Nucleotide sequence encoding the di-haem cytochrome c_{551} peroxidase from *Pseudomonas aeruginosa*

Christopher J. Ridout*, Richard James, Colin Greenwood

School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Received 18 April 1995

Abstract The nucleotide sequence of the gene encoding cytochrome c_{551} peroxidase from *Pseudomonas aeruginosa* is reported. The translated amino acid sequence differs from the sequence reported earlier by peptide mapping most significantly by the presence of a section containing an additional 20 residues. A number of minor differences are also evident. The new sequence translates to a protein containing 346 amino acids, the first 23 being typical of a hydrophobic leader peptide with a characteristic protease cleavage site.

Key words: Cytochrome *c* peroxidase; *Pseudomonas aeruginosa*; Nucleotide sequence; Amino acid sequence

1. Introduction

Cytochrome c_{551} peroxidase from *Pseudomonas aeruginosa* (PsCCP) catalyses the peroxidative oxidation of azurin and cytochrome c_{551} and contains two haem *c* moities on a single polypeptide chain [1]. The two *c*-haem groups are bound covalently at the Cys-X-Y-Cys-His motif characteristic of *c*-type cytochromes, with one high potential group functioning as an electron-accepting pole and the other a low-potential peroxidatic centre [2]. The interaction between the enzyme and the reducing substrates has been extensively investigated by a variety of spectroscopic and kinetic techniques [1,3].

The primary structure of PsCCP has been determined by aligning peptide sequences after cyanogen bromide fragmentation and is reported to comprise 302 amino acids giving a calculated molecular weight of 33,690 Da [4]. Recently, we have determined by electrospray mass spectrometry the molecular weight to be about 36,500 Da, some 2,500 Da heavier than that predicted from the primary amino acid sequence [5]. The molecular size has also been estimated by SDS-PAGE to be about 40 kDa, similar to the di-haem cytochrome *c* peroxidases recently isolated from *Paracoccus denitrificans* (42 kDa), *Nitrosomonas europaea* (44 kDa) and *Rhodobacter capsulatus* (44 kDa) [6,7,8].

The haem groups of bacterial *c*-type cytochromes are normally covalently attached following translocation of the apoprotein to the periplasm. Cytochrome *c* peroxidase can be isolated from the periplasm of *Pseudomonas stutzeri* and *Paracoccus denitrificans*, but inconsistent results have been obtained when determining the location of the enzyme in *P. aeruginosa*. In some experiments the enzyme could be detected in the cytoplasm or attached to the cell membranes, but such results could be be an artefact of the method used to fractionate the cells [6,9]. Periplasmic proteins are normally synthesised with a hydrophobic leader peptide required for translocation, which is subsequently cleaved during maturation of the holo-protein [10].

As part of an on-going programme to investigate the interaction between PsCCP and its substrates, we have cloned the gene so that we can attempt to produce large quantities of the enzyme by cloning into an expression vector and transformation to a suitable host. From the DNA sequence we can establish accurately the amino acid sequence enabling the correct assignments to be made for our crystallographic data [5] and allowing the three dimensional structure to be determined.

2. Materials and methods

2.1. Bacterial strains and plasmids

Bacterial strains used in the study were *E. coli* JM101, JM105, TA Cloning One Shot competent cells (Invitrogen, San Diego, CA, USA), and *Pseudomonas aeruginosa* NTCC 6750. *E. coli* was routinely grown in LB medium [11] and *P. aeruginosa* in a *Pseudomonas* growth medium [12]. Plasmids used were pUC 18, pUC 18 *Sma*I BAP (Pharmacia Biotech, Milton Keynes, UK) and pCR II from the TA Cloning kit (Invitrogen, San Diego, CA).

2.2. Enzymes and general cloning procedures

Isolation of genomic DNA from *P. aeruginosa* was as described in Current Protocols on CD-ROM (Greene Publishing Associates and John Wiley and Sons Inc., USA). Digestion with restriction enzymes (Boeringer-Mannheim GmbH, Germany), agarose gel electrophoresis, ligation with T4 Ligase (New England Biolabs, USA) and transformation were essentially as described by Maniatis et al. [11] and in accordance with the instructions of the supplier. DNA was purified from agarose gels by the Geneclean II kit (BIO 101, Vista, CA, USA) and routine analysis of plasmids by the Insta-Mini-Prep Kit (3′→5′ Boulder, CO, USA). DNA for sequencing was routinely purified by Quiagen Plasmid Kit (Qiagen Inc., Chatsworth, CA, USA).

