Catabolism of Phenylacetic Acid in Escherichia coli

CHARACTERIZATION OF A NEW AEROBIC HYBRID PATHWAY*

(Received for publication, May 8, 1998, and in revised form, July 16, 1998)

Abel Ferrández‡, Baltasar Miñambres†, Belén García†, Elías R. Olivera†, José L. García‡, and Eduardo Díaz‡‡

From the ‡Department of Molecular Microbiology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, 28006 Madrid and the §Department of Biochemistry and Molecular Biology, Facultad de Veterinaria, Universidad de León, 24074 León, Spain

The paa cluster of Escherichia coli W involved in the aerobic catabolism of phenylacetic acid (PA) has been cloned and sequenced. It was shown to map at min 31.0 of the chromosome at the right end of the mao region responsible for the transformation of 2-phenylethylamine into PA. The 14 paa genes are organized in three transcription units: paaZ and paaABCDEFGHIJK, encoding catabolic genes; and paaXY, containing the paaX regulatory gene. The paaK gene codes for a phenylacetyl-CoA ligase that catalyzes the activation of PA to phenylacetyl-CoA (PA-CoA). The paaABCDEFGE gene products, which may constitute a multicomponent oxygenase, are involved in PA-CoA hydroxylation. The PaaZ protein appears to catalyze the third enzymatic step, with the paaFGHIJK gene products, which show significant similarity to fatty acid β-oxidation enzymes, likely involved in further mineralization to Krebs cycle intermediates. Three promoters, Pz, Pa and Px, drive the expression of genes paaZ, paaABCDEFGHIJK, and paaX, respectively, have been identified. The Pa promoter is negatively controlled by the paaX gene product. As PA-CoA is the true inducer, PaaX becomes the first regulator of an aromatic catabolic pathway that responds to a CoA derivative. The aerobic catabolism of PA in E. coli represents a novel hybrid pathway that could be a widespread way of PA catabolism in bacteria.

Escherichia coli living in the animal gut encounters aromatic compounds such as phenylacetic acid (PA),1 phenylpropionic acid, and their hydroxylated derivatives, as a result of the action of intestinal microflora on plant constituents, the amino acids phenylalanine and tyrosine, fatty acids with a terminal phenyl substituent, and some of their metabolites (1–3). The aerobic catabolism of these aromatic compounds by E. coli could occur close to the epithelial cells in the guts of warm-blooded animals, as well as in soil, sediment, and water once E. coli is excreted from its intestinal residence (4). The ability of E. coli to mineralize 3- and 4-hydroxyphenylacetic acids (5), 3-phenylpropionic, 3-(3-hydroxyphenylpropionic), and 3-hydroxycinnamic acids (2, 6), and phenylacetic acid (2, 7) has been reported previously. Recently, the molecular characterization of these catabolic pathways, with the only exception of that for PA degradation, has been carried out (1, 8–11), demonstrating that E. coli is endowed with its own set of genes and enzymes for the catabolism of aromatic compounds, and that they are similar to those of other microorganisms more relevant in the environment such as bacteria of the genus Pseudomonas.

Although PA is a common source of carbon and energy for a wide variety of microorganisms, the bacterial catabolism of this natural aromatic compound is still poorly understood (12, 13). Earlier reports suggested that aerobic PA catabolism implicated the typical initial attack by hydroxylation of the aromatic ring with the formation of the corresponding 2,5- or 3,4-dihydroxyphenylacetate as intermediates (13). However, much of this evidence was circumstantial, and none of the typical aerobic routes that could explain PA degradation were responsible of this catabolism in different PA-degrading bacteria (13, 14). According to these data, it has been recently shown that Pseudomonas putida U mineralizes PA aerobically through a novel catabolic pathway, which does not follow the conventional routes for the aerobic catabolism of aromatic compounds and whose first step is the activation of PA to phenylacetyletylenzime A (PA-CoA) by the action of a PA-CoA ligase (12, 15). In this sense, the participation of a PA-CoA ligase in the aerobic catabolism of PA has been also inferred from its specific induction during growth on PA of different bacterial strains (13, 16).

Here we present the cloning, genetic characterization, mechanism of regulation, and a partial biochemical characterization of the PA biodegradation pathway from E. coli W. This work reveals that the PA degradation in E. coli follows an unusual route for the aerobic catabolism of aromatic compounds, which involves CoA derivatives. With the molecular characterization of the paa-encoded pathway, all aromatic catabolic routes so far reported in E. coli are now described at the molecular level.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Plasmids, and Growth Conditions—The E. coli strains used were E. coli W ATCC11105 (17); E. coli W14 (18); E. coli C (8); and the E. coli K-12 strains MV1190 (19), C600 (20), TG1 (20), ET8000 (21), W3110 (20), HB101 (20), DH1 (20), CC118 (1), DH5α (20), JM109 (20), and S17–1λpir (22). E. coli W14Rif (this work) is a spontaneous rifampicin-resistant mutant from E. coli W14; E. coli AF141 (this work) is a lacZ− mutant obtained from the E. coli
The paa Cluster of E. coli

25795

W14 Rif strain by N-methyl-N′-nitro-N-nitrosoguanidine treatment as described by Miller (23). For cloning and expression purposes we have used two chloramphenicol-resistant low copy number cloning vectors, plasmids pUCK01 (1) and pSJ19Not (like pUCK01 but containing the SphI restriction site (26)) as well as pUC18, pUC19 (26) and pUC18Not (2) vectors. The SphI vector is a pUJ9 derivative (22) bearing a 85-bp Smal BamHI fragment that contains the 5′-leader region of IS10 transposase (19). To integrate the lacZ fusions into the chromosome, plasmid pUTmini-Tn5 Km2 was used (22). Unless otherwise stated, bacteria were grown in Luria-Bertani (LB) medium (20) at 37 °C. When used as carbon sources, PA and/or glycerol were supplied at 5 and 20 mM, respectively, to M63 minimal medium (23) containing the corresponding necessary nutritional supplements, and the cultures were incubated at 30 °C. Where appropriate, antibiotics were added at the following concentrations: ampicillin (100 μg/ml), chloramphenicol (35 μg/ml), kanamycin (50 μg/ml), and rifampicin (50 μg/ml).

DNA Manipulations and Sequencing—Plasmid DNA was prepared by the rapid alkaline lysis method (20). Transformation of E. coli was carried out using the RhCl method (20). DNA manipulations and other molecular biology techniques were essentially as described (20). DNA fragments were purified by using low melting point agarose. Southern blot analyses were performed as previously reported (20), using as probes DNA fragments labeled with [32P]dCTP by the random primer method (26). Nucleotide sequence analyses were determined directly from plasmids by using the dideoxy chain termination method (25). Standard protocols of the manufacturer for T7 DNA polymerase-initiated cycle sequencing reactions with fluorescein-labeled dideoxynucleotide terminators (Applied Biosystems Inc.) were used. The sequencing reactions were analyzed using a model 377 automated DNA sequencer (Applied Biosystems Inc.). Sequences were extended by designing primers based on the previously determined sequence.

