Identification of Novel Benzoylformate Decarboxylases by Growth Selection^{▽†}

Helge Henning,^{1‡} Christian Leggewie,^{1‡} Martina Pohl,¹ Michael Müller,²
Thorsten Eggert,¹ and Karl-Erich Jaeger^{1*}

Institute of Molecular Enzyme Technology, Heinrich Heine University Duesseldorf, Research Centre Juelich, D-52426 Juelich, Germany,¹ and Institute of Pharmaceutical Sciences, Albert Ludwigs University, D-79104 Freiburg, Germany²

Received 4 July 2006/Accepted 23 September 2006

A growth selection system was established using *Pseudomonas putida*, which can grow on benzaldehyde as the sole carbon source. These bacteria presumably metabolize benzaldehyde via the β -ketoacid pathway and were unable to grow in benzoylformate-containing selective medium, but the growth deficiency could be restored by expression in *trans* of genes encoding benzoylformate decarboxylases. The selection system was used to identify three novel benzoylformate decarboxylases, two of them originating from a chromosomal library of *P. putida* ATCC 12633 and the third from an environmental-DNA library. The novel *P. putida* enzymes BfdB and BfdC exhibited 83% homology to the benzoylformate decarboxylase from *P. aeruginosa* and 63% to the enzyme MdlC from *P. putida* ATCC 12633, whereas the metagenomic BfdM exhibited 72% homology to a putative benzoylformate decarboxylase from *Polaromonas naphthalenivorans*. BfdC was overexpressed in *Escherichia coli*, and the enzymatic activity was determined to be 22 U/ml using benzoylformate as the substrate. Our results clearly demonstrate that *P. putida* KT2440 is an appropriate selection host strain suitable to identify novel benzoylformate decarboxylase-encoding genes. In principle, this system is also applicable to identify a broad range of different industrially important enzymes, such as benzaldehyde lyases, benzoylformate decarboxylases, and hydroxynitrile lyases, which all catalyze the formation of benzaldehyde.

White biotechnology uses microorganisms and (microbial) enzymes to manufacture a wide variety of different chemicals, including polymers, bulk chemicals, agrochemicals, and pharmaceuticals (18). The biocatalytic production of these chemicals offers a number of advantages, including usually mild reaction conditions, the avoidance of toxic wastes, and, in particular, access to a variety of enantiopure compounds due to the high substrate specificity and enantioselectivity of many enzymes. However, the identification of novel biocatalysts that possess desired properties is still a challenge. Classical strain-screening approaches that are used for the isolation of microorganisms are laborious and often unsuccessful. Recently, the metagenome approach was developed to access useful genetic information encoded by environmental DNA, which is directly isolated from various habitats, cloned, and expressed in appropriate microbial host strains (37). However, the identification of a clone producing a biocatalyst of interest within a large metagenomic library requires high-throughput screening technologies, which are difficult to set up and usually expensive to run. Selection provides an elegant solution to the problem. Here, the desired enzyme activity is linked to the survival of the respective clone. Novel selection strategies that use natural or modified transcriptional regulators that bind to the product of

an enzymatic reaction have recently been developed (22). Transcriptional control can also be accomplished by a modified eukaryotic nuclear receptor (39) or by a “riboswitch” RNA (3). Enzymes producing compounds such as prephenate (7, 28), pyruvate (8), ammonia (36), and 2-hydroxybenzaldehyde (48) and 1,2,4-trichlorobenzene (29) were identified. A similar approach, termed chemical selection or complementation (1), is based on the detection of small molecules in an *in vivo* assay of lyase- and ligase-type reactions with a modified yeast three-hybrid system.

In the present work, we report on the construction of a selection system useful to identify novel enzymes producing benzaldehyde as the reaction product, which can be used by the bacterial selection host as the sole carbon source. The respective strain must possess the enzyme benzaldehyde dehydrogenase and the β -ketoacid pathway to convert benzoate into the tricarboxylic acid cycle intermediates succinyl-coenzyme A (CoA) and acetyl-CoA (Fig. 1). The genes encoding enzymes of the β -ketoacid central pathway were detected in many bacteria, mainly belonging to the genera *Acinetobacter* and *Pseudomonas* (10). An example is *Pseudomonas putida* ATCC 12633, which is able to grow on aromatic compounds as the sole carbon source (13–15). In this strain, mandelate is converted to acetyl-CoA via the mandelate/ β -ketoacid pathway (Fig. 1), and the enzyme benzoylformate decarboxylase MdlC (synonym, BFD) (E.C. 4.1.1.7) catalyzes the formation of benzaldehyde from benzoylformate by decarboxylation. The structure of BFD was solved in the absence (11) and the presence (34) of mandelic acid as an inhibitor, confirming that the enzyme acts as a tetramer, and active-site residues have been elucidated by site-directed mutants (34, 40) and directed-evolution studies (26, 27).

