JOURNAL OF BACTERIOLOGY, Apr. 1997, p. 2573–2581 Vol. 179, No. 8
0021-9193/97/$04.00+0

Genetic Characterization and Expression in Heterologous Hosts of the 3-(3-Hydroxyphenyl)Propionate Catabolic Pathway of *Escherichia coli* K-12

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

Departamento de Microbiología Molecular, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain

Received 25 November 1996/Accepted 12 February 1997

We report the complete nucleotide sequence of the gene cluster encoding the 3-(3-hydroxyphenyl)propionate (3-HPP) catabolic pathway of *Escherichia coli* K-12. Sequence analysis revealed the existence of eight genes that map at min 8 of the chromosome, between the *lac* and *hemB* regions. Six enzyme-encoding genes account for a flavin-type monooxygenase (*mhpA*), the extradiol dioxygenase (*mhpB*), and the *meta*-cleavage pathway (*mhpCDFE*). The order of these catabolic genes, with the sole exception of *mhpF*, parallels that of the enzymatic steps of the pathway. The *mhpF* gene may encode the terminal acetaldehyde dehydrogenase (acylating) not reported previously in the proposed pathway. Enzymes that catalyze the early reactions of the pathway, MhpA and MhpB, showed the lowest level of sequence similarity to analogous enzymes of other aromatic catabolic pathways. However, the genes *mhpCDFE* present the same organization and appear to be homologous to the *Pseudomonas xyl*, *dmp*, and *nah meta*-pathway genes, supporting the hypothesis of the modular evolution of catabolic pathways and becoming the first example of this type of catabolic module outside the genus *Pseudomonas*. Two bacterial interspersed mosaic elements were found downstream of the *mhpABCDFE* locus and flank a gene, *orfT*, which encodes a protein related to the superfamily of transmembrane facilitators that might be associated with transport. All of the genes of the 3-HPP cluster are transcribed in the same direction, with the sole exception of *mhpR*. Inducible expression of the *mhp* catabolic genes depends upon the presence, in the *cis* or *trans* position, of a functional *mhpR* gene, which suggests that the *mhpR* gene product is the activator of the 3-HPP biodegradative pathway. The primary structure of MhpR revealed significant similarities to that of members of the IclR subfamily of transcriptional regulators. A 3-HPP catabolic DNA cassette was engineered and shown to be functional not only in enteric bacteria (*E. coli* and *Salmonella typhimurium*) but also in *Pseudomonas putida* and *Rhizobium meliloti*, thus facilitating its potential application to improve the catabolic abilities of bacterial strains for degradation of aromatic compounds.

Microbial catabolism of phenylpropanoid compounds resulting from the degradation of lignin and other aromatic constituents of plants plays an important role in the degradative cycle of both natural compounds and many industrial pollutants. In particular, degradation of cinnamic acid, phenylpropionic acid, and their hydroxylated derivatives has been reported in several bacteria, such as *Acinetobacter* sp. (originally described as *Achromobacter* sp.) (16), *Pseudomonas* sp. (2, 15, 68), *Arthrobacter* sp. (42), and *Escherichia coli* (9). *E. coli* living in the animal gut encounters phenylpropionic acid and its derivatives as a result of the action of intestinal microflora on plant constituents, the amino acids phenylalanine and tyrosine, and some of their metabolites (9). Aerobic catabolism of these aromatic compounds by *E. coli* could occur in the guts of warm-blooded animals, close to epithelial cells, where oxygen passes from blood through the epithelium to the microbial populations attached to it, as well as in soil, sediment, and water once *E. coli* is excreted from its intestinal residence (62).

A branched *meta*-fission pathway used for the aerobic catabolism of 3-phenylpropionate and 3-(3-hydroxyphenyl)propionate (3-HPP) by *E. coli* K-12 has been described (9). These compounds are converted by separate routes to 2,3-dihydroxyphenylpropionate, which is ultimately degraded to Krebs cycle intermediates. Mutants of *E. coli* defective in the catabolism of 3-phenylpropionate and 3-HPP have been isolated and characterized (12). The pathway for the catabolism of 3-HPP in *E. coli* K-12 is shown in Fig. 1. The first step is catalyzed by the MhpA hydroxylase, which inserts one atom of molecular oxygen into position 2 of the phenyl ring of 3-HPP (Fig. 1B, compound I) to give 2,3-dihydroxyphenylpropionate (compound II), which suffers an extradiol cleavage by a 2,3-dihydroxyphenylpropionate 1,2-dioxygenase (MhpB) between ring carbons 1 and 2, yielding the *meta* ring fission product (compound III). The *mhpB* gene encoding the MhpB dioxygenase has been cloned (8), and very recently its nucleotide sequence has been determined, revealing that the MhpB dioxygenase belongs to a new family of extradiol dioxygenases (67). Compound III is cleaved by the MhpC hydrolase to give succinate (compound V) and 2-ketopent-4-enoic acid (compound IV), which is hydrated to 4-hydroxy-2-ketovaleric acid (compound VI) by the MhpD hydratase. The MhpE aldolase then catalyzes the fission of compound VI to give pyruvate and acetaldehyde (compounds VII and VIII, respectively).

Although the 3-HPP biodegradation pathway of *E. coli* K-12 was elucidated several years ago and it was shown to be strongly induced by this aromatic compound (9, 11), nothing is known about its genetic organization and mechanism of regulation. Here we present the sequencing and subsequent sequence analysis of the complete 3-HPP biodegradation cluster of *E. coli*. A positive transcriptional regulator has been iden-

* Corresponding author. Mailing address: Departamento de Microbiología Molecular, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Velázquez 144, 28006 Madrid, Spain. Phone: 34-1-5611800. Fax: 34-1-5627518. E-mail: CIBDF4F @FRESNO.CSIC.ES.

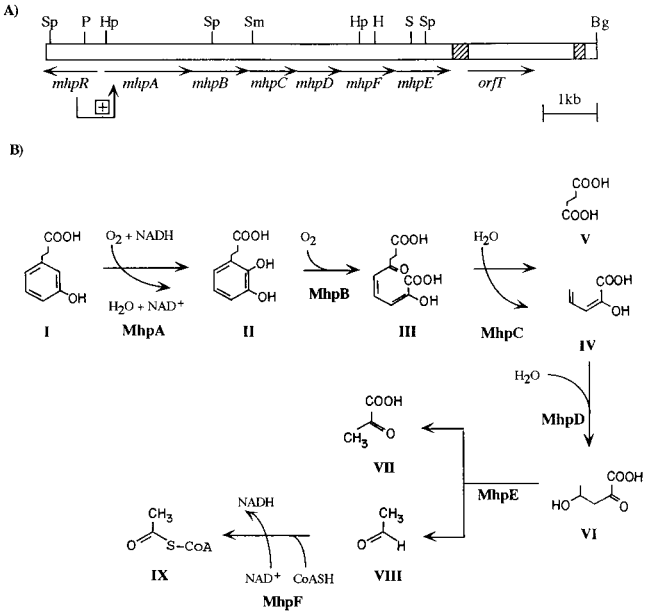

FIG. 1. Pathway for the catabolism of 3-HPP in *E. coli* and genetic organization of the corresponding structural and regulatory genes. (A) Physical and genetic map of the chromosomal *mhp* region. The locations of genes are shown relative to those of some relevant restriction endonuclease sites, i.e., *Bgl*II (Bg), *Hin*dIII (H), *Hpa*I (Hp), *Pst*I (P), *Sac*I (S), *Sma*I (Sm), and *Sph*I (Sp). The arrows indicate the directions of gene transcription. The boxed plus sign indicates stimulation of gene expression by the *mhpR* gene product in the presence of 3-HPP. Striped areas show the locations of BIME sequences. (B) Biochemistry of the 3-HPP catabolic pathway. MhpA to MhpF are the enzymes encoded by the corresponding *mhp* structural genes. The metabolites are 3-HPP (compound I), 3-(2,3-dihydroxyphenyl)propionate (compound II), 2-hydroxy-6-ketononadienedioate (compound III), 2-keto-4-pentenoate (compound IV), succinate (compound V), 4-hydroxy-2-ketovalerate (compound VI), pyruvate (compound VII), acetaldehyde (compound VIII), and acetyl coenzyme A (CoA) (compound IX). The enzymes are as follows: MhpA, 3-(3-hydroxyphenyl)propionate hydroxylase; MhpB, 3-(2,3-dihydroxyphenyl)propionate 1,2-dioxygenase; MhpC, 2-hydroxy-6-ketonona-2,4-dienedioate hydrolase; MhpD, 2-keto-4-pentenoate hydratase; MhpE, 4-hydroxy-2-ketovalerate aldolase; MhpF, acetaldehyde dehydrogenase (acylating).

