Characterization of the *hca* Cluster Encoding the Dioxygenolytic Pathway for Initial Catabolism of 3-Phenylpropionic Acid in *Escherichia coli* K-12

EDUARDO DÍAZ,* ABEL FERRÁNDEZ, AND JOSÉ L. GARCÍA

Department of Molecular Microbiology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain

Received 14 January 1998/Accepted 26 March 1998

We have identified, cloned, and sequenced the *hca* cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid (PP) in *Escherichia coli* K-12. This cluster maps at min 57.5 of the chromosome and is composed of five catabolic genes arranged as a putative operon (*hcaA1A2CBD*) and two additional genes transcribed in the opposite direction that encode a potential permease (*hcaT*) and a regulator (*hcaR*). Sequence comparisons revealed that while *hcaA1A2CD* genes encode the four subunits of the 3-phenylpropionate dioxygenase, the *hcaB* gene codes for the corresponding cis-dihydrodiol dehydrogenase. This type of catabolic module is homologous to those encoding class IIB dioxygenases and becomes the first example of such a catabolic cluster in *E. coli*. The inducible expression of the *hca* genes requires the presence of the *hcaR* gene product, which acts as a transcriptional activator and shows significant sequence similarity to members of the LysR family of regulators. Interestingly, the *hcaA1A2CD* and *HcaB* enzymes are able to oxidize not only PP to 3-(2,3-dihydroxyphenyl)propionate (DHPP) but also cinnamic acid (CI) to its corresponding 2,3-dihydroxy derivative. Further catabolism of DHPP requires the *mhp*-encoded meta fission pathway for the mineralization of 3-hydroxypropionate (3HPP) (A. Ferrández, J. L. García, and E. Díaz, J. Bacteriol. 179:2573–2581, 1997). Expression in *Salmonella typhimurium* of the *mhp* genes alone or in combination with the *hca* cluster allowed the growth of the recombinant bacteria in 3-hydroxycinnamic acid (3HCl) and CI, respectively. Thus, the convergent *mhp* and *hca*-encoded pathways are also functional in *S. typhimurium*, and they are responsible for the catabolism of different phenylpropanoid compounds (3HPP, 3HCl, PP, and CI) widely available in nature.

Phenylpropanoid compounds are widely available in natural environments, and they can originate from putrefaction of proteins in soil or as breakdown products of several constituents of plants, such as lignin, various oils, and resins (2, 6, 14, 20). Microbial catabolism of phenylpropanoid compounds plays an important role not only in the natural degradative cycle of these aromatic molecules but also in their industrial applications such as wine making, aging, and storage (13). In particular, degradation of cinnamic acid (CI), 3-phenylpropionic acid (PP), and their hydroxylated derivatives has been reported in several bacteria, including *Acinetobacter* sp. (14), *Pseudomonas* sp. (2, 51), *Arthrobacter* sp. (51), *Escherichia coli* (10), and *Rhodococcus globularus* (6). Although most of the intermediates of these pathways are known, there has been little genetic characterization of these degradative routes, with the exception of the 3-(3-hydroxyphenyl)propionate (3HPP) catabolic pathways of *E. coli* K-12 (20) and *R. globularus* PWD1 (6).

Biochemical studies and the isolation and characterization of mutants defective in the catabolism of PP and 3HPP (compounds I and IV in Fig. 1B, respectively) revealed that in *E. coli* the aerobic degradation of these compounds proceeds by two initially separate routes that converge into 3-(2,3-dihydroxyphenyl)propionate (DHPH) (compound III), which suffers an extradiol ring cleavage and is ultimately degraded to Krebs cycle intermediates (9–11) (Fig. 1B). The cloning, sequencing, and transcriptional regulation of the *meta* fission cluster for the catabolism of 3HPP in *E. coli* K-12 have been recently reported (20).

The catabolism of PP in *E. coli* is initiated by a dioxygenolytic pathway (10, 11) (Fig. 1B). The first step is catalyzed by a 3-phenylpropionate dioxygenase, which inserts both atoms of molecular oxygen into positions 2 and 3 of the phenyl ring of PP, yielding cis-3-(3-carboxyethyl)-3,5-cyclohexadiene-1,2-diol (PP-dihydrodiol; compound II), which is subsequently oxidized by the 3-phenylpropionate-dihydrodiol dehydrogenase to give DHPH (compound III) (10, 11) (Fig. 1B). Enzyme assays and respirometry showed that the synthesis of enzymes required to convert the two initial growth substrates, PP and 3HPP, into DHPH are inducible and under separate control (10, 11). Very recently, it has been shown that in batch cultures the utilization of PP was immediately repressed by glucose (30).