Sequence Data Analyses—Nucleotide sequence analyses were done with the DNA-Strider 1.2 program. Amino acid sequences were analyzed with the Wisconsin Sequence 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 made by using the BLASTP, BLASTN, and BLASTX programs (26) via the National Institute for Biotechnology Information server. Pairwise and multiple protein sequence alignments were made with ALIGN (27) and CLUSTAL W (28) programs, respectively, at the Bayreuth Protein Analysis Toolbox (http://www.bio.uni-giessen.de). A new sequence data base collection ECDC (29) was accessed via the Internet. 2

Cloning and Expression of the Pa, Pz, and Pp Promoter Regions—The Pa16 promoter region was PCR-amplified from plasmid pAAD by using primers Pp5 (5′-GGGTTGAATACAAACGGCTACTG-3′), which corresponds to nucleotides 2402–2420 in Fig. 2) and Pa5–1 (5′-CAATCTGGAAATCCGATGAGAGG-3′; the sequence corresponds to nucleotides 2292–2267 in Fig. 2). The resulting 455-bp DNA fragment was digested with KpnI and SpI and cloned as a 424-bp fragment into the double-digested KpnI–SpI pS3J vector to form plasmid pAAP1 (Fig. 5). The Pa16 promoter region was PCR-amplified by using primers PAPs1–2 (5′-CGGGCATCCAGTCCTGTGGCTCG-3′), the sequence corresponds to nucleotides 12456–12436 in Fig. 2; engineered BamHI site is double-underlined). The resulting 554-bp DNA fragment was digested with SphI and BamHI, and cloned as a 214-bp fragment into the double-digested SphI–BamHI pS3J vector to form plasmid pAAPX (Fig. 5). To integrate into the genome of E. coli, plasmids pAAP1 and pAAPX were treated with Tn1000 transposition mutagenesis (see Cautionary Note). To select for conjugative transfer of this plasmid from the recA F′ donor strain E. coli MG1063 to the E. coli recipient strain E. coli CC118. After mating, both parents were grown at 37 °C, without shaking, to an optical density of about 0.5. Two milliliters of donor cells and one milliliter of recipient cells were mixed and incubated at 37 °C without shaking for 2 h. After addition of 12 ml of LB medium and further incubation with vigorous shaking for 3 h, exconjugants were selected on LB medium containing rifampicin and chloramphenicol. The pAAD derivatives bearing Tn1000 insertions within the paa genes were further analyzed. The primer Tn5900 (5′-AAAAGGGGACTAGAGAGTTATCAAAATAGAGTGCG-3′) that hybridized with the 5′ terminus of transposon Tn1000 was used to sequence the insertion sites.

Construction of E. coli AF1411 and AF1412 Strains—By means of RP-4 mediated mobilization, the plasmids pAAP1T and pAAP2T, which contain mini-Tn5 hybrid transposons expressing Pa-lacZ fusions (Fig. 5), were transferred from E. coli S17–λpir to E. coli AF141. The recombinants were selected on LB plates containing 50 μg/ml chloramphenicol (35 g/l). EcoRI digestion of plasmids pCK01 (1) and pSJ19Not (like pCK01 but containing the SphI restriction site (26)) containing the corresponding necessary nutritional supplements, and the cultures were incubated at 30 °C. Where appropriate, antibiotics were added at the following concentrations: ampicillin (100 μg/ml), chloramphenicol (35 μg/ml), kanamycin (50 μg/ml), and rifampicin (50 μg/ml).

Cloning and Expression of the Pa, Pz, and Pp Promoter Regions—The Pa16 promoter region was PCR-amplified from plasmid pAAD by using primers Pp5 (5′-GGGTTGAATACAAACGGCTACTG-3′); the sequence corresponds to nucleotides 2402–2420 in Fig. 2) and Pa5–1 (5′-CAATCTGGAAATCCGATGAGAGG-3′; the sequence corresponds to nucleotides 2292–2267 in Fig. 2). The resulting 455-bp DNA fragment was digested with KpnI and SpI and cloned as a 424-bp fragment into the double-digested KpnI–SpI pS3J vector to form plasmid pAAP1 (Fig. 5). The Pa16 promoter region was PCR-amplified by using primers PAPs1–2 (5′-CGGGCATCCAGTCCTGTGGCTCG-3′), the sequence corresponds to nucleotides 12456–12436 in Fig. 2; engineered BamHI site is double-underlined). The resulting 554-bp DNA fragment was digested with SphI and BamHI, and cloned as a 214-bp fragment into the double-digested SphI–BamHI pS3J vector to form plasmid pAAPX (Fig. 5). The Pa promoter region was PCR-amplified by using primers Pp5,5f2 (5′-CGGGTATGAGTAGCTTCGAGG-3′; the sequence corresponds to nucleotides 11910–11928 in Fig. 2) and Pp5–2 (5′-GGGATCCGGAATATCACATAGAGAGG-3′) where the sequence corre-
French press (Aminco Corp.) operated at a pressure of 20,000 p.s.i. The cell debris was removed by centrifugation at 26,000 x g for 30 min. The clear supernatant fluid was carefully decanted and used as crude extract. Protein concentration was determined by the method of Bradford (23) using bovine serum albumin as standard. Phenylacetyl-CoA ligase was assayed as described previously (33). One unit of enzyme activity is defined as the catalytic activity leading to the formation of 1 nmol of phenylacetylhydroxamate in 1 min at 37 °C. The intracellular 14C-labeled metabolites accumulated in resting cell supernatants was extracted from culture supernatants with an equal volume of ethyl acetate and dried with sodium sulfate. Samples were derivatized with N, O-bis(trimethylsilyl)trifluoroacetamide and subjected to gas chromatography-mass spectrometry analysis as described elsewhere (34).

RESULTS

Identification of the paa Genes for the Catabolism of PA—We had recently reported the isolation of an E. coli W mutant, strain W14, unable to grow on PA and 2-phenylethylamine as the sole carbon and energy source (18). A recombinant plasmid, pFA2, which contains a 33.3-kb BamHI DNA insert from the chromosome of the wild-type E. coli W strain (Fig. 1A), had been also constructed and was able to confer to E. coli W14 the ability to grow on either PA or phenylethylamine as the sole carbon source (18). When the HindIII-digested total DNA from E. coli W14 and E. coli C, a strain also unable to grow on PA as the sole carbon source (2), was analyzed by Southern blot using the 33.3-kb BamHI fragment of pFA2 as probe, no hybridization bands were observed (data not shown). These results indicated that E. coli strains W14 and C have a deletion encompassing at least the 33.3-kb DNA fragment cloned in pFA2, and therefore it is likely that they lack the genes involved in PA catabolism (paa genes). Moreover, we have observed that, whereas the E. coli K-12 strains MV1190, C600, TG1, ET8000, W3110, and MG1655 grew on PA, the K-12 strains HB101, DH1, CC118, DH5a, and JM109 did not grow on this aromatic compound. A Southern blot analysis of the HindIII-digested total DNA from E. coli K-12 PA− (W3110, MG1655, and C600) and PA− (DH5a, HB101, and DH1) strains revealed the same pattern of hybridization bands, thus suggesting that the PA− phenotype in E. coli K-12, in contrast to that in E. coli W14 and E. coli C, could be produced by point mutations or small deletions or insertions. Interestingly, all E. coli PA− strains were able to grow on this aromatic compound when harboring plasmid pFA2, and the loss of this plasmid after several rounds of cultivation of the recombinant strains in the absence of selective pressure was also accompanied by the loss of the PA− phenotype.

