Genes Coding for a New Pathway of Aerobic Benzoate Metabolism in Azoarcus evansii

Johannes Gescher,1 Annette Zaar,1 Magdy Mohamed,1 Hermann Schägger,2 and Georg Fuchs1*

Mikrobiologie, Institut Biologie II, Universität Freiburg,1 and Zentrum der Biologischen Chemie, Klinikum der Johann Wolfgang Goethe-Universität Frankfurt, Frankfurt,2 Germany

Received 23 May 2002/Accepted 16 August 2002

A new pathway for aerobic benzoate oxidation has been postulated for Azoarcus evansii and for a Bacillus stearothermophilus-like strain. Benzoate is first transformed into benzoyl coenzyme A (benzoyl-CoA), which subsequently is oxidized to 3-hydroxyadipyl-CoA and then to 3-ketoadipyl-CoA; all intermediates are CoA thioesters. The genes coding for this benzoate-induced pathway were investigated in the β-proteobacterium A. evansii. They were identified on the basis of N-terminal amino acid sequences of purified benzoate metabolic enzymes and of benzoate-induced proteins identified on two-dimensional gels. Fifteen genes probably coding for the benzoate pathway were found to be clustered on the chromosome. These genes code for the following functions: a putative ATP-dependent benzoate transport system, benzoyl-CoA ligase, a putative benzoyl-CoA oxygenase, a putative isomerizing enzyme, a putative ring-opening enzyme, enzymes for β-oxidation of CoA-activated intermediates, thioesterase, and lactone hydrolase, as well as completely unknown enzymes belonging to new protein families. An unusual putative regulator protein consists of a regulator protein and a shikimate kinase 1-type domain. A deletion mutant with a deletion in one gene (boxA) was unable to grow with benzoate as the sole organic substrate, but it was able to grow with 3-hydroxybenzoate and adipate. The data support the proposed pathway, which postulates operation of a new type of ring-hydroxylating dioxygenase acting on benzoate-CoA and nonoxygenolytic ring cleavage. A β-oxidation-like metabolism of the ring cleavage product is thought to lead to 3-ketoadipyl-CoA, which finally is cleaved into succinyl-CoA and acetyl-CoA.

Aerobic metabolism of aromatic compounds, such as benzoate, has been studied in considerable detail in various microorganisms (45a; for a recent review, see reference 23). Catechol (1,2-dihydroxybenzene) and protocatechuate (3,4-dihydroxybenzoate) were identified as early intermediates, depending on the initial oxygenases catalyzing benzoate hydroxylation. Benzoate metabolism via 4-hydroxybenzoate and protocatechuate is common in fungi, whereas in bacteria the catechol pathway has been established and the protocatechuate route (45a) is uncertain. Both compounds serve as substrates for ring-cleaving dioxygenases, which in the case of the ortho-cleavage pathway cleave the aromatic ring between the hydroxyl groups. Catechol and protocatechuate ortho cleavage and the subsequent reactions lead to 3-ketoadipate, which is converted into succinyl coenzyme A (succinyl-CoA) and acetyl-CoA via 3-ketoadipyl-CoA (Fig. 1).

However, some observations could not be explained by the established mechanisms. Thus, cell extracts of some Bacillus spp. grown on benzoate or 3-hydroxybenzoate utilized gentisate but not catechol or protocatechuate. In an attempt to explain these findings, hydroxylation of benzoate resulting in 3-hydroxybenzoate and gentisate (2,5-dihydroxybenzoate) was proposed (9, 10, 12). However, direct evidence for the suggested intermediates and reactions has not been obtained so far.

More recently, it was shown that a gram-positive Bacillus stearothermophilus-like strain (26) and the facultatively denitrifying gram-negative bacterium Azoarcus evansii, belonging to the β-group of the Proteobacteria (2, 6), are able to utilize benzoate, 3-hydroxybenzoate, and gentisate aerobically as sole sources of carbon and energy (1, 26, 32). 2-Hydroxy- and 4-hydroxybenzoates, protocatechuate, catechol, and 3,4-dihydroxybenzoate did not support aerobic growth (1, 26). In conjunction with the presence of an aerobically inducible benzoate-CoA ligase (AMP forming) and gentisate 1,2-dioxygenase (1, 42), the degradation of benzoate was proposed to proceed via benzoyl-CoA and either 2-hydroxybenzoyl-CoA or 3-hydroxybenzoyl-CoA as intermediates (1, 26, 32). Further hydroxylation of either compound hypothetically could yield gentisyl-CoA, which might undergo thioester hydrolysis to gentisate (1, 26). However, the enzymatic reactions catalyzing the proposed pathway to gentisate have remained elusive. 3-Hydroxybenzoate was shown to be metabolized via 6-hydroxylation of 3-hydroxybenzoate to gentisate, and gentisate is cleaved by gentisate 1,2-dioxygenase to maleylpyruvate. Maleylpyruvate is isomerized to fumarylpyruvate, which is cleaved into fumarate and pyruvate (1).

A study of the conversion of 13C-labeled benzoyl-CoA by cell extracts of A. evansii and the B. stearothermophilus-like strain under aerobic conditions revealed unexpected intermediates. In contrast to earlier proposals, benzoate was not converted into hydroxybenzoate or gentisate. Under aerobic conditions benzoyl-CoA was an in vivo product of benzoate catabolism in both microbial species and was converted into various CoA thioesters by cell extracts in oxygen- and NADPH-dependent reactions. By using [13C]benzoyl-CoA as a substrate, cis-3,4-dehydroadipyl-CoA, trans-2,3-dehydroadipyl-CoA, the 3,6-lactone of 3-hydroxyadipyl-CoA, and 3-hydroxyadipyl-CoA were identified as products by nuclear magnetic
resonance spectroscopy. A protein mixture from \textit{A. evansii} transformed benzoyl-CoA in an NADPH- and oxygen-dependent reaction into 6-hydroxy-3-hexenoyl-CoA (48). The data suggested that there is a novel aerobic pathway of benzoate catabolism via CoA intermediates leading to $\beta$-ketoadipyl-CoA, an intermediate of the known $\beta$-ketoadipate pathway (23) (Fig. 2).

Hence, the committed first step in the new benzoate pathway is the activation of benzoate to benzoyl-CoA by a specifically induced benzoate-CoA ligase (AMP forming) that catalyzes the following reaction: benzoate $\rightarrow$ CoA + Mg-ATP $\rightarrow$ benzoyl-CoA + Mg-AMP + pyrophosphate. This enzyme was purified and was shown to differ from an isoenzyme that catalyzes the same reaction under anaerobic conditions (29). The second step is postulated to involve the hydroxylation of benzoyl-CoA to an unknown product by a novel benzoyl-CoA oxygenase, presumably a multicomponent enzyme system. A benzoate-induced iron-sulfur flavoprotein, BoxA, which may be a component of this system, was purified and characterized (29). This protein has a native molecular mass of 98 kDa (homodimer of 50-kDa subunits) and contains (per mole of native protein) 0.72 mol of flavin adenine dinucleotide (FAD), 10.4 to 18.4 mol of Fe, and 13.3 to 17.9 mol of acid-labile sulfur, depending on the method of protein determination. This enzyme catalyzes a benzoyl-CoA-, FAD-, and O$_2$-dependent NADPH oxidation, surprisingly without hydroxylation of the aromatic ring. However, H$_2$O$_2$ is formed as follows: NADPH + H$^+$ + O$_2$ → NADP$^+$ + H$_2$O$_2$. The gene coding for this enzyme (\textit{boxA}, for benzoyl-CoA oxidizing) was cloned and sequenced (29). This gene codes for a 46-kDa protein (414 amino acids) with two consensus amino acid sequences for two [4Fe-4S] centers at the N terminus. The deduced amino acid sequence shows homology with subunits of ferredoxin-NADP$^+$ oxidoreductase, nitric oxide synthase, NADPH-cytochrome P450 oxidoreductase, and phenol hydroxylase. Upstream of the \textit{boxA} gene another benzoate-induced gene, \textit{boxB}, encoding a 55-kDa protein (473 amino acids), was found. The \textit{boxB} gene exhibits the highest level of
homology to an open reading frame (ORF) in *Sulfobolus solfataricus* (44) which probably codes for a component of a putative aerobic phenylacetyl-CoA-oxidizing system (15). In the present work we tried to find and study the missing genes involved in this novel aerobic benzoate oxidation pathway in *A. evansii*. Here we show that up to 15 genes are involved, some of which belong to new enzyme families.

