Subcloning and nucleotide sequence of the 3,4-dihydroxyphenylacetate (homoprotocatechuate) 2,3-dioxygenase gene from *Escherichia coli* C

David L. Roper and Ronald A. Cooper

Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK

Received 17 September 1990

A cloned gene encoding the *Escherichia coli* C homoprotocatechuate (HPC) dioxygenase, an aromatic ring cleavage enzyme, was used to produce large amounts of the protein. Preparations of *E. coli* C HPC dioxygenase, whether expressed from the cloned gene or produced by the bacterium, lost activity very rapidly. The pure protein showed one type of subunit of M_r 33000. The first 21 N-terminal amino acids were sequenced and the data used to confirm that the open reading frame of 831 bp, identified from the nucleotide sequence, encoded HPC dioxygenase. Comparison of the derived amino acid sequence with those of other extradiol and intradiol dioxygenases showed no obvious similarity to any of them.

Homoprotocatechuate 2,3 dioxygenase; Protein purification; N-terminal sequence; Gene sequence; Derived primary structure; *Escherichia coli* C

1. INTRODUCTION

The aromatic nucleus is one of the most abundant units of chemical structure in the biosphere. Because it is strongly stabilised by resonance, the ring must be hydroxylated under aerobic conditions to make it sufficiently reactive to undergo degradation. Normally, two hydroxyl groups need to be present for ring opening. The hydroxyls are usually required on adjacent carbon atoms to give a *cis*-hydroxylated aromatic ring. The activated ring is then opened by dioxygenase action. In *ortho*-fission (intradiol cleavage) the carbon-carbon bond between the *cis*-hydroxyls is cleaved. In *meta*-fission (extradiol cleavage) a carbon-carbon bond immediately adjacent to the *cis*-hydroxyls is the one broken [1].

Sequence information is now available for a number of these intradiol and extradiol dioxygenases. The sequences of three intradiol enzymes have been compared and it was concluded that two distinct catechol 1,2-dioxygenases have a common ancestry which is shared, but more distantly, with a protocatechuate 3,4-dioxygenase [2]. Consideration of the sequences of 5 extradiol dioxygenases led to the conclusion that they were members of the same gene super family, which was distinct from the family of intradiol dioxygenases [3]. However, a recent report of the sequence of a sixth extradiol dioxygenase, protocatechuate 4,5-dioxygenase, concluded that it was unrelated to any of the other extradiol or intradiol dioxygenases [4].

We have determined the sequence of a seventh extradiol dioxygenase whose substrate, homoprotocatechuate (HPC; 3,4-dihydroxyphenylacetate), is the next higher chemical homolog of protocatechuate. As reported here the sequence of *Escherichia coli* HPC, 2,3-dioxygenase has no marked similarity to any other extradiol or intradiol dioxygenase yet described.

Correspondence address: R.A. Cooper, Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK

2. MATERIALS AND METHODS

2.1. *Bacteria, plasmids and growth conditions*

Cells were grown aerobically at 37°C in Luria broth [5] containing 100 $\mu g \cdot ml^{-1}$ ampicillin. The *E. coli* K12 strain 5K [6] was used as host for the plasmids used in this study. The construction of pDR1830, pDR1930 and pDR934 is described in the text. pDR1930 was used to construct a nested set of deletions (pDR9304-9315, pDR9315; Fig. 2) using a Pharmacia exonuclease III deletion kit.

2.2. *Enzyme assays*

HPC dioxygenase and 5-carboxymethyl-2-hydroxymuconate (CHM) isomerase were assayed as described previously [7,8]. For measurement of the dioxygenase pH optimum allowance was made for the pH dependence of the extinction coefficient of the substance monitored, 5-carboxymethyl-2-hydroxymuconic semialdehyde [9]. K_m values were obtained by non-linear regression analysis of v against s [10].