2.3. Construction of a gene probe by polymerase chain reaction and isolation of a fragment by Southern hybridisation

The following heterologous oligonucleotide primers were constructed from the amino acid sequence proposed by Ronnberg et al. [4], with *Sac*I and *Kpn*I restriction sites (italics), respectively, included for subsequent cloning: *H61* = 5′*GAGCTC*GA[C,T]GC[C,G,T,A]CA-[G,A]C3′ and, from the complementary DNA strand; *K34* = C*GGTA-CC*CAC[G,T,A]AT[G,A]TT[C,T]TC (Fig. 1). The primers were used (6.3 μM) in the polymerase chain reaction with *P. aeruginosa* genomic DNA (1 μg), Taq DNA polymerase (1.25 U) (Boehringer-Mannheim) and dNTPs (200 mM) (Pharmacia Biotech) in a DNA Thermal Cycler (Perkin Elmer Cetus, Norwalk, CN, USA) with 30 cycles of 94°C (1 min), 55°C (1 min) and 72°C (1.5 min). The resulting 340 base pair fragment was cloned into the pCR II vector, and the sequence verified to be part of the PsCCP gene by comparison of the translated amino acid sequence with the published sequence. The fragment was excised and labelled with ^{32}P-labelled gATP (Amersham, UK) with the Multiprime DNA labelling System (Amersham). Total genomic DNA from *P. aeruginosa* was restricted for 16 h with *Eco*RV, electrophoresed, blotted onto nucleic acid tranfer membrane (Hybond N$^+$) (Amersham) and probed with the labelled PCR fragment. Fragments of about 5.2 kb hybridising to the PCR probe were isolated by electrophoresis,

*Corresponding author. Fax: (44) (1603) 59-2250;
E-mail: c.ridout@uea.ac.uk

SSDI 0014-5793(95)00461-0

purified by the Geneclean II kit and cloned into pUC 18 *Sma*I BAP plasmid resulting in plasmid pCR01. The PCR fragment contained a *Pst*I site (Fig. 1) and restriction with this enzyme indicated that the entire gene must be contained within the 5.2 kb fragment of pCR01.

2.4. Sequencing and sequence analysis

The 5.2 kb insert was excised from from pCR01 by restriction with *Kpn*I and *Hin*dIII, purified by electrophoresis and sonicated for 10 s with a Soniprobe (Dawe, UK). Fragments of 400–600 bases were selected by electrophoresis, purified by the Geneclean II kit and ragged ends end-filled by T4 polymerase and polynucletide kinase (Pharmacia Biotech). The fragments were sub-cloned into pUC *Sma*I BAP and clones were selected randomly and sequenced by the universal forward and reverse primers with the AutoRead. Sequencing Kit (Pharmacia Biotech). Fluorescent oligonucletide primers were constructed for sequencing the regions not obtained by the random method. The region containing the entire PsCCP gene was sequenced at least twice in both directions and sequences were aligned by the LaserGene programme (DNAstar, Madison, WI, USA). Analysis of hydrophilicity [13] was performed with GeneJockey II programme (Biosoft, Cambridge, UK) and alignment to other proteins with the LaserGene Programme.

2.5. Nucleotide sequence accession number

The Genbank accession number for the gene sequence is U23766.

3. Results and discussion

The nucleotide and translated amino acid sequence of the PsCCP gene is shown in Fig. 1. Compared to the published amino acid sequence, the major difference is an insertion of 20 residues between G99 and D118. This insertion accounts for much of the discrepancy between the molecular weight previously reported [4] and our measurement by electrospray mass spectrometry [5]. Although there are no published gene sequences for any known di-haem cytochrome *c* peroxidase, the *Mau*G gene from *Methylobacterium extorquens* translates to a presumed di-haem protein with significant homology to the enzyme from *P. aeruginosa* [14]. A hypothetical open reading frame encoding a presumed tri-heam protein homologous to the *P. aeruginosa* enzyme has also recently been identified from *E. coli* (SwissProt; YHJA_ECOLI) [15]. The missing section of the PsCCP protein sequence is homologous to similar sections of the *E. coli* ORF and the *Mau*G protein (Fig. 2) indicating that this omission is a result of a protein sequencing error rather that a difference in the protein between two strains of *P. aeruginosa*. The missing section is very hydrophobic (Fig. 3) and is contained within a peptide spanning the region from K98 to K122 that would result from a tryptic digest. The mobility of hydrophobic peptides in reverse phase HPLC can be severly retarded unless the solvents are supplemented with organic acids such as formate [16], and it is possible that this peptide did not elute from the column during the purification of CNBr and tryptic cleavage products.

