JOURNAL OF BACTERIOLOGY, Jan. 1996, p. 111–120 Vol. 178, No. 1
0021-9193/96/$04.00+0

Molecular Characterization of the 4-Hydroxyphenylacetate Catabolic Pathway of *Escherichia coli* W: Engineering a Mobile Aromatic Degradative Cluster

MARÍA A. PRIETO,[1] EDUARDO DÍAZ,[2]† AND JOSÉ L. GARCÍA[1]*

Department of Molecular Microbiology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain,[1] and Division of Microbiology, GBF-National Research Centre for Biotechnology, D-38124 Braunschweig, Germany[2]

Received 2 August 1995/Accepted 24 October 1995

We have determined and analyzed the nucleic acid sequence of a 14,855-bp region that contains the complete gene cluster encoding the 4-hydroxyphenylacetic acid (4-HPA) degradative pathway of *Escherichia coli* W (ATCC 11105). This catabolic pathway is composed by 11 genes, i.e., 8 enzyme-encoding genes distributed in two putative operons, *hpaBC* (4-HPA hydroxylase operon) and *hpaGEDFHI* (*meta*-cleavage operon); 2 regulatory genes, *hpaR* and *hpaA*; and the gene, *hpaX*, that encodes a protein related to the superfamily of transmembrane facilitators and appears to be cotranscribed with *hpaA*. Although comparisons with other aromatic catabolic pathways revealed interesting similarities, some of the genes did not present any similarity to their corresponding counterparts in other pathways, suggesting different evolutionary origins. The cluster is flanked by two genes homologous to the *cstA* (carbon starvation protein) and *tsr* (serine chemoreceptor) genes of *E. coli* K-12. A detailed genetic analysis of this region has provided a singular example of how *E. coli* becomes adapted to novel nutritional sources by the recruitment of a catabolic cassette. Furthermore, the presence of the *pac* gene in the proximity of the 4-HPA cluster suggests that the penicillin G acylase was a recent acquisition to improve the ability of *E. coli* W to metabolize a wider range of substrates, enhancing its catabolic versatility. Five repetitive extragenic palindromic sequences that might be involved in transcriptional regulation were found within the cluster. The complete 4-HPA cluster was cloned in plasmid and transposon cloning vectors that were used to engineer *E. coli* K-12 strains able to grow on 4-HPA. We report here also the in vitro design of new biodegradative capabilities through the construction of a transposable cassette containing the wide substrate range 4-HPA hydroxylase, in order to expand the *ortho*-cleavage pathway of *Pseudomonas putida* KT2442 and allow the new recombinant strain to use phenol as the only carbon source.

Although most of our current knowledge about the general bacterial metabolic pathways has been derived from the analysis of *Escherichia coli*, very few data are available about the ability of this microorganism to grow on aromatic compounds other than amino acids. It has been shown that *E. coli* B, C, and W, but not K-12 strains, are able to degrade 4-hydroxyphenylacetic acid (4-HPA) and homoprotocatechuate (3,4-hydroxyphenylacetate) (HPC) via an inducible, chromosomally encoded *meta*-cleavage pathway (8, 10).

The HPC degradative operon of *E. coli* C has been partially cloned (25, 43), and some of its products have been characterized (16–18, 40–42, 44, 48). In addition, we have previously demonstrated that the first step in the 4-HPA degradation in *E. coli* W, i.e., the formation of HPC, is catalyzed by a two-component aromatic hydroxylase (38, 39). This enzyme is encoded by two genes which appear to be part of the same operon (38). The homologous 4-HPA hydroxylase operon of *E. coli* C has been also cloned and partially sequenced (38, 39). In spite of this information, some genes of the 4-HPA catabolic pathway as well as their genetic arrangement remained unknown. Hence, the aim of this work was to characterize the complete 4-HPA degradation pathway of *E. coli* W. This strain was selected because it is the only strain of *E. coli* able to use either 4-HPA or phenylacetic acid as a carbon source (8). In addition, this strain contains near the 4-HPA operon the *pac* gene (38, 39), encoding a penicillin G acylase, an enzyme able to hydrolyze a wide range of amides and esters of 4-HPA and phenylacetic acids and thus useful in expanding the range of catabolic substrates for the 4-HPA pathway.

* Corresponding author: Department of Molecular Microbiology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Velázquez 144, 28006 Madrid, Spain. Phone: 34-1-5611800. Fax: 34-1-5627518. Electronic mail address: CIBJL16@PINAR1.CSIC.ES.

† Present address: Department of Molecular Microbiology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain.

MATERIALS AND METHODS

Materials. Restriction endonucleases, T4 DNA ligase, the Klenow fragment of *E. coli* DNA polymerase I, and the T7 DNA polymerase sequencing kit were from Pharmacia Fine Chemicals. Fluorescence-labeled dideoxynucleotide terminators and *Taq* DNA polymerase were from Applied Biosystems Inc. [α-^{35}S]dCTP and [α-^{32}P]dCTP were from Amersham Corp. Culture media were from Difco. All other chemicals were of the highest grade available and were purchased from Sigma or Merck.

Strains, plasmids, media, and growth conditions. The bacterial strains used were *E. coli* W ATCC 11105 (6), *E. coli* K-12 strain DH1 (45), *E. coli* K-12 strain W3110 (39), *E. coli* W (provided by A. Garrido-Pertierra) (39), *E. coli* C (39), *E. coli* B/rK (39), *E. coli* ET8000 (29), *E. coli* S17-1λ*pir* (22), *Pseudomonas putida* KT2442 (22), *Klebsiella pneumoniae* M5a1 (39), and *Kluyvera citrophila* ATCC 21285 (4). Bacteria were grown in Luria broth or M9 minimal medium (45) at 30°C with shaking. The plasmids used were pUC18 (55); pUT mini-Tn*5*Km2 (13); pUC18Not (22); pCNB5 (12); pAG464, which contains the gene encoding HPC 2,3-dioxygenase from *K. pneumoniae* M5a1 (provided by A. Garrido-Pertierra) (31); pAJ19, which contains the operon encoding 4-HPA hydroxylase and the *pac* gene encoding penicillin G acylase (39); and pAJ221, which contains only the 4-HPA hydroxylase operon (39).

DNA manipulation, transformation, and sequencing. Isolation of plasmids and chromosomal DNA was carried out by standard procedures (45). Restriction endonucleases, alkaline phosphatase, and T4 DNA ligase were used according to

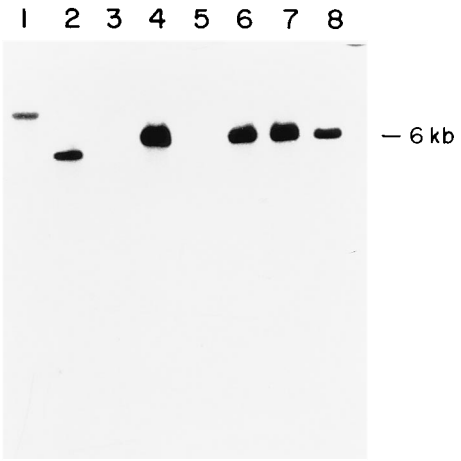

FIG. 1. Southern analysis of various chromosomal DNAs probed with the HPC 2,3-dioxygenase gene of *K. pneumoniae* M5a1. Lanes 1 to 8, chromosomal DNAs from *K. citrophila*, *K. pneumoniae* M5a1, *E. coli* W3110, *E. coli* W, *E. coli* DH1, *E. coli* C, *E. coli* B/rK, and *E. coli* W ATCC 11105, respectively, digested with *Bam*HI. The size of the band corresponding to the *Bam*HI fragment contained in pHCB1 (see Fig. 2) is indicated.

the manufacturer's instructions. Transformation of *E. coli* cells was carried out by the RbCl method (45) or by electroporation (Gene Pulser; Bio-Rad). Insertions of minitransposon elements into the chromosomes of the target strains were produced with a filter-mating technique (22). Southern blot and colony hybridization analyses were performed as previously described (39, 45). Nucleotide sequences were determined directly from plasmids by using either the T7 DNA polymerase kit from Pharmacia or the *Taq* DNA polymerase-initiated cycle sequencing reactions with fluorescence-labeled dideoxynucleotide terminators in an Applied Biosystems 373A automated DNA sequencer. Templates for sequencing were obtained by deletion subcloning. Synthetic oligonucleotides were used as primers when required to complete the sequence. DNA fragments were purified by standard procedures using Gene Clean (Bio 101, Inc.).