2.3. *Purification of HPC dioxygenase*

The enzyme was purified from *E. coli* 5K (pDR9304). An ultracentrifuged crude extract was prepared from 400 ml Luria broth-ampicillin grown cells as described previously [11]. The ultracentrifuged extract (4 ml per run) was applied to a Pharmacia HR 10/10 Mono Q anion exchange column and chromatographed using a Pharmacia fast protein liquid chromatography (FPLC) system. An 80 ml gradient of 0.0→0.5 M NaCl in 20 mM Tris-HCl buffer pH 7.5, with a flow rate of 4 $ml \cdot min^{-1}$ was used and the fractions with the highest specific activities, eluted at approx. 0.3 M NaCl, were pooled. The enzyme at this stage appeared to be about 80% pure as judged by SDS-PAGE, so further purification was attempted. A sample was applied to a Pharmacia Phenyl Superose HR 5/5 hydrophobic interaction column and eluted with a 10 ml decreasing gradient of 1.7→0.0 M $(NH_4)_2SO_4$ in 0.1 M sodium phosphate buffer pH 7.5 at a flow rate of 0.5 $ml \cdot min^{-1}$. The HPC dioxygenase eluted at 0.2 M $(NH_4)_2SO_4$ but virtually all the enzyme activity had been lost. A further sample from the Mono Q step was subjected to gel filtration using two Phar-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

macia HR10/30 Superose 12 columns connected in series. The columns were equilibrated with 0.05 M sodium phosphate/0.15 M NaCl buffer pH 7.5 and run at a flow rate of 0.4 ml·min^{-1}. The recovery of HPC dioxygenase activity in this case was much better but some activity was lost and the specific activity decreased. Details of the purification are given in Table I. Protein in the various fractions was measured as described previously [11].

2.4. *Molecular mass estimation*

Subunit molecular masses were estimated by SDS-polyacrylamide gel electrophoresis (PAGE) [12] using gradient gels of 7.5% to 20% acrylamide. The proteins used to calibrate the gels were bovine serum albumin (66000); ovalbumin (45000); glyceraldehyde 3-phosphate dehydrogenase (36000); carbonic anhydrase (29000); trypsinogen (24000); trypsin inhibitor (20100) and bovine lactalbumin (14200).

2.5. *Amino acid sequencing*

The peak fraction after Phenyl Superose chromatography was run on a 7.5–20% SDS-PAGE gradient gel and electroblotted onto a polyvinylidene difluoride (PVDF) membrane [13] using 50 mM glycine-50 mM Tris, pH 10, transfer buffer and stained with Coomassie blue R-250. The stained M_r 33000 protein band was excised and loaded into an Applied Biosystems 470 A gas-phase sequencer without polybrene.

2.6. *DNA manipulations*

Small- and large-scale preparations were carried out by standard procedures [14]. Restriction endonucleases, T4 DNA ligase and exonuclease III deletion kit were used according to the manufacturer's instructions. Isolation of DNA fragments from low melting point agarose gels was as described [15]. Transformations were carried out using the morpholinepropane sulphonic acid (MOPS)-RbCl method [16].

2.7. *Nucleotide sequencing and oligonucleotide synthesis*

Sequencing was carried out using T7 DNA polymerase according to Tabor and Richardson [17]. The nucleotide analog 7-methyldeazaguanosine was included in the nucleotide mixes during sequencing to overcome base compressions. Plasmid DNA for sequencing was prepared as described [18]. Oligonucleotides were synthesized with an Applied Biosystems 380 B DNA synthesizer using cyanoethylphosphoramidite chemistry.

2.8. *Chemicals*

HPC was purchased from Sigma and CHM was prepared as described [6]. Restriction endonucleases were from Gibco-BRL and Pharmacia, T7 DNA polymerase and the exonuclease III deletion kit were from Pharmacia. Nucleotides were from US Biochemicals and the ^{35}S-dATP (1000 mCi/mmol) was from Amersham. All other chemicals were of the highest grade commercially available.

