ELSEVIER

Biochimica et Biophysica Acta 1188 (1994) 447–449

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

Nucleotide sequence of the *petE* gene encoding plastocyanin from the photosynthetic prokaryote, *Prochlorothrix hollandica* ☆

Arulvathani Arudchandran, Dierk Seeburg [1], William Burkhart [2], George S. Bullerjahn *

Department of Biological Sciences, Center for Photochemical Sciences, Bowling Green State University, Bowling Green, OH 43403, USA

Received 29 August 1994

Abstract

We have determined the nucleotide sequence of the *petE* gene encoding plastocyanin from the chlorophyll *a*/*b*-containing photosynthetic prokaryote, *Prochlorothrix hollandica*. Comparison of the deduced amino acid sequence encoded by the gene with the N-terminal sequence of the purified protein revealed that plastocyanin is synthesized as a precursor bearing an N-terminal domain of 34 amino acids having some structural similarity to thylakoid lumenal transit peptides identified in other organisms. The mature protein has an apparent isoelectric point of 8.37 and a molecular mass of 10236 Da.

Keywords: Plastocyanin; Copper protein; *petE*; Prochlorophyte; Transit peptide; Photosynthesis; (*Prochlorothrix*)

Plastocyanin is a small (~ 11 kDa), copper-containing protein that functions as an electron donor to Photosystem I in both chloroplast systems and in some strains of cyanobacteria (for review see [1]). In some lower eukaryotes and most cyanobacteria, a small *c*-type cytochrome (cyt) serves this function [2,3], and many of these organisms can synthesize both components, replacing plastocyanin with cyt *c*-553 under conditions of copper limitation [4,5]. *Prochlorothix hollandica* is a chlorophyll *a*/*b*-containing photosynthetic prokaryote that has been shown to be a divergent member of the cyanobacteria, based on comparisons made through parsimony analysis of both protein-coding genes and 16S rRNAs ([6–9]; review, [10]). This report supports these findings, as the *P. hollandica petE* gene is shown to encode a plastocyanin having a primary structure divergent from other cyanobacterial and chloroplast homologs.

P. hollandica plastocyanin was purified to homogeneity by standard methods [11] employing anion exchange chromatography of soluble extracts from cells grown in BG-11 medium containing copper [12]. Following concentration by lyophilization, the blue-gray material was subjected to sequencing by automated Edman degradation. Sequences were obtained from both the N-terminus and from HPLC-purified peptides generated by CNBr treatment. From the available peptide sequence data, a 64-degenerate 20 residue oligonucleotide probe (5′-GARTTYGTNATGAA-YAARGT-3′) was synthesized that encoded amino acids 30–36 of the mature polypeptide. The probe identified 2 positive λEMBL3 plaques of 15000 screened; these clones both yielded identical patterns of restriction fragments. Subcloning appropriate restriction fragments into pBluescript SK^- allowed for nucleotide sequencing by the chain termination method [13]. Such sequencing identified a 393 nucleotide open reading frame encoding all of the peptide sequences determined by Edman degradation of the purified protein (Fig. 1, underlined). The start of transcription was determined by primer extension [14] to be 26 nucleotides prior to the ATG start codon (Fig. 1, asterisk). Sequence analysis of the 5′ and 3′ ends of the gene did not identify any consensus sequences that might represent a promoter or a transcriptional terminator, although 3′ to the coding region is a 48 nucleotide

☆ The nucleotide sequence reported in this paper has been deposited in the Genbank database under the accession number U13912.

* Corresponding author. Fax: +1 (419) 3722024.

Present addresses: [1] Department of Botany, Arizona State University, Tempe, AZ 85287, USA; [2] Glaxo Research Institute, 5 Moore Drive, Research Triangle Park, NC 27709, USA.