There are several minor differences between the translated

```
    Met Gln Ser Ser Gln Leu Leu Pro Leu Gly Ser Leu Leu Leu Ser Phe Ala Thr Pro Leu Ala Gln    22
  1 ATG CAG TCC TCG CAA CTG CTC CCG CTT GGC AGC CTG TTG CTG TCG TTC GCC ACG CCC CTC GCC CAG

    Ala Asp Ala Leu His Asp Gln Ala Ser Ala Leu Phe Lys Pro Ile Pro Glu Gln Val Thr Glu Leu    44
 67 GCC GAT GCC CTG CAC GAC CAG GCC AGC GCG CTG TTC AAA CCC ATC CCC GAG CAG GTC ACC GAG CTA

    Arg Gly Gln Pro Ile Ser Glu Gln Gln Arg Glu Leu Gly Lys Lys Leu Phe Phe Asp Pro Arg Leu    66
133 CGC GGC CAG CCC ATC AGC GAG CAA CAG CGC GAA CTC GGC AAG AAA CTG TTC TTC GAC CCG CGC CTG

    Ser Arg Ser His Val Leu Ser Cys Asn Thr Cys His Asn Val Gly Thr Gly Gly Ala Asp Asn Val    88
199 TCA CGC AGC CAT GTG CTC AGC TGC AAC ACC TGC CAC AAC GTC GGC ACC GGC GGC GCC GAC AAC GTA

    Pro Thr Ser Val Gly His Gly Trp Gln Lys Gly Pro Arg Asn Ser Pro Thr Val Phe Asn Ala Val   110
265 CCG ACG TCG GTC GGT CAC GGC TGG CAG AAG GGG CCA CGC AAT TCA CCG ACG GTC TTC AAC GCC GTG

    Phe Asn Ala Ala Gln Phe Trp Asp Gly Arg Ala Lys Asp Leu Gly Glu Gln Ala Lys Gly Pro Ile   132
331 TTC AAT GCC GCG CAG TTC TGG GAT GGC CGT GCC AAG GAC CTG GGG GAG CAG GCC AAG GGT CCG ATC

    Gln Asn Ser Val Glu Met His Ser Thr Pro Gln Leu Val Glu Gln Thr Leu Gly Ser Ile Pro Glu   154
397 CAG AAC AGC GTC GAG ATG CAC AGT ACC CCG CAG TTG GTC GAA CAG ACC CTG GGC AGC ATC CCG GAA

    Tyr Val Asp Ala Phe Arg Lys Ala Phe Pro Lys Ala Gly Lys Pro Val Ser Phe Asp Asn Met Ala   176
463 TAC GTG GAC GCC TTC CGC AAG GCC TTC CCC AAG GCC GGC AAG CCG GTC AGC TTC GAC AAC ATG GCG

    Leu Ala Ile Glu Ala Tyr Glu Ala Thr Leu Val Thr Pro Asp Ser Pro Phe Asp Leu Tyr Leu Lys   198
529 CTG GCC ATC GAG GCC TAC GAG GCG ACC CTG GTG ACT CCG GAC TCG CCC TTC GAC CTG TAT CTC AAG
                        H61 ------------>
    Gly Asp Asp Lys Ala Leu Asp Ala Gln Gln Lys Lys Gly Leu Lys Ala Phe Met Asp Ser Gly Cys   220
595 GGC GAC GAC AAG GCC CTC GAC GCA CAG CAG AAG AAA GGC CTC AAG GCA TTC ATG GAC AGT GGC TGC
                                                                          Pst I
    Ser Ala Cys His Asn Gly Ile Asn Leu Gly Gly Gln Ala Tyr Phe Pro Phe Gly Leu Val Lys Lys   242
661 AGC GCC TGC CAC AAC GGC ATC AAC CTG GGC GGC CAG GCC TAC TTC CCG TTC GGC CTG GTG AAG AAG

    Pro Asp Ala Ser Val Leu Pro Ser Gly Asp Lys Gly Arg Phe Ala Val Thr Lys Thr Gln Ser Asp   264
727 CCC GAC GCC AGC GTG CTG CCC AGC GGC GAC AAG GGC CGC TTC GCC GTG ACC AAG ACC CAG AGC GAC

    Glu Tyr Val Phe Arg Ala Ala Pro Leu Arg Asn Val Ala Leu Thr Ala Pro Tyr Phe His Ser Gly   286
793 GAG TAC GTA TTC CGC GCC GCG CCC CTG CGC AAC GTC GCC CTC ACC GCG CCG TAC TTC CAC AGC GGC

    Gln Val Trp Glu Leu Lys Asp Ala Val Ala Ile Met Gly Asn Ala Gln Leu Gly Lys Gln Leu Ala   308
859 CAG GTC TGG GAA CTC AAG GAC GCG GTG GCG ATC ATG GGC AAC GCC CAG CTC GGC AAG CAG TTG GCG
                  <------------ K34
    Pro Asp Asp Val Glu Asn Ile Val Ala Phe Leu His Ser Leu Ser Gly Lys Gln Pro Arg Val Glu   330
925 CCG GAC GAC GTG GAG AAC ATC GTC GCC TTC CTG CAC AGC CTG AGC GGC AAG CAG CCG CGC GTC GAA

    Tyr Pro Leu Leu Pro Ala Ser Thr Glu Thr Thr Pro Arg Pro Ala Glu Stop                      346
991 TAT CCG CTG CTG CCG GCC AGC ACG GAG ACC ACG CCG CGT CCC GCG GAA TAA
```