3. RESULTS AND DISCUSSION

3.1. *Subcloning of the HPC dioxygenase gene (*hpc B*)*

E. coli C genes involved in the catabolism of HPC have been cloned [6]. A 3.0 kbp *Bam*HI-*Sal*I fragment from one of these clones (pJJ801) was introduced into the polylinker region of pUC18 and pUC19 to give pDR1830 and pDR1930, respectively. When pDR1830 and pDR1930 were propagated in the *E. coli* K12 strain 5K (which itself is devoid of the *hpc* genes) two enzymes of the HPC pathway, HPC dioxygenase and CHM isomerase were present at high activity for 5K(pDR1930) but were barely detectable for 5K(pDR1830) (data not shown). When the two cell-free extracts were analysed by SDS-PAGE, that from 5K (pDR1930) showed two very strong bands (M_r 14000 and 33000) that were not seen in the 5K (pDR1830) extract. The M_r 14000 band is known to be CHM isomerase [11] so it seemed likely that the M_r 33000 band was HPC dioxygenase.

When the insert DNA of pDR1930 was shortened unidirectionally from the *Sal*I end by exonuclease III digestion a series of deletion subclones was obtained. The shortest subclone that still expressed HPC dioxygenase activity was pDR9304. The shorter subclones (see section 3.4) enabled the position of the HPC dioxygenase gene to be defined. Such shorter subclones like pDR9315 still expressed CHM isomerase activity. When the *Kpn*I site in the insert DNA was used along with the *Kpn*I site in the polylinker to delete DNA from the opposite (*Bam*HI) end of the insert DNA the resulting plasmid (dDR934; section 3.4) expressed neither HPC dioxygenase nor CHM isomerase activity. Since the *Kpn*I site in the insert DNA had been deleted in pDR9315 which still produced CHM isomerase it seemed that this *Kpn*I site was within the HPC dioxygenase gene.

3.2. *Enzyme purification*

The HPC dioxygenase produced by *E. coli* C cells grown on 4-hydroxyphenylacetate [7] was very unstable and the crude extract lost all its activity after storage for

Table I

A summary of the purification of *E. coli* C HPC dioxygenase

Purification Step	Volume (ml)	Total protein (mg)	Total units (μmol·min^{-1})	Specific activity (μmol·min^{-1}·mg protein^{-1})	Recovery %	Purification (Fold)
Ultracentrifuged extract	8	198	6903	37.8	100.0	0.0
Pooled Mono Q fractions	8	35	4584	130	66.0	3.4
*Pooled Superose 12 fractions	3	15	1500	100	22.0	2.7

* Only half the pooled Mono-Q material was used in this step so the figures have been adjusted, as necessary, to make allowance for this

1 10

Gly Lys Leu Ala Leu Ala Ala Lys Ile Thr

11 21

His Val Pro Ser Met Tyr Leu Ser Glu Leu Pro

Fig. 1. N-terminal amino acid sequence of *E. coli* C HPC dioxygenase.

24 h at 0°C. The addition of reagents such as glycerol, acetone, dithiothreitol or iron salts that are known to stabilize some dioxygenases had no effect on the *E. coli* enzyme. Because pDR9304 strongly expressed HPC dioxygenase, extracts prepared from 5K(pDR9304) have 100× higher initial specific activities than those from *E. coli* C and therefore facilitated purification of the enzyme. However, these preparations also lose activity quite rapidly so the emphasis in the purification was to monitor both the catalytic activity and the M_r 33 000 protein.

The major protein peak obtained after Phenyl Superose chromatography of the HPC dioxygenase from the Mono Q step had only slight catalytic activity but the M_r 33 000 protein accounted for more than 95% of the total protein in the fraction. No other fractions had HPC dioxygenase activity. The recovery of active HPC dioxygenase after Superose 12 gel filtration was higher than after Phenyl Superose chromatography and exactly matched the distribution of the M_r 33 000 protein. However, in the best fractions, the M_r 33 000 protein accounted for about 85% of the total protein and so was slightly less pure than after Phenyl Superose chromatography. The M_r 33 000 protein from the Phenyl Superose step was the one used for N-terminal sequence estimation. Using the enzyme from the Mono Q column, the effect of pH on the rate of reaction was measured in 0.1 M sodium phosphate buffer over the pH range 6.6–8.6. The optimum region was from 7.2 to 7.8. The K_m for HPC measured at pH 7.5 using solutions saturated with O_2 was 16 ± 3 μM.