tified, and the expression of an engineered 3-HPP catabolic cassette was analyzed in different gram-negative bacteria.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The *E. coli* K-12 strains used were MC4100 [F$^-$ *araD139* Δ(*argF-lac*)*U169 rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR*], CC118 [F$^-$ Δ(*ara-leu*)Δ*lacX74 thi-1 rpsE rpoB recA1*], DH5α [F$^-$ *endA1 hsdR17* (r_K^- m_K^+) *supE44 thi-1 recA1 gyrA relA1* Δ(*argF-lac*)*U169 deoR* ϕ80*dlac*Δ(*lacZ*)*M15*], and HB101 [F$^-$ *hsdS20* (r_B^- m_B^-) *recA13 ara-14 proA2 leuB lacY1 galK2 rpsL20 xyl-5 mtl-1 supE44*] (60). The other strains used in this work were *Salmonella typhimurium* LT-2 (kindly provided by J. Barbé), *Rhizobium meliloti* Rm1021Rif (21), and *Pseudomonas putida* KT2442 (21). Plasmid pLC20-30 belongs to the Clark-Carbon bank of hybrid plasmids (14). Plasmid pCK01 is a chloramphenicol resistance low-copy-number cloning vector (27). Plasmid pSJ33 is a broad-host-range cloning vector (34) that can be transferred from the donor strain to different recipient strains by mobilization with pRK600, an RK2-Tra$^+$ RK2-Mob$^+$ chloramphenicol resistance plasmid (18). Plasmid pUC18Not is a pUC18 derivative with two *Not*I sites flanking the polylinker region (18). Plasmid pPAR is a pUC18Not derivative containing the *mhpR* gene. Plasmids pPAD2 and pPADR2 are pSJ33 derivatives containing the complete *mhpABCDFE orfT* region and a truncated or wild-type *mhpR* gene, respectively. Unless otherwise stated, bacteria were grown in Luria-Bertani (LB) medium (60) at 37°C (*E. coli* and *S. typhimurium*) or 30°C (*P. putida*); *R. meliloti* was cultivated at 30°C in TY medium (21). When required, M63 minimal medium (45) was used, except for *S. typhimurium*, which was cultivated in AB minimal medium (13). Media were solidified with 1.5% Bacto Agar (Difco Laboratories, Detroit, Mich.). Where appropriate, antibiotics were added at the following concentrations: ampicillin, 100 μg/ml; chloramphenicol, 35 μg/ml; rifampin, 50 μg/ml; kanamycin, 50 μg/ml (100 μg/ml for *R. meliloti*).

Plasmid isolation, transformation, and conjugation. Plasmid DNA was prepared by the rapid alkaline lysis method (60). Transformation of *E. coli* was carried out by the RbCl method (60). Electroporation (Gene Pulser; Bio-Rad) was used for *S. typhimurium*. The transfer of plasmids pSJ33 and pPADR2 from *E. coli* to *P. putida* and *R. meliloti* was accomplished by mobilization through triparental filter matings by using as a helper strain *E. coli* HB101(pRK600) (18). *P. putida* and *R. meliloti* transconjugants were selected on LB or TY kanamycin-rifampin plates, respectively.

DNA manipulations and sequencing. DNA manipulations and other molecular biology techniques were done essentially as previously described (60). DNA fragments were purified by using low-melting-point agarose. Southern blot analyses were performed as previously reported (60), by using as probes 5′-end radioactively labeled oligonucleotides. The *mhpR* gene was PCR amplified from plasmid pLC20-30 by using primers MhpR-5 5′-GGGGATCCATTTCAGTAC CTCACGACTCGG-3′ and MhpR-3 5′-GGAAGCTTACTGGTTTCACATTCAC CACCC-3′ (engineered *Bam*HI and *Hin*dIII restriction sites are underlined and double underlined, respectively). The PCR product was cleaved with the appropriate restriction enzymes and gel purified prior to its ligation to the vector. 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 (61). 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. DNA bending analysis was performed by use of the DNASTAR computer programs (DNASTAR, Inc.). Amino acid sequences were analyzed with the Protein Analysis Tools at 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 done with the BLASTP, BLASTN, and BLASTX programs (1) via the National Institute for Biotechnology Information server. Pairwise and multiple protein sequence alignments were done with the ALIGN (77) and CLUSTAL W programs (72), respectively, at the Baylor College of Medicine-Human Genome Center server. The *E. coli* database collection ECDC (38) (release 27) was obtained on the Internet via a World Wide Web server.

Preparation of crude extracts and enzyme assays. *E. coli* MC4100 cells harboring pPAR, pSJ33, or its derivatives were cultured overnight at 30°C in antibiotic-containing LB medium in the absence or presence of 5 mM 3-HPP. Cells were harvested by centrifugation, washed, and resuspended in 0.1 volume of 50 mM sodium phosphate buffer, pH 7.5, containing 5 mM 2-mercaptoethanol prior to disruption by a single passage through a French press (Aminco Corp.) operated at a pressure of 20,000 lb/in^2. The cell debris was removed by centrifugation at 18,000 rpm for 30 min in an SS-34 rotor (Sorvall Instruments). The clear supernatant fluid was carefully decanted and used as crude extract. Protein concentration was determined by the method of Bradford (6) by using bovine serum albumin as the standard. Assays of MhpC activity in crude extracts were performed at 30°C in 50 mM sodium phosphate buffer, pH 7.5, with 42 μM substrate. Since 2,3-dihydroxyphenylpropionate was not commercially available to synthesize the natural substrate of MhpC, the ring fission product of 3-methylcatechol was used as the substrate. This substrate was prepared fresh daily in a 1-ml reaction mixture containing 600 μM 3-methylcatechol from a 10 mM stock solution, 0.1 M sodium phosphate buffer (pH 7.5), and 0.5 μl of a pure enzyme preparation of *P. putida* mt-2 XylE catechol 2,3-dioxygenase (kindly provided by L. Eltis); the reaction was allowed to proceed to completion (20). The reactions were monitored by measuring the decrease in A_{386} with a Shimadzu UV-260 spectrophotometer equipped with a thermojacketed cuvette holder and an Ultraterm circulating-water bath. One unit of enzymatic activity was defined as the amount of enzyme converting 1 μmol of substrate per minute with the molar absorption coefficient (15,000 M^{-1} cm^{-1}) reported previously (25).

Nucleotide sequence accession numbers. The nucleotide sequence reported in this study has been submitted to the GenBank/EMBL data bank (accession numbers Y09555 and X97543).

RESULTS AND DISCUSSION

Nucleotide sequence of the 3-HPP catabolic cluster from *E. coli* K-12. Previous studies indicated that genes *mhpA*, *mhpB*, and *mhpC* (Fig. 1A) map at min 8 of the *E. coli* K-12 chromosome, between the *lac* operon and the *hemB* gene, and are present in the 30-kb (from 7.9 to 8.5 min) chromosomal insert of Clarke-Carbon plasmid pLC20-30 (8, 11). Analysis of the *E. coli* database collection ECDC (38) revealed a gap in the sequence of about 6.6 kb between the *lacI* gene (accession no. J01636; referred to hereafter as J01636) and the *hemB* region (accession no. D64043; referred to hereafter as D64043). When sequence D64043 was analyzed in more detail at its 5′