**MATERIALS AND METHODS**

**Materials and bacterial strains.** Chemicals were obtained from Sigma-Aldrich (Deisenhofen, Germany), Merck (Darmstadt, Germany). Biomol (Hamburg, Germany), or Roth (Karlsruhe, Germany); biochemicals were obtained from Roche Diagnostics, and High-pressure liquid chromatography equipment was obtained from Waters (Milford, MA, USA). Primers were obtained from MBI Fermentas (St.Leon-Rot, Germany), Roche Diagnostics, and Roth (Karlsruhe, Germany); or Roth (Karlsruhe, Germany); biochemicals were obtained from Biozym (Hamburg, Germany), or Roth (Karlsruhe, Germany). Enzymes used for cloning experiments were purchased from MBI Fermentas (St.Leon-Rot, Germany); Roche Diagnostics, and Amersham Pharmacia Biotech (Freiburg, Germany).

*Escherichia coli* strains were grown at 37°C with ampicillin (100 µg/ml), kanamycin (50 µg/ml) and tetracycline (10 µg/ml) supplementation. *A. evansii* strains were grown aerobically at 37°C with kanamycin, spectinomycin, and gentamicin. *A. evansii* strains grow slowly at 37°C. In order to get sufficient material for cloning and expression experiments, *A. evansii* cultures were grown aerobically at 37°C at 100 liters in a 200-liter fermentor (air flow 100 liter/min; 200 rpm). For all experiments, *A. evansii* strain 11/3 (17) was used, which was isolated from hot springs in Jürgen Hilden, Germany. The strain was grown aerobically at 37°C in a medium containing 0.1% (w/v) yeast extract (high purity; Difco, USA) and 0.05% (w/v) tryptone (Difco, USA). The pH of the medium was adjusted to 7.2 before autoclaving and the medium was autoclaved at 121°C for 20 min. The shake flasks were incubated in a water bath shaker at 150 rpm and 37°C. For all experiments, *A. evansii* strain 11/3 was used, which was isolated from hot springs in Jürgen Hilden, Germany. The strain was grown aerobically at 37°C in a medium containing 0.1% (w/v) yeast extract (high purity; Difco, USA) and 0.05% (w/v) tryptone (Difco, USA). The pH of the medium was adjusted to 7.2 before autoclaving and the medium was autoclaved at 121°C for 20 min. The shake flasks were incubated in a water bath shaker at 150 rpm and 37°C.

**Polyacrylamide gel electrophoresis.** One microgram of purified DNA from cells grown aerobically on malate was used for reverse transcription-PCR (RT-PCR). RNA was isolated with an RNeasy kit (QIAGEN, Germany) and was separated from contaminating DNA by treatment with proteinase K and chloroform. The RNA was precipitated with 100% ethanol.

**Sodium dodecyl sulfate-polyacrylamide gel electrophoresis.** Two-dimensional gel electrophoresis was performed as described by Görg et al. (17) by using the Immobiline Dry Strips system (linear gradient from pH 3 to 10; Amersham Pharmacia Biotech) accord- ing to the manufacturer’s protocol. The second dimension (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was performed as described above.

**Alignments were generated by using the CLUSTLW program contained in the ZAP Express cloning kit instruction manual (Stratagene).**

**Screening of cosmid and AZAP Express gene banks.** PCR products were used as probes in screening analyses were labeled with digoxigenin-11-DUTP by PCR. They were detected by using anti-digoxigenin-AP Fab fragments, nitroblue tetrazolium chloride, and 5-chloro-4-bromo-3-indolylphosphate (toluidine salt) (Biotom, Hamburg, Germany). Probes were amplified with primers Mol and Mor, Probilífiser and Probilífiser, ABC-probebor and ABC-probebor, and bhabhor and hhabhor (Table 2).

**RT-PCR.** Total RNA from *A. evansii* cells grown aerobically on benzoate and from cells grown aerobically on malate was used for reverse transcription-PCR (RT-PCR). RNA was isolated with an RNeasy kit (QIAGEN, Germany) and was separated from contaminating DNA by treatment with proteinase K and chloroform. The RNA was precipitated with 100% ethanol.

**DNA techniques and purification of nucleic acids.** Standard protocols were used for DNA cloning, transformation, amplification, purification, and sequencing (4, 30). Plasmid DNA was purified by the method of Birnboim and Doly (5). Both strands of the cloned chromosomal DNA containing the gene cluster were sequenced.

**Computer analysis.** DNA and amino acid sequences were analyzed by using the BLAST network service at the National Center for Biotechnology Information. Alignments were generated by using the CLUSTLW program contained in the DNAman software package (Lynnont, Montreal, Canada).

**Synthesis of CoA esters.** Benzoyl-CoA was prepared by previously described procedures (19, 41). The yield was 65%.

**Purification of component A of benzoyl-CoA oxygenase (BoxA).** BoxA was purified by the method of Mohamed et al. (29).

**Enzyme assay for BoxA.** BoxA enzyme activity was monitored spectrophotometricaly at 365 nm by determining benzoyl-CoA-, oxygen-, and FAD-depen- dent oxidation of NADPH at 37°C. The standard assay mixture (0.5 ml) con- tained 100 mM Tris-HCl (pH 8.0), 0.1 mM FAD, 0.3 mM NADPH, and cell extract (10 µl of a 100,000 ×g supernatant). The test was started by adding 0.1 mM benzoyl-CoA.

**Sodium dodecyl sulfate-polyacrylamide gel electrophoresis.** Polyacrylamide (11.5%) gel electrophoresis was performed by the Laemmli method (11). Proteins were visualized by Coomassie blue staining (49).

**Two-dimensional gel electrophoresis.** For two-dimensional gel electrophoresis the first dimension (isoelectric focusing) was performed with cell extract (120 µg of protein) as described by Görg et al. (17) by using the Immobiline Dry Strip system (linear gradient from pH 3 to 10; Amersham Pharmacia Biotech) accord- ing to the manufacturer’s protocol. The second dimension (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was performed as described above.