3.3. *N-terminal amino acid sequence*

Although the PVDF blot was well washed to remove contaminating buffer, very small amounts of glycine are always retained. In the first sequencing cycle the amount of glycine detected was much greater than that of any amino acid in subsequent cycles. But no other amino acid was seen in the first cycle so glycine appeared to be the amino terminal residue. Unambiguous sequence information was obtained for the first 21 residues and this is given in Fig. 1.

3.4. *Nucleotide sequence of the HPC dioxygenase gene*

The sequencing strategy used is described in the legend to Fig. 2. Because the deletion subclones had an average size difference of 100 bp and the sequence of

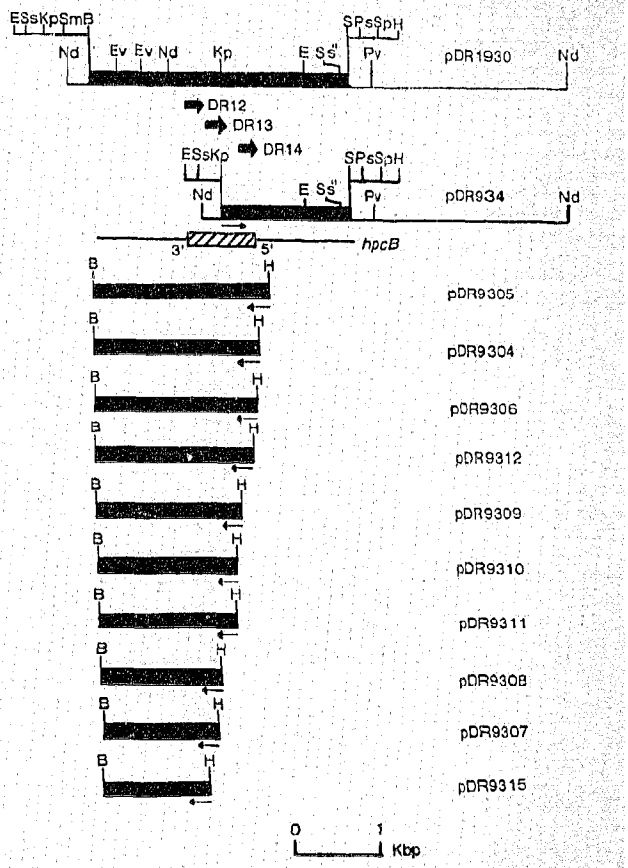

Fig. 2. Restriction map of subclones and sequencing strategy for the *E. coli* C HPC dioxygenase gene. The 3.0 kbp *Bam*HI-*Sal*I fragment (heavy line) cloned into the polylinker site of pUC19 to give pDR1930 is shown with major restriction sites. The restriction sites of the polylinker are not drawn to the same scale. Deletion of the 1.5 kbp *Bam*HI-*Kpn*I fragment of the insert DNA gives pDR934. The physical location of the HPC dioxygenase gene (*hpcB*) is shown by the hatched box. Unidirectional deletion subclones were created from the *Sal*I end of the insert DNA of pDR1930 by the exonuclease III digestion procedure. The coding strand encompassing HPC dioxygenase was sequenced using the deletion subclones pDR9304–9312 and pDR9315 as shown. The small arrows indicate the start, direction and extent of dideoxynucleotide sequencing carried out using the M13 reverse sequencing primer. The complementary strand was sequenced using the M13 universal primer. The broad arrows indicate the start, direction and extent of sequencing carried out using the synthetic oligonucleotides DR12–14 with the small arrow indicating, similarly, sequencing from the *Kpn*I end of the insert DNA of pDR934. Restriction sites are: Nd, *Nde*I; E, *Eco*RI; Ss, *Sst*I; Kp, *Kpn*I; Sm, *Sma*I; B, *Bam*HI; Ev, *Eco*RV; Ss'', *Sst*II; S, *Sal*I; Ps, *Pst*I; Sp, *Sph*I; H, *Hin*dIII; Pv, *Pvu*II.