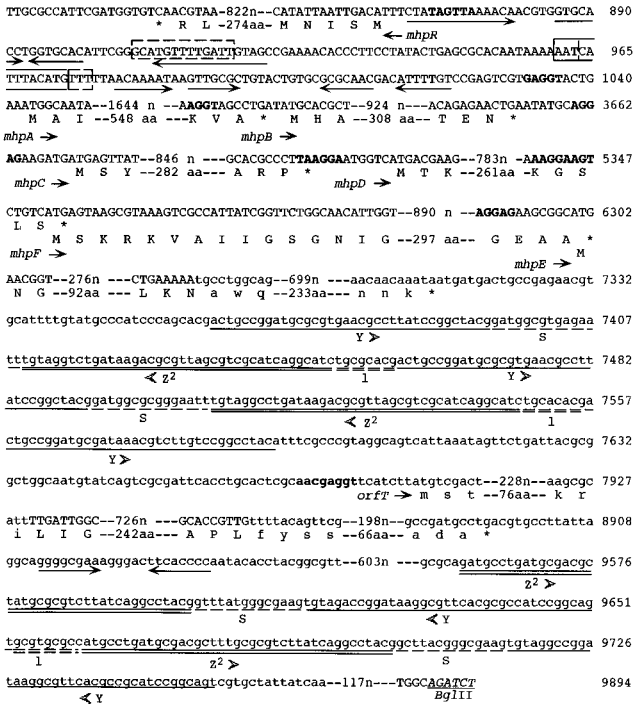

FIG. 2. Nucleotide and derived amino acid sequences of the 3-HPP catabolic pathway. The sequence data (uppercase letters) appear in the GenBank-EMBL data bank under accession number Y09555 (nucleotides 1 to 6592). Nucleotides 1 to 20 overlap nucleotides 1 to 20 of sequence J01636, and nucleotides 6583 to 6592 overlap nucleotides 1 to 10 of sequence D64043. Although the sequence from nucleotides 6593 to 9894 (lowercase letters) has already been reported (accession number D64043), we have detected five nucleotide changes from positions 7931 to 8674 (the sequence data appear in the GenBank-EMBL data bank under accession number X97543) that create *orfT*. Only the sequences of the 5′- and 3′-end coding regions of the *mhpRABCDF* genes are shown. The *Bgl*II restriction site at the 3′ end of the sequence is indicated. Short arrows show the direction of gene transcription. Asterisks indicate the stop codons. Boldface nucleotides represent potential Shine-Dalgarno sequences. Inverted repeats are marked with convergent arrows underneath the sequence. Sequences that show good similarity to the 5′-(A/T)ATCAANNNNTT(A/G)-3′ core consensus integration host factor-binding motif (19) are boxed with continuous (orientation, 5′ to 3′) or discontinuous (orientation, 3′ to 5′) lines. Motifs Y and Z^2 in BIME-2 sequences (3) are underlined and double underlined with continuous lines, respectively, and their orientations are indicated by arrowheads. The BIME-2 right (S) and left (l) internal motifs are underlined and double underlined with discontinuous lines, respectively. aa, amino acids; n, nucleotides.

end, we found an open reading frame (ORF) encoding a product that showed a high level of similarity to the COOH terminus of aldolases involved in the catabolism of aromatic compounds. This observation pointed to the location of the *mhp* genes at the DNA fragment located upstream of sequence D64043 in the *E. coli* K-12 chromosome. To determine the complete nucleotide sequence of the *mhp* cluster, we used plasmid pLC20-30. Sequencing reactions were initiated with primers whose sequences were based on those reported for the 5′ ends of sequences J01636 and D64043, respectively. Figure 2 shows the sequence of a 9,894-bp DNA fragment that carries the *mhp* cluster. Analyses of the ORFs and sequence comparisons (see below) suggested the existence of eight genes arranged as follows: (i) six enzyme-encoding genes corresponding to the initial monooxygenase (*mhpA*), the extradiol dioxygenase (*mhpB*), and the *meta*-cleavage (*mhpCDFE*) genes, (ii) one regulatory gene (*mhpR*), and (iii) a gene (*orfT*) that might encode a transporter and is flanked by two bacterial interspersed mosaic elements (BIMEs). All of the genes appear to be transcribed in the same direction, with the sole exception of *mhpR* (Fig. 1A and 2). Although the Shine-Dalgarno sequences of the *mhpR*, *mhpA*, and *mhpC* genes are located in intergenic regions, those of *mhpB*, *mhpD*, *mhpF*, and *mhpE* overlap the preceding ORFs, suggesting that the most common mechanism of translational coupling (28) may occur. Interestingly, the last gene of the *meta*-cleavage pathway, *mhpE*, is the only one that presents three consecutive stop codons and the TAAT translational termination signal (Fig. 2), the strongest reported in *E. coli* (71). The G+C content of the *mhp* coding regions averaged 55.76%, a value close to the mean G+C content of *E. coli* genomic DNA (50%) (46).

Sequence comparisons. Databases were searched for proteins similar to those of the 3-HPP pathway, and the ones showing the highest similarities were then retrieved and compared. Interestingly, enzymes that catalyze the early reactions of the pathway showed the lowest levels of sequence identity with the corresponding enzymes of other aromatic catabolic pathways (Table 1).

The *mhpA* gene encodes a protein of 62,049 Da (554-amino-acid length) that shows similarity to several bacterial flavin-type aromatic hydroxylases, such as TcmG for tetracenomycin C biosynthesis (17), PcpB (pentachlorophenol-4-monooxygenase) (50), PheA (phenol monooxygenase) (49), TfdB (2,4-dichlorophenol hydroxylase) (52), and PobA (*p*-hydroxybenzoate hydroxylase) (75) (Table 1). In these proteins, there are three regions of special relevance (Fig. 3). Near the NH_2 terminus of the enzymes (residues 17 to 56 in MhpA), a conserved region aligned fairly well with a proposed consensus fingerprint sequence, the so-called βαβ fold, which is involved in binding of the ADP moiety of flavin adenine dinucleotide (FAD) in flavoproteins (76). A second conserved region (residues 173 to 175 in MhpA) (Fig. 3) in PobA has been shown to be involved in FAD binding (23). Finally, a stretch of amino acids (residues 286 to 342 in MhpA) showed at its NH_2 terminus significant similarity to the consensus motif that binds the flavin moiety of FAD (23), whereas its COOH terminus was conserved among the six hydroxylases (Fig. 3). Thus, these data strongly suggest that the *mhpA* gene product is the monooxygenase that hydroxylates 3-HPP. Although MhpA requires NADH for its activity (10), no NADH-binding site could be inferred from the analysis of its amino acid sequence, a characteristic also observed with other flavin monooxygenases (23). It is worth noting that an isomer of 3-HPP, 3-(2-hydroxyphenyl)propionate (melilotate), is the substrate recognized by the flavoprotein melilotate hydroxylase isolated from *Pseudomonas* sp. (68). Although MhpA does not oxidize melilotate (10) and melilotate hydroxylase does not recognize 3-HPP (68), both enzymes give the same product and have similar M_rs and amino acid compositions. Hence, when the primary structure of melilotate hydroxylase becomes available, its comparison with that of MhpA will provide new insights into the structure-function relationships of these monooxygenases. Interestingly, the other initial monooxygenase of aromatic compounds analogous to HPP that has been described in *E. coli*, i.e., the two-component 4-(3-)hydroxyphenylacetate monooxygenase (57), does not show similarity to MhpA.

The nucleotide sequence of the *mhpB* gene encoding the 2,3-dihydroxyphenylpropionate 1,2-dioxygenase has been determined very recently (67). The inferred amino acid sequence showed significant similarity to that of other extradiol dioxygenases, such as the catechol 2,3-dioxygenase (MpcI) from *Alcaligenes eutrophus* (35), the β subunit (LigB) of protocatechuate 4,5-dioxygenase from *P. paucimobilis* (47), the homoprotocatechuate 2,3-dioxygenase from *E. coli* W (HpaD) (57) and C (HpcB) (58), and the human 3-hydroxyanthranilate dioxygenase (Hao) (43) (Table 1). This new family of enzymes

TABLE 1. 3-HPP pathway genes, gene products, and identities with other proteins[a]

Gene	Gene product	Deduced no. of residues	% Identity with other gene products (no. of residues)
mhpR	Activator of *mhp* cluster	281	26, IclR (274); 23, PobR (271); 23, PcaU (278); 21, PcaR (291)
mhpA	3-(3-Hydroxyphenyl)propionate hydroxylase	554	27, TcmG (572); 22, PcpB (537); 22, PheA (607); 22, TfdB (598); 20, PobA (394)
mhpB	3-(2,3-Dihydroxyphenyl)propionate 1,2-dioxygenase	314	55, MpcI (313); 23, LigB (302); 19, HpaD (283); 18, HpcB (276); 17, Hao (286)
mhpC	2-Hydroxy-6-ketonona-2,4-dienedioate hydrolase	288	53, BphD (286); 50, BphD_LB (286); 35, XylF (282); 34, TodF (276); 34, DmpD (283); 27, CmtE (288)
mhpD	2-Keto-4-pentenoate hydratase	269	50, TodG (268); 50, CmtF (262); 48, BphE (268); 40, XylJ (261); 39, DmpE (261); 39, BphH_LB (260); 34, HpaH (267); 33, HpcG (264)
mhpE	4-Hydroxy-2-ketovalerate aldolase	337	81, XylK (345); 80, DmpG (345); 79, TodH (352); 78, BphF (352); 76, CmtG (350); 56, NahM (346); 56, BphI_LB (346)
mhpF	Acetaldehyde dehydrogenase (acylating)	316	78, XylQ (312); 77, DmpF (312); 76, TodI (316); 75, BphG (316); 65, CmtH (314); 56, NahO (307); 55, BphJ_LB (304)
orfT	Potential transporter	403	34, PcaK (448); 28, PcaK_Ac (421); 25, PcaT (429); 20, Pht1 (451); 19, HpaX (458)