* Corresponding author. Mailing address: Institute of Molecular Enzyme Technology, Heinrich Heine University Duesseldorf, Research Centre Juelich, D-52426 Juelich, Germany. Phone: (49)-2461-613716. Fax: (49)-2461-612490. E-mail: karl-erich.jaeger@fz-juelich.de.

† This paper is dedicated to Maria-Regina Kula on the occasion of her 70th birthday.

‡ H.H. and C.L. contributed equally to this work.

▽ Published ahead of print on 29 September 2006.

FIG. 1. Mandelate and β -ketoadipate pathways in *P. putida*. The mandelate pathway in *Pseudomonas putida* ATCC 12633 is shown in dark gray, and the conversion of benzoylformate to benzaldehyde by a benzoylformate decarboxylase (MdlC) is highlighted. The β -ketoadipate pathway (gray) also exists in the majority of all pseudomonads. Additionally, *P. putida* KT2440 and *P. putida* DSM50198 possess a benzaldehyde dehydrogenase of unknown function. The metabolism of succinyl-CoA and acetyl-CoA via the tricarboxylic acid cycle is shown in light gray. MdlA, mandelate racemase; MdlB, S-mandelate dehydrogenase; MdlDE, NAD⁺- and NADP⁺-benzaldehyde dehydrogenases; BenABC, benzoate dioxygenase; BenD, 2-hydro-1,2-dihydroxybenzoate dehydrogenase; CatA, catechol-1,2-dioxygenase; CatB, *cis,cis*-muconate lactonizing enzyme (cycloisomerase); CatC, muconolactone isomerase; PcaD, β -ketoadipate enolactone hydrolase I; PcaIJ, β -ketoadipate succinyl-CoA transferase subunit; PcaF, β -ketoadipyl CoA thiolase; TCA, tricarboxylic acid.

We have newly constructed a *Pseudomonas*-based selection system to identify BFDs. These enzymes need thiamine diphosphate as a cofactor and have been identified in bacteria such as *Pseudomonas putida*, *Acinetobacter calcoaceticus*, and *Pseudomonas aeruginosa* (2, 2a, 12). They catalyze the decarboxylation of benzoylformate to benzaldehyde, but they also exhibit a carbolligase side activity enabling the conversion of aldehydes to chiral 2-hydroxy ketones, which are versatile building blocks for a variety of different fine chemicals (17, 33).

The functionality of the selection system was proven by the

isolation of three novel benzoylformate decarboxylases, one of them originating from a metagenomic library. These novel enzymes showed only low sequence similarity to presently known BFDs.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. The *Escherichia coli* strain DH5 α was used as a host for cloning and construction of recombinant plasmids. The strain *E. coli* DH5 α carrying the plasmid pRK2013 with the *tra* genes was used

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant characteristics ^a	Reference or source
<i>Pseudomonas</i> strains		
<i>P. putida</i> ATCC 12633	Wild type; catabolizes mandelate	ATCC (42, 43)
<i>P. aeruginosa</i> PAO1	Wild type	16
<i>P. putida</i> KT2440	Wild type; toluene-degrading <i>P. putida</i> mt-2	30, 35
<i>P. putida</i> DSM50198	Wild type; lack of mandelate pathway	DSMZ
<i>E. coli</i> strains		
DH5 α	λ^- ϕ 80dlacZ Δ M15 Δ (lacZYA-argF)U169 recA1 endA1 hsdR17(rK $^-$ mK $^-$) supE44 thi-1 gyrA relA1	9
BL21(DE3)	F $^-$ ompT hsdS _B (r _B $^-$ m _B $^-$) gal dcm (DE3)	45, 46
Plasmids		
pBBR1MCS-1	rep mob lacZ' Cm ^r	24
pBBR1MCS-2	rep mob lacZ' Km ^r	23
pBBR1MCS-5	rep mob lacZ' Gm ^r	23
pSUP202	pBR325 derivative; Amp ^r Cm ^r Tc ^r mob	41
pBluescript SK	P _{T7} P _{T3} P _{lac} lacZ' Cm ^r ColE1	Stratagene
pWKR202	pACY177 derivative; Gm ^r	4
pRK213	tra(RK2) ⁺ Km ^r ColE1	6
pET16b	P _{T7} P _{lac} Amp ^r ColE1; His tag	Novagen
pET22b	P _{T7} P _{lac} Amp ^r ColE1 pelB signal sequence; His tag	Novagen
pG-TF2	P _{Tc} groES groEL tig Cm ^r	Takara Bio Inc. (31)
pBBRBFDPp	Gm ^r ; pBBR1MCS-5 derivative carrying the <i>mdlC</i> gene from <i>P. putida</i> ATCC 12633	This study
pBBRBFDPa	Gm ^r ; pBBR1MCS-5 derivative carrying the <i>mdlC</i> gene from <i>P. aeruginosa</i> PAO1	This study
pSUP Δ MDLC	Plasmid based on pSUP202 for deletion of <i>mdlC</i> from <i>P. putida</i> ATCC 12633; Amp ^r Cm ^r Tc ^r Gm ^r	This study