**TABLE 1. DNA clones and vectors used**

<table>
<thead>
<tr>
<th>Clone or vector</th>
<th>Genes</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEX 6p-1</td>
<td>amp lacK</td>
<td>Amersham Pharmacia Biotech, Freiburg, Germany</td>
</tr>
<tr>
<td>SupercoI</td>
<td>amp neo</td>
<td>Stratagene, Amsterdam, The Netherlands</td>
</tr>
<tr>
<td>pIJ200SK</td>
<td>Gm, sacB ral</td>
<td>This study</td>
</tr>
<tr>
<td>pUC4-KSAC vector</td>
<td>kan ampK</td>
<td>This study</td>
</tr>
<tr>
<td>BK-CMV</td>
<td>kan</td>
<td>This study</td>
</tr>
<tr>
<td>BK-CMVc5</td>
<td>Includes bp 1 to 4370 of benzoate gene cluster</td>
<td>This study</td>
</tr>
<tr>
<td>BK-CMV17</td>
<td>Includes bp 2028 to 6862 of benzoate gene cluster</td>
<td>This study</td>
</tr>
<tr>
<td>BK-CMVcII</td>
<td>Includes bp 4115 to 9506 of benzoate gene cluster</td>
<td>This study</td>
</tr>
<tr>
<td>SupercoISGD1</td>
<td>Includes bp 7131 to 19189 of benzoate gene cluster</td>
<td>This study</td>
</tr>
</tbody>
</table>
and kanamycin (50 μg/ml) was plated on Gelrite (0.8%, wt/vol) minimal medium containing sucrose (5 mM) and transferred by conjugation into E. coli S17-1 and transferred by conjugation into vector pJQ200SK (38) (Table 1). The resulting plasmid was transformed into E. coli (strain H11003) and the ampicillin-resistant colonies were selected on 0.2% ampicillin, 0.3% sucrose, 0.8% gelrite, and 25 μg/ml kanamycin agar plates. A kanamycin resistance cassette was inserted into the pJQ200SK vector (38) by double recombination (37). Exconjugants which had lost the pJQ200SK vector due to double recombination were selected by screening for sucrose resistance. The presence of the desired boxA mutant was confirmed by colony PCR performed with the primers mentioned above and by Western immunoblotting, which showed that the mutant failed to produce BoxA.

**Protein contents.** Protein contents were determined by the method of Bradford (11) by using bovine serum albumin as the standard.

**Nucleotide sequence accession number.** The sequence data reported here have been deposited in the EMBL database under accession no. AF548005.

### RESULTS AND DISCUSSION

**Proteins induced by benzoate.** The ability to metabolize benzoate aerobically is induced by the substrate benzoate. Benzoate-grown cells contained 20-fold higher specific activities of benzoate-CoA ligase and BoxA enzyme activity than 3-hydroxybenzoate-grown cells, and in acetate- or malate-grown cells these enzyme activities were hardly detectable. In contrast, the 3-hydroxybenzoate pathway was induced to similar extents in 3-hydroxybenzoate-grown cells and in benzoate-grown cells. This is consistent with similar specific activities of characteristic enzymes of 3-hydroxybenzoate metabolism in the two types of cells (i.e., 3-hydroxybenzoate 6-hydroxylase and gentisate 1,2-dioxygenase activities) (32). We therefore used 3-hydroxybenzoate-grown cells as a reference for comparison of the protein patterns. Benzoate-grown cells should differ only by additional, benzoate-induced proteins when they are compared to 3-hydroxybenzoate-grown cells.

Two-dimensional gel electrophoresis of soluble cell extracts (100,000×g supernatant) revealed at least seven strongly benzoate-induced proteins in benzoate-grown cells, five of which

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT5' Enol-for</td>
<td>GACACGGCGGGATTGCAGGCAGTC</td>
<td>RT-PCR, fragment 0</td>
</tr>
<tr>
<td>Thio-for</td>
<td>CGACTTTTGGCGAAGCGAGCTATC</td>
<td>RT-PCR, fragment 0</td>
</tr>
<tr>
<td>RT-Thioesterase-for</td>
<td>CTCATGGCGGGATCTGAGGAGCA</td>
<td>RT-PCR, fragment 1</td>
</tr>
<tr>
<td>Hydro-for</td>
<td>GATTCAGGCCGAGCAGCAGAG</td>
<td>RT-PCR, fragment 2</td>
</tr>
<tr>
<td>RT-Hydro-for</td>
<td>CTGGGAGAATGCAGGCAGCACC</td>
<td>RT-PCR, fragment 2</td>
</tr>
<tr>
<td>hhb-for</td>
<td>GCCACATCTGCAAGGAGGC</td>
<td>RT-PCR, fragment 2</td>
</tr>
<tr>
<td>RTlgase-rev</td>
<td>GCGTTGGCGGAAAGACGCTG</td>
<td>RT-PCR, fragment 3</td>
</tr>
<tr>
<td>bindeprot-for</td>
<td>CTGCTCTCTGAGATCTGGATGG</td>
<td>RT-PCR, fragment 3</td>
</tr>
<tr>
<td>RT-ligase-for</td>
<td>CTGGGAGGAGCAGCAGGAG</td>
<td>RT-PCR, fragment 4</td>
</tr>
<tr>
<td>ORF3-rev</td>
<td>TGCGGAGGAGATTGGAGTGTGGCA</td>
<td>RT-PCR, fragment 4</td>
</tr>
<tr>
<td>paaZ-for</td>
<td>GATGCGAGGAGATTGGAGTGTGGCA</td>
<td>RT-PCR, fragment 5</td>
</tr>
<tr>
<td>Regulator-rev</td>
<td>CGTCTTGGCGAGGAGGGACGAGCAG</td>
<td>RT-PCR, fragment 5</td>
</tr>
<tr>
<td>Shiki-Fus-for</td>
<td>GCGCAGGCGGAGGAGGGAGGACGAG</td>
<td>RT-PCR, fragment 6</td>
</tr>
<tr>
<td>Shiki-Fus-rev</td>
<td>GGTGAACTTACCTGGATGATGG</td>
<td>RT-PCR, fragment 6</td>
</tr>
<tr>
<td>Fus-boxB-for</td>
<td>CTGGACTTTCAACCCGCGCCAGCA</td>
<td>RT-PCR, fragment 7</td>
</tr>
<tr>
<td>Fus-boxA-rev</td>
<td>GATCAGGCTCGAGGGAGGAGGAGGCAGCAG</td>
<td>RT-PCR, fragment 7</td>
</tr>
<tr>
<td>box b-boxA-rev</td>
<td>GCTCGGCTTCTCAGGAGGAGGACGAGCAG</td>
<td>RT-PCR, fragment 8</td>
</tr>
<tr>
<td>boxA-Thio-for</td>
<td>GCTACATCTCATACGCGAGCCGAG</td>
<td>RT-PCR, fragment 8</td>
</tr>
<tr>
<td>NMO4</td>
<td>GCGGACGCGGATGGTGGTGGT</td>
<td>RT-PCR, fragment 9</td>
</tr>
<tr>
<td>Thio-Hypo-for</td>
<td>CGTATCCGAGGAGGCTCTGCTG</td>
<td>RT-PCR, fragment 9</td>
</tr>
<tr>
<td>Thio-Hypo-rev</td>
<td>ATGGAACCSCSCGGCSCGGA</td>
<td>Screening probe 1</td>
</tr>
<tr>
<td>MoI</td>
<td>CCGAGAAGTGGAGTGGAGGTC</td>
<td>Screening probe 2</td>
</tr>
<tr>
<td>MorI</td>
<td>CGCCCGAGGGCATCGCTGGAC</td>
<td>Screening probe 3</td>
</tr>
<tr>
<td>Probelyfor</td>
<td>CGCCCGAGGGCATCGCTGGAC</td>
<td>Screening probe 3</td>
</tr>
<tr>
<td>Probelyrev</td>
<td>ACGGCTTCCAGGTTGCGAGCAG</td>
<td>Screening probe 4</td>
</tr>
<tr>
<td>ABC-probeAfor</td>
<td>GATTCAGAGGAGGAGGGGCAG</td>
<td>Screening probe 5</td>
</tr>
<tr>
<td>ABC-probeBrev</td>
<td>GATTCAGAGGAGGAGGGGCAG</td>
<td>Screening probe 6</td>
</tr>
<tr>
<td>hhabfor</td>
<td>GATCAGGCTTCTCGGAGCCGAG</td>
<td>Screening probe 7</td>
</tr>
<tr>
<td>hhabfor</td>
<td>GACACATCTGAAACCGAGCCGAG</td>
<td>Screening probe 8</td>
</tr>
<tr>
<td>BoxABamHI-forw</td>
<td>CACCGCGATGCCCACCATCACCCTGCTG</td>
<td>boxA mutant 100</td>
</tr>
<tr>
<td>BoxANot1-rev</td>
<td>ATAGAAATGCGGCGGCGAGCCGAGCAGCATTGCGGAGAC</td>
<td>boxA mutant 100</td>
</tr>
</tbody>
</table>