[a] Sequences included in this analysis are as follows. Bph and Bph_LB are from the biphenyl pathway of *Pseudomonas* sp. strains KKS102 (36) and LB400 (32), respectively; Cmt proteins are from the *p*-cumate pathway of *P. putida* F1 (26); Dmp proteins are from the phenol-dimethylphenol pathway of *Pseudomonas* sp. strain CF600 (64); Hpa and Hpc are from the 4-hydroxyphenylacetate and homoprotocatechuate pathways of *E. coli* W (57) and C (58, 59), respectively; Nah proteins are from the naphthalene pathway of *Pseudomonas* sp. strain NCIMB9816 (54); Tod proteins are from the toluene pathway of *P. putida* F1 (74); Xyl proteins are from the toluene-xylene pathway of *P. putida* mt-2 (29); PcaRK (30) and PcaT (U48776) are from the protocatechuate pathway of *P. putida* PRS2000; PcaU (U04359) and PcaK_Ac (37) are from the protocatechuate pathway of *A. calcoaceticus* ADP1; PobR and PobA are from the 4-hydroxybenzoate pathway of *A. calcoaceticus* ADP1 (24) and *P. fluorescens* (75), respectively; IclR is from the glyoxylate bypass operon of *E. coli* (70); TcmG, PcpB, PheA, and TfdB are hydroxylases of *Streptomyces glaucescens* ETH 22794 (17), *Flavobacterium* sp. strain ATCC 39723 (50), *Pseudomonas* sp. strain EST1001 (49), and *A. eutrophus* JMP134 (52), respectively; MpcI, LigB, and Hao are dioxygenases of *A. eutrophus* JM222 (35), *P. paucimobilis* (47), and humans (43), respectively; Pht1 is from the phthalate pathway of *P. putida* NMH102-2 (48).

comprises all of the extradiol dioxygenases which lack sequence similarity to the *Pseudomonas* catechol 2,3-dioxygenases. Amino acid sequence alignment of the NH_2- and COOH-terminal halves of MhpB suggests that the enzyme consists of two similar domains (67). Although the MhpB amino acid sequence reported here is almost identical to that described by Spence et al. (67), it is worth noting that amino acid residues 138-NKA-140, H-152, and T-157 have been replaced by ING, R, and A residues, respectively. These changes led to a better alignment between the NH_2- and COOH-terminal domains of MhpB and also with those of MpcI. Nevertheless, since both amino acid sequences were obtained from different *E. coli* K-12 strains, W3110 (67) and CS520 (this work), the existence of strain-specific amino acid substitutions cannot be ruled out.

The next gene of the *mhp* cluster, *mhpC*, encodes a protein of 31,937 Da whose deduced amino acid sequence shows significant identity to that of hydrolases cleaving carbon-carbon bonds of the *meta*-cleavage products of aromatic compounds (Table 1). Hence, we have ascribed the MhpC protein to the 2-hydroxy-6-ketonona-2,4-dienedioate hydrolase of the 3-HPP pathway. This observation is consistent with previous genetic and biochemical analyses that proposed an *mhpABC* gene order (8, 11). The isolation and characterization of the extradiol cleavage product (Fig. 1B, compound III) and the stereochemistry of its enzymatic hydrolysis have been previously reported (39). As shown in Table 2, the enzyme MhpC also hydrolyzes the 3-methylcatechol ring fission product, the natural substrate of the XylF hydrolase, which shows 35% sequence identity to MhpC (Table 1). Sequence alignments between MhpC and isofunctional enzymes from the *meta*-cleavage biphenyl (BphD), xylene (XylF), toluene (TodF), phenol (DmpD), and *p*-cumate (CmtE) biodegradative pathways strongly suggest that MhpC is a serine hydrolase that belongs to the α/β hydrolase fold family (20) (data not shown). These comparisons also point to residues S-114, D-239, and H-267 as the putative catalytic triad of MhpC (data not shown). Interestingly, the

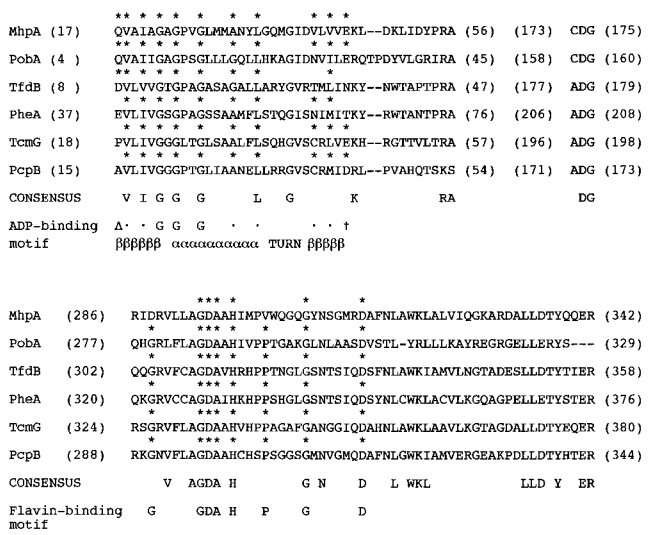

FIG. 3. Amino acid sequence alignments of three specific regions of MhpA and those of other bacterial hydroxylases. The other sequences and their sources are as follows: *P. fluorescens* PobA (*p*-hydroxybenzoate hydroxylase) (75), *Alcaligenes eutrophus* JMP134 TfdB (2,4-dichlorophenol hydroxylase) (52), *Pseudomonas* sp. strain EST1001 PheA (phenol-2-monooxygenase) (49), *Streptomyces glaucescens* ETH22794 TcmG (tetracenomycin polyketide synthesis hydroxylase) (17), and *Flavobacterium* sp. strain ATCC 39723 PcpB (pentachlorophenol-4-monooxygenase) (50). A consensus sequence was derived for those positions where less than three conservative replacements were observed. Conservative replacements are A, S, and T; D and E; N and Q; R and K; I, L, M, and V; and F, Y, and W. The numbers indicate the positions of the residues in the complete amino acid sequence of the protein. Gaps (indicated by dashes) were introduced to optimize the alignment. The fingerprint ADP-binding motif (76) is shown as follows: △ indicates K, R, H, S, T, Q, N, or D; ● indicates A, I, L, V, M, C, or Y; and † indicates D or E. α, β, and TURN represent the secondary-structure features α-helix, β-sheet, and turn, respectively. The flavin-binding motif that characterizes aromatic flavin monooxygenases (23) is shown at the bottom. Asterisks indicate agreement with the sequences of the ADP- or flavin-binding motifs.

TABLE 2. Growth and MhpC activities of different *E. coli* MC4100 derivatives

E. coli MC4100 plasmid(s)	MhpC activity (mU/mg of protein)[a]		Growth on 5 mM 3-HPP[b]
	Uninduced	Induced with 5 mM 3-HPP	
pPADR2	6	45	+
pPAD2	5	5	−
pSJ33, pPAR	4	5	−
pPAD2, pPAR	5	15	+

[a] Cells were grown overnight at 30°C in LB (uninduced) or LB containing 5 mM 3-HPP (induced). Preparation of cellular extracts and determination of MhpC activity were done as described in Materials and Methods. Results of one experiment are given; values were reproducible in three separate experiments.
[b] Cells were cultured at 30°C on M63 minimal medium containing 5 mM 3-HPP as the sole carbon source. +, growth; −, no growth.

sequence 112-GNSMGG-117 around the potential active-site serine of MphC fits very well in the nucleophilic motif present in all members of the α/β hydrolase fold family (20). Moreover, the conserved dipeptide 42-HG-43 at the NH_2 terminus of the MhpC molecule can be tentatively assigned as the central dipeptide that characterizes the oxyanion hole in α/β fold hydrolases (20). All of these assumptions, however, must be assessed by site-directed mutagenesis studies and by determination of the three-dimensional structure of the MhpC enzyme.

The *mhpD* gene codes for a protein of 28,889 Da (269-amino-acid length) that can be ascribed to a 2-keto-4-pentenoate hydratase since it shows similarity to analogous hydratases from other aromatic catabolic pathways (Table 1). Although the hydration catalyzed by MhpD is stereospecific (10), nothing is known about the residues involved in the reaction mechanism. Comparison of the eight hydratase sequences of Table 1 revealed that identical amino acid residues are fairly evenly spread over the entire length of the sequences (data not shown).