FIG. 1. Genetic organization of the paa genes responsible for the catabolism of PA in E. coli, and locations of Tn1000 insertions. A, physical and genetic map of the chromosomal region containing the paa genes, and different cloned DNA fragments. Locations of the genes are shown relative to those of some relevant restriction sites. Arrows indicate the direction of gene transcription. Enlarged is shown the paa cluster (black box). In plasmid pFA2 the nucleotide sequence of the vector (pBR322) is represented by the dotted box. Plasmids containing different subcloned DNA fragments (the sizes of the inserts are shown on the top) are indicated with continuous (pUC derivatives) or discontinuous (pCK01 derivatives) lines. The open arrowheads represent the Plac promoter. B, locations of transposon Tn1000 in pAA-derived mutant plasmids. The 15.4-kb DNA insert in plasmid pAAD is represented with a striped box. The γ- and δ-γ orientations of Tn1000 insertions are shown by > and < symbols, respectively. Insertions which affect (filled symbols) or do not affect (empty symbols) catabolic functions are indicated. Δ, a truncated gene. Cm′, the gene that confers chloramphenicol resistance. Restriction sites are: B, BglII; Ba, BamHI; E, EcoRI; Ev, EcoRV; H, HindIII; N, NotI; P, PstI; Sm, Smal.
The paa Cluster of E. coli

As several aromatic catabolic pathways are encoded by plasmids and E. coli W was shown to host cryptic plasmids (35), it was checked whether the paa genes were also located in a plasmid. When a pulse field electrophoresis of unrestricted total DNA from E. coli W was analyzed by Southern blot using the 33.3-kb DNA fragment as probe, hybridization signals were only found in the sample wells (data not shown), thus indicating a chromosomal location (36) of the paa genes.

To localize the paa genes within the cloned 33.3-kb DNA fragment, its physical map was determined and different subclones were constructed (Fig. 1A) and checked for their ability to restore the growth of E. coli W14 on PA-containing minimal medium. Interestingly, although plasmids pFCP142 and pFCH112 (Fig. 1A) did not confer to E. coli W14 the ability to grow on PA, plasmid pAAD restored the growth of strain W14 on this aromatic compound, thus locating the paa genes in a 15.4-kb DNA fragment at the right end of the previously identified mao region (Fig. 1A) responsible of the transformation of 2-phenylethylamine into PA (18, 37, 38).

**Sequencing and Gene Arrangement of the paa Cluster**—To genetically characterize the PA catabolic pathway of E. coli W, the 15.4-kb insert of plasmid pAAD (Fig. 1A) was sequenced. The nucleotide sequence of a 14,328-bp region of this insert is shown in Fig. 2. Computer analysis of this sequence revealed the presence of 14 ORFs, all of which are transcribed in the same direction with the sole exception of paaZ (Figs. 1 and 2). Data bases were searched for similar proteins to the paa genes we found inverted repeat potential –10 and –35 boxes are shown in italics. Inverted repeats in promoter region (Fig. 2), suggesting that the most common mechanism of translational coupling (39) may occur. Downstream of the paaZ, paaK, and paaY genes we found inverted repeat sequences (Fig. 2) predicted to form hairpin loops with ΔG values of –13.4, –25.3, and –15.7 kcal/mol, respectively, which could act as transcriptional terminators. The G+C content of the paa cluster averaged 52.5%, a value close to the mean G+C content of E. coli genomic DNA (51.5%) (40). At the 5′-end of the sequenced fragment, a partial ORF corresponded to the gene YdbC (Fig. 2), a gene of unknown function from E. coli W, transposon mutagenesis of the 15.4-kb insert of plasmid pAAD was carried out. A collection of Tn1000 cassette in pAAD was constructed (Fig. 1A) responsible of the transformation of 2-phenylethylamine into PA (18, 37, 38). Analysis of PA Pathway Intermediates—To identify possible aromatic compound, thus locating the paa genes in a 15.4-kb DNA fragment at the right end of the previously identified mao region (Fig. 1A) responsible of the transformation of 2-phenylethylamine into PA (18, 37, 38).
A. W, they appeared not to be intermediates in PA catabolism. To check whether the different pAAD::Tn1000 derivatives caused accumulation of PA pathway intermediates, E. coli W14 cells harboring these plasmids were grown in minimal medium containing PA and glycerol. Supernatants of the cultures were then analyzed by HPLC, revealing that only the clones containing plasmids with Tn1000 derivatives in genes paaX, paaY, and paaZ, showed PA consumption. However, although disruptions of genes paaX and paaY did not cause the accumulation of any compound, disruption of the paaZ gene caused the conversion of PA into a metabolite whose retention time in HPLC was identical to that of standard 2-HPA. Gas chromatography-mass spectrometry analysis confirmed this metabolite as 2-HPA.

The paa Cluster of E. coli

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene product (Db/aa)*</th>
<th>Prosite signature</th>
<th>Name</th>
<th>Function</th>
<th>Organism</th>
<th>% Identity/aa</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>paaZ</td>
<td>73,045/681</td>
<td></td>
<td>MaoC</td>
<td>Putative regulator</td>
<td>K. aerogenes</td>
<td>78.7 /335 (336-681)**</td>
<td>P49251</td>
</tr>
<tr>
<td>paaA</td>
<td>35,437/309</td>
<td></td>
<td></td>
<td></td>
<td>R. leguminosarum</td>
<td>28.3 /161 (545-681)</td>
<td>P08634</td>
</tr>
<tr>
<td>paaB</td>
<td>10,942/95</td>
<td></td>
<td></td>
<td></td>
<td>B. vulgaris</td>
<td>24.5 /500 (1-527)</td>
<td>P28337</td>
</tr>
<tr>
<td>paaC</td>
<td>27,878/248</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaD</td>
<td>18,606/167</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaE</td>
<td>39,333/356</td>
<td>2Fe-2S binding region</td>
<td>TdnB</td>
<td>Aniline dioxygenase reductase</td>
<td>P. putida</td>
<td>30.1 /337</td>
<td>D85415</td>
</tr>
<tr>
<td>paaF</td>
<td>27,164/255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaG</td>
<td>28,450/262</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaH</td>
<td>51,704/475</td>
<td>3-hydroxyacyl-CoA hydratase-isomerase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaI</td>
<td>14,850/140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaJ</td>
<td>42,276/401</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaK</td>
<td>48,966/437</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaX</td>
<td>35,266/316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>paaY</td>
<td>21,439/196</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* aa, amino acid.
** In parentheses is indicated the partial amino acid sequence of the gene product that has been used for comparison.

* DH means dehydrogenase.

