Characterization of the \textit{hca} Cluster Encoding the Dioxygenolytic Pathway for Initial Catabolism of 3-Phenylpropionic Acid in \textit{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 \textit{hca} cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid (PP) in \textit{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 (\textit{hca}A1A2CBD) and two additional genes transcribed in the opposite direction that encode a potential permease (\textit{hca}T) and a regulator (\textit{hca}R). Sequence comparisons revealed that while \textit{hca}A1A2CD genes encode the four subunits of the 3-phenylpropionate dioxygenase, the \textit{hca}B 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 module in \textit{E. coli}. The inducible expression of the \textit{hca} genes requires the presence of the \textit{hca}R gene product, which acts as a transcriptional activator and shows significant sequence similarity to members of the LysR family of regulators. Interestingly, the \textit{Hca}A1A2CD and \textit{Hca}B 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 \textit{mhp}-encoded meta fission pathway for the mineralization of 3-hydroxypropionic acid (3HPP) (A. Ferrández, J. L. García, and E. Díaz, J. Bacteriol. 179:2573–2581, 1997). Expression in \textit{Salmonella typhimurium} of the \textit{mhp} genes alone or in combination with the \textit{hca} cluster allowed the growth of the recombinant bacteria in 3-hydroxycinnamic acid (3HCl) and CI, respectively. Thus, the convergent \textit{mhp-} and \textit{hca}-encoded pathways are also functional in \textit{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 degradation 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 \textit{Acinetobacter} sp. (14), \textit{Pseudomonas} sp. (2, 51), \textit{Arthrobacter} sp. (51), \textit{Escherichia coli} (10), and \textit{Rhodococcus globerulus} (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-hydroxypropionate) (3HPP) catabolic pathways of \textit{E. coli} K-12 (20) and \textit{R. globerulus} 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 \textit{E. coli} the aerobic degradation of these compounds proceeds by two initially separate routes that converge into 3-(2,3-dihydroxyphenyl)propionate (DHP) (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 \textit{meta} fission cluster for the catabolism of 3HPP in \textit{E. coli} K-12 have been recently reported (20).

The catabolism of PP in \textit{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-dihydriodiol, compound II), which is subsequently oxidized by the 3-phenylpropionate-dihydriodiol dehydrogenase to give DHPP (compound III) (10, 11) (Fig. 1B). Enzyme assays and respirometry showed that the syntheses of enzymes required to convert the two initial growth substrates, PP and 3HPP, into DHPP 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 \textit{hca} genes encoding the complete dioxygenolytic pathway for the catabolism of PP in \textit{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 \textit{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 \textit{E. coli} K-12 strains used were MC1061 [F– hsdR meqB araD139 (λaraABC-leu)7679 λΔx74 galU galK recA1 thi-1 relA1 Δ(aarF-lacY1)U169 deoR480lacZΔM15] (46), DH5α [F– endA1 hsdR17 (K+ merr) supE44 thi-1 recA1 gyrA relA1 Δ(aarF-lacY1)U169 deoR480lacZΔM15] (46), and MG1655 (F– Δ lacY1) (41). \textit{E. coli} ED1061 is an \textit{hca}A1 mutant of \textit{E. coli} MC1061 (this study). The other strain
used in this study was *Salmonella typhimurium* LT-2 (20). For cloning and sequencing 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 pPAD2 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 (46) at 37°C (*E. coli*) or 30°C (*S. typhimurium*). 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*. 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 dideoxy chain termination method (47). Standard protocols of the manufacturer for Taq 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, on the Baylor College of Medicine Human Genome Center server. The *E. coli* database collection ECDC (31) was accessed via the Internet.