The last two genes of the *meta*-cleavage pathway, *mhpF* and *mhpE*, are the most closely related to analogous genes of other aromatic *meta*-cleavage pathways (Table 1). The proteins encoded by genes *mhpE* and *mhpF* were ascribed to a 4-hydroxy-2-ketovalerate aldolase and acetaldehyde dehydrogenase (acylating), respectively, since they were homologous to equivalent enzymes of the toluene-xylene (*xyl*), phenol (*dmp*), toluene (*tod*), biphenyl (*bph*), *p*-cumate (*cmt*), and naphthalene (*nah*) catabolic pathways (Table 1). Sequence similarities were spread over the entire length of the compared sequences (data not shown). The MhpE aldolase has a predicted M_r of 36,470 and a deduced isoelectric point of 5.6. Although aldolases are usually stereospecific in their action, the MhpE aldolase has been shown to attack both enantiomers of compound VI (Fig. 1B) (10). Acetaldehyde dehydrogenase (acylating) is the enzyme that catalyzes the terminal reaction in *meta*-cleavage pathways, that is, the transformation of acetaldehyde into acetyl coenzyme A (Fig. 1B). The MhpF acetaldehyde dehydrogenase has a predicted M_r of 33,442 (316-amino-acid length) and a theoretical isoelectric point of 5.3, and its presence in the *E. coli* 3-HPP degradative pathway had not been previously described. Interestingly, the metabolically equivalent succinic semialdehyde dehydrogenase of the homoprotocatechuate biodegradative pathway (66) is outside of the *hpc* and *hpa* clusters in *E. coli* C (69) and W (57), respectively. The NH_2 terminus of MhpF may encompass the NAD^+-binding site, since it contains (residues 5 to 38) a characteristic ADP-binding fingerprint sequence (76) (data not shown). Determination of whether MhpF and MhpE need to be physically associated in vivo to form an active complex, as has been shown for DmpGF (55) and suggested for the XylKQ and NahMO aldolase-acetaldehyde dehydrogenase pairs (54), requires further research.

The order of the catabolic genes in the *meta*-cleavage pathway, with the sole exception of *mhpF*, parallels that of the enzymatic steps in the 3-HPP catabolic pathway (Fig. 1). Genes *mhpCDFE* present the same organization and appear to be homologous to the *Pseudomonas xylFJQK*, *dmpDEFG*, and *nahNLOM* genes (54, 78). As these three *meta*-cleavage pathways have been shown to share a common ancestry (78), the *mhpCDFE* genes represent the first example of this type of catabolic module (*xyl*, *dmp*, and *nah*) outside of the genus *Pseudomonas*. Interestingly, while the three *Pseudomonas meta*-cleavage pathways are plasmid encoded, the *E. coli mhp* pathway is located in a chromosomal region that is rich in insertion elements such as IS*2* and IS*30* (5).

An ORF, *orfT*, which encodes a potential protein of 41,550 Da with a theoretical isoelectric point of 9.7 is located 0.4 kb downstream from the 3′ end of the *mhpE* gene (Fig. 2). A hairpin loop that has a free energy of −9.3 kcal/mol and could act as a transcription termination signal is located 16 bp downstream from the *orfT* stop codon (Fig. 2). OrfT shows a high level of similarity to the members of the major superfamily of transmembrane facilitators (MFS). All of the proteins of this superfamily are about 400 amino acids long and present a predicted secondary structure consisting of 12 membrane-spanning α helices (spanners) which are believed to form a channel for transport through the membrane (44). The hydrophilicity profile of OrfT revealed that the protein could be divided by a central hydrophilic region into two halves, each containing six transmembrane hydrophobic regions (data not shown). Moreover, OrfT presents some stretches of amino acids matching common motifs that characterize the members of MFS. Thus, (i) residues 74-DRYGRK-79 match perfectly the consensus motif (N/D)(R/K)XGR(R/K) that is predicted to occur in the hydrophilic region between spanners 2 and 3 (44), (ii) residues 111-MTGVGLGAALPNLIALTSEAAG PRFRG-137 might constitute PROSITE (PS00217) signature 2 of sugar transporters (4), and (iii) the PESPR sequence that characterizes the central hydrophilic region of sugar transport proteins (31) is also present as residues 189-PESAV-193 in OrfT. Interestingly, OrfT showed the highest level of similarity to PcaK (Table 1), a 4-hydroxybenzoate transporter and chemoreceptor in *P. putida* PRS2000 that is also a member of MFS (30). HpaX, the transporter of 4-hydroxyphenylacetic acid in *E. coli* (56), and other potential transporters of aromatic acids, such as PcaK from *Acinetobacter calcoaceticus* (37) and PcaT and Pht1 (48) from *P. putida*, also exhibited significant similarity to OrfT (Table 1). Hence, we believe that OrfT might be a new member of MFS likely to be involved in the transport of 3-HPP.

Analysis of the intergenic regions revealed the presence of two BIMEs flanking *orfT* (Fig. 1A and 2). BIME sequences are one of the five families of conserved repetitive sequences that have been identified in transcribed but untranslated DNA regions of enteric bacteria (3). The two BIME sequences found in the *mhp* region belong to the BIME-2 subfamily; i.e., they contain palindromic units (REP) alternating in both orientation and type of motif (Y and Z^2) (3). The BIME-2 sequence located 45 bp downstream of the *mhpE* stop codon consists of 233 bp with the structure YSZ^2lSZ^2lY, where S and l are the right and left internal motifs (3), respectively (Fig. 2). The 193-bp BIME-2 sequence located downstream of *orfT* has the structure Z^2SYlZ^2SY (Fig. 2). An interesting property of

BIMEs is their functional diversity, since they have been shown to participate in transcription termination, mRNA stabilization, control of translation, and genomic rearrangements (3). Although the role of BIMEs in the *mhp* region is unknown, it is tempting to speculate on their participation in transcription termination since no potential transcription termination signals have been found at the 3′ end of the putative *mhpABCDFE* operon and the REP sequences are predicted to form hairpin loops with high levels of free energy (data not shown). It is worth noting that REP sequences have also been found in the 4-hydroxyphenylacetate (*hpa*) and catechol (*cat*) catabolic pathways of *E. coli* (57) and *P. putida* (33), respectively.

Identification of *mhpR*, a regulatory gene required for *mhp* expression. The 5′ end of the *mhp* region contains an ORF, *mhpR*, that is oriented in the direction opposite to that of other ORFs (Fig. 1A). MhpR, the potential protein encoded by *mhpR*, is a 281-amino-acid polypeptide with a calculated M_r of 31,767. However, an additional methionine codon present at amino acid position 5 might also serve to initiate translation, since it is preceded by a potential ribosome-binding site (Fig. 2). The deduced amino acid sequence of MhpR revealed significant similarities with that of IclR, the repressor protein of the glyoxylate bypass operon from *E. coli* (70), that of PobR, the activator of the *pobA* gene encoding the *p*-hydroxybenzoate hydroxylase from *A. calcoaceticus* (24), and those of PcaU and PcaR, the activators of the *pca* operons for the catabolism of protocatechuate in *A. calcoaceticus* and *P. putida* (30), respectively (Table 1). Sequence comparisons between these bacterial regulatory proteins and MhpR showed two regions of high similarity. Near the NH_2 terminus of MhpR, the 36-ASVG LLAELSGLHRTTVRRLLET-58 (underlined are the amino acids most strongly conserved) stretch of residues matches the potential helix-turn-helix structural motif involved in binding to the target DNA in other transcriptional regulators (7). Near the COOH terminus of the MhpR protein, residues 201-G QNYRGWDQEEKIASIAVPLRS-222 partially match the PROSITE pattern (PS01051) (4) found in members of the IclR subfamily of regulators, such as IclR, PobR, and PcaR. Furthermore, the *mhpR* gene exhibits two common features of regulatory genes (24); i.e., its codon usage predicts a low level of expression, and the amino acid composition of the corresponding gene product reveals that arginine is overrepresented (29 residues) with respect to lysine (10 residues). All of these observations strongly suggest that MhpR is the transcriptional regulator of the 3-HPP degradative pathway from *E. coli.*

Experimental evidence that MhpR was required for *mhp* expression emerged from genetic studies involving complementation of *E. coli* mutants unable to grow on 3-HPP as the sole carbon source. *E. coli* CC118 (Δ*lacX74*) and MC4100 (Δ*lacU169*), two strains containing chromosomal deletions spanning the entire *lac* operon and its flanking regions (53), were unable to grow on 3-HPP-containing minimal media. Southern blot analyses of their *Hin*dIII- and *Hpa*I-digested chromosomal DNAs probed with oligonucleotides derived from the nucleotide sequence of the *mhp* cluster (data not shown) mapped the right end of the *lacX74* and *lacU169* chromosomal deletions between the 5′ end of the *mhpD* gene and the *Hin*dIII site of the *mhpF* gene (Fig. 1A). These results demonstrate that *E. coli* CC118 and MC4100 did not grow on 3-HPP because they lack, at least, the *mhpABCD* catabolic genes. *E. coli* MC4100 cells harboring plasmid pPAT or pPAD, two pCK01 derivatives containing a truncated *mhpR* gene and the complete *mhpABCDFE orfT* region, respectively (Fig. 4), were also unable to grow on 3-HPP as the sole carbon and energy source. On the basis of this evidence, the presence of *mhpR* appeared to be essential for expression of the *mhp*