Computer analysis. Nucleotide sequences were analyzed with the University of Wisconsin Molecular Biology Package (WIMP) available on the VAX cluster mainframe computer system at the Biotechnology National Center (Madrid, Spain). Sequence comparisons were made against the GenBank/EMBL and the Swissprot databases with the FASTA and TFASTA programs.

RESULTS AND DISCUSSION

Cloning and physical map of the 4-HPA *meta*-cleavage operon of *E. coli* W ATCC 11105. To isolate the genes encoding the *meta*-cleavage pathway of 4-HPA, we explored the possibility of using the gene encoding HPC 2,3-dioxygenase of *K. pneumoniae* M5a1 as a DNA probe. This gene has been found to be similar to that encoding the equivalent dioxygenase of *E. coli* C (31). Southern blot analyses of various chromosomal DNAs probed with this gene demonstrated that only the strains able to grow on 4-HPA showed hybridization bands (Fig. 1). Hence, several genomic libraries of *E. coli* W ATCC 11105 constructed in pUC18 were screened for the presence of the homologous HPC 2,3-dioxygenase gene. Two plasmids, pHCB1 and pHCB3, containing overlapping 6-kb *Bam*HI and 11-kb *Eco*RI fragments, respectively, were isolated (Fig. 2). Interestingly, the *E. coli* K-12 transformants harboring plasmid pHCB3 showed the typical black phenotype previously observed with cells carrying plasmid pAJ19, containing the 4-HPA hydroxylase operon of *E. coli* W (39). Further hybridization and biochemical analyses (data not shown) confirmed that pHCB3 contained this operon (Fig. 2).

Preliminary sequencing data around the left-hand *Bam*HI site in pHCB1 (Fig. 2) showed a truncated gene corresponding to the *hpcR* gene of the HPC operon of *E. coli* C. Using the 2.7-kb *Bam*HI-*Eco*RI fragment of pHCB1 as a DNA probe to screen the *Eco*RI library of *E. coli* W ATCC 11105, we isolated plasmid pHCR1, which contained a 6-kb *Eco*RI overlapping fragment (Fig. 2). Restriction enzyme analyses of plasmids pHCB1, pHCB3, and pHCR1 allowed us to determine the physical map of the complete 4-HPA catabolic cluster (Fig. 2).

Sequence of the 4-HPA cluster. The complete nucleotide sequence of the gene cluster encoding the 4-HPA pathway as well as part of the flanking regions was determined (Fig. 3). Analyses of the open reading frames (ORFs) and sequence comparisons (see below) suggested that the 4-HPA metabolic cluster is composed of 11 genes arranged as follows: (i) 8 enzyme-encoding genes organized in two putative operons, the 4-HPA hydroxylase operon (*hpaBC*) (38) and the HPC *meta*-cleavage operon (*hpaGEDFHI*), similar to that of *E. coli* C (43); (ii) 2 regulatory genes, *hpaR* and *hpaA*; and (iii) the gene of unknown function (*hpaX*) that is likely cotranscribed with *hpaA*. All the genes are transcribed in the same direction with the sole exception of *hpaR*.

The first gene of the *meta*-cleavage operon, *hpaG*, encodes a protein of 46,927 Da, almost identical to the product of *hpcE*, a bifunctional decarboxylase/isomerase enzyme in the HPC degradative pathway of *E. coli* C (42). However, HpaG is 24 amino acids longer than HpcE because of a 7-bp deletion in the 3′ end of *hpcE* producing a premature termination. It has been proposed that the bifunctional enzyme HpcE evolved from a gene duplication, since the N-terminal half of the protein is very similar to the C-terminal one (42). It is therefore interesting that the alignment of the N- and C-terminal parts of HpaG was even better than that of the two halves of HpcE (42), because of the contribution of the additional amino acids present in the C-terminal end of HpaG (data not shown). The HpaG decarboxylase is similar to other *meta*-fission catabolic pathway decarboxylases and hydratases which, in spite of catalyzing different reactions, have been previously suggested to have a common origin (42, 44). Especially remarkable was the high similarity found between the two similar halves of HpaG and a protein derived from a cDNA library of the nematode *Caenorhabditis elegans* (57 to 58% similarity; 30 to 35% identity) (54) and a truncated protein encoded by a gene of unknown function located upstream of the phosphatidylserine decarboxylase gene of *Saccharomyces cerevisiae* (58 to 65% similarity; 37 to 40% identity) (9). These proteins of unknown function might be the first representatives of a decarboxylase/hydratase family in eukaryotic cells.

Interestingly, a comparative search of the protein databases did not reveal any similarity between HpaG and other isomerases. This type of bifunctional decarboxylase/isomerase activity proposed for HpcE (HpaG in *E. coli* W) has been found only in the tryptophan pathway (11), but no equivalent 2-hydroxy-hept-2,4-diene-1,7-dioic acid (HHDD) isomerase activity was detected in other aromatic catabolic pathways, where the enol-keto isomerization appears to be a dispensable enzymatic step. Harayama et al. (20) have demonstrated that the unstable 2-hydroxypent-2,4-dienoate (enol form), but not its isomeric keto forms, is the actual substrate of the XylJ hydratase of the TOL pathway. It was surprising that the 5-oxo-pent-3-ene-1,2,5-tricarboxylic acid decarboxylase and HHDD isomerase activities of *E. coli* C were not found associated in the earliest biochemical analyses of HpcE (25). In fact, this finding allowed the isolation of the unstable product HHDD (enol form), which was further used to assay the isomerase and hydratase activities of HpcE and HpcG, respectively (17, 20, 25, 42, 43). Hence, we believe that there are not conclusive data to assume that 2-oxo-hept-3-ene-1,7-dioic acid is the substrate of HpcG hydratase (HpaH in *E. coli* W), and we propose that the transformation of HHDD into 2,4-dihydroxy-hept-2-ene-1,7-dioic