Here we present the cloning, genetic characterization, and mechanism of regulation of the *hca* genes encoding the complete dioxygenolytic pathway for the catabolism of PP in *E. coli* K-12. This work constitutes the first genetic characterization of such a pathway and represents the first report of a gene cluster encoding a phenyl ring hydroxylating dioxygenase from *E. coli*. Moreover, we provide experimental evidence that 3HPP and PP catabolic pathways are also responsible for the catabolism of 3-hydroxycinnamic acid (3HCl) and CI, respectively.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The *E. coli* K-12 strains used were MC1061 (F− hsdR mcrB araD139 (l-arabinose trcDNA)7679 Δ:: lacU galU galK trpL thi) (46), DH5α (F− endA1 hsdR17 (rK− mK−) supE44 thi-1 recA1 gyrA relA1 Δ(argF-1acv)U169 deoR dcm lacZΔM15) (46), and MG1655 (F− λ−) (41). *E. coli* ED1061 is an *hcaA1* mutant of *E. coli* MC1061 (this study). The other strain
used in this study was *Salmonella typhimurium* LT-2 (20). For cloning and expression purposes we have used two chloramphenicol (CM) resistance low-copy-number cloning vectors, plasmids pCK01 (20) and pVTR-B (41), as well as the pUC18, pUC19 (46), and pUC18Not (20) vectors. Plasmid pPADR2 is an RSF1010-based promiscuous plasmid containing the complete *mhp* pathway (20). Plasmids pUC4K (Pharmacia) and pMAK700 (24) were used for insertional inactivation of the *hcaA1* gene (former *orfA*) and construction of the strain *E. coli* ED1061, respectively. Unless otherwise stated, bacteria were grown in Luria-Bertani (LB) medium at 37°C. When used as carbon sources, aromatic acids were supplied at 1 mM (CI) or 5 mM (PP, 3HPP, and 3HCI) to M63 minimal medium. *S. typhimurium* LT-2 was produced on LB medium containing 0.5% glucose and 5 mM PP. Where appropriate, antibiotics were added at the following concentrations: ampicillin (AP), 100 μg/ml; kanamycin (KM), 50 μg/ml; chloramphenicol (CM), 25 μg/ml; and kanamycin (KM), 50 μg/ml.

**DNA manipulations and sequencing.** Plasmid DNA was prepared by the rapid alkaline lysis method (46). Transformation of *E. coli* was carried out by the RbCl method (46). Electroporation (Gene Pulser, Bio-Rad) was used for *S. typhimurium* plasmid DNA manipulations and other molecular biology techniques were essentially as described elsewhere (46). DNA fragments were purified by using low-melting-point agarose. Oligonucleotides were synthesized on an Oligo-1000M nucleotide synthesizer (Beckman Instruments, Inc.). Nucleotide sequences were determined directly from plasmids by using the dideoxynucleotide chain termination method (47). Standard protocols of the manufacturer for T7u DNA polymerase-initiated cycle sequencing reactions with fluorescently labeled dideoxynucleotide terminators (Applied Biosystems Inc.) were used. The sequencing reactions were analyzed with a 377 automated DNA sequencer (Applied Biosystems Inc.). Sequences were extended by designing primers based on the already-determined sequence.

**Sequence data analyses.** Nucleotide sequence analyses were done with the DNA-Strider 1.2 program. Amino acid sequences were analyzed with Protein Analysis Tools on the ExPASy World Wide Web molecular biology server of the Geneva University Hospital and the University of Geneva. Nucleotide and protein sequence similarity searches were made by using the BLASTP, BLASTN, and BLASTX programs (1) via the National Institute for Biotechnology Information server. Pairwise and multiple protein sequence alignments were made with the ALIGN (59) and CLUSTAL W (56) programs, respectively. DNA sequence similarity searches were made by using the BLASTP, BLASTN, and BLASTX programs (1) via the National Institute for Biotechnology Information server. Pairwise and multiple protein sequence alignments were made with the ALIGN (59) and CLUSTAL W (56) programs, respectively.

**FIG. 1.** Convergent pathways for the catabolism of PP (CI) and 3HPP (3HCI) in *E. coli*. (A) Physical and genetic map of the chromosomal *hca* region. The locations of the genes are shown relative to those of some relevant restriction endonuclease sites, i.e., *BamH* (B), *ClaI* (C), *KpnI* (K), *PstI* (P), and *SphI* (Sp). Arrows indicate the directions of gene transcription. The boxed plus sign indicates stimulation of gene expression by the *hcaR* gene product in the presence of PP. Genes with similar shadings encode subunits of the same protein. (B) Proposed biochemistry of the PP (3HPP) and CI (3HCI) catabolic pathways. *HcaA1A2CD* and *HcaB* are the enzymes encoded by the corresponding *hca* structural genes. *MhpA* to *MhpF* are the enzymes for the catabolism of 3HPP (3HCI) and DHPP (DHCI). The metabolites are PP (compound I), cis-3-(3-carboxyethyl)-3,5-cyclohexadiene-1,2-diol (compound II), DHPP (compound III), 3HPP (compound IV), 2-hydroxy-6-ketommatrideioidate (compound V), CI (compound VI), cis-3-(3-carboxyethyl)-3,5-cyclohexadiene-1,2-diol (compound VII), DHCI (compound VIII), 3HCI (compound IX), and 2-hydroxy-6-ketommatrideioidate (compound X). Enzymes: *HcaA1A2CD*, 3-phenylpropionate dioxygenase; *HcaB*, 3-phenylpropionate-dihydrodiol dehydrogenase; *MhpA*, 3-(3-hydroxyphenyl)propionate hydroxylase; *MhpB*, 3-(2,3-dihydroxyphenyl)propionate hydroxylase; *MhpC*, 2-hydroxy-6-ketommatrideioidate dehydrogenase; *MhpD*, 2-keto-4-pentenoate hydratase; *MhpE*, 4-hydroxy-2-ketovalerate aldolase; *MhpF*, acetaldehyde dehydrogenase (acylating).
ogous recombination into the chromosome of *E. coli* MC1061 by selecting for KM resistance at 44°C. Since replication from the plasmid origin is deleterious to the host cell, when the cointegrates were subsequently grown at 30°C a second recombination event occurred, regenerating free plasmid in the cell and a disrupted orfA in the chromosome. The resident plasmid was then cured from the mutant strain by growing the cells at 44°C in LB medium without the antibiotic resistance marker of the vector, i.e., CM. A KM-resistant and CM-sensitive mutant strain was selected and named *E. coli* ED1061.

