaaccttaacttggagcctctgcccc atg agt att atc aaa

6   Lys Pro Asp Leu Ser Asp Pro Asp Leu Arg
917 aag ccg gat ctt agc gat ccc gat tta cgg
```

Fig. 1. Nucleotide sequence of the noncoding strand of the *petB* gene and part of the *petD* gene from *Synechocystis* PCC 6803. The deduced amino acid sequence for cytochrome b_6 and for the aminoterminal part of subunit IV of the cytochrome b_6f complex is indicated. Amino acids as determined by N-terminal sequencing of the purified protein are underlined.

sequenced PCR product is not an artefact. Both ORFs are in close proximity, resembling the *petB/petD* clusters found in other organisms [4].

Comparison of the amino acid sequence of the present *petB* gene with those of other plants and cyanobacteria [5–16,18] reveals a high overall similarity (70–80%, Fig. 2). Hydropathy analysis predicts four transmembrane helices (I: C42-Y65, II: R90-F109, III: L124-D148, IV: Y191-I213) and one amphiphilic helix (I157-M176). By analogy to bc_1 complexes, four invariant His-residues (H93, H107, H194, H209) are probable the ligands to the two heme groups. Different from green plants and algae the ORF for cytochrome b_6 from *Synechocystis* predicts seven additional amino acids at the amino terminus. Higher plants instead contain an intron, ranging from 752 to 812 bp in size, after the exon coding for the first two amino acids. As introns were recently also found in prokaryotic organisms like cyanobacteria [19,20], we wanted to exclude the existence of an intron in the *petB* gene of *Synechocystis* by sequencing the amino terminus of the isolated cytochrome b_6 protein (Fig. 1). This amino acid sequence shows an aminoterminal extension of the cytochrome b_6 consisting of only four residues. It is therefore most likely that the first three amino acids of the aminoterminal extension are posttranslationally removed. Although an additional ATG codon can be found 66 bp upstream of the ORF, the start of translation from this codon is unlikely: Similar to *Synechocystis*, the gene sequence of the cyanobacterium *Synechococcus* PCC 7002 [9] and the prochlorophyte *Prochlorothrix hollandica* [13] predicts an aminoterminal extension of seven amino acids, but in these organisms no ATG codon in the corresponding upstream region could be found, indicating that translation starts indeed at the indicated ATG codon. As in these organisms no sequencing of the protein was done the aminoterminal sequence of the mature protein is unknown. Our data now indicate that the aminoterminal extension of these organisms is not the consequence of an unknown intron but is rather a critical feature of the mature protein. This view is supported by the high degree of similarity between the extensions of these organisms: A characteristic pattern of six amino acids follows the first Met residue: aromatic (F) – hydrophilic (T,S) – positively charged (K) – negatively charged (E)/neutral (Q) – hydrophobic (V) – hydrophilic (T)/negatively charged (Q). The function of this aminoterminal extension remains unknown at the moment. However, as this extension is not found in the closely related filamentous cyanobacterium *Nostoc* [5], a role in protein targeting seems unlikely.

Fig. 3 shows a dendrogram of all available cytochrome b_6 sequences. As expected, the established

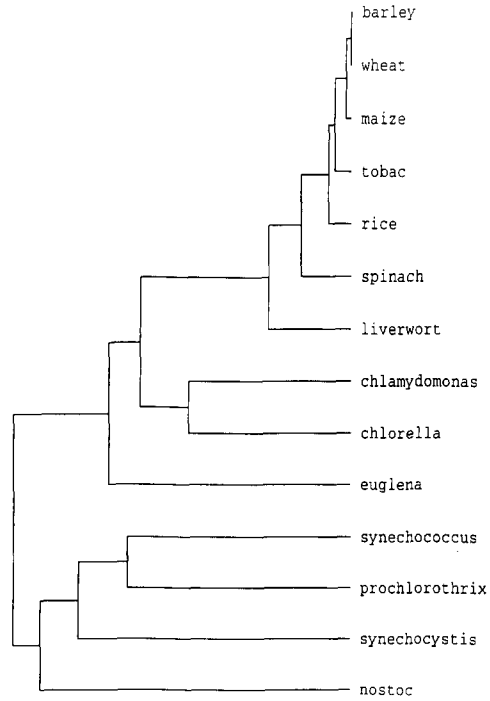

Fig. 3. Phylogenetic tree for cyanobacteria and eukaryotic phototrophs based on cytochrome b_6 sequences. The dendrogram was calculated with the PileUp program from the HUSAR program package (DKFZ Heidelberg, Germany). Distance along the horizontal axis is proportional to the difference between sequences; distance along the vertical axis has no significance at all.