**Comparison of extracts of cells grown on benzoate and 3-hydroxybenzoate allowed identification of benzoate-induced proteins.**

Electrophoretic transfer of protein and determination of N-terminal amino acid sequences. Proteins of benzoate-grown cells were separated by two-dimensional gel electrophoresis and transblotted onto an Immobilon-P membrane (Millipore, Bedford, Mass.) by using the Nova Blot system (Multiphor II; Pharmacia LKB, Freiburg, Germany) (11). Transblotted proteins were detected by Ponceau S staining (11). Polyclonal antibodies raised against BoxA of A. evansi were used to test for the presence of BoxA. BoxA was detected by Coomassie blue staining (49). Benzoate-induced proteins were excised and sequenced by using an Applied Biosystems 473A sequencer.

**Immunodetection.** Cell extracts of benzoate-grown and malate-grown cells were separated by two-dimensional gel electrophoresis and transblotted onto nitrocellulose filters (pore size, 0.45 μm) (11). The transblotted proteins were detected by Ponceau S staining (11). Polyclonal antibodies raised against BoxA of A. evansi were used to test for the presence of BoxA. BoxA was detected by Coomassie blue staining (49). Benzoate-induced proteins were excised and sequenced by using an Applied Biosystems 473A sequencer.

**Protein contents.** Protein contents were determined by the method of Bradford (11) by using bovine serum albumin as the standard.

**Nucleotide sequence accession number.** The sequence data reported here have been deposited in the EMBL database under accession no. AF548005.

### RESULTS AND DISCUSSION

**Proteins induced by benzoate.** The ability to metabolize benzoate aerobically is induced by the substrate benzoate. Benzoate-grown cells contained 20-fold higher specific activities of benzoate-CoA ligase and BoxA enzyme activity than 3-hydroxybenzoate-grown cells, and in acetate- or malate-grown cells these enzyme activities were hardly detectable. In contrast, the 3-hydroxybenzoate pathway was induced to similar extents in 3-hydroxybenzoate-grown cells and in benzoate-grown cells. This is consistent with similar specific activities of characteristic enzymes of 3-hydroxybenzoate metabolism in the two types of cells (i.e., 3-hydroxybenzoate 6-hydroxylase and gentisate 1,2-dioxygenase activities) (32). We therefore used 3-hydroxybenzoate-grown cells as a reference for comparison of the protein patterns. Benzoate-grown cells should differ only by additional, benzoate-induced proteins when they are compared to 3-hydroxybenzoate-grown cells.

Two-dimensional gel electrophoresis of soluble cell extracts (100,000×g supernatant) revealed at least seven strongly benzoate-induced proteins in benzoate-grown cells, five of which...
were N terminally sequenced (Fig. 3). The induced protein spots were labeled protein spots 1 to 7. The proteins of three of the spots, protein spots 1, 3, and 5, could be sequenced, and two others (protein spots 2 and 4) gave no sequence, possibly due to an N-terminal block. In addition, the N-terminal sequences of two purified proteins, benzoate-CoA ligase and BoxA, were known (see above). All N-terminal amino acid sequences obtained, as well as the estimated sizes and isoelectric points of benzoate-induced proteins, are summarized in Table 3. Antibodies raised against BoxA were used to determine by Western blotting whether one of the benzoate-induced proteins was BoxA. Cells grown on malate were used as a negative control. Malate-grown cells instead of 3-hydroxybenzoate-grown cells were used in this experiment since malate-grown cells contained virtually no BoxA activity, whereas 3-hydroxybenzoate-grown cells contained approximately 5% of the fully benzoate-induced enzyme activity. This low level was still detected by the sensitive immunoassay. A faint protein spot next to benzoate-induced protein spot 1, which was not identical to protein spot 1, specifically reacted with the serum. This protein spot did not appear when cells were grown on malate. This confirmed that BoxA is another benzoate-induced protein which is present at only a low concentration.

**Cloning and sequencing of the genes coding for benzoate-induced proteins.** Before this study, only the genes for BoxA and BoxB had been sequenced; the order of these genes is *boxBA*. As determined by inference from the high number of benzoate-induced soluble proteins, the complexity of the pathway is much greater, and several genes must be considered still missing. A cosmid gene library of chromosomal DNA was generated, and DNA probes derived from *boxA* were used for screening. The cloning strategy is summarized in Fig. 4.

One positive clone was obtained, which contained a 37-kb insert carrying *boxBA*. Upstream of *boxBA* the DNA sequence contained two additional ORFs which were oriented in the same direction as *boxBA*. The N-terminal sequence deduced from the 5′ end of one of these ORFs (ORF 11) was identical to the N-terminal sequence of induced protein 3. Further upstream, three ORFs oriented in the opposite direction were found, and the last of these ORFs was incomplete. The N-terminal amino acid sequence deduced from the 5′ end of this

<table>
<thead>
<tr>
<th>Purified and/or induced protein (PRF)</th>
<th>Estimated molecular mass (kDa)</th>
<th>Estimated isoelectric point</th>
<th>N-terminal amino acid sequence (^b)</th>
<th>Deduced N-terminal amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzoate-CoA ligase (ORF 7)</td>
<td>56</td>
<td>5.7</td>
<td>MTTLSAADHSTSPHTLPRQYNAAD</td>
<td>MTTLSAADHSTSPHTLPRQYNAAD</td>
</tr>
<tr>
<td>BoxA (ORF 13)</td>
<td>46</td>
<td>5.7</td>
<td>MNAPAEHANLRQHLIDPE</td>
<td>MNAPAEHANLRQHLIDPEICIRCNT</td>
</tr>
<tr>
<td>Protein 1 (BoxB; ORF 12)</td>
<td>50</td>
<td>5.8</td>
<td>XINYSERIPN</td>
<td>MINYSERIPNNVL</td>
</tr>
<tr>
<td>Protein 2</td>
<td>55</td>
<td>6.2</td>
<td>No sequence</td>
<td></td>
</tr>
<tr>
<td>Protein 3 (ORF 11)</td>
<td>60</td>
<td>5.6</td>
<td>XqAVANKPVAELvDYrEPEs</td>
<td>MQAVANKPVAELvDY RTEPS</td>
</tr>
<tr>
<td>Protein 4</td>
<td>57</td>
<td>5.7</td>
<td>No sequence</td>
<td></td>
</tr>
<tr>
<td>Protein 5 (ORF 6)</td>
<td>40</td>
<td>8.7</td>
<td>AEKIKVLMLPYTGTYAALG</td>
<td>MKNARMRRTLMOAMLGVIGALVPLGAQAOQ</td>
</tr>
<tr>
<td>Protein 6 (possibly ORF 3)</td>
<td>28</td>
<td>8.6</td>
<td>ND (^c)</td>
<td></td>
</tr>
<tr>
<td>Protein 7 (possibly ORF 1 or ORF 8)</td>
<td>19</td>
<td>5.5</td>
<td>ND (^c)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The experimentally determined N-terminal amino acid sequences are compared to the N-terminal amino acid sequences deduced from the presumed genes of the benzoate oxidation pathway. X indicates an unidentified amino acid, lowercase letters indicate uncertain amino acids, and underlining indicates a leader peptide.