**Construction of plasmids.** To isolate the *hca* cluster, we constructed an EcoRI DNA library of *E. coli* MC1061 into pUC18 using as host the mutant strain ED1061, and then the transformants were screened for their ability to recover the dark red phenotype of *E. coli* MC1061 due to DHPP accumulation when growing on PP-containing LB medium. All red colonies harbored the plasmid pHCAES containing a 13.2-kb EcoRI insert. However, since plasmid pHCAES was shown to be highly unstable after several rounds of cultivation, the 13.2-kb EcoRI fragment was subcloned into the low-copy-number vector pUC19. The resulting plasmid pCKES (Fig. 2) also conferred on *E. coli* ED1061 the ability to produce the dark red color when growing on PP-containing LB medium, and this phenotype was stably maintained when the cells were grown in the presence of CM.

The *hcaR* gene was isolated from plasmid pCKES as a KpnI-SphI 2.1-kb fragment and subcloned into the KpnI-SphI double-digested pUC19 cloning vector to form pHCAR (Fig. 2). To delete genes *hcaRT* from the *hca* cluster, an *hcaA1* gene truncated at its 3’ end was PCR amplified from plasmid pCKES by using primers HCAR (5’-CCCTGCAGGTAAGCGGCGG-3’) and HCA3 (5’-CCCCCGGGCCGTAGTTCCATCACCTTC-3’) (the sequence corresponds to nucleotides 2319 to 2338 in Fig. 3; the engineered Smal restriction site is underlined) and HCA3 (5’-CCCTGCAGGTAAGCGGCGG-3’) (the sequence corresponds to nucleotides 3790 to 3815 in Fig. 3). The 1.5-kb PCR product was digested with *Smal* and *XhoI*, gel purified, and ligated to the *Smal*-XhoI double-digested pCKES plasmid to form pCKET (Fig. 2). Since we have observed constitutive expression of the *hca* catabolic genes in *E. coli* ED1061 (pCKER) cells, to avoid the possible transcriptional readthrough from the promoter of the CM resistance gene of the vector (Fig. 2), we subcloned the 6.0-kb EcoRI-HindIII fragment carrying the *hcaA1AXCDorfB* genes into the *EcoRI*-HindIII double-digested low-copy-number pVTR-B vector. The resulting plasmid pCKET contained the catabolic *hca* genes together with its potential promoter region, downstream of the strong *T3, T2* transcriptional terminators of the *E. coli* *rrnB* operon (41) (Fig. 2), thus excluding additional expression signals from the vector.

**Resting-cell reactions.** Phenylpropionate dioxygenase and phenylpropionate-dihydrodiol dehydrogenase activities were checked by analyzing the formation of DHPP in resting-cell assays. Thus, cultures of *E. coli* or *S. typhimurium* were grown overnight in LB medium and then diluted into fresh medium in the presence or absence of 1 mM aromatic inducer (PP or CI) to an optical density at 600 nm of about 0.08. Growth was resumed at 30°C (*E. coli*) or 37°C (*S. typhimurium*) until the cultures reached an optical density at 600 nm of about 0.8. The cell cultures were then centrifuged at 3,000 × *g* for 10 min at 20°C, and cells were washed and resuspended in a 0.05 volume of M63 minimal medium. The resting-cell reaction was performed in a final volume of 5 ml containing 4.5 ml of M63 minimal medium supplemented with 1 mM glucose and 0.5 ml of the cell suspension. The reaction was started by the addition of 0.5 mM PP or CI, and the tubes were incubated on a rotary shaker platform at a temperature of 30°C. Samples of 0.5 ml were taken at different times and centrifuged for 5 min at 10,000 × *g* to remove the cells. Products accumulated in the supernatant were analyzed with Gilson high-pressure liquid chromatography (HPLC) equipment using a Lichrosphere 5 RP-8 column (150 by 4.6 mm) and an isocratic flow of a 40% methanol-H₂O mobile phase pumped at a flow rate of 1 ml/min. Peaks with retention times of 22.3, 15.9, 6.20, and 4.63 min, corresponding to those of authentic standard CI, PP, 2,3-dihydrodiacetic acid (DHICA), and DHPP, respectively, were monitored at 210 nm.

To confirm the formation of DHICA, ¹H nuclear magnetic resonance (NMR) spectra were recorded in CD₃OD at 30°C on a Varian Unity 500 spectrometer. ¹H chemical shifts were referenced to internal residual CHD₃OD.

**Nucleotide sequence accession numbers.** The nucleotide sequences reported in this study have been submitted to the GenBank/EMBL database under accession numbers Y11070 and Y11071.