Fig. 1. Nucleotide and translated amino acid sequence of PsCCP. The amino acids comprising the presumed leader peptide are shown in italics. The location of the oligonucleotide primers (H61 andK34) used for the PCR reaction are shown by arrows and the location of the *Pst*I site used to verify that the 5.2 kB *Eco*RV fragment contained the entire gene is also shown.

```
Conserved    ...P...P.V.N.........WDGR
Majority     XXGPRNXPTVFNXVFNXAQFWDGR

Ps.a PsCCP   QKGPRNSPTVFNAVFNAAQFWDGR  120
E.co ORF     VPLPRRTPPVLNLAWGTAFQWDGR  134
M.ex MauG    AVGPINAPTVFNSVFNVEQFWDGR  253
```

Fig. 2. Alignment of the previously ommited section (G99–D119) of PsCCP with the same region from the the presumed *Mau*G protein from *M. extorquens* [14] and the open reading frame from *E. coli* [15].

gene sequence and the previously published amino acid sequence (Fig. 4). Four of these differences were recorded as variants in the original sequence, possibly because the laboratory isolate used contained more than one strain producing slightly different forms of the protein. The other differences recorded in the present study could also be due to strain differences or possibly to protein sequencing errors.

A hydrophobic leader peptide is evident in the translated sequence (Fig. 3), with a characteristic protease recognition site [17,18] at Ala-21 and Ala-23, the cleavage site being at Ala-23 (Fig. 1). The leader peptide and protease cleavage site indicate that PsCCP is likely to be directed to the periplasm for maturation and covalent attachment of *c*-type haems. The PsCCP apoprotein is therefore initally comprised of 346 amino acids residues giving a calculated molecular weight of 37,408 Da. After cleavage of the leader peptide, the processed PsCCP apoprotein comprises 323 amino acid residues with a calculated molecular weight of 35,038 Da. The inconsistent association of PsCCP with the membranes [6,9] could therefore be due to the inconsistent cleavage of the leader peptide.