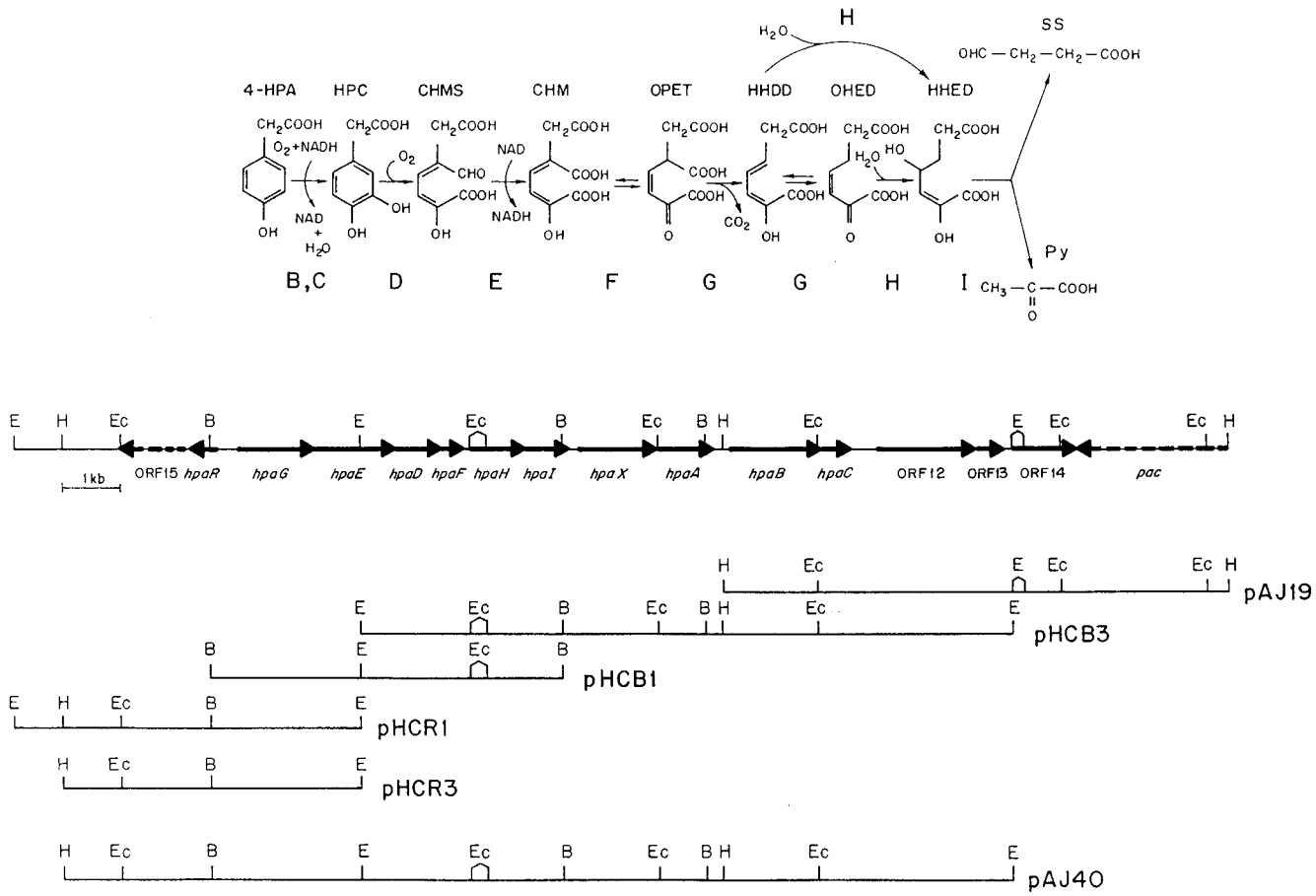

FIG. 2. Restriction map of the 4-HPA catabolic gene cluster and plasmid derivatives. Arrows indicate the direction of gene transcription. Dashed arrows indicate the genes which have been only partially sequenced in this work. Restriction sites are as follows: B, *Bam*HI; E, *Eco*RI; Ec, *Eco*RV; H, *Hin*dIII. The proposed 4-HPA catabolic pathway is indicated at the top. Abbreviations: HPC, homoprotocatechuic acid; CHMS, 5-carboxymethyl-2-hydroxy-muconic semialdehyde; CHM, 5-carboxymethyl-2-hydroxy-muconic acid; OPET, 5-oxo-pent-3-ene-1,2,5-tricarboxylic acid; OHED, 2-oxo-hept-3-ene-1,7-dioic acid; HHED, 2,4-dihydroxy-hept-2-ene-1,7-dioic acid; Py, pyruvic acid; SS, succinic semialdehyde. B, C, D, E, F, G, H, and I represent the enzymes encoded by the respective *hpa* genes. The DNA fragments contained in plasmids pAJ19, pHCB3, pHCB1, pHCR1, pHCR3, and pAJ40 are indicated at the bottom.

acid occurs without a previous enzymatic isomerization step (Fig. 2). To demonstrate this hypothesis, we are trying to synthesize chemically HHDD, which should allow us to determine the specificity of the hydratase by following the procedure described by Harayama et al. (20).

The gene *hpaE* encodes a protein of 53,011 Da that presents all the motifs that characterize the aldehyde dehydrogenase superfamily (23). The HpaE protein shows a 98.6% identity to its putative counterpart, 5-carboxymethyl-2-hydroxy-muconic semialdehyde dehydrogenase of *E. coli* C (16, 44), and about 40% identity to other *meta*-cleavage pathway aldehyde dehydrogenases, such as the DmpC and XylG proteins from the phenol pathway of *Pseudomonas* sp. strain CF600 and the TOL pathway of *P. putida*, respectively (23, 44, 47).

The gene *hpaD* encodes a protein of 32,018 Da homologous to the HPC 2,3-dioxygenase of *E. coli* C encoded by the *hpcB* gene (41). The most relevant difference between both enzymes was observed in their C-terminal ends, where a 2-bp insertion produces a premature termination in HpcB. Surprisingly, both dioxygenases did not present any similarity to other extradiol or intradiol dioxygenases sequenced so far. However, it is worth noting that the HPC 2,3-dioxygenase of *K. pneumoniae* M5a1, which has the peculiarity of requiring magnesium for activity (19), seems to be similar to HpaD and HpcB (Fig. 1). Hence, we suggest that these enzymes constitute a new family of dioxygenases.

The protein encoded by *hpaF* is identical to the 5-carboxymethyl-2-hydroxy-muconic acid isomerase of *E. coli* C encoded by *hpcD* (40). As in the case of the dioxygenases, both *E. coli* isomerases do not present any similarity to equivalent enzymes of other aromatic degradative pathways. Hence, the isomerases and dioxygenases of the HPC *meta*-cleavage pathway appear to have evolved from a different origin.

The *hpaH* gene codes for a protein of 29,714 Da that we have ascribed to 2-oxo-hept-3-ene-1,7-dioic acid hydratase, since it is homologous to the HpcG hydratase from *E. coli* C (17, 44). The HpaH hydratase is also similar to other bacterial hydratases and decarboxylases, reinforcing the previous suggestion about the common origin for these enzymes (see above).

According to Roper et al. (43), the last enzyme of the HPC *meta*-cleavage pathway, the 2,4-dihydroxy-hept-2-ene-1,7-dioic acid aldolase of 28,072 Da, is encoded by the *hpaI* gene. The equivalent *hpcH* gene of *E. coli* C has not yet been sequenced, and the only significant similarity (44% identity) was observed with a truncated ORF in a gene of unknown function located upstream from two ORFs related to the gluconate metabolism of *E. coli* (27) (Fig. 4).