\(^b\) The N-terminal amino acid sequences of the ORF 7 and 13 products were reported in references 1 and 29 and are included for comparison.

\(^c\) ND, not determined.
incomplete ORF was identical to the N-terminal amino acid sequence of benzoate-CoA ligase. This suggested that all these ORFs are likely to be involved in benzoate metabolism. Sequencing of the ORFs downstream of \textit{boxBA} revealed two other ORFs oriented in the same direction as \textit{boxBA}, which were separated from each other by only short intergenic regions. Then 74 bp farther downstream there was a putative ORF without a ribosome-binding site, which exhibited similarity only to putative ORFs in the database. Hence, the whole 11.8-kbp DNA sequence contained one incomplete and eight complete putative genes.

Further screening of a \textit{\lambda}ZAP Express gene bank led to the discovery of a total of 15 ORFs which were found to be clustered and which are likely to be involved in benzoate metabolism; 126 bp downstream of ORF 1 an incomplete ORF coding for another putative enoyl-CoA hydratase was found. The order of these ORFs, their orientations, their sizes, and their secondary DNA structures are shown in Fig. 4. Table 4 summarizes the properties of these ORFs.

**Gene organization and induction.** The orientation and organization of the 15 ORFs indicate that the benzoate metabolic genes are organized in at least two different operons. Several DNA duplex structures appear to be present (Fig. 4); the function of these secondary structures is unknown. The lengths of the intergenic regions vary (Table 4); the largest intergenic region is between ORFs 9 and 10 and is the presumed promoter and operator region of two divergently transcribed operons. This region contains several direct and inverted repeats that are more than 7 bp long. All ORFs but ORF 1 contain ribosome-binding sites which are very similar to the consensus sequence (AGGAGG).

Induction of ORFs during growth on benzoate was studied by performing RT-PCR experiments with mRNA from benzoate-grown cells and comparing the results to the results obtained with cells grown on malate (Fig. 5 and Table 5). The amplified DNA fragments contained the intergenic regions between adjacent genes. The data indicate that there was induction of all ORFs but ORF 1. Clearly, as expected, no transcript between ORFs 9 and 10 could be obtained. ORFs 2 to 9 and 11 to 13 were cotranscribed and detected even in highly (100- to 1,000-fold) diluted samples. At high dilutions there was no amplification of the intergenic region upstream of ORF 1 and of the regions between ORFs 1 and 2, ORFs 10 and 11, ORFs 13 and 14, and ORFs 14 and 15. Basal expression of ORFs 2 to 9 and ORFs 11 to 13 could be part of a global regulation strategy because ORFs 3 to 6 code for putative benzoate transport and ORFs 7, 11, 12, and 13 code for the proven or putative first steps in benzoate metabolism (see below). A basal level of expression of these genes could help if there were a change in carbon source. Cells grown on malate did not show any BoxA in immunodetection experiments, although RT-PCR detected \textit{boxA} transcription. This may indicate that BoxA production is also regulated at the translation stage. In summary, these data suggest that ORFs 2 to 13 function in benzoate metabolism. The putative regulator gene ORF 10 may be transcribed separately. The role of ORFs 1 and 15 is questionable.

Of the products of the 15 ORFs of the putative benzoate degradation gene cluster, 5 proteins were directly identified as being part of this functional unit (Table 3). The products of ORFs 7 and 13 were purified as benzoate-induced enzymes or proteins, and their N-terminal amino acid sequences agreed well with the deduced sequences. The products of ORFs 6, 11, and 12 were identified as benzoate-induced proteins by N-terminal sequencing. In addition, the product of ORF 13 was identified as a benzoate-induced protein by Western blotting performed with antibodies raised against BoxA.

**Characterization of the genes of benzoate metabolism and**
their putative functions. The putative functions of the sequenced genes and deduced gene products are summarized in Table 4. The first gene cluster consists of the following nine ORFs.

ORF 1 encodes a protein that has similarity to thioesterases, notably 4-hydroxybenzoyl-CoA thioesterase involved in the conversion of 4-chlorobenzoate to 4-hydroxybenzoate (4). The enzyme may function in benzoate metabolism as a security valve, cleaving CoA thioesters (or derived dead-end products) that accumulate and trap CoA if subsequent steps (e.g., oxygen-dependent steps) become limiting. This may easily occur in bacteria that profit from their ability to change quickly from anoxic to anoxic, denitrifying life style, which is characteristic of A. evansi. Otherwise, CoA trapping would be life threatening. The missing ribosome-binding site could then be part of a downregulation strategy for this ORF, with the aim of securing just a low level of the corresponding protein.

The ORF 2 product has similarity to lactone hydrolases. Its secondary structure is similar to the conserved structure of α/β-hydrolase fold enzymes (31), and the amino acid sequence fits well with the consensus motif (Sm-X-Nu-X-Sm-Sm, where Sm is a small amino acid, X is any amino, and Nu is a nucleophilic amino acid) for the nucleophile member (the nucleophile elbow) of the catalytic triad (34). Since the lactone of 3-hydroxyadipyl-CoA was observed to be product of benzoate transformation by cell extracts, this enzyme may function in hydrolyzing this lactone.

The product of ORF 3 has similarity to a putative ATP-binding subunit of an ABC transporter system. It has the typical Walker motifs of ATP-binding proteins (Walker A, GRNGMKTT_69; Walker B, LLILDE_184) (24). The ORF 4 product has similarity to a membrane-spanning subunit of an ABC transporter system (24). A prediction for transmembrane helices indicates that there are nine possible membrane-spanning α-helices. Furthermore, this protein has Walker consensus motifs for an ATP-binding site (Walker A, GPN-GAGKST_71; Walker B, LLILDE_130). Therefore, this protein seems to be a two-domain protein. The protein encoded by ORF 5 has similarity to a putative membrane-spanning protein of an ABC transporter system. It seems to contain six transmembrane helices. The ORF 6 product has similarity to a putative substrate-binding protein of an ABC transporter system. This protein corresponds to benzoate-induced protein 5 (Table 3). It contains a 31-amino-acid leader peptide. The sequence of this leader peptide is typical of the sequences found in the Sec transport system (37). It contains an N domain with a net positive charge (amino acids 1 to 10) and a hydrophobic H domain (amino acids 11 to 24). In summary, the products of ORFs 3 to 6 are likely to represent the four components of an ATP transporter system responsible for the highly efficient uptake of benzoate.

ORF 7 codes for the aerobically induced benzoate-CoA ligase, which differs from the corresponding isoenzyme induced during anaerobic growth on aromatic substrates. The deduced N-terminal amino acid sequence is identical to the 20-amino-acid N-terminal sequence determined for the purified enzyme. The purified enzyme is a homodimer which has a native molecular mass of 130 kDa and subunits with molecular masses of approximately 56 kDa (deduced molecular mass, 58 kDa). The enzyme acts on benzoate but not on 3-hydroxybenzoate (1).

The protein encoded by ORF 8 shows low similarity to the enzyme 4-hydroxylaminobenzoate lyase, which catalyzes an odd reaction, the hydrolytic transformation of 4-hydroxylaminobenzoate to protocatechuic (3,4-dihydroxybenzoate) and ammonia during 4-nitrobenzoate degradation in Pseudomonas putida TW3 (25). There are similar entries in the database (18) which show only very low similarity to other gene products, indicating that these proteins may form a new family. The role of the ORF 8 product in benzoate metabolism is enigmatic. In any case, the substrate of this enzyme seems to be a benzoate derivative in both types of metabolism (i.e., 4-nitrobenzoate degradation and benzoate degradation).