SSDI 0005-2728(94)00164-2

```
GAGGTCTCAGATCCCCCCCTGAACCCAAAAAATAAAAAAATCCTACGCCAGGGGAACTCA 60

                                                          *
CCGATACATTAACCGTTGTATAGTCTTACTACGGCTTGTCGAAACGACAGCCCGTCAAAA 120

TTTTTATTTCAAAGAACCATGAAATTTTTTGCATCCCTGTCTAAGCGTTTTGCTGCGGTT 180
                  M  K  F  F  A  S  L  S  K  R  F  A  A  V  14

TTATCTCTCGTCGTGCTGGTGGCTGGCACCCTTCTGCTGTCTGCCgctccCGCTTCTGCG 240
 L  S  L  V  V  L  V  A  G  T  L  L  L  S  A  A  P  A  S  A  34

GCTACGGTTCAAATCAAAATGGGCACCGATAAGTATGCCCCCCTCTATGAACCCAAGGCC 300
 A  T  V  Q  I  K  M  G  T  D  K  Y  A  P  L  Y  E  P  K  A  54

TTGTCCATCAGCGCCGGTGATACCGTTGAGTTCGTGATGAACAAGGTTGGTCCCCACAAC 360
 L  S  I  S  A  G  D  T  V  E  F  V  M  N  K  V  G  P  H  N  74

GTGATCTTTGATAAGGTTCCCGCCGGTGAGAGCGCCCCTGCTCTGTCCAACACCAAGTTG 420
 V  I  F  D  K  V  P  A  G  E  S  A  P  A  L  S  N  T  K  L  94

CGTATCGCTCCGGGTTCGTTCTACAGCGTCACCCTAGGAACCCCCGGTACCTACAGCTTC 480
 R  I  A  P  G  S  F  Y  S  V  T  L  G  T  P  G  T  Y  S  F  114

TATTGCACGCCCCACCGTGGCGCTGGCATGGTCGGCACCATCACCGTTGAATAATTGCAA 540
 Y  C  T  P  H  R  G  A  G  M  V  G  T  I  T  V  E  ***       131

TCATTGTTAGAGACTACGCTCTGTTAAGTCTGGCTTTAAGTCTGGCTTTAAGTCTGGTTT 600

TAAGTCTGGTTTCAGCAGAGTTTAGATGCTAAATGGGCGCCATGGGTACC 650
```

Fig. 1. Nucleotide sequence of the *P. hollandica petE* gene and the deduced primary structure of its product. The amino acid sequences determined by peptide sequencing of the processed N-terminus and CNBr-generated peptides are underlined. The transcriptional start site is indicated by the asterisk. Underlined nucleotide sequences are the putative ribosome binding site (nucleotides 131–135), the oligonucleotide probe annealing site (nucleotides 328–347) and a direct repeat region following the coding sequence.

A

```
          +       ++
NH2-MKFFASLSKRFAAVLSLVVLVAGTLLLSAAPASA/ATVQI...
```

B

```
                  *          *          *          *          *50
P. holl    ATVQIKMGTDKYAPLYEPKALSISAGDTVEFVMNKVGPHNVIFDKVPAGE
Anabaena   ETYTVKLGSDKGLLVFEPAKLTIKPGDTVEFLNNKVPPHNVVFDAALNPA
Syn 6803    NATVKMGSDSGALVFEPSTVTIKAGEEVKWVNNKLSPHNIVFAAD---G
Entero      AAIVKLGGDDGSLAFVPNNITVGAGESIEFINNAGFPHNIVFDEDAVPA
Scen         ANVKLGADSGALVFEPATVTIKAGDSVTWTNNAGFPHNIVFDEDAVPA
Chlamy      DATVKLGADSGALEFVPKTLTIKSGETVNFVNNAGFPHNIVFDEDAIPS
Poplar       IDVLLGADDGSLAFVPSEFSISPGEKIVFKNNAGFPHNIVFDEDSIPS

                  *          *                     *          *         97
P. holl    SAPALSNTKLRIAPGSFYSVTLGT-------PGTYSFYCTPHRGAGMVGTITVE
Anabaena   KSADLAKSLSHKQLLMSPGQSTSTTFPADAPAGEYTFYCEPHRGAGMVGKITVAG
Syn 6803   VDADTAAKLSHKGLAFAAGASFTSTFTE---PGTYTYYCEPHRGAAMVGKVVVE
Entero     GVDADAISA-E-DYLNSKGQTVVRKLTT---PGTYGVYCDPHSGAGMKMTITVQ
Scen       GVNADALSH-D-DYLNAPGESYTAKFDT---AGEYGYFCEPHQGAGMVGKVIVQ
Chlamy     GVNADAISR-D-DYLNAPGETYSVKLTA---AGEYGYYCEPHQGAGMVGKIIVQ
Poplar     GVDASKISMSEEDLLNAKGETFEVALSN---KGEYSFYCSPHQGAGMVGKVTVN

percent identity/similarity:   Anabaena        41/55
                               Synechocystis   35/56
                               Enteromorpha    33/48
                               Scenedesmus     31/47
                               Chlamydomonas   36/46
                               Populus nigra   31/44
```

Fig. 2. The *P. hollandica* plastocyanin primary structure. Panel A: amino acid sequence of the putative thylakoid transit peptide, indicating the basic amino acid residues and the major hydrophobic domain (underlined). The processing site is indicated by the slash (/). Panel B: alignment of the amino acid sequence of *P. hollandica* plastocyanin with plastocyanins from eukaryotic and cyanobacterial sources [16–22]. Shown in boldface are residues that are either identical or conservative replacements in all the sequences; for each homolog, the percent identity and similarity are also listed. His-39, Cys-83, His-85 and Met-90 are conserved residues involved in forming the copper binding site [1].