around 250 residues was obtained for each construct there was very significant repetition of sequence measurement. The single-strand sequence was thus obtained with a high degree of accuracy. This sequence information was used to design two synthetic oligonucleotides (DR12 and 13) that were used in sequencing part of the opposite strand. The remaining part of the opposite strand was sequenced using the deletion subclone pDR934. Sequencing from the *Kpn*I site of pDR934 utilising the universal primer site of pUC19

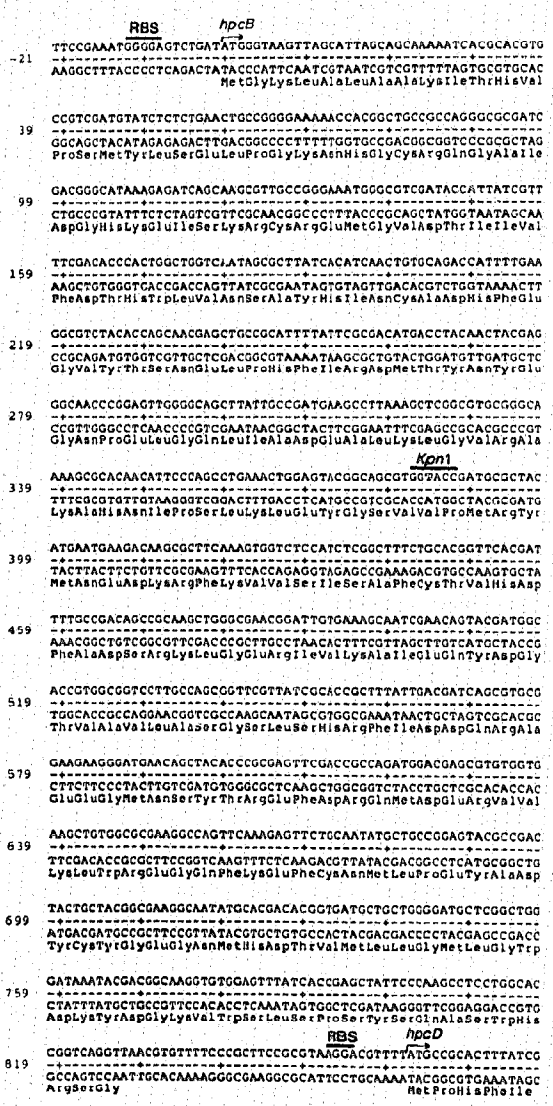

Fig. 3. Nucleotide sequence and deduced amino acid sequence of the *E. coli* C HPC dioxygenase gene (*hpcB*). The sequence is written in the 5′→3′ direction of the coding strand with the deduced amino acid sequence below. The amino terminus of the gene was identified by matching the predicted sequence to the sequence obtained from the purified protein. The *Kpn*I site within the gene and the ribosome binding site (RBS) are indicated. The start of the CHM isomerase gene (*hpcD*) is 32 bases after the end of *hpcB* as shown. The numbers refer to the nucleotide positions with the A of the ATG start codon numbered 1.

gave information for about 250 bases and this sequence was extended past the start of the dioxygenase gene by use of a further synthetic oligonucleotide, DR14, Fig. 2.