The *hpaR* gene encodes a protein of 17,235 Da, identical to

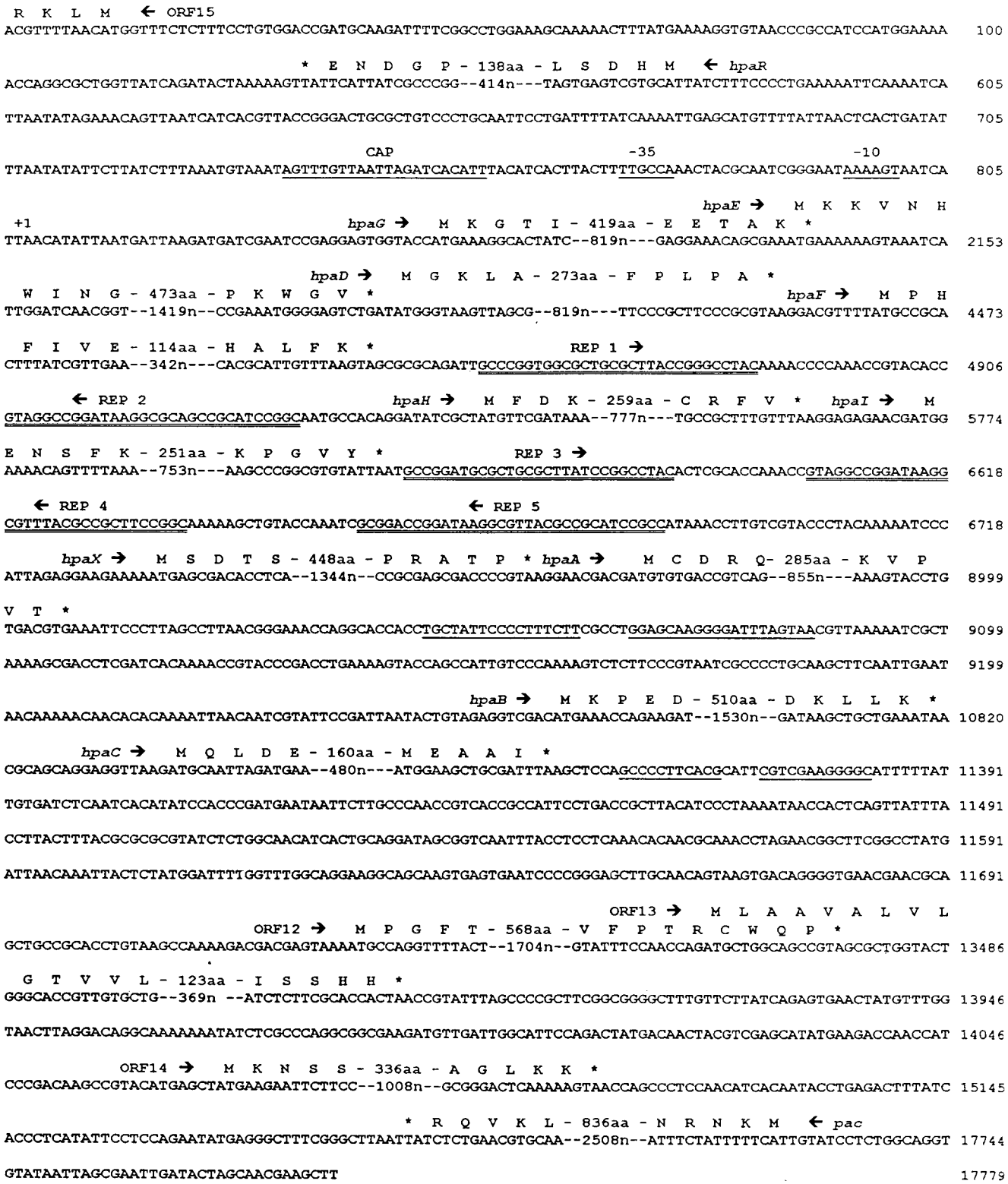

FIG. 3. Nucleotide and derived amino acid sequences of the 4-HPA catabolic pathway. Only the sequences of the 5′- and 3′-end coding regions of the genes are shown. The sequence data appear in the GenBank/EMBL Data Bank with accession numbers Z37980 (nucleotides 1 to 14852) and M17609 (nucleotides 14856 to 17779) (36). Arrows indicate the direction of gene transcription. The five REP sequences are double underlined. Putative hairpin loops and catabolite activator protein (CAP) and promoter sequences are underlined. +1, the adenine at which the transcription of the *hpaGEDFHI* operon might be initiated. Stars indicate the stop codons.

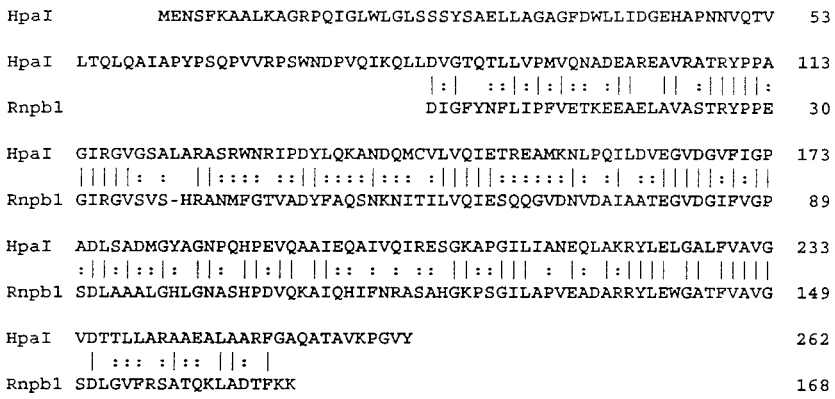

FIG. 4. Amino acid sequence alignment of HpaI aldolase with an ORF of *E. coli*. The truncated ORF of *E. coli* is encoded by a gene of unknown function located near the *mpB* gene (27). Lines indicate identity; colons indicate similarity.

the putative regulatory *hpcR* gene of the HPC *meta*-cleavage pathway of *E. coli* C. HpcR has been proposed to function as a repressor and appears to be unrelated to any other regulator (43).

The *hpaA* gene, which encodes a protein of 34,129 Da, was previously located upstream from the 4-HPA hydroxylase operon (38). On the basis of the homology of HpaA with the AraC/XylS family of regulators, we had postulated that this protein plays a regulatory role in the 4-HPA hydroxylase operon (38). This hypothesis has been recently confirmed by genetic and biochemical analyses (unpublished data). Interestingly, *hpaA* appears to be cotranscribed with the gene *hpaX*, which encodes a protein of 50,568 Da. A comparative search in the data banks revealed that HpaX is similar (34% identity; 57% similarity) to the protein encoded by the *pht1* gene of the phthalate degradative pathway of *P. putida* (35) (Fig. 5). The Pth1 protein was suggested to function either as a positive regulator for the expression of the *pth* genes or as a phthalate transporter (35). This last assumption was made because of its similarity to the glycerol-3-phosphate transporter (GlpT) of *E. coli* (15). Hence, HpaX might be a new member of the fourth cluster or family of the major superfamily of transmembrane facilitators (MFS) (29). HpaX presents significant similarity to the members of this cluster, i.e., the glycerol-3-phosphate antiporter (GlpT) of *E. coli* (49% similarity; 17% identity), the phosphoglycerate antiporter (PgtP) of *Salmonella typhimurium* (52% similarity; 21% identity), and the hexose-phosphate antiporter (UhpT) of *E. coli* (46% similarity; 20% identity) and its regulatory protein (UhpC) (50% similarity; 22% identity). All the proteins of the MFS are about 400 amino acids long

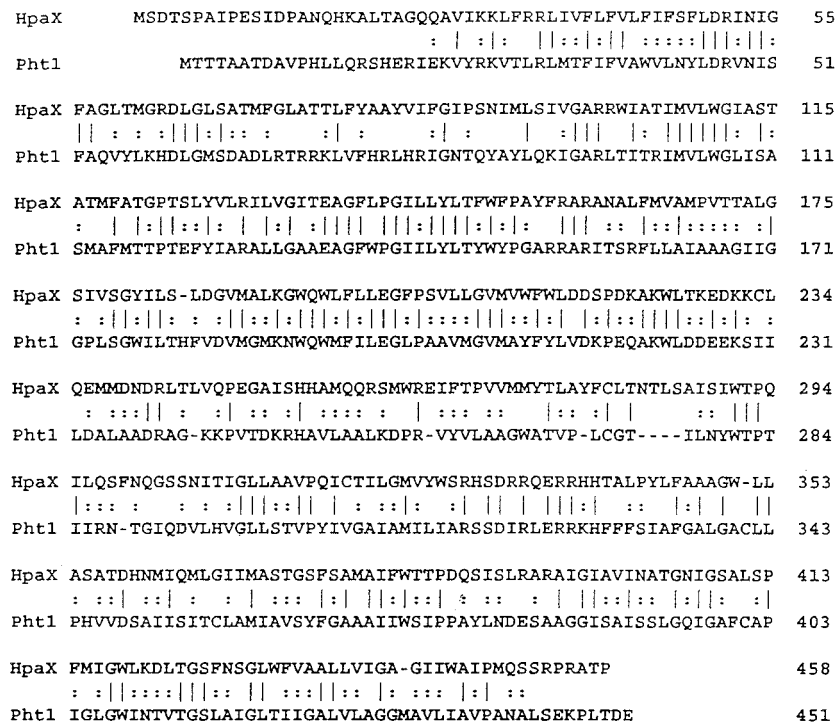

FIG. 5. Amino acid sequence alignment of HpaX with Pht1. The Pht1 protein is encoded by the *pht1* gene of the phthalate degradative pathway of *P. putida* (35). Lines indicate identity; colons indicate similarity.