Insertional inactivation of the *hcaA1* gene (*orfA*) and construction of *E. coli* ED1061. The 3′-end-truncated *hcaA1* gene (former *orfA*) was PCR amplified from the chromosome of *E. coli* MC1061 by using oligonucleotides HCA5 (5′-CCGAATTCACATATTAGCAACCAACCAGC-3′ [the sequence corresponds to nucleotides 3790 to 3815 in Fig. 3; the engineered *CiaI* restriction site is underlined]) and HCA3 (5′-CCCTGCAGGTAAGCGGCGGTTTTATC-3′ [the sequence corresponds to nucleotides 2384 to 2407 in Fig. 3; the engineered *PstI* site is underlined]) as primers. The 1.4-kb amplified fragment was digested with EcoRI-PstI and cloned into the *EcoR*I double-digested pUC19 vector to form pPHCA. The *orfA* in pHCA was inactivated by the insertion at its *CiaI* restriction site of a 1.3-kb *AccI* KM resistance cassette from plasmid pUC4K. The disrupted *orfA* was then subcloned as a 1.9-kb *EcoRI*-SpeI blunt-ended fragment into the blunt-ended-*SpeI*-digested pMAK700 plasmid, a pSC101-derived temperature-sensitive replicon (24), to form pHCA700. Because this plasmid replicates at 30°C but not at 44°C, it was possible to identify its integration through homol-
E. COLI hca-ENCODED PATHWAY

Vol. 180, 1998

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 O157 and 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 re suspended 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 shaking 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 LichroSpher 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-dihydroxypropiolic acid (DPP), and DHPP, respectively, were monitored at 210 nm.

To confirm the formation of DHCI, 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 data bank 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-phenylpropiolic 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-

FIG. 2. Schematic representation of the subcloning and expression of the regulatory elements and catabolic hca genes. The subcloning strategies are described in detail in Materials and Methods. The relevant elements and restriction sites are indicated. The thick line represents the DNA fragment whose sequence is shown in Fig. 3. Vector-derived sequences are indicated by dashed lines. The Plac and the Ptec promoters and direction of transcription are indicated (arrows). Δ truncated gene, T1 and T2, are the transcriptional terminators of the E. coli mmb operon (41). K and D, oligonucleotides HCAR and HCA3, which were used as primers for the PCR to construct plasmid pCK3. The region encoding the replication (ori) function is also indicated. B, BamHII; Bg, BglII; E, EcoRI; H, HindIII; K, KpnI; N, NotI; P, PstI; S, SmaI; Sp, SpII; X, XhoI. Ap6 and Cm6, genes conferring resistance to AP and CM, respectively.
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) and 3-phenylpropionate-dihydriodiol dehydrogenase (hcaB), (ii) a regulatory gene (hcaR), and (iii) a gene (hcaT) that might encode a transporter. The hca genes are located on the chromosome in the order hcaTRAI2CBD (Fig. 1A and 2). Downstream of the hcaD gene, the closely linked orfX (Fig. 1A and 3) could also be a member of the hca cluster. Interestingly, all five catabolic genes and orfX appear to be transcribed in the same direction. The Shine-Dalgarno sequences of hcaA1, hcaA2, hcaC, hcaB, and hcaD overlap the preceding ORFs (Fig. 3), suggesting that translational coupling occurs (23). Further-

<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, pHCA)</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>

* Plasmids pCKET, pHCA, 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.

---

![Image](http://ftp.asm.org/DownloadedFrom)
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 hcaA1A2CDorfX, 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 hca-coding 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 55.7 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 supraoperonic clustering of the aromatic catabolic genes has been observed in a limited region of the chromosome (15, 16), 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 that were compared with entries in the databases, and the ones that