phylogeny with the grouping of the cyanobacteria, higher plants and green algae can be seen. However – in analogy with the occurrence of the aminoterminal extension – the prochlorophyte *Prochlorothrix* seems to be more related to unicellular cyanobacteria like *Synechocystis* and *Synechococcus* than to the filamentous cyanobacterium *Nostoc* or green plants. This points to an interesting border position for *Prochlorothrix* in the evolution of photosynthesis: on the one hand it resembles higher plant photosynthesis due to the occurrence of a light harvesting complex containing chlorophyll *b* [21] – on the other hand, the sequence of *petB* and other genes [22] suggests a close relationship with unicellular cyanobacteria.

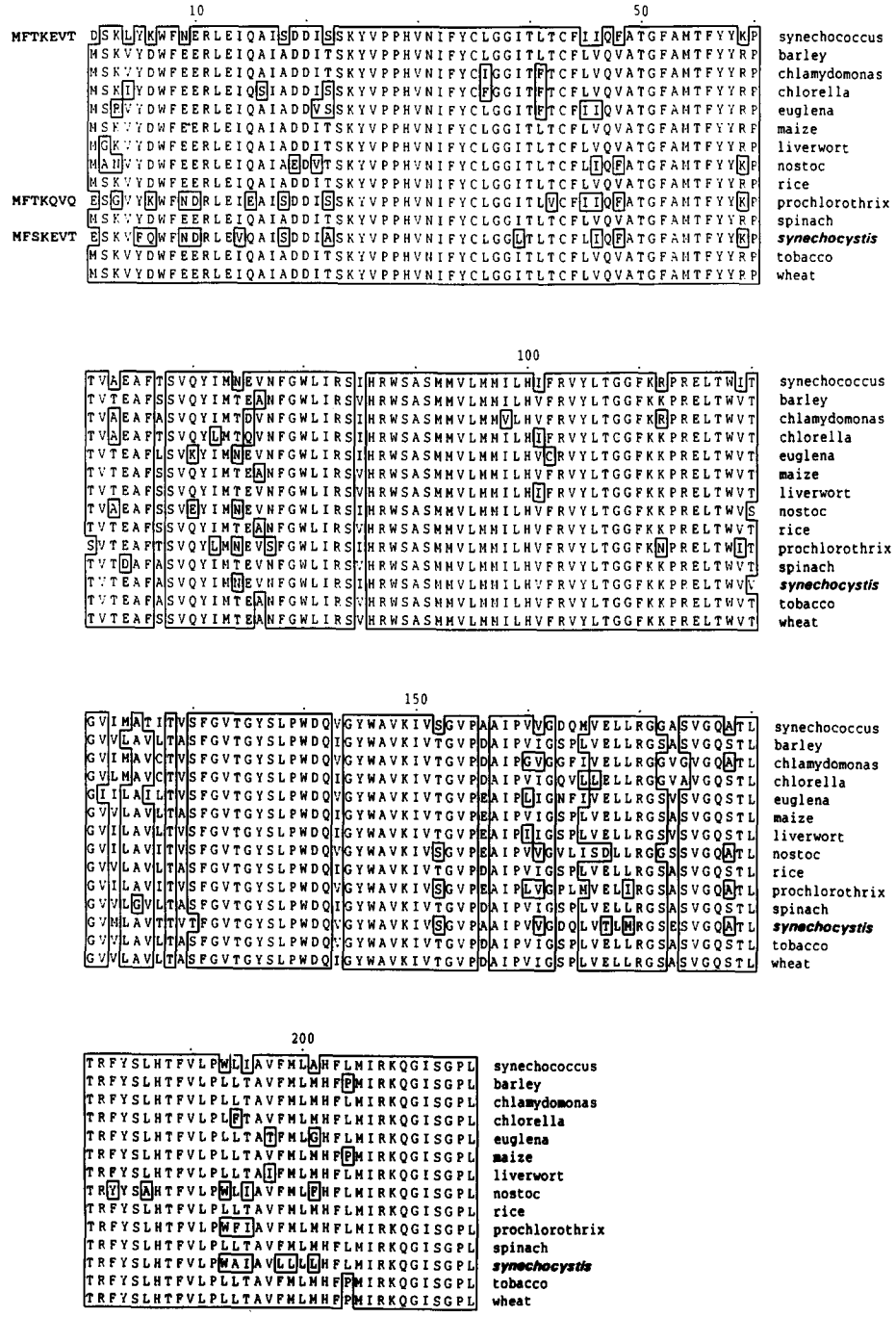

Fig. 2. Alignment of the *Synechocystis* PCC 6803 PetB protein sequence with those of other organisms reported to date. References: barley [11], wheat [14], maize [12], tobacco [8], rice [18], spinach [10], liverwort [6], *Chlamydomonas reinhardtii* [15], *Chlorella protothecoides* [7], *Euglena gracilis* [16], *Synechococcus* PCC 7002 [9], *Prochlorothrix hollandica* [13], *Nostoc* PCC 7906 [5]. Homology computation was done using the HUSAR program package (DKFZ Heidelberg, Germany).