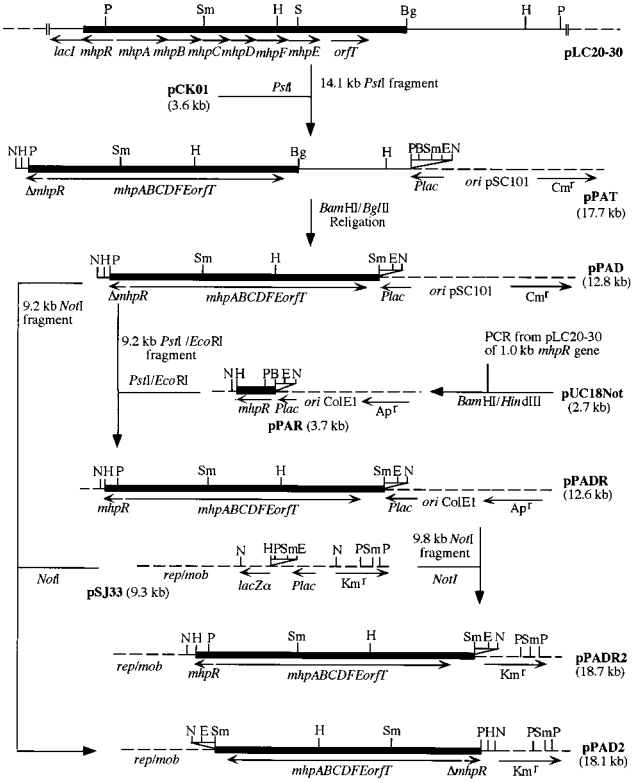

FIG. 4. Schematic representation of the construction of a *Not*I-DNA cassette expressing the *mhp* cluster. Plasmids are drawn with the relevant elements and restriction sites indicated. The black block represents the DNA fragment whose sequence is shown in Fig. 2. Vector-derived sequences are represented by discontinuous lines. Arrows show the *Plac* promoter and the direction of transcription of the genes. Δ*mhpR* means a truncated *mhpR* gene. The regions of the plasmids encoding the replication (*ori* and *rep*) and mobilization (*mob*) functions are indicated. Restriction sites: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; N, *Not*I; P, *Pst*I; S, *Sac*I; Sm, *Sma*I. Ap^r, Cm^r, and Km^r indicate the genes conferring resistance to ampicillin, chloramphenicol, and kanamycin, respectively.

catabolic genes. To demonstrate this assumption, the *mhpR* gene was isolated by PCR from plasmid pLC20-30 and subcloned into the pUC18Not cloning vector (Fig. 4). The resulting plasmid, pPAR, was used to reconstruct the original *mhp* cluster in plasmid pPADR (Fig. 4). When plasmid pPADR was introduced into *E. coli* MC4100, the transformants acquired the ability to grow on 3-HPP. *E. coli* MC4100 cells harboring broad-host-range plasmid pPAD2 (Δ*mhpR mhpABCDFE orfT*) or pPADR2 (*mhpR mhpABCDFE orfT*) (Fig. 4) gave similar results; i.e., cells grew on 3-HPP as the sole carbon source only in the presence of a functional *mhpR* gene (Tables 2 and 3). Moreover, enzymatic assays carried out with crude extracts of cells cultured in Luria broth without (uninduced) or with 5 mM 3-HPP (induced) demonstrated that *mhpR* fostered inducible expression of the *mhpC* gene (Table 2) and suggest that MhpR is a transcriptional activator of the *mhp* catabolic genes. To confirm that the *mhpR* gene could also function in *trans*, plasmid pPAR was introduced into *E. coli* MC4100 containing pPAD2 and the resulting strain was screened for the ability to grow on minimal medium supplemented with 3-HPP (Table 2). Surprisingly, when *mhpR* was expressed in *trans* from high-copy-number plasmid pPAR, the MhpC activity detected in cellular extracts was lower than that found in extracts from cells expressing *mhpR* in *cis* on medium-copy-number plasmid pPADR2 (Table 2). A plausible explanation for this effect, also observed with other activators of aromatic catabolic

TABLE 3. Expression of the *mhp* cluster in several gram-negative bacteria[a]

Strain	Plasmid	No. of CFU/ml (doubling time [h])	
		Glycerol (20 mM)	3-HPP (5 mM)
E. coli MC4100	pSJ33	2.0×10^9 (2)	1.2×10^8 (ND[b])
	pPADR2	2.0×10^9 (2)	1.0×10^9 (3.5)
S. typhimurium LT-2	pSJ33	2.1×10^9 (1.5)	1.1×10^8 (ND)
	pPADR2	2.0×10^9 (1.5)	1.2×10^9 (2)
R. meliloti Rm1021Rif	pSJ33	7.5×10^8 (8)	1.5×10^8 (ND)
	pPADR2	7.0×10^8 (8)	9.2×10^8 (4)
P. putida KT2442	pSJ33	1.1×10^9 (3)	1.9×10^8 (ND)
	pPADR2	1.4×10^9 (3)	8.1×10^8 (6)

[a] Cells were cultured in minimal medium containing 20 mM glycerol and 5 mM 3-HPP. Cultures were then washed three times and diluted to 1×10^8 to 2×10^8 CFU ml^{-1} in fresh medium supplemented with 20 mM glycerol or 5 mM 3-HPP as the sole carbon source. Growth was resumed, and the maximal growth of the cultures was monitored after 24 h (*E. coli* and *S. typhimurium*) or 48 h (*R. meliloti* and *P. putida*) of incubation.

[b] ND, not detected. With *R. meliloti* and *P. putida*, slight growth was observed during the first 24 h of incubation, although the number of CFU per milliliter after 48 h of incubation was similar to that reported at time zero.

pathways (19), could be that overexpression of MhpR titers out a cellular factor(s) required for full expression of the *mhp* genes, a process mechanistically similar to the squelching of eukaryotic promoters (41).

A 178-bp intergenic region with a higher A+T content (58%) than the *mhp* coding regions (44.24%) is located between the translational start sites of the divergently transcribed *mhpR* and *mhpA* genes (Fig. 2). This observation suggests that the *mhpR* promoter is located near or overlaps the regulated promoter of the gene encoding the MhpA hydroxylase. A similar genetic arrangement has been observed with the *pobR* and *pobA* genes in the *A. calcoaceticus* chromosome (22) and resembles that of the members of the LysR subfamily of transcriptional activators (63). Secondary-structure analyses of the intergenic region showed the presence of several inverted repeats (Fig. 2), a common feature of divergent transcriptional regulators. Remarkably, the 7-bp inverted repeat sequence 886-GTGCACCTGGTGCAC-900 is equivalent to palindromic repeats that have been shown to be the binding site (operator) for the IclR and PobR regulatory proteins (22, 51). Moreover, three 13-bp segments in the intergenic region present similarity to the integration host factor-binding consensus motif (19) and are associated with potential bent DNA regions according to Trifonov's criteria (73) (Fig. 2). All of these observations suggest that expression of the *mhp* cluster is the subject of a complex regulatory mechanism that awaits further research.

Expression of the 3-HPP catabolic cassette in heterologous hosts. To demonstrate that the *E. coli* chromosomal DNA fragment analyzed as described above encodes the complete 3-HPP catabolic pathway, we constructed plasmid pPADR2 (Fig. 4), which contains the genes *mhpR*, *mhpABCDFE*, and *orfT* cloned into pSJ33, an RSF1010-based promiscuous plasmid (34), as a 9.8-kb *Not*I-DNA cassette. Expression of the *Not*I cassette in pPADR2 depends on its own regulatory elements, since the vector does not provide any additional expression signals. As already mentioned above, this catabolic plasmid conferred on *E. coli* MC4100 cells the ability to grow on 3-HPP (Tables 2 and 3), and when it was introduced into *S. typhimurium* LT-2 (which lacks the DNA region corresponding to the *E. coli lac* operon and its flanking genes) (40), *R. meliloti* Rm1021, or *P. putida* KT2442, all of the recombinant strains were able to grow efficiently on minimal medium containing 3-HPP as the sole carbon and energy source (Table 3). These results indicated that the 3-HPP catabolic cassette was expressed and functional not only in enteric bacteria (*E. coli* and *S. typhimurium*) but also in *Pseudomonadaceae* (*P. putida*) and even in a representative of the α subgroup of *Proteobacteria* (*R. meliloti*). Moreover, these data strongly suggest that the cloned 3-HPP cluster contains all of the catabolic genes necessary for the metabolism of this aromatic compound into Krebs cycle intermediates and also the regulatory elements of the pathway.