As analyses of culture supernatants did not reveal any true intermediate in PA catabolism, intracellular accumulation of possible metabolites was monitored by thin-layer chromatography of resting cell assays of E. coli W14 (pAAD::Tn1000 derivative 3) in resting cell assays. To identify the radioactive products accumulated, extracts of E. coli W14 (pAAD::Tn1000 derivative 3) were analyzed by HPLC and the detected radioactive peaks were shown to cochromatograph with authentic PA and PA-CoA (data not shown). Furthermore, radioactive 2-HPA was detected by HPLC analysis of the supernatants derived from E. coli W14 (pAAD::Tn1000 derivative 3) in resting cell assays.
The paa Cluster of E. coli

FIG. 3. HPLC chromatograms of the formation of 2-HPA and PA-CoA from PA in resting cell assays. A, time course of PA consumption and 2-HPA formation in the supernatants of the resting cell assays from E. coli W14 (pAAD::Tn1000 derivative). Assays were performed as described under “Experimental Procedures.” Samples (20 μl) were retrieved after 0, 2, 8, and 24 h, and analyzed by HPLC as indicated under “Experimental Procedures.” B, conversion of [14C]PA to [14C]PA-CoA in E. coli W14 (pAFK5) resting cell assays. The elution profile of [14C]-labeled compounds separated by HPLC is shown. For details, see “Experimental Procedures.”

Therefore, all these data taken together suggest that 2-HPA is secreted to the culture medium when the catabolism of PA is blocked, CoA derivatives such as PA-CoA being the true intermediates of the pathway.

The paaK Gene Encodes a PA-CoA Ligase—The formation of PA-CoA during the catabolism of PA in E. coli W suggests the existence of a PA-CoA ligase activity involved in the activation of PA to its CoA derivative. As the derived amino acid sequence of the paaK gene product revealed a putative AMP-binding consensus motif (Table I), and showed a high identity to that of the PhaE, former Pcl (65.6%), and PaaK_Y2 (67.3%) proteins responsible for the conversion of PA to PA-CoA in P. putida U (12) and Pseudomonas sp. Y2 (16), respectively, we assumed that PaaK could be the PA-CoA ligase of E. coli W. To experimentally demonstrate that paaK encoded a PA-CoA ligase, this gene was subcloned in plasmid pUC19 under the control of the lac promoter resulting in plasmid pAFK5 (Fig. 4A). Crude extracts of E. coli W14 (pAFK5) cells grown in glycerol-containing minimal medium showed a high level of PA-CoA ligase activity (450 units/mg protein), and this activity was dependent on the presence of ATP, CoA, and PA. SDS-polyacrylamide gel electrophoresis analysis of crude lysates from these cells revealed the presence of an intense band corresponding to a protein with an apparent molecular mass of 49 kDa (data not shown), which is in good agreement with the predicted molecular mass for the PaaK protein (Table I). The N-terminal sequence, MITNTK, of the overproduced protein corresponded with that deduced from the nucleotide sequence of the paaK gene, thus confirming it as the paaK gene product and showing that no processing of its N-terminal end occurs.

To demonstrate that the product of the reaction catalyzed by PaaK was PA-CoA, we performed resting cell assays of E. coli W14 (pAFK5) in the presence of [14C]PA. Two radioactive peaks that cochromatographed with authentic PA and PA-CoA were observed (Fig. 3B). To confirm that the second radioactive peak corresponded to PA-CoA, we performed the hydrolytic removal of the CoA moiety by treatment of the sample with NaOH at 65 °C for 30 min, and, as expected, the resulting [14C]-labeled product was shown to elute in HPLC as standard PA.

Analyses of crude extracts of E. coli W14 (pAAD::Tn1000) mutants grown in minimal medium containing PA and glycerol revealed that insertions of transposon Tn1000 in genes paaB, paaC, paaD, paaE, paaF, paaG, and paaJ caused a significant decrease in the PA-CoA ligase activity, such activity being below detection limits when the Tn1000 insertions were located within the paaK gene (Table II). These results may indicate that genes paaABCDEFGHIJK constitutive an operon and that insertions of transposon Tn1000 within this operon exert strong polar effects on the genes located downstream of the insertion site. In agreement with this hypothesis, Tn1000 insertions in genes paaZ, paaX, and paaY did not reduce the PA-CoA ligase activity of the corresponding mutants (Table II).

Functional Organization of the paa Genes—To study the functional organization of the paa genes within the PA catabolic pathway, plasmids pAFK5 (paaK), pAFZ1 (paaZ), pFB67 (paaZpaaABCDE), pAFAF1 (paaABCDE), and pAFFK (paaFGHIJKpaaXY) (Figs. 1 and 4), were constructed. E. coli W14 cells harboring these plasmids were grown in minimal medium containing both glycerol and PA, and the supernatants of the cells were analyzed by HPLC. We could only observe PA consumption when genes paaABCDE were expressed simultaneously to the paaK gene. Thus, E. coli W14 (pAFK5, pAFAF1) cells removed PA with the concomitant accumulation of 2-HPA in the culture medium (data not shown). However, PA remained unaltered in the culture medium when E. coli W14 cells containing plasmid pAFK5, pAFAF1, pAFFK, pAFZ1, or pFB67 were analyzed. Therefore, these results indicate that all or some of the paaABCDE genes are involved in 2-HPA formation and that this hydroxylation reaction requires the expression of the paaK gene responsible of PA-CoA formation. As 2-HPA is not a true intermediate in the PA catabolic pathway, it can be assumed that after the first catabolic step in the PA degradation, i.e. activation to PA-CoA, a hydroxylation reaction occurs leading to the formation of a hydroxylated derivative of PA-CoA. A blockade in the PA degradation pathway preventing further catabolism of the hydroxylated PA-CoA derivative could cause the secretion of the latter to the culture medium as 2-HPA.

Although the simultaneous expression of the paaABCDEFGHIJKpaaK genes gave rise to 2-HPA, the additional presence of the paaZ gene did not reveal accumulation of 2-HPA in the supernatants of E. coli W14 (pFB67, pAFAF1) cultures. Therefore, these data suggest that the paaZ gene product could be responsible of the third enzymatic step in the aerobic catabolism of PA in E. coli W, genes paaFGHIJK being likely involved in further catabolism of PA to Krebs cycle intermediates.

Regulation of the paa Cluster—Analysis of the paaK expression showed that the paa-encoded pathway was inducible. Thus, although crude extracts of E. coli W14 (pAAD) cells grown in glycerol-containing minimal medium in the absence of PA did not reveal PA-CoA ligase activity, a significant level of PaaK activity was observed when the cells were grown in minimal medium containing glycerol and PA (Table III). Interestingly, extracts from E. coli W14 (pAAD::Tn1000 derivatives
6 and 32) cells grown in the presence of PA showed levels of PA-CoA ligase activity similar to those of extracts from the same cells grown in the absence of PA (Table III), thus indicating that Tn1000 insertions in gene paaX cause a constitutive expression of the paa-encoded pathway. However, no constitutive expression of the paaK gene was observed when paaY was disrupted by Tn1000 insertion (Table III). Therefore, these data suggest that the paaX gene product behaves as a negative regulator of the paa catabolic genes.

The arrangement of the paa genes and the polar effects derived from the Tn1000 insertions in the paa cluster suggest the existence of three transcription units, two of them, paaZ and paaABCDEFGHJK, responsible of catabolic functions, and a third one, paaXY, involved in regulation. To study the promoter regions of the paa cluster, DNA fragments containing the potential Pz, Pa, and Px promoters of genes paaZ, paaABCDEFGHJK, and paaXY, respectively, were PCR-isolated and ligated into the lacZ gene of the promoter-probe vector pSJ3. The resulting translational fusion plasmids pAPPA1 (PaaZ-lacZ), pAPPA2 (PaaABCDEFGHJK-lacZ), and pAPPA3 (PaaXY-lacZ) (Fig. 5), conferred to the host strain E. coli CC118 the ability to produce blue colonies on media containing the β-galactosidase indicator 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside, thus indicating the presence of a functional promoter in the four cloned fragments.