**RESULTS AND DISCUSSION**

**Cloning of the *hca* catabolic cluster of *E. coli* K-12.** Although during the course of this work the complete genome sequence of *E. coli* K-12 was reported (8), at the beginning of this research the analysis of the current *E. coli* database collection (ECDC release 27) (31) revealed the existence of an unmapped 4.6-kb sequence (accession no. Z37966), containing a 1.5-kb open reading frame (ORF) (orfA) that coded for a product showing significant similarity to the large terminal subunit of some multicomponent aromatic-ring initial dioxygenases (12). Since it was reported that the catabolism of PP in *E. coli* proceeds via dioxygenolytic attack of the ring (10, 11) (Fig. 1B), we assumed that orfA could encode a component of the 3-phenylpropionic acid (hydrocinnamic acid) initial dioxygenase (*Hca* dioxygenase). To test this assumption, we constructed *E. coli* MC1061 mutants by the insertion of a KM resistance cassette within orfA (see Materials and Methods). The selection of the mutant strains was based on the previous observation that accumulation of DHPP on rich medium gen-
erates a reddish-brown color due to autooxidation of this aromatic compound to the corresponding quinones and semiquinones (11). Thus, an *E. coli* strain such as MC1061, which contains a chromosomal deletion (ΔlacX74) spanning the initial genes of the *mhp* cluster responsible for the catabolism of DHPP (20), formed a dark red color when grown on PP, containing LB medium due to DHPP accumulation. Interestingly, the MC1061 KM-resistant mutants did not show the PP-dependent color reaction, indicating that the disrupted orfA was involved in the initial dioxygenation of PP. One of these mutants was selected, and it is referred to hereafter as *E. coli* ED1061.

To genetically characterize the complete Hca dioxygenase and the following enzyme of the dioxygenolytic pathway, i.e., the 3-phenylpropionate-dihydrodiol dehydrogenase (*HcaB*) (Fig. 1B), we constructed plasmid pCKES (Fig. 2) (see Materials and Methods), which contains a 13.2-kb *EcoRI* fragment from *E. coli* MC1061 that confers on *E. coli* ED1061 the ability to produce the dark red color during growth on PP-containing LB medium. Interestingly, *S. typhimurium* LT-2, which does not attack PP (see below) and lacks the *mhp* genes (20), showed the typical red color when transformed with pCKES on PP-containing LB medium. To directly assay the oxidation of PP to DHPP, resting-cell reactions were performed in the presence of PP, and the resultant products accumulated in the supernatants were analyzed by HPLC. *E. coli* mutant ED1061 and *S. typhimurium* LT-2 cells grown in LB medium in the presence of 1 mM PP were used as control bacteria since they did not attack this aromatic compound (Table 1). In contrast, resting cells of *S. typhimurium* LT-2(pCKES) and *E. coli* ED1061(pCKES) bacteria grown in LB medium in the presence of 1 mM PP rapidly consumed this aromatic compound and produced a metabolite that was eluted by HPLC as standard DHPP (Table 1). Therefore, all these data indicated that the 13.2-kb *EcoRI* DNA fragment encoded the complete dioxygenolytic pathway for the conversion of PP into DHPP; this pathway is also functional in *Salmonella*.

### Structural analysis of the hca genes

The nucleotide sequence of a 7,259-bp DNA fragment that carries the *hca* cluster was determined (Fig. 3). Analyses of the ORFs and sequence comparisons (see below) suggested the existence of seven genes arranged as follows: (i) five genes encoding the 3-phenylpropionate dioxygenase (*hcaA1A2CD*), 3-phenylpropionate-dihydroidiol dehydrogenase (*hcaB*), (ii) a regulatory gene (*hcaR*), and (iii) a gene (*hcaT*) that might encode a transporter.

### Expression of the hca cluster

Expression of the *hca* catabolic genes was monitored in resting-cell assays by measuring PP consumption and DHPP formation in HPLC. Cells were grown in LB medium in the presence or absence of 1 mM PP (inducer), and the resting-cell assays were performed for 60 min (*S. typhimurium* LT-2 cells) or 10 min (*E. coli* ED1061 cells) as described in Materials and Methods with 0.5 mM PP as substrate. BD, below detection limits.

### Products

<table>
<thead>
<tr>
<th>Strain</th>
<th>Inducer</th>
<th>Conversion of PP to DHPP (mol%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT-2</td>
<td>PP</td>
<td>BD</td>
</tr>
<tr>
<td>LT-2(pCKET, pHCAR)</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>LT-2(pCKET, pHCAR)</td>
<td>PP</td>
<td>10</td>
</tr>
<tr>
<td>ED1061</td>
<td>PP</td>
<td>BD</td>
</tr>
<tr>
<td>ED1061(pCKES)</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>ED1061(pCKES)</td>
<td>PP</td>
<td>80</td>
</tr>
</tbody>
</table>

### Nucleotide and derived amino acid sequences

TTTGCAGACAGGCCTCT--120n--CTCAGATCTCAGATGAGGAAAC--1113n--TTGCAAAAACATGAGTC 1287

<table>
<thead>
<tr>
<th>Gene</th>
<th>Start</th>
<th>Stop</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>hcaA1A2CD</td>
<td>129n--</td>
<td>--864n--</td>
<td>735n--</td>
</tr>
<tr>
<td>hcaB</td>
<td>864n--</td>
<td>--951n--</td>
<td>97n--</td>
</tr>
<tr>
<td>hcaR</td>
<td>951n--</td>
<td>--1056n--</td>
<td>105n--</td>
</tr>
<tr>
<td>hcaT</td>
<td>1056n--</td>
<td>--1080n--</td>
<td>24n--</td>
</tr>
</tbody>
</table>