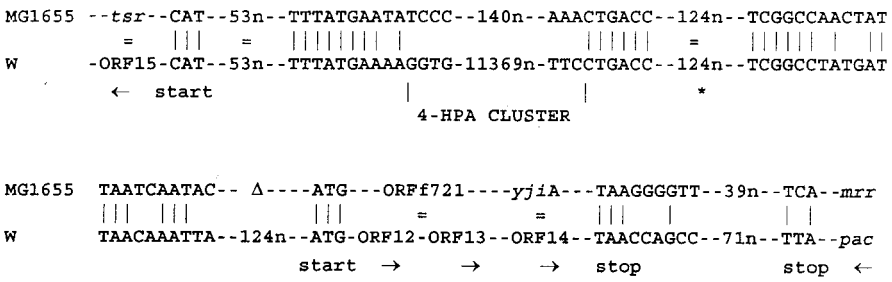

FIG. 6. Comparison of the flanking sequences of the 4-HPA cluster with the corresponding region of the chromosome of *E. coli* K-12 strain MG1655. The region located between min 92.8 and 0.1 of the chromosome of *E. coli* K-12 strain MG1655 (7) is compared with the flanking sequences of the 4-HPA cluster of strain W (Fig. 3). Arrows indicate the direction of gene transcription. |, identical nucleotides; Δ, sequence deletion; =, highly conserved sequences; *, region containing the rare tetramer CTAG in both strains (7); n, nucleotides.

and have a common structural motif of 12 transmembrane α-helices (spanners). In addition, most of these proteins contain a common motif, (R/K)XXX(R/K), between spanners 2 and 3 and spanners 8 and 9. The hydropathic plot of HpaX revealed the presence of six hydrophobic segments (spanners 1 to 6), followed by a large central hydrophilic loop; two hydrophobic segments (spanners 7 and 8); a short, highly hydrophilic loop; and four additional hydrophobic segments (spanners 9 to 12) (data not shown). A similar distribution of the 12 spanners has been published for the LacY permease of *E. coli* and for the bacterial tetracycline antiporters (29). Two (R/K)XXX (R/K) motifs are found in tandem (RHSDRRQERR) between spanners 8 and 9 of HpaX, creating the highly hydrophilic region mentioned above. However, between spanners 2 and 3 of HpaX we found the sequence VGARR, which, although it is not a consensus motif, still contains two positively charged residues. The functional significance of this motif has yet to be defined, but the loop regions appear to be essential for normal transport function (29). More recently, it has been shown that the *pcaRKF* gene cluster of *P. putida* PRS2000, involved in the metabolism of 4-hydroxybenzoate via β-ketoadipate, encodes a protein of 47 kDa (PcaK) which is required for 4-hydroxybenzoate transport and chemotaxis and that can be considered a new member of MFS (21). This protein is also similar to HpaX (50% similarity; 25% identity). Interestingly, only the fourth cluster of MFS contains members that act as transcriptional regulators (UhpC). The direct involvement of transport proteins in transcriptional regulation has been well documented (29), and it has been proposed that UhpC controls transcription by bringing inducers into the cell, by directly interacting with the transcriptional regulatory complex, or by a still uncharacterized mechanism (29). On the other hand, it has been shown that some transporters share the same locus with other regulatory proteins, e.g., *pth12345* (35), *uhpABCT* (32), and *pcaRKF* (21). This could be also the case for *hpaX*, which appears to form an operon with the *hpaA* transcriptional regulator. Recently, an inducible transport system for 4-HPA has been described for *K. pneumoniae* M5a1 (1).

Analysis of the flanking regions and localization of the 4-HPA cluster in the *E. coli* chromosome. The analysis of the regions flanking the 4-HPA cluster showed the presence of four ORFs (Fig. 2 and 3). The truncated ORF15 product was similar to the serine chemoreceptor encoded by the *tsr* gene of *E. coli* (2). This gene has been also found immediately downstream from the *hpcR* regulatory gene of the *E. coli* C 4-HPA pathway (43). The ORF12 product is homologous to the carbon starvation CstA protein of *E. coli* MC4100 (46) and to the first 567 amino acids of a putative Cst protein encoded by ORFf721 of *E. coli* MG1655 (accession number U14003). The ORF13 product is homologous to a protein of unknown function (the ORF2 product) that belongs to the same operon as CstA (46) and to the C-terminal part of the Cst-like protein of *E. coli* MG1655 (accession number U14003). It is worth noting that *cstA* is located 600 bp downstream from the iron-regulated *entCEBA*-P15 cluster, which encodes the enzymatic activities responsible for the synthesis and activation of 2,3-dihydroxybenzoic acid during the formation of the catechol siderophore enterobactin, a cyclic trimer of 2,3-dihydroxybenzoylserine (33). Whether CstA could play a role in the metabolism of aromatic compounds is unknown, but it has been demonstrated that carbon starvation proteins are induced by different aromatic pollutants (5). The ORF14 product is very similar to a protein of unknown function encoded by a gene, *yjiA*, located upstream from the *mrr* gene of different *E. coli* K-12 strains (7, 52) and to the P47K protein of *Pseudomonas chlororaphis* B23, which is involved in the metabolism of nitrile compounds (34).

Interestingly, the gene *cstA* maps at min 14 of the *E. coli* K-12 chromosome, whereas the *tsr* and *mrr* genes map at min 99 and 98.5, respectively (3). This finding suggests either that the acquisition of the 4-HPA cluster is concomitant with a rearrangement of these genes in the chromosome of *E. coli* W or that the genomic map of this strain is very different from that of K-12. In this sense, it has been shown very recently that the genes *tsr*, ORFf721 (*cst*-like), *yjiA*, and *mrr* map contiguously in the chromosome of *E. coli* K-12 strain MG1655 between min 92.8 and 0.1 (7) (Fig. 6). According to the new map of this region, the 4-HPA cluster might have been inserted between nucleotides 282281 and 282429, which correspond to nucleotides 73 and 11450 of Fig. 3 (Fig. 6). These nucleotides are located 57 and 101 bp downstream of the stop codons of the *hpaR* and *hpaC* genes, respectively. In addition, a comparison of the sequences of strains W and MG1655 revealed the existence of a short insertion just upstream of the ORF12 start codon (nucleotides 11604 to 11727 in Fig. 3) (Fig. 6). These observations suggest that the 4-HPA cluster has been recruited by *E. coli* W as a catabolic cassette. Since we have not detected the presence of a recombinase in the 4-HPA cassette, we can assume that this cluster is not a typical transposable element. These results open new insights into the evolutionary mechanisms that enhance bacterial adaptability.

Finally, it is important to notice that the 4-HPA cluster in *E. coli* W ATCC 11105 is located in the vicinity of the *pac* gene that codes for a penicillin G acylase, one of the most important industrial enzymes able to hydrolyze a great variety of amides and esters of 4-HPA and phenylacetic acids (36, 51). In contrast, Southern blot analysis carried out using the *pac* gene as a DNA probe indicated that *E. coli* C, B, and K-12 do not contain this gene (data not shown) and that only *K. citrophila*

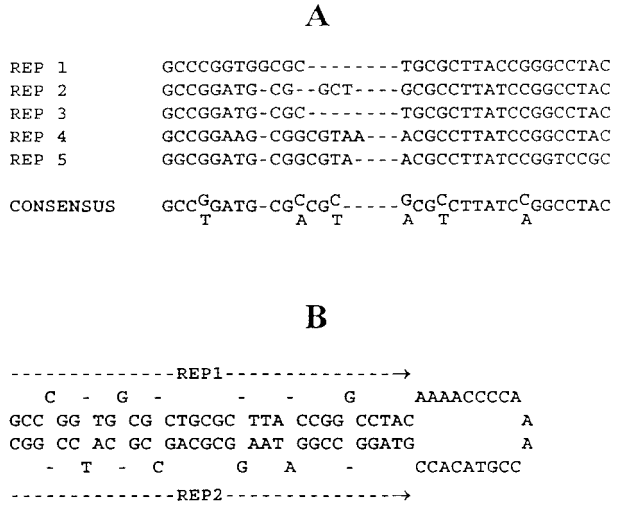

FIG. 7. Alignment of REP sequences. (A) REP sequences found in the intercistronic regions of the 4-HPA gene cluster aligned with the REP consensus sequence (49); (B) putative hairpin loop structure of the region containing the REP1 and REP2 sequences.