To determine the transcription initiation sites in the Pa, Pz, and Px promoters, primer extension analyses were performed with total RNA isolated from E. coli CC118 cells containing plasmids pAPPA1, pAPFX, and pAPFZ (Fig. 6). The transcription initiation sites of the paaA gene were mapped utilizing two different primers (Fig. 6, A and B), and potential −10 (TGGTAA) and −35 (TGTGTA) boxes typical of ρ0-dependent promoters were identified in the Pa promoter region (Fig. 2). The same results were obtained when plasmid pAPPA2 (Pa255-lacZ) was used as source of RNA for the primer extension analyses (data not shown). The transcription initiation site in the Pz promoter (Fig. 6C) was located 27 nucleotides upstream of the ATG translation initiation codon of the paaZ gene, showing a putative −10 box (TGTGTA) but lacking a consensus −35 sequence (Fig. 2). Analysis of the 194-bp region between the transcription start sites of the paaZ and paaA genes showed a high A+T content (70%), and revealed several inverted repeat sequences and a putative integration host factor-binding consensus motif (Fig. 2), which might be involved in the control of gene expression. Although these potential regulatory elements are present in Pa254, they are absent in Pa255. Transcription of paaX (Fig. 6D) starts 29 nucleotides upstream of the ATG translation initiation codon (Fig. 2). The presence in the Pz promoter of a putative extended −10 box (TGCTATGAT) could explain the absence of a consensus −35 hexamer (Fig. 2). It is worth noting that the putative stem-loop structure that could act as a transcriptional terminator of the paaABCDEFGHJK operon is located only 15 bp upstream from the extended −10 box of the Pz promoter (Fig. 2).

We have shown above that expression of the paa catabolic operon is controlled by the paaX gene product; therefore, to
Enzymatic activities were assayed using crude extracts of E. coli W14 containing different plasmids. Cells were grown in glycerol-containing minimal medium in the presence of 5 mM PA. Enzymatic assays were performed as described under “Experimental Procedures.” Results of one experiment are given; values were reproducible in three separate experiments.

<table>
<thead>
<tr>
<th>E. coli W14 containing plasmid</th>
<th>paa genes</th>
<th>PA-CoA ligase activity units/mg protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>paa cluster absent</td>
<td>BD*</td>
</tr>
<tr>
<td>pAAD: Tn1000 40</td>
<td>paaZ</td>
<td>22</td>
</tr>
<tr>
<td>pAAD: Tn1000 5</td>
<td>paaZ</td>
<td>30</td>
</tr>
<tr>
<td>pAAD: Tn1000 21</td>
<td>paaZ</td>
<td>55</td>
</tr>
<tr>
<td>pAAD: Tn1000 35</td>
<td>paaB</td>
<td>5</td>
</tr>
<tr>
<td>pAAD: Tn1000 59</td>
<td>paaC</td>
<td>5</td>
</tr>
<tr>
<td>pAAD: Tn1000 24</td>
<td>paaD</td>
<td>5</td>
</tr>
<tr>
<td>pAAD: Tn1000 83</td>
<td>paaE</td>
<td>5</td>
</tr>
<tr>
<td>pAAD: Tn1000 86</td>
<td>paaF</td>
<td>5</td>
</tr>
<tr>
<td>pAAD: Tn1000 16</td>
<td>paaG</td>
<td>5</td>
</tr>
<tr>
<td>pAAD: Tn1000 45</td>
<td>paaJ</td>
<td>6</td>
</tr>
<tr>
<td>pAAD: Tn1000 46</td>
<td>paaK</td>
<td>BD</td>
</tr>
<tr>
<td>pAAD: Tn1000 6</td>
<td>paaX</td>
<td>27</td>
</tr>
<tr>
<td>pAAD: Tn1000 52</td>
<td>paaX</td>
<td>20</td>
</tr>
<tr>
<td>pAAD: Tn1000 84</td>
<td>paaY</td>
<td>16</td>
</tr>
</tbody>
</table>

* BD, below detection limits.

Enzymatic activities were assayed using crude extracts of E. coli W14 cells containing different plasmids were grown in glycerol-containing minimal medium in the absence (uninduced) or presence of 5 mM PA induced). Preparation of cellular extracts and determination of PA-CoA ligase activity were done as described under “Experimental Procedures.” Results of one experiment are given; values were reproducible in three separate experiments.

<table>
<thead>
<tr>
<th>E. coli W14 plasmid</th>
<th>Tn1000 insertion</th>
<th>PA-CoA ligase activity units/mg protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAAD</td>
<td>None</td>
<td>BD*</td>
</tr>
<tr>
<td>pAAD: Tn1000 6</td>
<td>paaX</td>
<td>20</td>
</tr>
<tr>
<td>pAAD: Tn1000 52</td>
<td>paaX</td>
<td>15</td>
</tr>
<tr>
<td>pAAD: Tn1000 84</td>
<td>paaY</td>
<td>BD</td>
</tr>
</tbody>
</table>

* BD, below detection limits.

The paa Cluster of E. coli

The repressor effect of PaaX on Pa could not be significantly alleviated by growing the cells in the presence of 5 mM PA (Table IV), which indicates that this aromatic compound is not the true inducer of the pathway. Interestingly, when the paaK and paaX genes were simultaneously expressed in the reporter strains, β-galactosidase activities were shown to be inducible by 5 mM PA (Table IV), suggesting that the reaction product of the PaaK enzyme, i.e. PA-CoA, is the inducer of the Paa promoter. Furthermore, it is worth noting that P0424 and P0255 responded similarly to the PaaX-mediated regulation, thus suggesting that the 169-bp region that is absent in Paa5 is not directly involved in the promoter-operator sites of the paa catabolic operon.

The paa genes from E. coli W were located in a chromosomal 15.4-kb DNA fragment cloned in plasmid pAAD, and they mapped at the right end of the mao region (Fig. 1A), which is involved in the transformation of 2-phenylethylamine into PA (18, 37, 38). As the equivalent mao genes in E. coli K-12 have been mapped at min 31.0 on the chromosome (43), and two PA* mutants of E. coli K-12 had been located in this chromosomal region (7), a similar location of the paa genes in the chromosome of E. coli W can be suggested.