The gene sequence presented is the first for a known di-haem cytochrome *c* peroxidase and corrects the previously reported amino acid sequence. The new sequence will now enable the three-dimensional structure to be resolved from our crystallographic data [5], and permit the gene to be cloned into an expression vector for possible production of large quantities of the enzyme. There have recently been several successes with heterologous over-expression of proteins containing *c*-type heams both in *E. coli* and *Pseudomonas putida* [19,20,21]. An expression system will also enable us to embark on studies

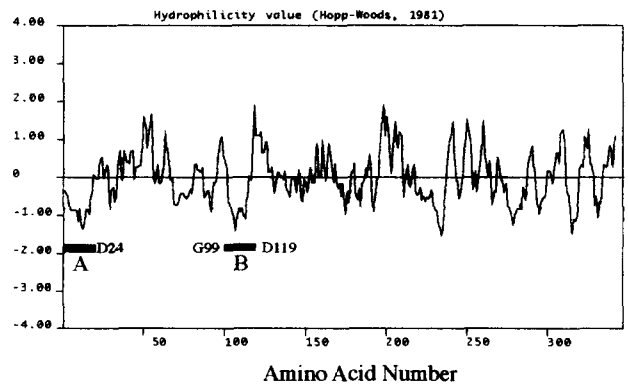

Fig. 3. Hydrophilicty plot [13] of the amino acid sequence from PsCCP showing the hydrophobic nature of the leader peptide (A) and the previously ommited section between G99 and D118 (B).

Type of Difference	Location
Substitution	G81→ D81*
	G83→ D83*
	D86→ G86
	E137→ Q137
	P164→ A164*
	K198→ I198*
	E290→ Q290
Inverted Position	G93↔ H94
	G286↔ Q287
Insertion	W96
	N228

Fig. 4. Minor differences between the amino acid sequence previously published (SwissProt; CCPR_PSESP) and the new sequence. * = recorded as variants in the original publication.

involving site-directed mutagenesis to further probe the interaction between the enzyme and its substrates.

Acknowledgements: The financial support of the Biotechnology and Biological Sciences Research Council is acknowledged.

References

[1] Foote, N., Turner, R., Brittain, T. and Greenwood, C. (1992) Biochem. J. 283, 839–843.
[2] Rönnberg, M., Osterlund, K. and Ellfolk, N. (1980) Biochim. Biophys. Acta 626, 23–30.
[3] Rönnberg, M. and Ellfolk, N. (1975) Acta Chem. Scand. B 29, 719–727.
[4] Rönnberg, M., Kalkkinen, N. and Elffolk, N. (1989) FEBS Lett. 250, 175–178.
[5] Fülop, V., Little, R., Thompson, A., Greenwood, C. and Hajdu, J. (1993) J. Mol. Biol. 232, 1208–1210.
[6] Goodhew, C.F., Wilson, I.B.H., Hunter, D.J.B. and Pettigrew, G.W. (1990) Biochem. J. 271, 707–712.
[7] Arciero, D.M. and Hooper, A.B. (1994) J. Biol. Chem. 269, 11878–11886.
[8] Hanlon, S.P., Holt, R.A. and McEwan, A.G. (1992) FEMS Microbiol. Lett. 97, 283–288.
[9] Soininen, R., Sojonen, H. and Ellfolk, N. (1970) Acta Chem. Scand. 24 2314–2320.
[10] Thöny-Meyer, L., Ritz, D. and Hennecke, H. (1994) Mol. Microbiol. 12, 1–9.
[11] Maniatis, T.E., Fritsch, F. and Sambrook, J. (1982) Molecular Cloning: a Laboratory Manual, Cold Spring Harbour Laboratory, Cold Spring Harbor, NY, USA.
[12] Lenhoff, H.M. and Kaplan, N.O. (1956) J. Biol. Chem. 220, 967–981.
[13] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.
[14] Chistoserdov, A.Y., Chistoserdova, L.V., McIntire, W.S. and Lidstrom, M.E. (1994) J. Bacteriol. 176, 4052–4065.
[15] Sofia, H.J., Burland, V., Daniels, D.L., Plunket III, G. and Blattner, F.R. (1994) Nucleic Acids Res. 22, 2576–2586.
[16] Hermodson, M. and Mahoney, W.C. (1983) Methods Enzymol. 91, 352–359.
[17] von Heijne, G. (1983) Eur J. Biochem. 133, 17–21
[18] Randall, L.L. and Hardy, S.J.S. (1989) Science 243, 1156–1159
[19] Silvestrini, M.C., Cutruzzola, F., D'Alessandro, R., Brunori, M., Fochesato, N. and Zennaro, E. (1992) Biochem. J. 285, 661–666.
[20] Pollock, W.B.R. and Voordouw, G. (1994) Microbiology 140, 879–887.
[21] Kim, J., Fuller, J.H., Cecchini, G. and McIntire, W.S. (1994) J. Bacteriol. 176, 6349–6361.