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,i and JOSÉ L. GARCÍA1,ii

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, Germany2

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 of 11 genes, i.e., 8 enzyme-encoding genes distributed in two putative operons, hpaBC (4-HPA hydroxylase operon) and hpaGDEFHI (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 ctsA (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 biodegradable 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 un-known. 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.

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 deoxyx nucleotide terminators and Taq DNA polymerase were from Applied Biosystems Inc. [α-35S]dCTP and [α-32P]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 S17-1pir (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-Tn5Km2 (13); pUC18Not (22); pCNBS (12); pAG464, which contains the gene encoding HPC 2,5-dioxygenase from K. pneumoniae M5a1 (provided by A. Garrido-Pertierra) (31); pA19, which contains the operon encoding 4-HPA hydroxylase and the pac gene encoding penicillin G acylase (39); and pA2122, 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. Restriction endonucleases, alkaline phosphatase, and T4 DNA ligase were used according to

---

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

ii 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: CIBIL16@PINARI.ES.CSIC.ES.
Carpeta carrying plasmid pAJ19, containing the 4-HPA hydroxylase gene, the typical black phenotype previously observed with cells of E. coli containing overlapping 6-kb HPC2,3-dioxygenase gene. Two plasmids, pHCB1 and pHCB3, in pUC18 were screened for the presence of the homologous operon in pHCB1 (Fig. 2) showed a truncated gene corresponding to the hpaR gene of the HPC operon of E. coli C. Using the 2.7-kb BamHI-EcoRI fragment of pHCB1 as a DNA probe to screen the EcoRI library of E. coli W ATCC 11105, we isolated plasmid pHCR1, which contained a 6-kb EcoRI 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 (hpaGDEFH)I, 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 XyI 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 HpaG, 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 HpaG hydratase (HpaH in E. coli W), and we propose that the transformation of HHDD into 2,4-dihydroxy-hept-2-ene-1,7-dioic acid should be catalyzed by HpaH.
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 that presents all the motifs that characterize the aldehyde dehydrogenase superfamily (23). The HpaD protein shows a 98.6% identity to its putative counterpart, 5-carboxybenzyl-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 *hpaF* 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 *hpaH* encodes 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 that the common origin for these enzymes (see above).

Hence, we suggest that these enzymes constitute a new family of dioxygenases.

The protein encoded by *hpaF* is identical to the 5-carboxybenzyl-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 *hpaI* 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 nucleotide 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 hpaGEDFH operon might be initiated. Stars indicate the stop codons.
The putative regulatory 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 *phi1* gene of the phthalate degradative pathway of *P. putida* (35) (Fig. 5). The *Phi1* protein was suggested to function either as a positive regulator for the expression of the *phi* 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 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 transporter (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.
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 hydrophilic 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 (RHSDDRRQERR) between spanners 8 and 9 of HpaX, creating the highly hydrophobic region mentioned above. However, between spanners 2 and 3 of HpaX we found the sequence VGGARR, 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-ketoacidipate, 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., phl12345 (35), ubpABCT (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 hprR 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 ORF721 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 ctsA 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-dihydroxybenzoic acid (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 (3). 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 ctsA 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, ORF721 (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 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. 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. $\Delta$, identical nucleotides; $\Delta$, sequence deletion; $\rightarrow$, highly conserved sequences; $\leftarrow$, region containing the rare tetramer CTAG in both strains (7); n, nucleotides.
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 hpeECDBGH 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 pA340 by inserting the 11-kb EcoRI fragment of pHC3 into the EcoRI site of plasmid pHCR3, obtained after a HindIII deletion of plasmid pHC1 (Fig. 2 and 8). Plasmid pA340 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 mini transposon delivery vector pUTmini-Tn5Km2 (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 Ppre 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

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 draws 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 hpaGEDFH operon between the genes hpaF and hpaI. 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 loop structures 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 catalolism of benzoate and toluates (30). However, we have shown that two putative hairpin loops that might act as transcription terminators are located downstream of the hpaA gene and hpaBC operon (38) (Fig. 3).
FIG. 8. Construction of plasmids and minitransposons carrying the 4-HPA gene cluster. Abbreviations: Ap, ampicillin resistance; B, BamHI; Cm, chloramphenicol resistance; E, EcoRI; H, HindIII; Km, kanamycin resistance; N, NcoI; Pir, ac promoter; tsp, Tn5 transposase. The 19-bp I and O Tn5 ends, oriTRP4, and oriR6K are indicated.
TABLE 1. Induction of 4-HPA hydroxylase in different E. coli strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Caetehol (nmol/ml)*</th>
<th>-4-HPA</th>
<th>+4-HPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-12 strain ET8000</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>329</td>
</tr>
<tr>
<td>K-12 strain ET4025</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>695</td>
</tr>
<tr>
<td>W ATCC 11105</td>
<td>&lt;1</td>
<td>65</td>
<td></td>
</tr>
</tbody>
</table>

*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).