The nucleotide sequence of the paa cluster revealed the presence of 14 ORFs, paaZpaaABCDCEFHIKpaayY (Figs. 1 and 2), that corresponded with those of unknown function whose Protein Identification Database accession numbers are g178653–g178664, g178665, and g178666, and that have been recently sequenced in E. coli K-12 (accession numbers AE000236, AE000237, D90777, and D90778) (41). Although the left end of the paa cluster was near to the maoA gene both in E. coli W and K-12, the right end of the paa cluster was different in the two strains. Thus, although the paaY stop codon was found 231 bp upstream of the ATG start codon of the ydbC gene in E. coli W (Fig. 2), a 9.2-kb sequence encoding a long ORF (ydbA) disrupted by two insertion sequences (IS2 and IS30) was found between paaY (Protein Identification Database accession number g178667) and ydbC in E. coli K-12 (29). The presence of insertion sequences near the paa cluster and the location of this cluster in a nonessential region of the chromosome (44) provide some clues on the possible mechanisms of gene mobilization of a catabolic cassette that would enhance bacterial adaptability, and could explain the heterogeneity observed among different E. coli strains with respect to their ability to mineralize PA. It is also noteworthy that the mao genes for the metabolism of 2-phenylethylamine, an aromatic amine whose degradation gives rise to PA, lie adjacent to the paa cluster responsible for the further catabolism of PA. This association between genes belonging to the same catabolon (15), i.e. genes involved in convergent degradative routes, could be considered as an important evolutionary and adaptive advantage. Another example of such association within a PA catabolon has been recently described in the path-

Further analyze faithfully this regulatory system, we have engineered the reporter Pa-lacZ fusion within a mini-Tn5 vector (Fig. 5). The resulting constructions, pAPFA1T (P0424-lacZ) and pAPFA2T (P0255-lacZ) were used to detect by transposition the corresponding translational fusions into the chromosome of E. coli AF2014, a rifampicin-resistant E. coli W14 (lacZ) mutant strain devoid of β-galactosidase activity, giving rise to the reporter strains E. coli AF1414 and AF1412, respectively. The presence of a strong T7 phage transcriptional terminator downstream of the lacZ fusions and their orientation within the mini-Tn5 elements (Fig. 5), prevented read-through transcription from nearby chromosomal promoters after insertion, thus facilitating the regulatory studies. To check the influence of the PaaX protein on the expression of the reporter fusions, paaX was cloned in plasmid pAPFX2 (Fig. 7) and expressed into the reporter strains. The β-galactosidase assays of permeabilized E. coli AF1411 and AF1412 cells harboring control plasmid pCK01 showed a similar and constitutive expression of the reporter fusions (Table IV). When the gene paaX was expressed in trans, we observed a drastic decrease (more than 2 orders of magnitude) in the β-galactosidase levels of E. coli AF1411 (pAPFX2) and E. coli AF1412 (pAPFX2) cells (Table IV), thus indicating that PaaX behaves as a transcriptional repressor of Pa both in the Pa424 and Pa255 promoter fragments.
way for styrene degradation in *Pseudomonas* sp. Y2, where the *sty* genes responsible of the oxidation of styrene to PA are in tight association with the genes involved in PA degradation (16).

The genetic arrangement of the *paa* cluster and the mutagenesis of pAAD with transposon Tn1000 revealed that the 14 *paa* genes are organized in three transcriptional units, two of them, *paaZ* and *paaABCDEFGHIJK*, essential for the catabolism of PA, and a third one, *paaXY*, that contains the *paaX* regulatory gene. An overall sequence comparison analysis of the *paa* gene products showed that they were homologous to the recently described *pha* genes responsible of the catabolism of PA in *P. putida* U (15) (Fig. 8B). Here, we have presented experimental evidence that the *paaK* gene product is the PA-CoA ligase of *E. coli* W (Fig. 3B), an activity that had been

The paa Cluster of E. coli

25983
catabolism in bacteria. The participation CoA ligases in the initial step of the aerobic catabolism of 2-aminobenzoate (47) and benzoate (48) in Azoarcus evansii KB740 (formerly Pseudomonas sp. KB740), ferulate in P. putida (49) and Pseudomonas fluorescens (50), and 2-furoic acid in P. putida Fu1 (51) has been also reported, and the existence of a CoA ligase has been suggested for the aerobic catabolism of salicylate in Rhodococcus sp. strain B4 (52) and thiophen-2-carboxylate (53). Moreover, some dehalogenation mechanisms of aromatic compounds also involve CoA thioester formation in aerobiosis (54). Although the rationale for utilizing such hybrid pathways, i.e. aerobic catabolic pathways endowed with typical features of an anaerobic catabolism, is not known, it has been suggested that they could represent a strategy of facultative microorganisms to cope with the fluctuations of oxygen supply (55). In this sense, the existence of a hybrid pathway for the catabolism of PA in E. coli could reflect the facultative anaerobe character of this bacterium.

All or some of the paaABCDE genes appear to be responsible of the second enzymatic step in the catabolism of PA in E. coli. Thus, the expression of paaK and paaABCDE genes in E. coli W14 caused the consumption of PA and the accumulation of 2-HPA in the culture medium. However, 2-HPA appears not to be a true intermediate in the PA catabolic pathway as it does not support growth of E. coli W and is not consumed even when E. coli W cells are growing also in the presence of PA. Interestingly, a similar lack of growth on 2-HPA and accumulation of this compound after adding PA to some cultures of PA mutant strains from E. coli K-12 (7) and P. putida U (15), has been also observed. Although the possibility that exogenous 2-HPA does not enter the cells cannot be ruled out, the fact that 2-HPA formation requires the simultaneous expression of the paaK and paaABCDE genes strongly suggests that 2-HPA is not a true intermediate in PA degradation but derives from the accumulation of a hydroxylated PA-CoA intermediate that cannot be further degraded. The excretion to the culture medium of a hydroxylated aromatic compound as a dead-end product derived from the intracellular accumulation of a hydroxylated CoA derivative has been also reported in the hybrid pathway for the catabolism of 2-aminobenzoate (47), and could be a general strategy of the cells to prevent the possible metabolic risk of depletion of the intracellular pool of CoA (15, 56).

The second catabolic step in PA degradation in E. coli seems to be, therefore, the hydroxylation of PA-CoA. Although we could not detect a hydroxylated CoA derivative in E. coli W14 (pAAD::Tn1000 derivative) cells, intracellular accumulation of 2-HPA-CoA has been observed during the catabolism of PA by a P. putida U mutant strain (15). Sequence comparison analyses of the paaABCDE gene products revealed that the PaaE protein (356-amino acid length) showed significant similarity with the class IA-like reductases (Table I). These enzymes are members of the ferredoxin-NAD+ reductase (FNR) family and they contain a FNR-like domain consisting of a FMN(PAD)- and a NAD(P)-binding region (57). The residues 55RCYS58 in PaaE fit the RXXY consensus motif for binding of the isoalloxazine ring of the flavin cofactor, and residues 121GSGITP126 and 216CGPAM221 match the MXGX(G)2-P and CGXG(M)2 sequences for the binding of the NAD(P) ribose and NAD(P)-pyrophosphate-nicotinamide moieties of the nicotinamide cofactor, respectively (58). At the C terminus of the FNR-like domain, residues 299–337 in PaaE correspond to the CXXCXXXC24–34 motif of the plant-type ferredoxin [2Fe–2S] binding domain (58). Other members of the extended FNR family are the reductase components of the methane, alkene, phenol, and toluene diiron monoxygenases (59–63), a group of bacterial hydrocarbon oxidation enzymes that com-

![Diagram of the paa Cluster of E. coli](http://www.jbc.org/)

**FIG. 7. Subcloning and expression of the paaX regulatory gene.** Plasmids are drawn with the relevant elements and restriction sites indicated. Plasmid pH55-P has been described in Fig. 1A. Vector-derived sequences are indicated with continuous (pUC derivatives) or discontinuous (pCR01 derivatives) lines. Arrows show the Plac promoter and the direction of transcription of the genes. The bent arrows indicate the Px promoter. Ap’ and Cm’, the genes that confer ampicillin and chloramphenicol resistance, respectively. Δ, a truncated gene. Restriction sites are: Bc, BamHI; Bc, BclI; E, EcoRI; Et, Eco47III; H, HindIII; He, HinflI; N, NotI; P, PstI.

detected in this strain when it was grown in PA-containing medium (13). Analysis of the primary structure of PaaK (Fig. 2) revealed that residues 100SSGTGKPTV112 match the AMP-binding site consensus sequence T(SG)-S(G)-G-(ST)-T(SE)....(15, 56). The sequences 236DIYGLSE242 and 302YRTRD306 showing that this residue does not assume a major role in ATP risk of depletion of the intracellular pool of CoA (15, 56).