### Control experiments

To genetically characterize the complete Hca dioxygenase and the following enzyme of the dioxygenolytic pathway, i.e., the 3-phenylpropionate-dihydroidiol dehydrogenase (*HcaB*) (Fig. 1B), we constructed plasmid pCKES (Fig. 2) (see Materials and Methods), which contains a 13.2-kb *EcoRI* fragment from *E. coli* MC1061 that confers on *E. coli* ED1061 the ability to produce the dark red color during growth on PP-containing LB medium. Interestingly, *S. typhimurium* LT-2, which does not attack PP (see below) and lacks the *mhp* genes (20), showed the typical red color when transformed with pCKES on PP-containing LB medium. To directly assay the oxidation of PP to DHPP, resting-cell reactions were performed in the presence of PP, and the resultant products accumulated in the supernatants were analyzed by HPLC. *E. coli* mutant ED1061 and *S. typhimurium* LT-2 cells grown in LB medium in the presence of 1 mM PP were used as control bacteria since they did not attack this aromatic compound (Table 1). In contrast, resting cells of *S. typhimurium* LT-2(pCKES) and *E. coli* ED1061(pCKES) bacteria grown in LB medium in the presence of 1 mM PP rapidly consumed this aromatic compound and produced a metabolite that was eluted by HPLC as standard DHPP (Table 1). Therefore, all these data indicated that the 13.2-kb *EcoRI* DNA fragment encoded the complete dioxygenolytic pathway for the conversion of PP into DHPP; this pathway is also functional in *Salmonella*.

### Structural analysis of the hca genes

The nucleotide sequence of a 7,259-bp DNA fragment that carries the *hca* cluster was determined (Fig. 3). Analyses of the ORFs and sequence comparisons (see below) suggested the existence of seven genes arranged as follows: (i) five genes encoding the 3-phenylpropionate dioxygenase (*hcaA1A2CD*), 3-phenylpropionate-dihydroidiol dehydrogenase (*hcaB*), (ii) a regulatory gene (*hcaR*), and (iii) a gene (*hcaT*) that might encode a transporter. The *hcaR* gene is also shown. Short arrows, direction of gene transcription; asterisks, stop codons. Potential Shine-Dalgarno sequences are shown. The putative binding motif of LTTRs (49) is doubly underlined, and the characteristic T and A residues are boxed. The nucleotide sequence present in oligonucleotide HCAR, used for PCR amplification of *hcaA1*, is italicized.

### Control experiments

TABLE 1: hca expression in *S. typhimurium* LT-2 and *E. coli* ED1061

<table>
<thead>
<tr>
<th>Strain</th>
<th>Inducer</th>
<th>Conversion of PP to DHPP (mol%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT-2</td>
<td>PP</td>
<td>BD</td>
</tr>
<tr>
<td>LT-2(pCKET, pHCAR)</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>LT-2(pCKET, pHCAR)</td>
<td>PP</td>
<td>10</td>
</tr>
<tr>
<td>ED1061</td>
<td>PP</td>
<td>BD</td>
</tr>
<tr>
<td>ED1061(pCKES)</td>
<td>None</td>
<td>2</td>
</tr>
<tr>
<td>ED1061(pCKES)</td>
<td>PP</td>
<td>80</td>
</tr>
</tbody>
</table>

### Control experiments

* Plasmids pCKET, pHCAR, and pCKES are diagrammed in Fig. 2.

* Expression of the hca catabolic genes was monitored in resting-cell assays by measuring PP consumption and DHPP formation in HPLC. Cells were grown in LB medium in the presence or absence of 1 mM PP (inducer), and the resting-cell assays were performed for 60 min (*S. typhimurium* LT-2 cells) or 10 min (*E. coli* ED1061 cells) as described in Materials and Methods with 0.5 mM PP as substrate. BD, below detection limits.

### Control experiments

* Plasmids pCKET, pHCAR, and pCKES are diagrammed in Fig. 2.

* Expression of the hca catabolic genes was monitored in resting-cell assays by measuring PP consumption and DHPP formation in HPLC. Cells were grown in LB medium in the presence or absence of 1 mM PP (inducer), and the resting-cell assays were performed for 60 min (*S. typhimurium* LT-2 cells) or 10 min (*E. coli* ED1061 cells) as described in Materials and Methods with 0.5 mM PP as substrate. BD, below detection limits.

### Control experiments

* Plasmids pCKET, pHCAR, and pCKES are diagrammed in Fig. 2.

* Expression of the hca catabolic genes was monitored in resting-cell assays by measuring PP consumption and DHPP formation in HPLC. Cells were grown in LB medium in the presence or absence of 1 mM PP (inducer), and the resting-cell assays were performed for 60 min (*S. typhimurium* LT-2 cells) or 10 min (*E. coli* ED1061 cells) as described in Materials and Methods with 0.5 mM PP as substrate. BD, below detection limits.
more, immediately downstream of orfX there is an inverted-repeat sequence (Fig. 3) predicted to form a hairpin loop (ΔG, −23.9 kcal/mol) that could act as a transcriptional terminator of a potential operon. Genes hcaR and hcaT are located upstream of hcaA1A2CBD-orfX, but they are transcribed in the opposite direction (Fig. 1A and 3). Although the intergenic spacing between genes hcaR and hcaT was 159 bp, we could not detect in this DNA fragment typical transcriptional terminator and promoter sequences. The G+C content of the hcaA1A2CBD regions averaged 53.8%, a value close to the mean G+C content of the E. coli genomic DNA (51.5%) (37).

The hca cluster maps immediately downstream of gene csiE, which encodes the stationary-phase inducible protein CsiE, at min 57.5 of the E. coli chromosome (8) and therefore far from min 8, where the mph cluster responsible for DHPP degradation is located (20). Interestingly, the third aromatic catabolic pathway characterized so far in E. coli at the molecular level, i.e., the hpa cluster for 4-hydroxyphenylacetate degradation, was shown to map at min 98 (43). Thus, while in some Pseudomonas and Acinetobacter species a superoperonic clustering of the aromatic catabolic genes has been observed in a limited region of the chromosome (15, 16, 60), in E. coli the aromatic catabolic clusters are dispersed throughout the genome.