The ORF 9 product has similarity to proteins which are assumed to play a role in aerobic phenylacetate metabolism in E. coli (PaaZ) (14, 15), P. putida (PhaL) (33), A. evansi (PaaZ) (30), and presumably other bacteria (30). Phenylacetate metabolism also proceeds via CoA thioesters, and the products of the ORFs related to ORF 9 are thought to be involved in hydrolytic or acylcin (or aldol) cleavage of the ring and/or in aldolization. The hypothetical substrate for this kind of ring cleavage is a nonaromatic product (a cis-dihydrodiol) formed by a dioxygenase/reductase acting on phenylacetyl-CoA and benzyol-CoA. Ring cleavage may be preceded by isomerization of the double bonds (see below). The protein encoded by ORF 9 may be involved in further oxidation of the intermediate formed after C—C bond cleavage (see below). The N-terminal amino acid sequence of the ORF 9 product (amino acids 10 to 434) shows similarity to the sequences of aldehyde dehydrogenases. The conserved domains and_268 and cysteine-302 homologues (underlined), which in mammal aldehyde dehydrogenases are involved in the active center (14).

The second gene cluster, which is oriented in a different direction, consists of six genes with the following properties. The product of ORF 10 represents a two-domain protein. A similar ORF was found in a gene cluster of the β-proteobacterium Thauera aromatica, which contains genes for BoxA, BoxB, and benzoate-CoA ligase (K. Schuhle and G. Fuchs, unpublished data). The N-terminal domain (amino acids 32 to 87) has similarity to various regulatory proteins of the HTH family and seems to have the typical primary and secondary structures of helix-turn-helix proteins (7, 47). The first putative helix is between amino acids 44 and 51; it is separated from the second helix by a typical turn sequence (_GSVaS). The second helix is between amino acids 55 and 62. The C-terminal domain (amino acids 137 to 285) has similarity to shikimate kinase I of E. coli. Shikimate kinase I does not function in shikimate phosphorylation and in the biosynthesis of aromatic amino acids, although it phosphorylates shikimate in vitro, but with very low affinity for its substrate (K_m, 20 mM). The actual shikimate kinase II has a 100-fold-lower K_m (13). The ORF 10 product may play a role as an unprecedented regulator protein that becomes activated or inactivated by ATP-dependent phosphorylation in response to benzoate.

The protein encoded by ORF 11 corresponds to benzoate-induced protein spot 3. It has a mosaic structure. An N-terminal domain (amino acids 47 to 162) and a C-terminal domain
TABLE 4. Properties of genes and gene products assumed or proven to be involved in benzoate oxidation

<table>
<thead>
<tr>
<th>Gene</th>
<th>G + C content (mol %)</th>
<th>Length of intergenic region to next ORF (bp)</th>
<th>Putative function of gene product</th>
<th>Molecular mass (kDa)</th>
<th>Isoelectric point (pH)</th>
<th>Putative cellular localization</th>
<th>Similar proteins in databases</th>
<th>% Identity</th>
<th>% Similarity</th>
<th>E value</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF 1</td>
<td>65.1</td>
<td>6</td>
<td>Thioesterase</td>
<td>16</td>
<td>5.72</td>
<td>Cytosolic</td>
<td>Putative 4-hydroxybenzoyl-CoA thioesterase (Ralstonia solanacearum)</td>
<td>41</td>
<td>60</td>
<td>7e-20</td>
<td>NC_003296</td>
</tr>
<tr>
<td>ORF 2</td>
<td>68.5</td>
<td>1</td>
<td>Lactonase</td>
<td>28</td>
<td>5.22</td>
<td>Cytosolic</td>
<td>Putative hydrolase-related protein (Ralstonia solanacearum)</td>
<td>46</td>
<td>59</td>
<td>6e-48</td>
<td>NC_003295</td>
</tr>
<tr>
<td>ORF 3</td>
<td>66.9</td>
<td>2</td>
<td>ABC transporter, ATP binding</td>
<td>28</td>
<td>9.56</td>
<td>Cytosolic</td>
<td>Putative ABC transporter subunit HbaH (Rhodopseudomonas palustris)</td>
<td>52</td>
<td>68</td>
<td>1e-60</td>
<td>AAC13363</td>
</tr>
<tr>
<td>ORF 4</td>
<td>67.2</td>
<td>17</td>
<td>ABC transporter, ATP-binding membrane-spanning protein</td>
<td>62</td>
<td>6.88</td>
<td>Membrane</td>
<td>Putative ABC transporter subunit HbaG (Rhodopseudomonas palustris)</td>
<td>39</td>
<td>54</td>
<td>1e-104</td>
<td>AAC13364</td>
</tr>
<tr>
<td>ORF 5</td>
<td>65.1</td>
<td>77</td>
<td>ABC transporter, membrane-spanning protein</td>
<td>31</td>
<td>6.18</td>
<td>Membrane</td>
<td>Putative ABC transporter subunit HbaF (Rhodopseudomonas palustris)</td>
<td>45</td>
<td>59</td>
<td>1e-53</td>
<td>AAC13365</td>
</tr>
<tr>
<td>ORF 6</td>
<td>65.2</td>
<td>140</td>
<td>ABC transporter, substrate-binding protein</td>
<td>41 (with leader peptide), 39 (without leader peptide)</td>
<td>9.03 (with leader peptide), 8.52 (without leader peptide)</td>
<td>Periplasm</td>
<td>Putative substrate-binding protein (Azoarcus evansii)</td>
<td>80</td>
<td>86</td>
<td>1e-171</td>
<td>AAL02070</td>
</tr>
<tr>
<td>ORF 7</td>
<td>66.6</td>
<td>218</td>
<td>Benzoate-CoA ligase</td>
<td>58</td>
<td>5.61</td>
<td>Cytosolic</td>
<td>Benzoate-CoA ligase (Thauera aromatica)</td>
<td>76</td>
<td>84</td>
<td>&lt;1e-170</td>
<td>CAD21683</td>
</tr>
<tr>
<td>ORF</td>
<td>% Identity</td>
<td>% Similarity</td>
<td>E Value</td>
<td>Location</td>
<td>Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>66.0</td>
<td>98</td>
<td>17</td>
<td>4.97</td>
<td>Cytosolic</td>
<td>Benzoate-CoA ligase (Azoarcas evansi)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>70.4</td>
<td>327</td>
<td>54</td>
<td>8.62</td>
<td>Cytosolic</td>
<td>p-Hydroxylaminobenzoate lyase (Pseudomonas sp. strain Y102)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>66.5</td>
<td>40</td>
<td>34</td>
<td>6.37</td>
<td>Cytosolic</td>
<td>Aldehyde dehydrogenase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>68.1</td>
<td>159</td>
<td>61</td>
<td>5.44</td>
<td>Cytosolic</td>
<td>Unknown regulatory two-domain protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>64.3</td>
<td>163</td>
<td>55</td>
<td>5.62</td>
<td>Cytosolic</td>
<td>Enoyl-CoA hydratase/isomerase and possible ring-cleaving enzyme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>66.9</td>
<td>245</td>
<td>46</td>
<td>5.59</td>
<td>Cytosolic</td>
<td>Benzoyl-CoA oxygenase component B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>71.0</td>
<td>179</td>
<td>42</td>
<td>5.68</td>
<td>Cytosolic</td>
<td>Benzoyl-CoA oxygenase component A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>67.2</td>
<td>Unknown</td>
<td>21</td>
<td>9.51</td>
<td>Cytosolic</td>
<td>beta-Ketoadipyl-CoA thiolase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Similarity searches were done with the program blastp (http://www.ncbi.nlm.nih.gov/BLAST/). The percentage of identity was defined as the percentage of amino acids that are identical in two proteins. The percentage of similarity was defined as the percentage of amino acids that are identical or conserved in two proteins. The E value is an estimate of the statistical significance of the match, which specifies the number of matches with a given score that is expected in a search of a database of this size absolutely by chance.
enoyl-CoA isomerization, and a variety of different reactions, including enoyl-CoA hydration, function as the oxygenase part of benzoyl-CoA oxygenase in E. coli (PaaA) (15) and P. putida (PhaF) (33). The phenylacetate-CoA oxygenase is thought to consist of four subunits (PaaABCD) and to require a one-subunit NAD(P)H oxidoreductase component (PaaE). Similar ORFs are found in phenylacetate metabolism gene clusters of other bacteria, including A. evansii (30) and S. solfataricus (44). The primary structure of BoxB shows the two repeats of residues EX2H separated by 86 amino acids (120EGRH153 and 239EEAH242) that characterize the dinuclear iron-binding site of the large oxygenase subunit of methane, phenol, and toluene diiron monooxygenase (16, 36). We postulate that BoxB functions as the oxygenase part of benzoyl-CoA oxygenase in conjunction with BoxA, the reductase component.