The detection of radiolabeled PA-CoA inside E. coli W14 (pAAD::Tn1000 derivative) cells, indicates that disruption of the paaZ gene causes a blockade of the PA catabolic pathway leading to the accumulation of this CoA derivative, and confirms the physiological role of PaaK in the catabolism of PA in this microorganism. Assuming that the paaK gene product catalyzes the first enzymatic step of the PA catabolic pathway, the polar effects caused by the Tn1000 insertions within the potential paa catabolic operon containing the paaK gene at its 3’-end, can explain why pathway intermediates did not accumulate in E. coli W14 cells expressing the corresponding paaD::Tn1000 derivatives. The degradation of PA in P. putida U also appears to require PA-CoA as the first intermediate of the pathway (12), and a similar situation could be inferred in other bacteria that are able to use aerobically PA as the sole carbon source (13, 16). The aerobic catabolism of aromatic compounds via their initial activation to CoA derivatives constitutes an unusual strategy that resembles anaerobic degradation mechanisms (46), and could be a widespread way of PA
The paa Cluster of E. coli

Regulation of expression from the Pa promoter by PaaX

E. coli strains were grown in glycerol-containing minimal medium in the absence (uninduced) or in the presence of 5 mM PA (induced). β-Galactosidase activities were measured with permeabilized cells as described under “Experimental Procedures.” Results of one experiment are shown; values were reproducible in three separate experiments.

<table>
<thead>
<tr>
<th>E. coli strain</th>
<th>Plasmid(s)/gene(s)</th>
<th>Relevant background</th>
<th>Uninduced</th>
<th>Induced (5 mM PA)</th>
<th>-Fold induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF141 None</td>
<td>lacZ</td>
<td>BD</td>
<td></td>
<td>BD</td>
<td></td>
</tr>
<tr>
<td>AF141 pCK01</td>
<td>Pa424::lacZ</td>
<td>3111</td>
<td>4156</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>AF141 pFX2/paaX</td>
<td>Pa424::lacZ</td>
<td>10</td>
<td>46</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td>AF141 pAFK5/paaK</td>
<td>Pa424::lacZ</td>
<td>12</td>
<td>1114</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>AF1412 pCK01</td>
<td>Pa255::lacZ</td>
<td>3441</td>
<td>3877</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>AF1412 pFX2/paaX</td>
<td>Pa255::lacZ</td>
<td>16</td>
<td>59</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>AF1412 pAFK5/paaK</td>
<td>Pa255::lacZ</td>
<td>25</td>
<td>1737</td>
<td>69</td>
<td></td>
</tr>
</tbody>
</table>

* BD, below detection limits.

FIG. 8. Comparison of the paa cluster of E. coli W with thepha cluster of P. putida U. A, comparison of the geneticorganization of the paa and pha clusters. Blocks with similar shading or hatching indicate homologous regions encoding potential functional units in both gene clusters. The location and size of the intergenic regions, are also indicated. Bent arrows represent the promoters. B, percentages of amino acid sequence identity between the analogous paa and pha gene products. Note that genes phaK do not have counterparts in the paa cluster, and that genes paaB and paaE have not been described in the pha cluster.

prises an evolutionarily related protein family (60). These soluble multicomponent monooxygenases contain, in addition to the reductase component, a heteromultimeric (αβγ) oxygenase component, a low molecular weight activator protein (61–63), and, in some cases, a Rieske-type ferredoxin (59, 60). Interestingly, the primary structure of the PaaA protein (309-amino acid length) shows the two repeats of residues EX2H separated by approximately 100 amino acids (positions 155–158 and 249–252) that characterize the dinuclear iron binding-site of the large (α) oxygenase subunit of the methane, phenol, and toluene diiron monooxygenases (60). Moreover, the amino acid sequence of PaaB (95-amino acid length) shows the strictly conserved residues found in the low molecular weight dissociative activator protein that is required for optimal turnover of the oxygenase component in multicomponent diiron monooxygenases (64). Therefore, these sequence comparison analyses suggest that genes paaABCDE may encode the five subunits of a diiron multicomponent oxygenase, with PaaB being the effector protein and PaaE the reductase that mediates electron transfer between NAD(P)H and the PaaCD oxygenase component. It is worth noting that the paaE gene product can constitute the first example of a reductase subunit from a multicomponent oxygenase that shows a reversed domain order, i.e. a FNR-like N-terminal domain and a plant-type ferredoxin C-terminal domain, which supports the previous hypothesis that class IA-like reductases may have been recruited for a variety of aromatic ring oxidation reactions (65). Moreover, the putative PaaABCDE oxygenase, and its counterpart encoded by thephaFGHI operon of P. putida U (Fig. 8B), may represent the first reported multicomponent oxygenase acting on a CoA-activated aromatic acid.

The paaZ gene product appears to be responsible of the third enzymatic step of the PA catabolic pathway. The putative PaaZ protein (681-amino acid length) presents an N-terminal region (residues 1–527) whose primary structure shows similarity with that of aldehyde dehydrogenases (Table I). In this sense, the PaaZ residues 229FGSAATG236 and 281GQKTAIR288, respectively, match the consensus NAD(P) binding site and the active site motif spanning the catalytic cysteine (underlined) of aldehyde dehydrogenases (18, 66). Moreover, the sequence 253MEADSLN260 in PaaZ encompasses the potential catalytic glutamic acid residue (italicized) of aldehyde dehydrogenases (18, 66). The amino acid sequence of the C-terminal region of PaaZ shows similarity to that of the maoC and nodN gene products of unknown function (Table I). As has been suggested for the analogous PhaL protein of P. putida U (15) (Fig. 8B), the paaZ gene product in E. coli might catalyze the aromatic ring cleavage of the hydroxylated CoA derivative formed during PA degradation. Nevertheless, the formation by PaaZ of a non-aromatic CoA cyclic intermediate, similar to that described as the product of the reaction catalyzed by the aminobenzoyl-CoA monooxygenase-reductase during the aerobic catabolism of 2-aminoxenzoate (67), cannot be ruled out.