Financial support by the Deutsche Forschungsgemeinschaft (DFG, Ro 858/2-2) is gratefully acknowledged by J.K. and M.R.

[1] Cramer, W.A. and Knaff, D.B. (1990) Energy Transduction in Biological Membranes. Springer Verlag, New York.
[2] Hope, A.B. (1993) Biochim. Biophys. Acta 1143, 1–22.
[3] Scherer, S. (1990) Trends Biochem. Sci. 15, 458–462.
[4] Cramer, W.A., Martinez, S.E., Huang, D., Tae, G.-S., Everly, R.M., Heymann, J.B., Cheng, R.H., Baker, T.S. and Smith, J.L. (1994) J. Bioenerg. Biomembr. 26, 31–47.
[5] Kallas, T., Spiller, S. and Malkin, R. (1989) J. Biol. Chem. 263, 14334–14342.
[6] Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesons, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.
[7] Reimann, A. and Kück, U. (19) Plant Mol. Biol. 13, 255–256.
[8] Shinzaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hyashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torozawa, K., Meng, B.Y., Sugita, M., Deno, H, Komogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
[9] Brand, S.N., Tan, X. and Widger, W.R. (1992) Plant Mol. Biol. 20, 481–491.
[10] Heinemeyer, W., Alt, J. and Herrmann, R.G. (1984) Curr. Genet. 8, 543–549.
[11] Reverdatto, S.V., Andreeva, A.V., Buryakova, A.A., Chakhmakhcheva, O.G. and Efimov, V.A. (1989) Nucleic Acids Res. 17, 2859–2860.
[12] Rock, C.D., Barkan, A. and Taylor, W.C. (1987) Curr. Genet. 12, 69–77.
[13] Greer, K.L. and Golden, S.S. (1992) Plant Mol. Biol. 19, 355–365.
[14] Hird, S.M., Webber, A.N., Wilson, R.J., Dyer, T.A. and Gray, J.C. (1991) Plant Mol. Biol. 16, 745–747.
[15] Bueschlen, S., Choquet, Y., Kuras, R. and Wollman, F.A. (1991) FEBS Lett. 284, 257–262.
[16] Hallick, R.B., Hong, L., Drager, R.G., Favreau, M., Monfort, A., Orsat, B., Spielmann, A. and Stutz, E. (1993) Nucleic Acids Res. 21, 3537–3544.
[17] Osiewacz, H.D. (1992) Arch. Microbiol. 157, 336–343.
[18] Cote, J.C., Wu, N.H. and Wu, R. (1988) Plant Mol. Biol. 11, 873–874.
[19] Ferat, J.-L. and Michel, F. (1993) Nature 364, 358–361.
[20] Xu, M.-Q., Kathe, S.D., Goodrich-Blairt, H., Nierzwicki-Bauer, S.A. and Shub, D.A. (1990) Science 250, 1566–1569.
[21] Lockhart, P.J., Penny, D., Hendy, M.D. and Larkum, A.W.D. (1993) Photosynth. Res. 37, 61–68.
[22] Palenik, B. and Haselkorn, R. (1992) Nature 355, 265–267.