^a Abbreviations for antibiotics: Ap, ampicillin; Cm, chloramphenicol; Km, kanamycin; Gm, gentamicin; Tc, tetracycline.

for triparental conjugation. *E. coli* strains were cultured at 37°C and *Pseudomonas* strains at 30°C in solid or liquid Luria-Bertani medium. *Pseudomonas* strains displaying benzoylformate decarboxylase activity were selected by growth in liquid or solid minimal medium M9 (38) without glucose supplemented with 10 mM benzoylformate (purchased from Sigma-Aldrich) as the sole carbon source. The bacteria were incubated for at least 2 days at 30°C. When necessary, antibiotics were added at the following concentrations: chloramphenicol, 600 μ g/ml; kanamycin, 50 μ g/ml; and gentamicin, 30 μ g/ml for *Pseudomonas* strains, and chloramphenicol, 50 μ g/ml; kanamycin, 25 μ g/ml; gentamicin, 10 μ g/ml; and ampicillin, 100 μ g/ml for *E. coli* strains.

Construction of the deletion mutant *P. putida* Δ *mdlC*. Genomic DNA from *P. putida* ATCC 12633 was used as a template for the amplification of a 500-bp fragment located at either the 5' (*mdlC'*) or the 3' (*'mdlC*) end of the 1,584-bp *mdlC* gene coding for benzoylformate decarboxylase. The following primers were used: BFDanfu (5'-ATA TGG ATC CAA GCT TAC ATG GCG ATC AAA AAG GTG G-3'), BFDanfdw (5'-ATA TCT CGA GGT CGA CGG CAC CAC ATA CGA ACT CTT-3'), BFDendup (5'-ATA TCC TAG AGT CGA CGG CCT TTG GCA GAA AGC GCT T-3'), and BFDenddw (5'-ATA TGG ATC CCC CGC GAA GGT TGA CCA AGA CGC TGG C-3').

A gentamicin resistance cassette from plasmid pWKR202 was inserted between these two 500-bp fragments of *mdlC* by using the restriction site HindIII. The cloning steps were carried out in *E. coli* DH5 α by using the plasmid pBluescript SK(+). The mutagenesis cassette consisted of three different fragments in the order *mdlC'*, gentamicin resistance cassette, and *'mdlC*. Subsequently, the mutagenesis cassette was cloned into the suicide vector pSUP202, and the resulting plasmid, pSUP Δ MDLC, was transferred via triparental conjugation (6) into *P. putida* ATCC 12633. The benzoylformate decarboxylase gene, *mdlC*, was interrupted by homologous recombination, and the correct insertion of the mutagenesis cassette was confirmed by PCR analysis.

Construction of a genomic-DNA library from *P. putida* ATCC 12633 and *P. putida* ATCC 12633 Δ *mdlC*. The genomic DNAs from *P. putida* ATCC 12633 and *P. putida* ATCC 12633 Δ *mdlC* were isolated by a standard procedure (38) and partially digested with the restriction endonuclease Sau3AI. The broad-host-range vector pBBR1MCS was restricted with BamHI and used for cloning of genomic-DNA fragments of 2 to 8 kb. Finally, the genomic-DNA library was transferred into *P. putida* KT2440 by conjugation. The average insert size and the

frequency of clones harboring only vector DNA within each library were determined by restriction analysis of 24 clones.

Cloning and expression of the *mdlC* genes from *P. putida* ATCC 12633 and *P. aeruginosa* PAO1. Both *mdlC* genes were amplified using genomic DNA from *P. putida* ATCC 12633 and *P. aeruginosa*, respectively. The following primers were used: BFDpup (5'-ATA TCCATG GCT TCG GTA CAC GGC ACC ACA TAC-3'), BFDpdpw (5'-ATA TCT CGA GTC ACT TCA CCG GGC TTA CGG TGC TTA C-3'), BFDpup (5'-ATA TCA TAT GAA AAC CGT CCA TTC CGC G-3'), and BFDpadw (5'-ATA TAA GCT TTC AGG GTT CGA TGG TTT GCG-3'). The primers were designed for cloning into plasmid pBBR1MCS, which was used for subsequent screening of genomic or metagenomic libraries. Both PCR products were ligated into SmaI-restricted pBBR1MCS-2, conferring kanamycin resistance. The plasmids were designated pBBRBFDPp for the *mdlC* gene from *P. putida* ATCC 12633 and pBBRBFDPa for the *mdlC* gene from *P. aeruginosa*. The correct integration of the genes was confirmed by DNA sequencing. Subsequently, the recombinant plasmids were transferred to *P. putida* KT2440 and analyzed for growth on agar plates containing 10 mM benzoylformate as the sole carbon source.