ATCC 21285 has both the 4-HPA cluster and a homologous *pac* gene (data not shown). A comparison of the *pac* flanking sequences with the *tsr-cst-yjiA-mrr* region of strain MG1655 revealed that the *pac* gene could have been inserted just downstream of the stop codon of the *yjiA* gene, which corresponds to the stop codon of ORF14 (nucleotide 15120 in Fig. 3) (Fig. 6). Hence, the presence of the *pac* gene in the proximity of the 4-HPA cluster of *E. coli* W suggests that it is a recent acquisition to improve the ability of this strain to metabolize a wider range of substrates (39, 51). These findings draw attention to selective forces that may favor clustering of physiologically interdependent genes.

Analysis of the intergenic sequences. The analysis of the intergenic regions revealed the presence of five repetitive extragenic palindromic (REP) sequences (49) (Fig. 3 and 7A). Two REP sequences are located in the largest intercistronic region of the *hpaGEDFHI* operon between the genes *hpaF* and *hpaH*. The other three REP sequences were found downstream of *hpaI*, the last gene of the operon. The function of the REP sequences is still controversial, but it has been postulated that they play a role in the control of gene expression and in mRNA stability (49). Nevertheless, we cannot rule out that in our case the REP3, REP4, or REP5 sequences might act as transcription termination signals, since no potential hairpin loops have been found at the 3′ end of the putative *hpaGEDFHI* operon. Whether the inverted intercistronic REP sequences, REP1 and REP2, which could form a hairpin loop of high free energy (−65.9 kcal/mol [ca. −276 kJ/mol]) (Fig. 7B), also act as a transcription termination signal, dividing the *hpaGEDFHI* operon in two polycistrons, is still an open question. A potential secondary structure similar to this one has also been found within the chromosomal *cat* cluster for the degradation of catechol in two strains of *P. putida* (24). So far we have been unable to establish a precise transcriptional map of the *meta*-cleavage operon of the 4-HPA pathway in *E. coli* W, a finding similar to that observed with the *meta*-cleavage operon of the TOL pathway of *P. putida* for the catabolism of benzoate and toluates (30). However, we have shown that two putative hairpin loops that might act as transcriptional terminators are located downstream of the *hpaA* gene and *hpaBC* operon (38) (Fig. 3).

The transcription initiation site of the 4-HPA *meta*-cleavage operon may be located at nucleotide 808 by comparison with the corresponding site of the *hpcECBDGH* operon of *E. coli* C, and it is preceded, as in this strain, by typical −35 and −10 promoter sequences as well as by a putative catabolite activator protein binding region (43) (Fig. 3). Genetic and biochemical analyses have revealed the existence of two alternative transcription initiation sites for *hpaA*, suggesting that a very complex mechanism regulates the expression of the 4-HPA pathway (unpublished data).

Expression of the 4-HPA metabolic pathway in the heterologous host. To ascertain that the cloned genes encoded the complete 4-HPA catabolic pathway, we constructed plasmid pAJ40 by inserting the 11-kb *Eco*RI fragment of pHCB3 into the *Eco*RI site of plasmid pHCR3, obtained after a *Hin*dIII deletion of plasmid pHCR1 (Fig. 2 and 8). Plasmid pAJ40 allowed *E. coli* K-12 strains DH1 and ET8000 to grow on 4-HPA as the sole carbon source. In addition, the 4-HPA cluster was cloned into the minitransposon delivery vector pUTmini-Tn*5*Km2 (Fig. 8) and used to stably introduce a single copy of this cluster in the chromosome of strain ET8000. This new strain, ET4025, as in the case of the original W strain, was able to grow on 4-HPA as the sole carbon and energy source, and when cultured in glycerol as a carbon source, it transformed phenol to catechol after induction by 4-HPA (Table 1). These results strongly suggest that the cloned 4-HPA cluster contains not only the catabolic genes necessary for the mineralization of 4-HPA but also the regulatory genes of the pathway. Moreover, they indicate that the regulation of this pathway in the new recombinant K-12 strain is similar to that in the parental W strain. As far as we know, only a similar transposable cassette containing the complete sequenced *bph* operon for the degradation of chlorinated biphenyls from *Pseudomonas* sp. strain LB400 has been engineered so far (14).

A second mobile catabolic segment carrying the 4-HPA hydroxylase operon (*hpaBC*) under the control of the P*trc* promoter was engineered with plasmid pCNB5 (Fig. 8). This cassette was then introduced into the chromosome of *P. putida* KT2442, a strain unable to metabolize phenol, since it does not contain the gene(s) required to hydroxylate this compound. In contrast, the new recombinant strain of *P. putida*, named KTH2, was able to grow on phenol as the sole carbon and energy source, showing the typical black phenotype observed in the *E. coli* strains that express the 4-HPA hydroxylase operon (data not shown). Phenol is converted by the constitutively produced 4-HPA hydroxylase to catechol, which is further mineralized through the chromosomally encoded *ortho*-cleavage pathway of *P. putida* (37). This result illustrates the utility of a broad-substrate-range catabolic enzyme, the 4-HPA hydroxylase of *E. coli*, to increase the ability of heterologous hosts for degrading new aromatic compounds and is an example of in vitro pathway evolution by vertical expansion of a natural existing catabolic route, e.g., the *ortho*-cleavage pathway of *P. putida* KT2442.

Summarizing, the results presented here represent an original example of a completely sequenced, chromosomally encoded aromatic catabolic pathway that has been engineered as transposable cassettes that can be easily and stably inserted into the chromosome of a variety of gram-negative bacteria. Only a few aromatic catabolic pathways have been completely sequenced so far (23, 26, 47, 53), but only the one presented here proceeds exclusively via a dehydrogenase/decarboxylase branch. Moreover, an overall analysis of this cluster reveals that six proteins, HpaR, HpaB, HpaC, HpaD, HpaF, and HpaI, do not present any similarity to functionally analogous proteins of other degradative pathways, suggesting that they