The deduced amino acid sequences of the hca gene products were compared with entries in the databases, and the ones showing the highest similarities were then retrieved and analyzed in terms of deduced amino acid sequence identity with other gene products.
**hca** catabolic genes. Sequence comparison analyses of the *hcaA1*, *hcaA2*, *hcaC*, and *hcaD* gene products revealed significant similarities with the corresponding four protein subunits of the three-component class IIB ring-activating dioxygenases (12), mainly with the analogous *iph*, *cum*, and *bph* gene products (Table 2). The *hcaA1A2CD* genes are suggested, therefore, to encode the HcaA1A2CD initial dioxygenase of the PP catabolic pathway (Fig. 1).

The *hcaA1* gene encodes a protein of 51,109 Da (453 amino acids) that shows significant similarity with the large (∈) subunit of the terminal oxygenase component of multicomponent dioxygenases (Table 2). It is worth noting that although residues 85-CRHRMRVSVYACNTRAITFCPYH-108 in the HcaA1 protein match the binding site of a [2Fe-2S] Rieske-type iron-sulfur cluster (12) (the putative iron-sulfur ligands are underlined), the highly conserved G and S/T residues (12, 48) are replaced in HcaA1 by A and P (italicized), respectively. Residues 205-EQFASDOYHALFIS-218 in the HcaA1 primary structure match perfectly the mononuclear Fe(II) ligand at the site of oxygen activation (26). It should be noted that the C terminus of HcaA1 differs from that deduced from the reported sequence of *orfA* (accession no. Z37966), since *orfA* lacks a nucleotide leading to a change in the reading frame.

The *hcaA2* and *hcaC* genes encode proteins of 20,579 Da (172 amino acids) and 11,328 Da (106 amino acids) whose deduced amino acid sequences show significant identity to those of the small (∈) subunit of the terminal oxygenase and the ferrodoxin component of multicomponent dioxygenases, respectively (Table 2) (3). Residues 42-CSHGNASMEGYLE DDAVTECPLH-65 in HcaC are likely to be involved in the coordination of a Rieske-type [2Fe-2S] cluster (the putative iron-sulfur ligands are underlined); the unusual proline residue which is also present in the Rieske-type cluster of HcaA1 is italicized.

The next gene of the *hca* cluster, *hcaB*, encodes a protein of 28,498 Da (270 amino acids) that shows significant identity with cis-dihydriodiol dehydrogenases that participate in pathways involving class IIB dioxygenases (Table 2) and convert the stable cis-dihydriodiol formed by the initial dioxygenases into the corresponding dihydroxy derivatives with regeneration of NADH (12). Therefore, HcaB is postulated to be the 3-phenylpropionate-dihydriodiol dehydrogenase (Fig. 1). The length of HcaB falls within the average of 270 amino acids for members of the short-chain alcohol dehydrogenase (type II) superfamily, which includes all dihydriodiol dehydrogenases in Table 2 with the exception of BedD (21). Residues 13-GGGSGLGL-19 and 156-YTASKHAATGL-166 in HcaB fit the NAD⁺-binding domain (21) and the consensus pattern for short-chain alcohol dehydrogenases (18, 53), respectively. The highly conserved aspartate 92 of HcaB has been also implicated in the catalytic activity of analogous enzymes (38).

The *hcaD* gene encodes a 43,978-Da protein (400 amino acids) that is homologous to the ferrodoxin reductase subunit of other dioxygenases (Table 2). Multiple sequence alignments revealed the three conserved motifs in the same relative locations found in other reductase components of class IIB dioxygenases (12). Thus, residues 10-GGGQA-LPAMAAASLRQ DDD-26 and 151-GAGTIGLEELAASATQRKCVTVIE-174 of HcaD match the consensus sequence postulated to be involved in binding of the ADP moiety of flavin adenine dinucleotide and NAD⁺ (amino acids in italics indicate a replacement of a consensus residue (12), and the sequence 265-TCDPAIFAG GD-275 fits with the consensus motif that has been postulated to bind the O-3 group of the ribityl chain of the flavin moiety of flavin adenine dinucleotide (12).

The gene organization within the *hca* catabolic cluster is similar to that of the analogous *iph*, *cum*, *bph*, *tod*, *bed*, *tch*, *bnz*, and *tec* clusters encoding class IIB dioxygenases and consisting of the large subunit of the terminal oxygenase, small subunit, ferrodoxin, and reductase (57). However, the *hcaD* gene, although physically linked to the other three genes encoding the HcaA1A2CD dioxygenase, is separated from them by the *hcaB* gene (Fig. 1A and 3). An unusual location of the gene encoding the reductase component has been also observed for *bphA4* in *Pseudomonas* sp. strain KKS102 (28), *bphG* in *Comamonas testosteroni* B-356 (54), and *cmtAa* in *Pseudomonas putida* F1 (15). It has been reported that reductases have diverged more than the other components of the dioxygenases (57), and indeed the degrees of identity between HcaD and its orthologs in other clusters are lower than those observed for the other three HcaA1A2C subunits (Table 2). HcaD showed the highest level of identity to the CmtAa reductase component of the p-umate dioxygenase (Table 2); HcaA1A2CD and CmtAabdc are the only class IIB dioxygenases described so far that attack carboxylated aryls.