<table>
<thead>
<tr>
<th>Gene</th>
<th>% G+C content</th>
<th>Gene product</th>
<th>Deduced no. of residues (kDa)</th>
<th>% Identity with other gene products (no. of residues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hcaA1</td>
<td>52.9</td>
<td>Large terminal subunit of phenylpropanoate dioxygenase</td>
<td>453 (51.1)</td>
<td>47.8, TcbAa (450); 47.6, IpbA1_BD2 (460); 47.4, BphA1_A1 (460); 47.3, Tcda1 (450); 47.0, Tcda1 (449); 46.8, Bedc1 (450); 44.3, BphA1_P6 (461); 44.2, BphA1_KF (458); 44.2, BphA_B (457); 44.1, Bpda1 (461); 43.9, Bpha_LB (459); 42.9, Bnza (448); 42.9, Bnza (459); 42.9, IpbA1_JR (459); 42.9, Bpha_KKS (458); 27.5, CmtAb (434)</td>
</tr>
<tr>
<td>hcaA2</td>
<td>52.2</td>
<td>Small terminal subunit of phenylpropanoate dioxygenase</td>
<td>172 (20.5)</td>
<td>37.7, Bphe_B (186); 34.4, Camu2 (186); 34.4, IpbA2 JR (186); 34.2, Bphe_LB (188); 32.8, Bpha2_A1 (187); 32.3, IpbA2_BD2 (187); 32.0, Bpha2_KKS (193); 31.2, Bedc2 (187); 30.7, Tcda2 (187); 30.5, Bnza (187); 30.5, Tcda2 (187); 30.2, CmtAb (180); 29.6, BphA22 (186); 28.1, Bpha3_A1 (107); 47.2, IpbA3_BD2 (107); 44.5, Camu3 (109); 43.6, IpbA3_JR1 (109); 43.5, Bteca (107); 43.1, Bpha3_KKS (109); 42.6, Teca3 (107); 41.8, Bphf_LB (109); 41.3, Bcd (107); 40.0, Bpha3_P6 (108); 40.0, Bpha3_KF (108); 40.0, Bpda (108); 39.8, Toda (107); 39.0, CmtAd (118); 37.4, Bnca (91); 36.7, Bpfb (109)</td>
</tr>
<tr>
<td>hcaC</td>
<td>54.0</td>
<td>Ferredoxin subunit of phenylpropanoate dioxygenase</td>
<td>106 (11.3)</td>
<td>48.1, Bpha3_A1 (107); 47.2, IpbA3_BD2 (107); 44.5, Camu3 (109); 43.6, IpbA3_JR1 (109); 43.5, Bteca (107); 43.1, Bpha3_KKS (109); 42.6, Teca3 (107); 41.8, Bphf_LB (109); 41.3, Bcd (107); 40.0, Bpha3_P6 (108); 40.0, Bpha3_KF (108); 40.0, Bpda (108); 39.8, Toda (107); 39.0, CmtAd (118); 37.4, Bnca (91); 36.7, Bpfb (109)</td>
</tr>
<tr>
<td>hcaD</td>
<td>54.0</td>
<td>Ferredoxin reductase subunit of phenylpropanoate dioxygenase</td>
<td>400 (43.9)</td>
<td>34.6, CmtAa (402); 31.5, IpbA4 JR (411); 31.5, Camu4 (411); 30.9, Bpfb (400); 30.8, Bpfb (408); 30.8, Bpha4_KF (408); 29.7, BpdA (412); 29.7, BpdA (412); 29.5, Tcda4 (410); 29.5, Tcda5 (410); 28.8, Toda (410); 28.7, Bpha4_KKS (410); 28.5, IpbA4_BD2 (412); 27.5, Beda (410); 27.3, Bpha4_A1 (413); 27.2, Bnca (409)</td>
</tr>
<tr>
<td>hcbA</td>
<td>55.9</td>
<td>2,3-Dihydroxy-2,3-dihydroxyphenylpropanoate dehydrogenase</td>
<td>270 (28.5)</td>
<td>48.9, Bpba_B1 (263); 46.2, Bpha2_KKS (276); 45.7, Bpba_LB (277); 45.7, Bpba_KF (277); 45.1, Cmba (276); 44.3, Bpba_B (281); 43.7, IpbA2 JR (276); 43.1, Toda (275); 43.1, Toda (275); 43.1, Blda (320); 42.7, Bpba_P6 (280); 41.7, TcdaB (280); 42.2, Cteca (280); 41.7, Bnca (279); 17.4, Bnca (279)</td>
</tr>
<tr>
<td>hcaR</td>
<td>49.8</td>
<td>Activator of hca cluster</td>
<td>296 (32.8)</td>
<td>35.6, CtrA (289); 35.2, AlsR_B (302); 33.4, AlsR_E (D90801), regulators of the acetoin synthesis in Bacillus subtilis and E. coli, respectively; BphR, putative regulatory protein of biphenyl catabolism in Pseudomonas sp. strain KKS (39); AlsR_E (302); 33.4, AlsR_B (302); 32.8, CatM (303); 28.9, CtrB (314); 28.6, Traf (295); 27.9, Tcda (294); 26.3, Cmca (294); 25.0, Cld (228); 24.1, Peqa (311); 19.0, NahR (374)</td>
</tr>
<tr>
<td>hcaT</td>
<td>56.3</td>
<td>Potential transporter</td>
<td>379 (41.6)</td>
<td>35.7, BhaT (308); 35.6, Mada (394); 23.4, LacY (417); 21.6, CscB (415); 21.4, MhpC (403); 20.6, HppK (453); 20.4, HpaX (458); 20.1, HpaX (458); 19.5, TfdK (460); 19.3, Pht1 (451); 19.3, MopB (449); 18.9, Pea (429); 18.5, Pea (448); 18.0, BenK (466); 17.7, Pea (421)</td>
</tr>
</tbody>
</table>
**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 (α) subunit of the terminal oxygenase component of multicomponent dioxygenases (Table 2). It is worth noting that although residues 85-35RHR4MV5YSADNCRTAFICTCPYH-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-EQFASQDQYHALFSH-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 (β) subunit of the terminal oxygenase and the ferredoxin component of multicomponent dioxygenases, respectively (Table 2) (3). Residues 42-CISKIGNASMEGYLEDDATVECPLH-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 (italicized) 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*-dihydriodiols 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-GGGSGGLG19 and 156-YTASKHAATGL-166 in HcaC 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 ferredoxin 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-GGGQA3LH-65 and 85-CRHRMRV3YSADNCRTAFICTCPYH-108 in *bph* (accession no. M76990) are involved in the 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*, *tcb*, *bnz*, and *tec* clusters encoding class IIB dioxygenases and consisting of the large subunit of the terminal oxygenase, small subunit, ferredoxin, 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 KK5102 (28), *bphG* in *Comamonas testosteroni* B-356 (54), and cmt*A/a* 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 Cmt*A/a* reductase component of the p-umate dioxygenase (Table 2); HcaA1A2CD and CmtAabc were the only class IIB dioxygenases described so far that attack carbonylated aryls.