The results presented here represent the first genetically characterized 3-HPP catabolic pathway. It is worthwhile to note that several genes of the cluster, mainly those encoding the enzymes of the *meta*-cleavage pathway, maintain the same genetic arrangement and are homologous to the corresponding genes of different biodegradation pathways from nonenteric bacteria, mostly *Pseudomonas* species. This observation supports the hypothesis of the modular evolution of catabolic pathways (78). It also reveals that *E. coli* is not an "empty box" for cloning and expression of genes involved in the catabolism of aromatic compounds; on the contrary, it has its own set of genes which, when appropriately induced in the recombinant cells, lead to functionally merodiploid constructs. Although there is some experimental evidence that in *Pseudomonas* species aromatic catabolic genes are clustered in a limited region of the chromosome (26), this supraoperonic clustering is not observed in *E. coli*. Thus, the two aromatic catabolic pathways so far characterized in *E. coli* at the molecular level, i.e., those for 4-hydroxyphenylacetate (57) and 3-HPP degradation, map at min 98 and 8 of the *E. coli* chromosome, respectively. Since the engineered 3-HPP catabolic cassette is functional in microorganisms of environmental importance, such as *R. meliloti* and *P. putida*, its application to improve the catabolic abilities of strains for degradation of pollutants, e.g., oxidation of phenyl carboxylate intermediates produced during the catabolism of linear alkylbenzenes present in commercial detergents (65), can be of biotechnological interest.

ACKNOWLEDGMENTS

We thank E. García for critical reading of the manuscript. We are indebted to J. Barbé and C. Chiapella for sending *S. typhimurium* LT-2 and to M. K. B. Berlyn for providing plasmid pLC20-30. We also thank E. Cano and M. Carrasco for excellent technical assistance. The help of A. Díaz and G. Porras in the sequencing is gratefully acknowledged.

This work was supported by CICYT grant AMB94-1038-CO2-02. A. Ferrández 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

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Andreoni, V., and G. Bestetti.** 1986. Comparative analysis of different *Pseudomonas* strains that degrade cinnamic acid. Appl. Environ. Microbiol. **52:**930–934.
3. **Bachellier, S., W. Saurin, D. Perrin, M. Hofnung, and E. Gilson.** 1994. Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). Mol. Microbiol. **12:**61–70.
4. **Bairoch, A., P. Bucher, and K. Hofmann.** 1995. The PROSITE database, its status in 1995. Nucleic Acids Res. **24:**189–196.
5. **Birkenbihl, R. P., and W. Vielmetter.** 1989. Complete maps of IS1, IS2, IS3, IS4, IS5, IS30, and IS50 locations in *Escherichia coli* K12. Mol. Gen. Genet. **220:**147–153.
6. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
7. **Brennan, R. G., and B. W. Matthews.** 1989. The helix-turn-helix DNA bind-

ing motif. J. Biol. Chem. **264:**1903–1906.
8. **Bugg, T. D. H.** 1993. Overproduction, purification and properties of 2,3-dihydroxyphenylpropionate 1,2-dioxygenase from *Escherichia coli*. Biochim. Biophys. Acta **1202:**258–264.
9. **Burlingame, R., and P. J. Chapman.** 1983. Catabolism of phenylpropionic acid and its 3-hydroxy derivative by *Escherichia coli*. J. Bacteriol. **155:**113–121.
10. **Burlingame, R., and P. J. Chapman.** 1983. Stereospecificity in *meta*-fission catabolic pathways. J. Bacteriol. **155:**424–426.
11. **Burlingame, R. P.** 1983. Ph.D. thesis. University of Minnesota, Minneapolis.
12. **Burlingame, R. P., L. Wyman, and P. J. Chapman.** 1986. Isolation and characterization of *Escherichia coli* mutants defective for phenylpropionate degradation. J. Bacteriol. **168:**55–64.
13. **Clark, D. J., and O. Maaloe.** 1967. DNA replication and the division cycle of *Escherichia coli*. J. Mol. Biol. **23:**99–112.
14. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9:**91–99.
15. **Coulson, C. B., and W. C. Evans.** 1959. Microbiological degradation of *trans*-cinnamic acid by soil *Pseudomonas*. Chem. Ind. **17:**543–544.
16. **Dagley, S., P. J. Chapman, and D. T. Gibson.** 1965. The metabolism of β-phenylpropionic acid by an *Achromobacter*. Biochem. J. **97:**643–650.
17. **Decker, H., H. Motamedi, and C. R. Hutchinson.** 1993. Nucleotide sequences and heterologous expression of *tcmG* and *tcmP*, biosynthetic genes for tetracenomycin C in *Streptomyces glaucescens*. J. Bacteriol. **175:**3876–3886.
18. **de Lorenzo, V., and K. N. Timmis.** 1994. Analysis and construction of stable phenotypes in gram-negative bacteria with Tn*5*- and Tn*10*-derived mini-transposons. Methods Enzymol. **235:**386–405.
19. **de Lorenzo, V., M. Herrero, M. Metzke, and K. N. Timmis.** 1991. An upstream XylR- and IHF-induced nucleoprotein complex regulates the σ^{54}-dependent Pu promoter of TOL plasmid. EMBO J. **10:**1159–1167.
20. **Díaz, E., and K. N. Timmis.** 1995. Identification of functional residues in a 2-hydroxymuconic semialdehyde hydrolase. A new member of the α/β hydrolase-fold family of enzymes which cleaves carbon-carbon bonds. J. Biol. Chem. **270:**6403–6411.
21. **Díaz, E., M. Munthali, V. de Lorenzo, and K. N. Timmis.** 1994. Universal barrier to lateral spread of specific genes among microorganisms. Mol. Microbiol. **13:**855–861.
22. **DiMarco, A. A., and L. N. Ornston.** 1994. Regulation of *p*-hydroxybenzoate hydroxylase synthesis by PobR bound to an operator in *Acinetobacter calcoaceticus*. J. Bacteriol. **176:**4277–4284.
23. **DiMarco, A. A., B. A. Averhoff, E. E. Kim, and L. N. Ornston.** 1993. Evolutionary divergence of *pobA*, the structural gene encoding *p*-hydroxybenzoate hydroxylase, in an *Acinetobacter calcoaceticus* strain well-suited for genetic analysis. Gene **125:**25–33.
24. **DiMarco, A. A., B. Averhoff, and L. N. Ornston.** 1993. Identification of the transcriptional activator *pobR* and characterization of its role in the expression of *pobA*, the structural gene for *p*-hydroxybenzoate hydroxylase in *Acinetobacter calcoaceticus*. J. Bacteriol. **175:**4499–4506.
25. **Duggleby, C. J., and P. A. Williams.** 1986. Purification and some properties of the 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase (2-hydroxymuconic semialdehyde hydrolase) encoded by the TOL plasmid pWW0 from *Pseudomonas putida* mt-2. J. Gen. Microbiol. **132:**717–726.
26. **Eaton, R. W.** 1996. *p*-Cumate catabolic pathway in *Pseudomonas putida* F1: cloning and characterization of DNA carrying the *cmt* operon. J. Bacteriol. **178:**1351–1362.
27. **Fernández, S., V. de Lorenzo, and J. Pérez-Martín.** 1995. Activation of the transcriptional regulator XylR of *Pseudomonas putida* by release of repression between functional domains. Mol. Microbiol. **16:**205–213.
28. **Gold, L.** 1988. Posttranscriptional regulatory mechanisms in *Escherichia coli*. Annu. Rev. Biochem. **57:**199–233.
29. **Harayama, S., and M. Rekik.** 1993. Comparison of the nucleotide sequences of the *meta*-cleavage pathway genes of TOL plasmid pWW0 from *Pseudomonas putida* with other *meta*-cleavage genes suggests that both single and multiple nucleotide substitutions contribute to enzyme evolution. Mol. Gen. Genet. **239:**81–89.
30. **Harwood, C. S., N. N. Nichols, M.-K. Kim, J. L. Ditty, and R. E. Parales.** 1994. Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. J. Bacteriol. **176:**6479–6488.
31. **Henderson, P. J. F.** 1990. The homologous glucose transport proteins of prokaryotes and eukaryotes. Res. Microbiol. **141:**316–328.
32. **Hofer, B., S. Backhaus, and K. N. Timmis.** 1994. The biphenyl/polychlorinated biphenyl-degradation locus (*bph*) of *Pseudomonas* sp. LB400 encodes four additional metabolic enzymes. Gene **144:**9–16.
33. **Houghton, J. E., T. M. Brown, J. A. Andrew, E. J. Hughes, and L. N. Ornston.** 1995. Discontinuities in the evolution of *Pseudomonas putida cat* genes. J. Bacteriol. **177:**401–412.
34. **Jaenecke, S., V. de Lorenzo, K. N. Timmis, and E. Díaz.** 1996. A stringently controlled expression system for analysing lateral gene transfer between bacteria. Mol. Microbiol. **21:**293–300.
35. **Kabisch, M., and P. Fortnagel.** 1990. Nucleotide sequence of metapyrocatechase I (catechol 2,3-oxygenase I) gene *mpcI* from *Alcaligenes eutrophus* JMP222. Nucleic Acids Res. **18:**3405–3406.
36. **Kikuchi, Y., Y. Yasukochi, Y. Nagata, M. Fukuda, and M. Takagi.** 1994. Nucleotide sequence and functional analysis of the *meta*-cleavage pathway involved in biphenyl and polychlorinated biphenyl degradation in *Pseudomonas* sp. strain KKS102. J. Bacteriol. **176:**4269–4276.
37. **Kowalchuk, G. A., G. B. Hartnett, A. Benson, J. E. Houghton, K.-L. Ngai, and L. N. Ornston.** 1994. Contrasting patterns of evolutionary divergence within the *Acinetobacter calcoaceticus pca* operon. Gene **146:**23–30.
38. **Kröger, M., and R. Wahl.** 1996. Compilation of DNA sequences of *Escherichia coli* K12 (ECD and ECDC; update 1995). Nucleic Acids Res. **24:**29–31.
39. **Lam, W. W. Y., and T. D. H. Bugg.** 1994. Chemistry of extradiol aromatic ring cleavage: isolation of a stable dienol ring fission intermediate and stereochemistry of its enzymatic hydrolytic cleavage. J. Chem. Soc. Chem. Commun. **1994:**1163–1164.
40. **Lampel, K. A., and M. Riley.** 1982. Discontinuity of homology of *Escherichia coli* and *Salmonella typhimurium* DNA in the *lac* region. Mol. Gen. Genet. **186:**82–86.
41. **Levine, M., and J. Manley.** 1989. Transcriptional repression of eukaryotic promoters. Cell **59:**405–408.
42. **Levy, C. C.** 1967. Melilotate hydroxylase. Purification of the enzyme and the nature of the prosthetic group. J. Biol. Chem. **242:**747–753.
43. **Malherbe, P., C. Köhler, M. Da Prada, G. Lang, V. Kiefer, R. Schwarcz, H. W. Lahm, and A. M. Cesura.** 1994. Molecular cloning and functional expression of human 3-hydroxyanthranilic acid dioxygenase. J. Biol. Chem. **269:**13792–13797.
44. **Marger, M. D., and M. H. Saier, Jr.** 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem. Sci. **18:**13–20.
45. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. **Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84:**166–169.
47. **Noda, Y., S. Nishikawa, K. I. Shiozuka, H. Kadokura, H. Nakajima, K. Yoda, Y. Katayama, N. Morohoshi, T. Haraguchi, and M. Yamasaki.** 1990. Molecular cloning of the protocatechuate 4,5-dioxygenase genes of *Pseudomonas paucimobilis*. J. Bacteriol. **172:**2704–2709.
48. **Nomura, Y., M. Nakagawa, N. Ogawa, S. Harashima, and Y. Oshima.** 1992. Genes in PHT plasmid encoding the initial degradation pathway of phthalate in *Pseudomonas putida*. J. Ferment. Bioeng. **74:**333–344.
49. **Nurk, A., L. Kasak, and M. Kivisaar.** 1991. Sequence of the gene (*pheA*) encoding phenol monooxygenase from *Pseudomonas* sp. EST1001: expression in *Escherichia coli* and *Pseudomonas putida*. Gene **102:**13–18.
50. **Orser, C. S., C. C. Lange, L. Xun, T. C. Zahrt, and B. J. Schneider.** 1993. Cloning, sequence analysis, and expression of the *Flavobacterium* pentachlorophenol-4-monooxygenase gene in *Escherichia coli*. J. Bacteriol. **175:**411–416.
51. **Pan, B., I. Unnikrishnan, and D. C. LaPorte.** 1996. The binding site of the IclR repressor protein overlaps the promoter of *aceBAK*. J. Bacteriol. **178:**3982–3984.
52. **Perkins, E. J., M. P. Gordon, O. Caceres, and P. F. Lurquin.** 1990. Organization and sequence analysis of the 2,4-dichlorophenol hydroxylase and dichlorocatechol oxidative operons of plasmid pJP4. J. Bacteriol. **172:**2351–2359.
53. **Perkins, J. D., J. D. Heath, B. R. Sharma, and G. M. Weinstock.** 1993. *Xba*I and *Bln*I genomic cleavage maps of *Escherichia coli* K-12 strain MG1655 and comparative analysis of other strains. J. Mol. Biol. **232:**419–445.
54. **Platt, A., V. Shingler, S. C. Taylor, and P. A. Williams.** 1995. The 4-hydroxy-2-oxovalerate aldolase and acetaldehyde dehydrogenase (acylating) encoded by the *nahM* and *nahO* genes of the naphthalene catabolic plasmid pWW60-22 provide further evidence of conservation of *meta*-cleavage pathway gene sequences. Microbiology **141:**2223–2233.
55. **Powlowski, J., L. Sahlman, and V. Shingler.** 1993. Purification and properties of the physically associated *meta*-cleavage pathway enzymes 4-hydroxy-2-ketovalerate aldolase and aldehyde dehydrogenase (acylating) from *Pseudomonas* sp. strain CF600. J. Bacteriol. **175:**377–385.
56. **Prieto, M. A., and J. L. García.** Unpublished data.
57. **Prieto, M. A., E. Díaz, and J. L. García.** 1996. Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli* W: engineering a mobile aromatic degradative cluster. J. Bacteriol. **178:**111–120.
58. **Roper, D. I., and R. A. Cooper.** 1990. Subcloning and nucleotide sequence of the 3,4-dihydroxyphenylacetate (homoprotocatechuate) 2,3-dioxygenase gene from *Escherichia coli* C. FEBS Lett. **275:**53–57.
59. **Roper, D. I., J. M. Stringfellow, and R. A. Cooper.** 1995. Sequence of the *hpcC* and *hpcG* genes of the *meta*-fission homoprotocatechuic acid pathway of *Escherichia coli* C: nearly 40% amino-acid identity with the analogous enzymes of the catechol pathway. Gene **156:**47–51.
60. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
61. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.