The ORF 13 product is identical to the benzoate-induced enzyme BoxA. A similar ORF was found in the related organism T. aromatica next to the gene for BoxB (Schühle and Fuchs, unpublished). The similarity of BoxA to ferredoxin-NADP+ oxidoreductases and the benzoyl-CoA-dependent oxidation of NADPH catalyzed by the enzyme suggest that BoxA functions as the reducing component of benzoyl-CoA oxygenase, which, upon binding of benzoyl-CoA, transfers two electrons to the ring in the course of dioxygenation.

The protein encoded by ORF 14 has strong similarity to 3-ketoacidpyl-CoA thiolase. This is probably also the function of the protein in benzoate metabolism, which is assumed to lead to 3-ketoacidpyl-CoA. This implies that acetyl-CoA, succinyl-CoA, and CO2 are the products of the benzoate oxidation pathway.

### TABLE 5. Induction of genes of the benzoate degradation gene cluster, as determined by using cDNA from cells grown on benzoate and from cells grown on malate

<table>
<thead>
<tr>
<th>Amplified fragment</th>
<th>Fragment length (bp)</th>
<th>Templates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Benzoate-grown cells</td>
</tr>
<tr>
<td>0</td>
<td>500</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>402</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>801</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>1,008</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>862</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>618</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>500</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>593</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>721</td>
<td>+</td>
</tr>
<tr>
<td>9</td>
<td>836</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>259</td>
<td>+</td>
</tr>
</tbody>
</table>

*See Fig. 5.

a +, DNA fragment detected; --, no fragment detected.
ORF 15 is the last ORF of the cluster. Its product has no similarity to proteins in the database. Interestingly, a similar ORF was found in the gene cluster of *T. aromatica* mentioned above, which may code for benzoate oxidation in this related bacterium. The occurrence in similar gene clusters is an argument for a role for the ORF 15 product in the benzoate pathway, although no function can be ascribed to this protein yet.

Genes encoding a homologue of β-hydroxyacyl-CoA dehydrogenase and possibly another enoyl-CoA hydratase/isomerase are missing in the gene cluster sequenced. A β-oxidation dehydrogenase is needed to catalyze dehydrogenation of the detected intermediate 3-hydroxyadipyl-CoA to 3-oxoadipyl-CoA. Possibly the benzoyl-CoA pathway enzyme(s) is supplemented through the homologous protein(s) of primary metabolism.

Two-dimensional gel electrophoresis revealed at least seven benzoate-induced soluble proteins, and Western blotting revealed another protein. Of the proteins encoded by the 15 ORFs, 2 are likely to be membrane-bound components of an ABC transport system and therefore are not soluble; only small amounts of the regulator protein and the thioesterase are likely to be present. Hence, one would expect there to be 11 soluble benzoate-induced proteins. Thus, the complex protein induction pattern corresponds to the complex set of benzoate metabolic genes.

**BoxA**<sup>−</sup> mutant and expression of BoxA under different growth conditions. A boxA mutant was constructed by homologous recombination between the wild-type chromosome and an insertionally inactivated version of the gene carried on plasmid pJQ200MK. The boxA mutant failed to grow on benzoate as a sole carbon source. This shows that BoxA is essential for growth on benzoate.

To find a growth-supporting substrate (for further mutant analysis and biochemical studies of the pathway) which allowed expression of the benzoate pathway, regulation of boxA expression and therefore induction of the benzoate genes were tested with different substrates. Wild-type cells were grown on different substrates and combinations of substrates along with benzoate as an inducer for the benzoate genes. Benzoate gene expression was assessed by spectrophotometric measurement of BoxA activity (i.e., NADPH oxidation in the presence of benzoyl-CoA, FAD, and O<sub>2</sub> under H<sub>2</sub>O<sub>2</sub> production conditions). The substrates used, alone and in combination, included benzoate, 3-hydroxybenzoate, phenylacetate, adipate, malate, and acetate. No BoxA activity was observed with extracts of cells grown on acetate, malate, and adipate, while some activity was detected with cells grown on phenylacetate (10%) and 3-hydroxybenzoate (5%). Aromatic compounds resembling the aromatic substrate benzoate, like phenylacetate and 3-hydroxybenzoate, may act as poor gratuitous inducers. A combination of benzoate and 3-hydroxybenzoate resulted in 12% BoxA activity, a combination of benzoate and phenylacetate resulted in 50% BoxA activity, and a combination of benzoate and
Adipate resulted in 82% BoxA activity. No activity was observed with benzoate and acetate. This suggests that acetate acts as a strong catabolite repressor, whereas adipate does not. Also, 3-hydroxybenzoate and phenylacetate seemed to repress activity.

Proposal for the benzoate pathway and putative role of ORFs. We propose a benzoate pathway (Fig. 6) which is induced by benzoate and may be regulated by the product of ORF 10, a two-domain protein. An ABC transporter system consisting of the products of ORFs 3 to 6 may be responsible for effective benzoate uptake. The ORF 7 product activates benzoate to benzoyl-CoA. A protein mixture from *A. evansii* transformed benzoyl-CoA in an NADPH- and oxygen-dependent reaction into 6-hydroxy-3-hexenoyl-CoA. The proteins encoded by ORFs 9, 11, 12, and 13 may be involved in this complex reaction. Benzoyl-CoA is attacked by a putative 2,3-dioxygenase to obtain the cis-dihydrodiol product (unpublished results); BoxA (encoded by ORF 13) is thought to deliver electrons from NADPH to BoxB (encoded by ORF 12), which interacts with substrate and oxygen. BoxA has erroneously been reported to be benzoyl-CoA 3-monooxygenase (32), but a detailed study of the purified protein revealed that FAD-dependent oxidation of NADPH occurs in the presence of benzoyl-CoA without hydroxylase activity (29). The ORF 1 product may act as a thioesterase. This protein may be required to release CoA from the intermediates of the pathway if enzymes downstream of benzoyl-CoA become limiting, which would lead to trapping of CoA.