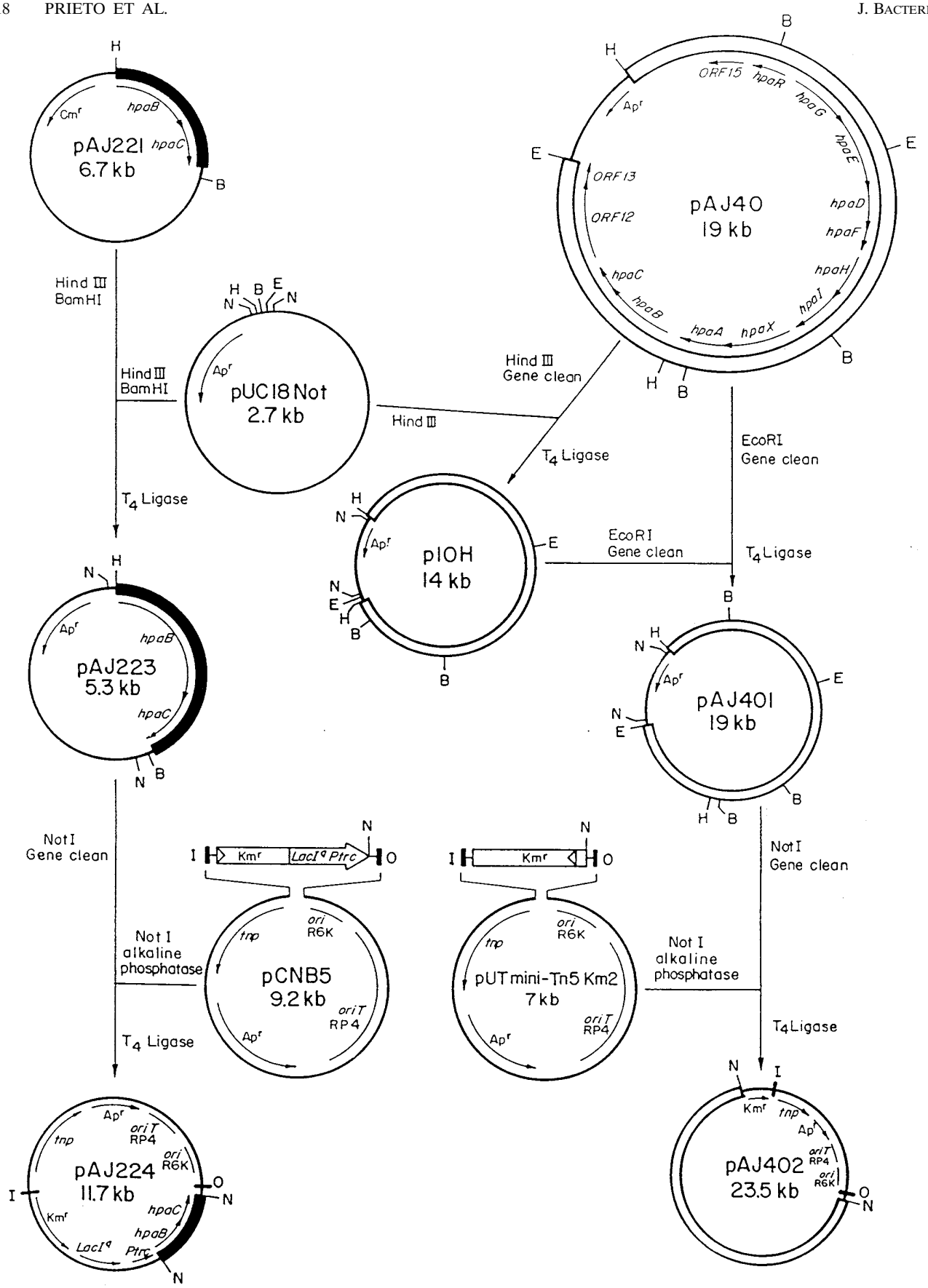

FIG. 8. Construction of plasmids and minitransposons carrying the 4-HPA gene cluster. Abbreviations: Ap^r, ampicillin resistance; B, *Bam*HI; Cm^r, chloramphenicol resistance; E, *Eco*RI; H, *Hin*dIII; Km^r, kanamycin resistance; N, *Not*I; P*trc*, *trc* promoter; tnp, Tn*5* transposase. The 19-bp I and O Tn*5* ends, *ori*TRP4, and *ori*R6K are indicated.

TABLE 1. Induction of 4-HPA hydroxylase in different *E. coli* strains

Strain	Catechol (nmol/ml)[a]	
	−4-HPA	+4-HPA
K-12 strain ET8000	<1	<1
K-12 strain ET4025	<1	329
W ATCC 11105	<1	695

[a] The induction of 4-HPA hydroxylase activity by 4-HPA was determined by measuring the in vivo transformation of phenol into catechol. *E. coli* cells were cultured overnight at 30°C in minimal medium containing 20 mM glycerol and 1 mM phenol in the presence or absence of 1 mM 4-HPA. Catechol in culture supernatants was determined as previously described (38).

evolved from a different origin. The characterization of the 4-HPA cluster provides novel catabolic tools and opens new possibilities for the use of *E. coli* as a host for biodegradation purposes. In fact, it has been argued that, in some instances, recombinant *E. coli* strains containing catabolic operons equipped with efficient foreign expression signals and which can be grown to very high cell densities on simple carbon sources will provide the best performance (50). The use of a well-characterized host such as *E. coli* might also minimize the regulatory concerns about releasing genetically modified organisms into the environment (50).

ACKNOWLEDGMENTS

We thank P. García, E. García, and R. López for helpful comments. We are indebted to K. Timmis for his generous support and encouragement during the stay of M. A. Prieto at the GBF. D. McKay is gratefully acknowledged for correcting the English version of the manuscript. We also thank E. Cano, M. Carrasco, and A. Hurtado for excellent technical assistance. We are grateful to V. de Lorenzo, J. L. Ramos, M. Vicente, A. Gibello, and A. Garrido-Pertierra for some plasmids and strains used in this work. We thank E. Moore, A. Arnscheidt, and A. Krüger for valuable help with the sequencing.