62. **Savage, D. C.** 1977. Microbial ecology of the gastrointestinal tract. Annu. Rev. Microbiol. **31:**107–133.
63. **Schell, M. A.** 1993. Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. **47:**597–626.
64. **Shingler, V., J. Powlowski, and U. Marklund.** 1992. Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. J. Bacteriol. **174:**711–724.
65. **Simoni, S., S. Klinke, C. Zipper, W. Angst, and H.-P. E. Kohler.** 1996. Enantioselective metabolism of chiral 3-phenylbutyric acid, an intermediate of linear alkylbenzene degradation, by *Rhodococcus rhodochrous* PB1. Appl. Environ. Microbiol. **62:**749–755.
66. **Skinner, M. A., and R. A. Cooper.** 1982. An *Escherichia coli* mutant defective in the NAD-dependent succinate semialdehyde dehydrogenase. Arch. Microbiol. **132:**270–275.
67. **Spence, E. L., M. Kawamukai, J. Sanvoisin, H. Braven, and T. D. H. Bugg.** 1996. Catechol dioxygenases from *Escherichia coli* (MhpB) and *Alcaligenes eutrophus* (MpcI): sequence analysis and biochemical properties of a third family of extradiol dioxygenases. J. Bacteriol. **178:**5249–5256.
68. **Strickland, S., and V. Massey.** 1973. The purification and properties of the flavoprotein melilotate hydroxylase. J. Biol. Chem. **248:**2944–2952.
69. **Stringfellow, J. M., B. Turpin, and R. A. Cooper.** 1995. Sequence of the *Escherichia coli* C homoprotocatechuic acid degradative operon completed with that of the 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase-encoding gene (*hpcH*). Gene **166:**73–76.
70. **Sunnarborg, A., D. Klumpp, T. Chung, and D. C. LaPorte.** 1990. Regulation of the glyoxylate bypass operon: cloning and characterization of *iclR*. J. Bacteriol. **172:**2642–2649.
71. **Tate, W. P., and S. A. Mannering.** 1996. Three, four or more: the translational stop signal at length. Mol. Microbiol. **21:**213–219.
72. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
73. **Trifonov, E.** 1985. Curved DNA. Crit. Rev. Biochem. **19:**89–106.
74. **Wang, Y., M. Rawlings, D. T. Gibson, D. Labbé, H. Bergeron, R. Brousseau, and P. C. K. Lau.** 1995. Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in *Pseudomonas putida* F1. Mol. Gen. Genet. **246:**570–579.
75. **Weijer, W. J., J. Hofsteenge, J. M. Vereijken, P. A. Jekel, and J. J. Beintema.** 1982. Primary structure of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. Biochim. Biophys. Acta **704:**385–388.
76. **Wierenga, R. K., P. Tepstra, and W. G. J. Hol.** 1986. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. J. Mol. Biol. **187:**101–107.
77. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80:**726–730.
78. **Williams, P. A., and J. R. Sayers.** 1994. The evolution of pathways for aromatic hydrocarbon oxidation in *Pseudomonas*. Biodegradation **5:**195–217.