Computer analysis of the nucleotide sequence revealed an open reading frame (ORF) of 831 bp with associated ribosome binding site [19] 10 bp before the ATG start codon. This ORF was confirmed as that encoding HPC dioxygenase by its predicted amino acid sequence for residues 2–22 which corresponded exactly to the amino terminal amino acid sequence obtained from the purified protein. The methionine at position one of the ORF was absent from the purified protein. There are 34 bp between the end of the HPC dioxygenase gene and the start of the CHM isomerase gene whose sequence has already been described [11]. The nucleotide sequence which shows a single *Kpn*I site and the predicted amino acid sequence for HPC dioxygenase are shown in Fig. 3. A *Kpn*I site was known from the deletion subcloning (see section 3.1) to be within the HPC dioxygenase gene. The molecular mass calculated from the predicted amino acid sequence was 31 332 which agrees well with that of 33 000 for the purified protein measured by SDS-PAGE.

The predicted amino acid sequence of HPC dioxygenase was compared to those of the various extradiol and intradiol dioxygenases already reported [2–4] using the GAP programme of the University of Wisconsin Molecular Biology package. The extent of similarity varied between 8% and 14% indicating that there was no striking relationship between HPC dioxygenase and any of the other aromatic ring-fission dioxygenases. No obvious similarity to any other protein was apparent when the HPC dioxygenase sequence was compared to the NBRF (release 20) and the SWISSPROT (release 10) protein data bases. There was no apparently significant sequence similarity between HPC dioxygenase and CHM isomerase [11], two constituent enzymes of the HPC catabolic pathway. So the relationship of HPC dioxygenase to other ring-fission dioxygenases or to other HPC pathway enzymes is still an open question.

Acknowledgements: We are indebted to Dr K.S. Lilley for the amino acid sequencing, Mr John Keyte for oligonucleotide synthesis, and Michelle Kelly for typing the manuscript. D.I.R. thanks the SERC for a research studentship.

REFERENCES

[1] Dagley, S. (1975) in: Essays in Biochemistry, vol. 11 (Campbell, P.N. and Aldridge, W.N. eds) pp. 81–138, Academic Press, London.
[2] Neidle, E.L., Hartnett, C., Bonitz, S. and Ornston, L.N. (1988) J. Bacteriol. 170, 4874–4880.
[3] Harayama, S. and Rekik, M. (1989) J. Biol. Chem. 264, 15328–15333.
[4] Noda, Y., Nishikawa, S., Shiozuka, K.-I., Kadokura, H., Nakajima, H., Yoda, K., Katayama, Y., Morohoshi, N., Haraguchi, T. and Yamasaki, M. (1990) J. Bacteriol. 172, 2704–2709.
[5] Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
[6] Jenkins, J.R. and Cooper, R.A. (1988) J. Bacteriol. 170, 5317–5324.
[7] Cooper, R.A. and Skinner, M.A. (1980) J. Bacteriol. 143, 302–306.
[8] Garrido-Pertierra, A. and Cooper, R.A. (1981) Eur. J. Biochem. 117, 581–584.
[9] Fawcett, T., Garrido-Pertierra, A. and Cooper, R.A. (1989) FEMS Microbiol. Lett. 57, 307–312.
[10] Walmsley, A.R. and Lowe, A.G. (1985) Comput. Methods Programmes Biomed. 21, 113–119.

[11] Roper, D.I. and Cooper, R.A. (1990) FEBS Lett. 266, 63–66.
[12] Hames, B.D. (1981) in: Gel Electrophoresis of Proteins – a Practical Approach (Hames, B.D. and Rickwood, D. eds) pp. 1–91, IRL Press, Oxford.
[13] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.
[14] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
[15] Perbal, B. (1988) A Practical Guide to Molecular Cloning, pp. 366–374, Wiley, New York.
[16] Kushner, S.R. (1978) in: Genetic Engineering (Boyer, H.B. and Nicosia, S. eds) pp. 17–23, Elsevier, Amsterdam.
[17] Tabor, S. and Richardson, C.C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767–4771.
[18] Kraft, R., Tardiff, J., Krauter, K.S. and Leinwand, L.A. (1988) Bio Techniques 6, 544–547.
[19] Shine, J. and Dalgarno, L. (1975) Nature 254, 34–38.