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

REFERENCES

1. **Allende, J. L., A. Gibello, M. Martín, and A. Garrido-Pertierra.** 1992. Transport of 4-hydroxyphenylacetic acid in *Klebsiella pneumoniae*. Arch. Biochem. Biophys. **292:**583–588.
2. **Ames, P., and J. Parkinson.** 1994. Constitutively signaling fragments of Tsr, the *Escherichia coli* serine chemoreceptor. J. Bacteriol. **176:**6340–6348.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
4. **Barbero, J. L., J. M. Buesa, G. G. Buitrago, E. Méndez, A. Pérez-Aranda, and J. L. García.** 1986. Complete nucleotide sequence of the penicillin acylase from *Kluyvera citrophila*. Gene **49:**69–80.
5. **Blom, A., W. Harder, and A. Matin.** 1992. Unique and overlapping pollutant stress proteins of *Escherichia coli*. Appl. Environ. Microbiol. **58:**331–334.
6. **Burkholder, P. R.** 1951. Determination of a vitamin B_{12} with a mutant strain of *Escherichia coli*. Science **114:**459–460.
7. **Burland, V., G. Plunkett II, H. J. Sofia, D. L. Daniels, and F. R. Blattner.** 1995. Analysis of the *Escherichia coli* genome. VI. DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res. **23:**2105–2119.
8. **Burlingame, R., and P. J. Chapman.** 1983. Catabolism of phenylpropionic acid and its 3-hydroxy derivate by *Escherichia coli*. J. Bacteriol. **155:**113–121.
9. **Clancey, C. J., S.-C. Chang, and W. Dowhan.** 1993. Cloning of a gene (*psd1*) encoding phosphatidylserine decarboxylase from *Saccharomyces cerevisiae* by complementation of an *Escherichia coli* mutant. J. Biol. Chem. **268:**24580–24590.
10. **Cooper, R. A., and M. A. Skinner.** 1980. Catabolism of 3- and 4-hydroxyphenylacetate by the 3,4-dihydroxyphenylacetate pathway in *Escherichia coli*. J. Bacteriol. **143:**302–306.
11. **Creighton, T. E.** 1970. *N*-(5′-Phosphoribosyl)anthranilate isomerase/indol-3-ylglycerol phosphate synthetase of tryptophan biosynthesis. Relationship between the two activities of the enzyme from *Escherichia coli*. Biochem. J. **120:**699–707.
12. **de Lorenzo, V., L. Eltis, B. Kessler, and K. N. Timmis.** 1993. Analysis of *Pseudomonas* gene products using *lacI*^{q}/*Ptrp-lac* plasmids and transposons that confer conditional phenotypes. Gene **123:**17–24.
13. **de Lorenzo, V., M. Herrero, U. Jakubzik, and K. N. Timmis.** 1990. Mini-Tn*5* transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. J. Bacteriol. **172:**6568–6572.
14. **Dowling, D. N., R. Pipke, and D. F. Dwyer.** 1993. A DNA module encoding *bph* genes for the degradation of polychlorinated biphenyls (PCBs). FEMS Microbiol. Lett. **113:**149–154.
15. **Eiglmeier, K., W. Boos, and S. T. Cole.** 1987. Nucleotide sequence and transcriptional startpoint of the *glpT* gene of *Escherichia coli*: extensive sequence homology of the glycerol-3-phosphate transport protein with components of the hexose 6-phosphate transport system. Mol. Microbiol. **1:** 251–258.
16. **Fawcett, T., A. Garrido-Pertierra, and R. A. Cooper.** 1989. 5-Carboxymethyl-2-hydroxymuconic semialdehyde dehydrogenases of *Escherichia coli* C and *Klebsiella pneumoniae* M5a1 show very high N-terminal sequence homology. FEMS Microbiol. Lett. **57:**307–312.
17. **Ferrer, E., and R. A. Cooper.** 1988. Studies with a cloned *Escherichia coli* C 2-oxo-hept-3-ene-1,7-dioate hydratase gene. FEMS Microbiol. Lett. **52:**155–160.
18. **Garrido-Pertierra, A., and R. A. Cooper.** 1981. Identification and purification of distinct isomerase and decarboxylase enzymes involved in the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli*. Eur. J. Biochem. **117:** 581–584.
19. **Gibello, A., E. Ferrer, M. Martín, and A. Garrido-Pertierra.** 1994. 3,4-Dihydroxyphenylacetate 2,3-dioxygenase from *Klebsiella pneumoniae*, a Mg^{2+}-containing dioxygenase involved in aromatic catabolism. Biochem. J. **301:**145–150.
20. **Harayama, S., M. Rekik, K.-L. Ngai, and L. N. Ornston.** 1989. Physically associated enzymes produce and metabolize 2-hydroxy-2,4-dienoate, a chemically unstable intermediate formed in catechol metabolism via *meta* cleavage in *Pseudomonas putida*. J. Bacteriol. **171:**6251–6258.
21. **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.
22. **Herrero, M., V. de Lorenzo, and K. N. Timmis.** 1990. Transposon vector containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign DNA in gram-negative bacteria. J. Bacteriol. **172:**6557–6567.
23. **Horn, J. M., S. Harayama, and K. N. Timmis.** 1991. DNA sequence determination of the TOL plasmid *xylGFH* genes of *Pseudomonas putida*: implications for the evolution of aromatic catabolism. Mol. Microbiol. **5:**2459–2474.
24. **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.
25. **Jenkins, J., and R. A. Cooper.** 1988. Molecular cloning, expression, and analysis of the genes of the homoprotocatechuate catabolic pathway of *Escherichia coli* C. J. Bacteriol. **170:**5317–5324.
26. **Kikuchi, Y., Y. Yasukochi, Y. Nagata, M. Fukuda, and M. Takagi.** 1994. Nucleotide sequence and functional analysis of the *meta*-cleavage pathway involved in the biphenyl and polychlorinated biphenyl degradation in *Pseudomonas* sp. strain KKS102. J. Bacteriol. **176:**4269–4276.
27. **Komine, Y., and H. Inokuchi.** 1991. Precise mapping of the *rnpB* gene encoding the RNA component of RNase P in *Escherichia coli* K-12. J. Bacteriol. **173:**1813–1816.
28. **MacNeil, D.** 1981. General method, using Mu-Mud1 dilysogens, to determine the direction of transcription of and generate deletions in the *glnA* region of *Escherichia coli*. J. Bacteriol. **146:**260–268.
29. **Marger, M. D., and M. H. Saier, Jr.** 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem. **18:**13–20.
30. **Marqués, S., J. L. Ramos, and K. N. Timmis.** 1993. Analysis of the mRNA structure of the *Pseudomonas putida* TOL *meta* fission pathway operon around the transcription initiation point, the *xylTE* and the *xylFJ* regions. Biochim. Biophys. Acta **1216:**227–236.
31. **Martín, M., A. Gibello, J. Fernández, E. Ferrer, and A. Garrido-Pertierra.** 1991. Catabolism of 3- and 4-hydroxyphenylacetic acid by *Klebsiella pneumoniae*. J. Gen. Microbiol. **132:**621–628.
32. **Merkel, T. J., D. M. Nelson, C. L. Brauer, and R. J. Kadner.** 1992. Promoter elements required for positive control of transcription of the *Escherichia coli uhpT* gene. J. Bacteriol. **174:**2763–2770.
33. **Nahlik, M. S., T. P. Fleming, and M. A. McIntosh.** 1987. Cluster of genes controlling synthesis and activation of 2,3-dihydroxybenzoic acid in production of enterobactin in *Escherichia coli*. J. Bacteriol. **169:**4163–4170.
34. **Nishiyama, M., S. Horinouchi, M. Kobayashi, T. Nagasawa, H. Yamada, and T. Beppu.** 1991. Cloning and characterization of genes responsible for metabolism of nitrile compounds from *Pseudomonas chlororaphis* B23. J. Bacteriol. **173:**2465–2472.
35. **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.

36. **Oh, S.-J., Y.-C. Kim, Y.-W. Park, S.-Y. Min, I.-S. Kim, and H.-S. Kang.** 1987. Complete nucleotide sequence of the penicillin G acylase gene and the flanking regions, and its expression in *Escherichia coli*. Gene **56:**87–97.
37. **Parsek, M. R., D. L. Shinabarger, R. K. Rothmel, and A. M. Chakrabarty.** 1992. Roles of CatR and *cis*,*cis*-muconate in activation of the *catBC* operon, which is involved in benzoate degradation in *Pseudomonas putida*. J. Bacteriol. **174:**7798–7806.
38. **Prieto, M. A., and J. L. García.** 1994. Molecular characterization of 4-hydroxyphenylacetate 3-hydroxylase of *Escherichia coli*. A two-protein component enzyme. J. Biol. Chem. **269:**22823–22829.
39. **Prieto, M. A., A. Pérez-Aranda, and J. L. García.** 1993. Characterization of an *Escherichia coli* aromatic hydroxylase with a broad substrate range. J. Bacteriol. **175:**2162–2167.
40. **Roper, D. I., and R. A. Cooper.** 1990. Purification, some properties and nucleotide sequence of 5-carboxymethyl-2-hydroxymuconate isomerase of *Escherichia coli* C. FEBS Lett. **266:**63–66.
41. **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.
42. **Roper, D. I., and R. A. Cooper.** 1993. Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of *Escherichia coli* C. Eur. J. Biochem. **217:**575–580.
43. **Roper, D. I., T. Fawcett, and R. A. Cooper.** 1993. The *Escherichia coli* C homoprotocatechuate degradative operon: *hpc* gene order, direction of transcription and control of expression. Mol. Gen. Genet. **237:**241–250.
44. **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.
45. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. **Schultz, J. E., and A. Matin.** 1991. Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*. J. Mol. Biol. **218:**129–140.
47. **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.
48. **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.
49. **Stern, M. J., G. Ferro-Luzzi, N. J. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37:**1015–1026.
50. **Timmis, K. N., R. J. Steffan, and R. Unterman.** 1994. Designing microorganisms for the treatment of toxic wastes. Annu. Rev. Microbiol. **48:**525–557.
51. **Valle, F., P. Balbás, E. Merino, and F. Bolivar.** 1991. The role of penicillin amidases in nature and industry. Trends Biochem. **16:**36–40.
52. **Waite-Rees, P. A., C. J. Keating, L. S. Moran, B. E. Slatko, L. J. Hornstra, and J. S. Benner.** 1991. Characterization and expression of the *Escherichia coli* Mrr restriction system. J. Bacteriol. **173:**5207–5219.
53. **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.
54. **Wilson, R., et al.** 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature (London) **368:**32–38.
55. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phaged cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.