sequence comprising a complex region of direct repeats (Fig. 1, underlined). The deduced primary translation product bears an N-terminal transit peptide of 34 amino acids; immediately upstream from the methionine start codon is a putative ribosome binding site (AAAGA). Analysis of the transit peptide sequence shares some structural similarity to thylakoid transit peptides from other organisms (Fig. 2A), as the N-terminus is highly basic and followed by a hydrophobic core domain [15]. The mature polypeptide has an apparent molecular mass of 10236 Da and a basic isoelectric point of 8.37. The deduced amino acid sequence of the mature protein is structurally more similar to the plastocyanins of cyanobacterial origin than to the chloroplast homologs from green algae and higher plants ([16–22]; Fig. 2B), as the *P. hollandica* protein is 41% identical and 55% similar to *Anabaena* plastocyanin, and only 31% identical and 44% similar to plastocyanin from poplar (Fig. 2B). Unsurprisingly, the greatest amount of divergence in the *P. hollandica* primary structure is seen between residues 45 and 74, a region that is not highly conserved across the protein family, although a negatively charged patch of amino acids is shared among the eukaryotic homologs [1]. As is seen in the cyanobacterial proteins, the *P. hollandica* polypeptide lacks both the hydrophobic and negatively charged patches shared by the chloroplast plastocyanins; such domains comprise the reaction partner binding sites in the chloroplast proteins (review, Ref. [1]). Overall, these sequence comparisons support the conclusions of others stating that *P. hollandica* is a deeply branched member of the cyanobacterial lineage [6–10].

This work was supported by National Science Foundation grant MCB-9204672 to G.S.B.

References

[1] Redinbo, M.R., Yeates, T.O. and Merchant, S. (1994) J. Bioenerg. Biomembr. 26, 49–88.
[2] Wood, P.M. (1978) Eur. J. Biochem. 87, 9–19.
[3] Ho, K.K. and Krogmann, D.W. (1982) Biochim. Biophys. Acta 766, 310–316.
[4] Merchant, S. and Bogorad, L. (1986) Mol. Cell Biol. 6, 462–469.
[5] Sandmann, G. (1986) Arch. Microbiol. 145, 76–79.
[6] Turner, S., Burger-Wiersma, T., Giovanonni, S.J., Mur, L.R. and Pace, N.R. (1988) Nature 337, 380–383.
[7] Palenik, B. and Haselkorn, R. (1992) Nature 355, 265–267.
[8] Urbach, E., Robertson, D.L. and Chisholm, S.W. (1992) Nature 335, 267–270.
[9] Greer, K.L. and Golden, S.S. (1992) Plant Mol. Biol. 19, 355–365.
[10] Bullerjahn, G.S. and Post, A.F. (1993) CRC Crit. Rev. Microbiol. 19, 43–59.
[11] Lightbody, J.J. and Krogmann, D.W. (1967) Biochim. Biophys. Acta 131, 508–515.
[12] Allen, M. (1968) J. Phycol. 4, 1–3.
[13] Sanger, F.S., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[14] Calzone, F.J., Britten, R.J. and Davidson, E.H. (1987) Methods Enzymol. 152, 611–623.
[15] Mackle, M.M. and Zilinskas, B.A. (1994) J. Bacteriol. 176, 1857–1864.
[16] Aitken, A. (1975) Biochem. J. 149, 675–683.
[17] Chazin, W.J. and Wright, P.E. (1988) J. Mol. Biol. 202, 623–636.
[18] Moore, J.M., Chazin, W.J., Powls, R. and Wright, P.E. (1988) Biochemistry 27, 7806–7816.
[19] Van der Plas, J., Bovy, A., Kruyt, F., De Vrieze, G., Dassen, E., Klein, B. and Weisbeek, P. (1989) Mol. Microbiol. 3, 275–284.
[20] Briggs, L.M., Pecoraro, V.L. and McIntosh, L. (1990) Plant Mol. Biol. 15, 633–642.
[21] Collyer, C.A., Guss, J.M., Sigimura, Y., Yoshizaki, F. and Freeman, H.C. (1990) J. Mol. Biol. 211, 617–632.
[22] Merchant, S., Hill, K., Kim, J.H., Thompson, J., Zaitlin, D. and Bogorad, L. (1990) J. Biol. Chem. 265, 12372–12379.
[23] Burger-Wiersma, T. and Matthijs, H.C.P. (1990) in Autotrophic Microbiology and One-Carbon Metabolism (Codd, G.A. et al., eds.), pp. 1–24.