At the 3′ end of the *hca* 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-Dalgaro 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), isoproplbenzene (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-FTTRAELKTOSPSLSQIQRDE 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)X2GXG(V/I)XV(L/V)P (49) fits with the sequence (232-VGMLGVTLLP-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, ClcR, 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*, *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 pPADDR2 (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(pPADDR2, 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 previous work and by analogy to other systems, the HcaB activity per se has not been demonstrated.

A 135-bp intergenic region is located between the potential translational start sites of the divergently transcribed hcaR and hcaA1 genes (Fig. 3), suggesting that the hcaR promoter is located near or overlaps the regulated promoter of the putative hca catabolic operon. As it has been noted with other LysR-type regulatory targets (40), the A+T content (66%) of the hcaR-hcaA1 intergenic region is higher than that of the hca genes (47%). LTTRs characteristically bind to a consensus T-N11-A DNA binding motif, with the T and A being part of a short inverted repeat, positionally conserved upstream of the regulated promoter (49). In this sense, within the hca intergenic region and located 85 nucleotides upstream of the putative hcaA1 translation start site, we have found the sequence TAG-N11-CTA that matches the binding motif of LysR-type regulators (Fig. 3). Underlined are the guanine and cytosine of the dyad, which have been shown to be involved in the binding of some LTTRs (45). Demonstration of these assumptions and elucidation of the mechanisms of hca repression by glucose (30) and hcaR expression will require further research.