At the 3′ end of the *hcaD* gene is located *orfX* (Fig. 1A and 3), which has two potential translational start codons at positions 6710 (ATG) and 6737 (GTG), although only the latter shows a putative Shine-Dalgarno sequence (GAGGT) at a reasonable distance (Fig. 3), and codes for a 155-amino-acid product of unknown function. It is worth noting that an additional ORF of unknown function has also been found in other gene clusters encoding biphenyl (19, 22, 55), isopropylbenzene (42), and benzoate (accession no. M76990) dioxygenases.

**Regulation of the *hca* cluster.** The 5′ end of the *hca* region contains two genes, *hcaR* and *hcaT*, that are oriented in the direction opposite to that of the other *hca* genes (Fig. 1A and 3). The *hcaR* gene encodes a protein of 32,838 Da (296 amino acids) that shows a size and an amino acid sequence similar to those of LysR-type transcriptional regulators (LTTRs) (49) (Table 2). The majority of the genes encoding LTTRs are transcribed divergently from the genes that they regulate (49).

In this sense, *hcaR* is transcribed divergently from the catabolic genes *hcaA1A2CBD* (Fig. 1A and 3). LTTRs show a high degree of similarity in the N-terminal domain, where the helix-turn-helix DNA-binding region is located (49). Thus, HcaR possesses a sequence (18-FTTRAELKHTOPSISQSQRIDLE NCV-43) that matches the LTTR helix-turn-helix motif (Prosite signature PS00044) (5). The C-terminal domain of LTTRs seems to be involved in multimerization, and its consensus motif (V/L)X,GXQ(V/I)XY(L/V)P (49) fits with the sequence (232-VGMLGVTLIP-242) found in HcaR. Within the LysR family, HcaR shows the highest degrees of identity with the AlsR regulator of acetoin synthesis and with a select group of regulators from other biodegradative operons (Table 2). This group constitutes the Cat subfamily and includes the CatR, CatM, TfdR, TcbR, CclR, and TfdT regulators that activate the genes encoding muconate- or chloromuconate-lactonizing enzymes and/or genes encoding oxygenases that act on catechol or chlorinated aromatic compounds from *Acinetobacter* sp., *Pseudomonas*, and *Ralstonia* species (33). Additionally, HcaR shows significant identity with the putative *bphR* gene product, which supposedly would be involved in regulation of biphenyl catabolism (27). Therefore, all of these observations strongly suggest that HcaR is the transcriptional regulator of the *hca* cluster of *E. coli*.

To study the regulation of the *hca* cluster, we performed complementation studies of strains lacking Hca activity. While resting-cell assays of *E. coli* ED1061(pCKES) bacteria grown in LB medium in the absence of PP did not show significant removal of this aromatic compound, resting cells of these bacteria grown in the presence of 1 mM PP revealed that removal
of PP was concomitant with the appearance of a product which cochromatographed with authentic DHPP in HPLC (Table 1). These data, therefore, confirmed that the _hca_-encoded pathway was inducible. To demonstrate that HcaR was required for _hca_ expression and to determine its mechanism of action, the _hcaR_ and the _hcaA1A2CBD-orfX_ genes were independently expressed from the compatible plasmids pHCAR and pCKET (Fig. 2), respectively. Since _S. typhimurium_ LT-2 is unable to attack PP (Table 1), we used this strain as host for studying the regulation of the _hca_ genes. Resting cells of _S. typhimurium_ LT-2(pCKET) bacteria grown in the presence of 1 mM PP did not reveal the formation of DHPP, and PP remained unaltered. However, when the gene _hcaR_ was provided in trans from plasmid pHCAR, the resulting strain _S. typhimurium_ LT-2(pCKET, pHCAR) showed a significant conversion of PP into DHPP in a resting-cell assay when the bacteria were grown in the presence of 1 mM PP (Table 1). Furthermore, while _S. typhimurium_ LT-2 harboring simultaneously the compatible plasmids pPADR2 (containing the _mhp_ genes for the catabolism of 3HPP) (20) and pCKES grew efficiently (doubling time, 5 h) on minimal medium containing 5 mM PP as the sole carbon and energy source, _S. typhimurium_ LT-2(pPADR2, pCKET) cells did not grow in this aromatic compound. Thus, all these data indicated that HcaR fostered inducible expression of the _hca_ catabolic genes, behaving as a transcriptional activator. It should be mentioned that although the formation of PP-dihydrodiol (compound II in Fig. 1B) is assumed by the activator, it should be noted that although the formation of PP-dihydrodiol (compound II in Fig. 1B) is assumed by the activator, it should be noted that the _hcaR_ gene is not required for the formation of PP-dihydrodiol (compound II in Fig. 1B) because the _hcaR_-encoded pathway (20) was not able to grow with 3HPP (compound IX in Fig. 1B) as the sole carbon and energy source. Hence, these data provided experimental demonstration that the _mhp_ genes are also responsible for the mineralization of 3HCl.

Although _E. coli_ cannot grow on CI (compound VI in Fig. 1B) as the sole carbon source, whole cells grown on PP rapidly oxidized CI, suggesting that HcaB enzymes are also able to attack this aromatic compound (10). To confirm this assumption, _E. coli_ ED1061(pCKES) bacteria were grown in LB medium containing 1 mM PP for the induction of the _hca_ genes, and then they were used in a resting-cell assay with 0.5 mM CI as substrate. After 30 min of incubation at 30°C, CI was enzymatically converted to a product which cochromatographed in HPLC with authentic DHCI (compound VIII in Fig. 1B). NMR spectroscopy of the purified product confirmed it as DHCl (data not shown). Interestingly, the mutant strain _E. coli_ ED1061 (control cells) grown in PP-containing LB medium did not attack CI in a resting-cell assay. The conversion of CI to DHCI was also observed with _S. typhimurium_ LT-2(pCKES) cells. Hence, we concluded that the _hcaA1A2CD_ dioxygenase and _hcaB_ dihydrodiol dehydrogenase were responsible for CI oxidation and that DHCI was the final product of these reactions.