The next steps are different from what one would expect from conventional pathways involving a non-CoA-activated free cis-dihydrodiol intermediate. In conventional pathways, the cis-diol undergoes oxidation and rearomatization to a dihydroxy aromatic product (Fig. 1). Yet a putative cis-diol dehydrogenase gene could not be found in the gene cluster. We propose that the CoA thioester grouping of the activated diol allows isomerization of the conjugated double-bond system of the cis-diol, leading to an enol which tautomerizes to the more stable unsaturated cyclic 3-ketoacyl-CoA. The postulated intermediate formed contains an α-hydroxycarbonyl (acyloin) group. Two types of hydrolytic COC cleavage reactions can be envisaged (Fig. 6), followed by decarboxylation. These reactions may be catalyzed by the ORF 11 product, a complex protein which contains an enoyl-CoA hydratase/isomerase domain.

Oxidation of the alcohol group of 6-hydroxy-3-hexenoyl-CoA to the carboxyl group results in cis-3,4-dehydrodipetyl-CoA, which was also detected. The protein encoded by ORF 9 may be involved in this four-electron oxidation. The order of events involved in transformation of cis-3,4-dehydrodipetyl-CoA, possibly via trans-2,3-dehydrodipetyl-CoA and/or the 3,6-lactone of 3-hydroxydipetyl-CoA, to 3-hydroxydipetyl-CoA cannot be determined on theoretical grounds; all these intermediates were observed, but the order of formation could not be inferred. The lactone may form spontaneously and/or be a preparation artifact. The product of ORF 2 may be involved in lactone hydrolysis. In any case, it seems logical that 3-hy-
droxyadipyl-CoA is oxidized to 3-ketoadipyl-CoA, which is thioytically cleaved by the ORF 14 product into succinyl-CoA and acetyl-CoA. Hence, the pathway converges with the classical \(-/H\)-ketoadipate pathway of benzoate oxidation at the last intermediate, 3-ketoadipyl-CoA. So far, the roles of the proteins encoded by ORF 8 and ORF 15 are completely unknown.

**Similar genes in other bacteria.** Genes similar to those described here are found in very recent database entries for proteobacterial genomes, but their roles are unknown (Fig. 7). The bacteria examined include *Burkholderia fungorum* LB400, which contains two very similar sets of genes, *Ralstonia metaltilidurans* CH34, *Magnetospirillum magnetotacticum* MS-1, and *Rhodopseudomonas palustris* CGA009. Similar genes have also been found in *T. aromatica* K172 (Schühle and Fuchs, unpublished).

Interestingly, in most cases homologs of ORFs 11 to 13 occur in the same order as in *A. evansii*. In *T. aromatica* sequencing is incomplete, and the ORF 11 homologue is expected to be upstream of the ORF 12 homologue. In *R. palustris*, ORF 13 is still missing. We postulate that the products of ORFs 11 to 13 in these cases represent the core enzymes that act on benzoyl-CoA and its first product, the 2,3-dihydrodiol. In most cases the regulator protein (encoded by ORF 10) and the putative lactone hydrolase (encoded by ORF 2) are also present. The ABC transporter (encoded by ORFs 3 to 6) is generally missing. The putative thioesterase (encoded by ORF 1) could be found only in *M. magnetotacticum*. The product of ORF 15, whose function is not known, could be found only in *T. aromatica*. This may suggest that ORFs 1, 3 to 6, and 15 are not crucial for the new benzoate pathway. The thiolase (encoded by ORF 14) is missing in all other organisms. The function of this enzyme may be performed by common \(-/H\)-ketothiolases; this argument also holds true for 3-hydroxyacyl-CoA dehydrogenase.
The *R. palustris* strain which has been sequenced may grow aerobically on benzoate, although other strains reportedly do not do this (22). In this case the benzoate-CoA ligase of the anaerobic benzoate degradation cluster could function in both anaerobic and aerobic metabolism of benzoate, although there would be two completely different strategies for dearomatization under the two conditions. There are indications that the benzoate-CoA ligase (encoded by an ORF 7 homologue) in *T. aromatica* not only is involved in benzoate transformation to benzyol-CoA under aerobic conditions but also is involved in activation of benzoate and 2-aminobenzoate under anaerobic, degrading conditions (Schülle anduchs, unpublished).

Comparison with other aromatic degradation pathways that proceed via CoA thioesters and working hypothesis for future experiments. In *A. evansi*, aerobic metabolism of 2-aminobenzoate, benzoate, and phenylacetate proceeds via the CoA esters of the individual substrates. While the pathway for 2-aminobenzoate has been studied so far only in *A. evansi* (8, 21, 27, 28, 43, 50), the new benzoate pathway has also been found in the related organism *T. aromatica* (unpublished results) and in a gram-positive thermophilic *B. stearothermophilus*-like strain (48). The phenylacetate pathway is widely distributed (reviewed in reference 30). The three metabolic pathways seem to use a novel principle. Our working hypothesis is schematically presented in Fig. 8. The assumptions are as follows. (i) The pathways start with activation of the substrates to CoA thioesters, and the intermediates are further processed in this form. (ii) The aryl-CoA thioesters are attacked by dioxygenases and reductases (benzoate, phenylacetate) to obtain the corresponding cis-dihydrodiol, which is not re aromatized by oxidation but is isomerized to an hydroxylated nonaromatic product with an acylion structure. In the case of anthranilate, a monoxygenase/reductase forms a monohydroxylated nonaromatic product. The CoA thioester allows reactions that would not be feasible in the case of nonactivated aromatic acids and their intermediates. (iii) The early reactive intermediates are labile and may rearomatize spontaneously, in the case of 2-aminobenzoate to 5-hydroxy-2-aminobenzoyl-CoA reactions of the suicide vector and J. Alt-Moeller, unpublished). Of the three expressions of the individual substrates. While the pathway for 2-aminobenzoate proceeds via the CoA esters of the individual substrates. While the pathway for 2-aminobenzoate has been studied so far only in *A. evansi* (8, 21, 27, 28, 43, 50), the new benzoate pathway has also been found in the related organism *T. aromatica* (unpublished results) and in a gram-positive thermophilic *B. stearothermophilus*-like strain (48). The phenylacetate pathway is widely distributed (reviewed in reference 30). The three metabolic pathways seem to use a novel principle. Our working hypothesis is schematically presented in Fig. 8. The assumptions are as follows. (i) The pathways start with activation of the substrates to CoA thioesters, and the intermediates are further processed in this form. (ii) The aryl-CoA thioesters are attacked by dioxygenases and reductases (benzoate, phenylacetate) to obtain the corresponding cis-dihydrodiol, which is notrearomatized by oxidation but is isomerized to a dihydroxylated nonaromatic product with an acylion structure. In the case of anthranilate, a monoxygenase/reductase forms a monohydroxylated nonaromatic product. The CoA thioester allows reactions that would not be feasible in the case of nonactivated aromatic acids and their intermediates. (iii) The early reactive intermediates are labile and may rearomatize spontaneously, in the case of 2-aminobenzoate to 5-hydroxy-2-aminobenzoyl-CoA (by isomerization), in the case of benzoyl-CoA to 3-hydroxybenzoyl-CoA (by isomerization), and in the case of benzoyl-CoA to 3-hydroxybenzoyl-CoA (by water elimination), and in the case of phe- nylacetyl-CoA to 2-hydroxyphenylacetyl-CoA (by water elimination), and in the case of anthranilate, a monooxygenase/reductase forms a monohydroxylated nonaromatic product. In the case of anthranilate, the gene for shikimate kinase II in *Escherichia coli* K12-12, *J. Bacteriol.* 165:226–232.


las PK1, and purification of gentisate 1,2-dioxygenase. Biol. Fertil. Soils 23:307–313.