The hcaT gene encodes a protein of 41,619 Da (379 amino acids) that shows significant identity with several members of the major facilitator superfamily (MFS) of transport proteins (34) (Table 2). Interestingly, HcaT was smaller than other MFS members (about 400 amino acids), and some common amino acid sequences that characterize this superfamily (20) were not found in the primary structure of the hcaT gene product. However, analysis of the predicted secondary structure of HcaT revealed the characteristic 12 membrane-spanning α-helices which are found in other MFS permeases and are believed to form a channel for transport through the membrane (34). Moreover, the hydrophility profile of HcaT showed that the protein could be divided by a central hydrophilic region into two halves, each containing six transmembrane domains (data not shown). Therefore, the putative HcaT protein might be involved in the uptake of PP in E. coli and could be another member of the rapidly expanding family of transporters for the catabolism of aromatic compounds (6).

It has been proposed that permeases for aromatic compounds are indirectly involved in the regulation of the catabolic pathways by bringing these substrates (inducers) inside the cell, leading to the induction of their respective regulatory proteins (43). This idea would agree with the close association between the hcaR and hcaT genes. Similar gene arrangements have been found for the permeases HpaX (4-hydroxyphenylacetate), PcaK (4-hydroxybenzoate), and HppR (3HPP) and the corresponding transcriptional regulators HpaA, PcaR, and HppR, respectively (6, 25, 39, 44). Interestingly, the catabolism of 3HPP in E. coli also involves a gene encoding a putative 3HPP permease (MhpT) that shows similarity with HcaT (Table 2) and a gene encoding a transcriptional activator (MhpR) which belongs to a family of regulators different from that of HcaR; these two genes are not transcriptionally coupled in the mhp operon (20).

Catabolism of CI and 3HCl. It is known that E. coli K-12 is also able to grow with 3HCl (compound IX in Fig. 1B) as the sole carbon and energy source (10). Since growth with 3HCl induces the synthesis of enzymes MhpA and MhpB, responsible for initial attack upon 3HPP, it was suggested that the same enzymes are used for catabolizing these two compounds (10). To demonstrate this hypothesis, plasmid pPADDR2 was introduced into S. typhimurium LT-2, a strain unable to grow on 3HPP (20) and 3HCl. The recombinant strain S. typhimurium LT-2(pPADDR2) acquired the ability to grow not only on 3HPP (20) but also on minimal medium containing 3HCl 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 Hca 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 DHCl (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 DHCl 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 DHCl 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, pPADDR2) 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 a catalytic 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 DHCl, 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 DHCl...
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 DHCl 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 Cl and the culture supernatants were analyzed by HPLC, we observed that Cl depletion was not accompanied by the accumulation of DHCl (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 DHCl. These data are in agreement with previous observations showing that whole cells of E. coli grown in PP or 3HPP were able to oxidize DHCl 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 DHHP (Fig. 1B) (10). Furthermore, it has been shown recently that although DHCl is a good ring fission product but differing from those of compound V, the ring fission product of this compound is hydrolyzed by the MhpC enzyme (36-fold less efficiently than the ring fission product of DHHP (32). All these data taken together may suggest that Cl cannot support the growth of E. coli because its oxidation to DHCl generates toxic levels of the corresponding ring fission product. Although the toxicity of the ring fission products has been reported (10), the possibility that Cl 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. typhimurium 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 anthropogenic compounds which are more peripheral to the natural carbon cycle (6). Since the whole structure and organization of the hca cluster resemble those of clusters responsible for initial dioxygenation of the highly recalcitrant polychlorinated biphenyls (PCBs) (bph) and chlorinated benzenes (teb and tec), it seems likely that these peripheral pathways have 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 RA1, 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 DHCl, M. Vicente and M. Aldea for the plasmid pMAK700 and strain MGI655, and A. Díaz 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. Fernández was the recipient of a predoctoral fellowship from the Plan Nacional de Formación de Personal Investigador-MEC. E. Díaz was the recipient of a Contrato Temporal de Investigadores from the CSIC.

REFERENCES