ThephaF, paaG, paaH, and paaI gene products show significant sequence similarities to fatty acid β-oxidation enzymes (Table I), and therefore can tentatively constitute a β-oxidation-like pathway involved in the successive oxidation reactions of the non-aromatic CoA intermediate. Interestingly,
The paa Cluster of E. coli

β-oxidation-like mechanism is another typical feature of the anaerobic catabolism of aromatic compounds (46). The primary structure of the putative PaaF (255-amino acid length) and PaaG (262-amino acid length) proteins shows similarity with that of members of the enoyl-CoA hydratase/isomerase superfamily (54, 68) (Table I). The paaH gene encodes a protein (475-amino acid length) that shares the signature sequence motives of 3-hydroxyacyl-CoA dehydrogenases (69) (Table I), thus suggesting that it could attack the product of the reaction catalyzed by the PaaF and PaaG enzymes. Although the paaI gene product (140-amino acid length) did not show a high level of sequence similarity with other proteins in the data bases, the paaJ gene product (401-amino acid length) presented a significant sequence similarity with the PcaF and CatF β-ketoacyl-CoA thiøxes (Table I), residues 90 and 386 in PaaJ being the putative catalytic cysteines. As PcaF and CatF catalyze the last step in the ortho-cleavage pathway for the aerobic degradation of protocatechuate and catechol, respectively (70), it is tempting to speculate that PaaJ and its analogous PhaD protein in P. putida U (Fig. 8) are also responsible for the last enzymatic step in PA degradation.

In the paa cluster, we have identified three promoters, Pz, Pa, and Px, which drive the expression of genes paaZ, paaAB-CDEFHIJK, and paaXY, respectively (Figs. 6 and 8A). The expression of the paa-encoded catabolic pathway is inducible, and it has been shown that the Pa promoter is negatively controlled by the paaX gene product (Table III and IV). The PaaX protein (316-amino acid length) contains a stretch of 25 residues at amino acids 39–64 that shares similarity with the helix-turn-helix motif predicted to be important for DNA recognition and binding in other transcriptional repressors such as GntR (71) and FadR (72). The GntR and FadR binding sites within the respective promoters contain a region of dyad symmetry, which is located very close to the transcription initiation sites (71–73). Interestingly, a region of dyad symmetry can also be found centered near the transcription initiation sites in the Pa promoter (Fig. 2). As the repression caused by PaaX was only alleviated by PA in the presence of the PaaK protein (Table IV), PA-CoA appears to be the true inducer of the paa-encoded pathway. In this sense, gel retardation assays have confirmed PA-CoA as the effector molecule.4 Therefore, PaaX constitutes the first reported transcription factor regulated by CoA derivatives that controls the catabolism of aromatic compounds. It is worth noting that the FadR transcriptional repressor, which is regulated by acyl-CoA compounds and shows local similarity to PaaX, is also controlling the expression of genes involved in β-oxidation mechanisms (73).

Overlapping the 3′-end of paaX, we have found the putative ATG translation initiation codon of the paaY gene. A palindromic sequence (ΔG value of −15.7 kcal/mol) followed by a (T)7 tract is located 42 bp downstream of the TAA stop codon of paaY (Fig. 2), and may act as a ρ-independent transcriptional terminator of the putative paaXY operon. Although the primary structure of the paaY gene product (196-amino acid length) and its analogous PhaM protein from P. putida U (Fig. 8B) show several repeats of the hexapeptide (LIV/GX2 motif) that characterizes the members of the bacterial transferases family, e.g. the CaiE protein from the carnitine operon of E. coli and the Fbp ferrirypochelin-binding protein of P. aeruginosa (Table I), the physiological role of these proteins in PA catabolism is still unknown.

Comparative studies of the whole structure and organization of the paa and pha clusters from E. coli and P. putida U, respectively (Fig. 8A), revealed interesting functional and evolutionary data. Thus, although the pha genes appear to be cotranscribed in four discrete DNA segments or modules encoding the six different functional units for the catabolism of PA, i.e. the β-oxidation and activation (phaABCDE), hydroxylation (phaFGHI), transport and dearomatization (phaKL), and regulation (phaMN) units, the paa cluster showed the transcriptional coupling of the hydroxylation-β-oxidation-activation functional units into the single operon paaABCDE-GHLLJK (Fig. 8A). As there is good evidence that operons coding for the catabolism of aromatic compounds are assembled in a stepwise manner from existing catabolic genes (74), it is tempting to speculate that the paa cluster from E. coli arose by the fusion of some gene blocks that are contiguous but separately regulated in the pha cluster of P. putida U, and therefore it could be considered as a further step in the evolution toward a single regulon of a common ancestral gene cluster involved in PA catabolism. Moreover, the differences in gene order within some of the DNA modules, and the relative locations of these modules in the paa and pha clusters, suggest that various DNA rearrangements have occurred during their evolution. As the G+C content of the paa (52.5%) and pha (63.5%) genes averaged a value close to the mean G+C content of E. coli (51.5%) and P. putida (60%) genomic DNA (40), it could be thought that these two set of genes have been imprisoned within each host over a long period of evolution. Especially remarkable is the observation that the phaJ and phaK genes of P. putida U, encoding a permease and a specific-channel-forming protein for the uptake of PA, respectively (15), are absent in the paa cluster from E. coli W (Fig. 8A). Interestingly, the phad gene product shows significant amino acid sequence identity (62.1%) with the product of the yjcG gene that is located at min 92.2 of the E. coli K-12 chromosome (29). Whether a permease, such as the putative YjcG protein, and a channel-forming protein are required for the catabolism of PA in E. coli is still an open question.

The identification and genetic characterization of the hybrid paa-encoded pathway complete our knowledge on the pathways so far described for the aerobic catabolism of aromatic compounds in E. coli. Although, in some Pseudomonas and Acinetobacter species, a supraperon clustie of the aromatic catabolic genes has been observed in a limited region of the chromosome, the aromatic catabolic clusters are dispersed throughout the genome in E. coli, with cluster mhp (3-3-hydroxyphenylpropionate and 3-hydroxycinnamate) at min 8 (1, 10), paa at min 31, hca (3-phenylpropionate) at min 57.5 (11), and hpa (3- or 4-hydroxyphenylacetate) at min 98 (8). These data also indicate that E. coli is not an "empty box" for the catabolism of aromatic compounds; on the contrary, it is endowed with typical aerobic degradation routes as well as with a novel hybrid pathway, which are considered among the most ubiquitous aromatic compound catabolic systems and therefore are thought to be closer to the central catabolism than those involved in the degradation of xenobiotic compounds (75).

The results presented in this work provide a framework for additional studies to determine the role and properties of the enzymes involved in PA catabolism through a hybrid aerobic pathway that is likely to be a widespread route for the metabolism of this aromatic compound. In this sense, the cloned paa genes should be useful as probes to identify homologous genes from distinct groups of bacteria. Moreover, we anticipate that the unique features of the aerobic paa-encoded pathway will reveal novel catabolic activities that can be of great biotechnological interest to improve some microorganisms for the degradation of PA-related aromatic environmental pollutants (e.g., styrene), and for the synthesis of pathway intermediates that

---

4 A. Ferrández, J. L. García, and E. Díaz, manuscript in preparation.
can be useful for the production of new or modified antibiotics and plastics (15).

Acknowledgments—We thank M. K. B. Berlyn for strains MG1063 and MG1655, S. Jaenecke for plasmids pSJ3 and pSJ19Not, A. Prieto and J. Varela for assistance with the gas chromatography-mass spectrometry and N-terminal amino acid sequence analyses, respectively, and A. Diaz and G. Porras for their help with the sequencing. The excellent technical assistance of E. Cano, M. Carrasco, and F. Morante is gratefully acknowledged. We are indebted to K. N. Timmis and D. Pipher for facilitating the short visit of A. Fernandez to their laboratory.

REFERENCES