To analyze whether CI is an inducer of the _hca_ genes, we performed a resting-cell assay using PP as substrate and _E. coli_ ED1061(pCKES) bacteria grown in CI-containing LB medium. Since conversion of PP into DHPP was observed, we concluded that CI can also induce the _hca_ genes. Moreover, we were able to demonstrate that the _hca_ and _mhp_ genes are responsible for CI mineralization by showing that _S. typhimurium_ LT-2(pCKES, pPADR2) cells grew on minimal medium containing 1 mM CI as the sole carbon and energy source. It is worth noting that while in some soil _Pseudomonas_ species and in _Lactobacillus pastorianus_ the catabolism of CI could be accomplished by an initial reduction of the double bond of the side chain with the formation of PP (2, 7, 58), the catabolism of CI by the _Hca_ enzymes produces DHCI, which, via the _mhp_-encoded pathway (20), will be finally mineralized to pyruvate, acetyl coenzyme A, and fumarate (Fig. 1B).

Since CI can induce the _hca_ genes and is converted to DHCI
by E. coli cells, it is difficult to explain the lack of growth of this bacterium on this aromatic compound. A possible explanation for this behavior could be that the DHCI generated in the reactions catalyzed by the Hca enzymes or other intermediates further down the mhp-encoded pathway accumulate to a toxic level that prevents the normal metabolic flux of the cell. When E. coli MG1655(pCKES, pPADR2) cells were grown on minimal medium containing 20 mM glycerol plus 1 mM CI and the culture supernatants were analyzed by HPLC, we observed that CI depletion was not accompanied by the accumulation of DHCI (data not shown). However, these supernatants acquired a yellow coloration that disappeared after acidification with HCl, thus suggesting accumulation of the ring fission product of DHCI. These data are in agreement with previous observations showing that whole cells of E. coli grown in PP or 3HPP were able to oxidize DHCI with the transient formation of a yellow compound having the typical characteristics of a ring fission product but differing from those of compound V derived from ring cleavage of DHPP (Fig. 1B) (10). Furthermore, it has been shown recently that although DHCI is a good substrate for the MhpB dioxygenase (50), the ring fission product of this compound is hydrolyzed by the MhpC enzyme 36-fold less efficiently than the ring fission product of DHPP (32). All these data taken together may suggest that CI cannot support the growth of E. coli because its oxidation to DHCI generates toxic levels of the corresponding ring fission product. Although the toxicity of the ring fission products has been reported (10), the possibility that CI can cause on E. coli toxic effects that are not directly related to its catabolism cannot be ruled out. Why this toxicity is not observed in S. niphurimum is still an open question.

In conclusion, the results presented here constitute the first genetic characterization of a dioxygenolytic pathway for the initial catabolism of PP and CI. It has been suggested that pathways for the catabolism of aromatic compounds widely available in nature, such as PP, HPP, phenylacetate, and hydroxyphenylacetate, are among the most ubiquitous aromatic-compound catabolic systems, and they are closer to central metabolism than those involved in the degradation of xenobiotic compounds (6). These pathways, which occupy central positions within secondary metabolism, may have been one of the most common sources for the initial recruitment of genes for many of the routes involved in the degradation of aromatic compounds. It seems likely that these peripheral pathways evolved from a central one, such as hca, through mutation, recombination, and gene transfer events. Interestingly, PCB degraders have been found to be associated with plant lignin degraders (29), and the breakdown of lignin is one of the major natural sources of phenylpropanoid compounds (2). Moreover, it has been postulated that in the PCB degrader Rhodococcus sp. strain RHA1, the meta cleavage pathway genes could have evolved from the same ancestor as hydroxyphenylacetate meta cleavage pathway genes (hpa and hpc) of E. coli (35). Thus, it is tempting to speculate that the catabolic pathways for the mineralization of PCBs may derive from aromatic-compound central pathways through the assembling of hca-like clusters with mhp- or hpa-like clusters. On the other hand, the characterization of the hca-encoded dioxygenolytic pathway of E. coli confirms that this bacterium is endowed with genetic systems (hca, mhp, and hpa) highly similar to those in environmentally relevant bacteria such as those of the genus Pseudomonas, and this fact should be taken into consideration when aromatic catabolic clusters are cloned and expressed in this enterobacterium. There are several reports on the cloning and expression of aromatic dioxygenases in E. coli claiming that equivalent enzymes from the host could explain partial activities observed when some of the subunits of the cloned dioxygenase were missing in the recombinant bacteria (17, 52). The expression of the hca genes in E. coli might explain these reported observations.

ACKNOWLEDGMENTS

We thank T. Bugg for providing DHPH and DHCI, M. Vicente and M. Aldea for the plasmid pMAK700 and strain MC1061, M. K. B. Berlyn for strain MG1655, and A. Diaz and G. Porras for assistance with the sequencing. The help of J. F. Espinosa with the NMR studies is gratefully acknowledged.

This work was supported by grants from CICYT (AMB94-1038-C02-02 and AMB97-063-C02-02). A. Ferrández was the recipient of a predoctoral fellowship from the Plan Nacional de Formación de Personal Investigador-MEC. E. Diaz was the recipient of a Contrato Temporal de Investigadores from the CSIC.

REFERENCES