Cloning and expression of novel *bfd* genes from *P. putida* ATCC 12633. Putative *bfd* genes were amplified using genomic DNA from *P. putida* ATCC 12633. The following primers for cloning into expression plasmids pET16b and pET22b (Novagen) were used: BFD3NdepET (5'-ATA TCA TAT GAA AAC TGT TCA CGG CGC CAC-3'), BFD3BampET (5'-ATA TGG ATC CGG GCT CGA TGG TCT GGG TCG-3'), and BFD2BamHioSTC (5'-TGG CCT TGA GGA TCC GCG GCT GCT-3'). The PCR products were cloned into the NdeI/BamHI-digested expression vectors pET16b and pET22b (Novagen). The correct integration of *bfdB* and *bfdC* was confirmed by DNA sequencing. The heterologous expression of both genes was accomplished in *E. coli* BL21(DE3), with coexpression of chaperones from plasmid pG-TF2 (31), as recommended by the manufacturer (Takara Bio Inc.).

Isolation and cloning of DNA from soil. A soil sample was collected from a meadow near Juelich, Germany, and the metagenomic DNA was isolated from 5 g soil based on the direct-lysis method of Zhou et al. (49). Additionally, the isolated metagenomic DNA was purified using the DNeasy Tissue Kit (QIAGEN), starting the protocol with the washing step. The purified metagenomic DNA was partially digested with the restriction endonuclease Sau3AI.

Fragments of 2 to 9 kb were isolated and ligated into the BamHI-restricted broad-host-range vector pBBR1MCS. The resulting metagenomic-DNA library was transferred to *P. putida* KT2440 by conjugation. The average insert size and the frequency of clones harboring only vector DNA within each library were determined by restriction analysis of 24 plasmids isolated from randomly chosen clones.

Cloning and expression of the novel BFD gene *bfdM* from a metagenomic library. The plasmid DNA from a clone exhibiting decarboxylase activity was sequenced by genome walking, and the following primers were used for the amplification of the novel *bfdM* gene and subsequent cloning into the expression vector pET22b (Novagen): BFDM1Nde (5'-ATA TCA TAT GCA AGA GAC AAC CCC CCA GAA T-3') and BFDM1BamH (5'-ATA GGA TCC GGC CAC TTC GAC GAG CAC GGG C-3'). The PCR product was cloned into the NdeI/BamHI-digested expression vector pET22b (Novagen), and the correct integration of *bfdM* was confirmed by DNA sequencing. The heterologous expression was accomplished as described for *bfdB* and *bfdC*.

Coupled decarboxylase assay. The decarboxylase activity toward benzoylformate was determined as described previously (17). In brief, the assay mixture was prepared from the following stock solutions in standard buffer: benzoylformate solution (100 μ l, 50 mM, adjusted to pH 6.0), NADH (100 μ l, 3.5 mM), horse liver alcohol dehydrogenase (50 μ l, 10 U; Sigma), and potassium phosphate buffer (700 μ l, 50 mM, pH 6.0). The components were mixed in a 1.7-ml cuvette and incubated for 3 min at 30°C, and the reaction was started by the addition of sonicated cell extracts from overexpression cultures. These extracts were prepared as follows: 100 μ l (equal to an optical density of 20 at 580 nm) supernatant of extracts centrifuged at 13,000 rpm for 20 min was sonicated twice for 3 min each time (50 cycles; 80% power; Sonoplus HD2070; Bandelin, Berlin, Germany). The slope was calculated from the linear part of the descending absorption curve determined at 340 nm and within a reaction time of 15 to 90 s.

Sequence analysis. The similarity of newly identified genes to known DNA or amino acid sequences was determined by Blastn or Blastp (<http://www.ncbi.nlm.nih.gov/BLAST/>). The sequence alignment was performed with ClustalW (<http://www.ebi.ac.uk/clustalw/>).

Nucleotide sequence accession numbers. The sequences of the three novel benzoylformate decarboxylase genes described here have been deposited in the EMBL database. They are AM284966 for *bfdB*, AM284967 for *bfdC*, and AM284968 for *bfdM*.

RESULTS AND DISCUSSION

Construction of a *Pseudomonas putida* selection strain. The elimination of a functional *mdlC* gene in *P. putida* ATCC 12633 interrupts the mandelate pathway and prevents the corresponding knockout mutant, *P. putida* ATCC 12633 Δ *mdlC*, from growing on mandelate as the sole carbon source. Consequently, this strain should be usable as a selection host to identify novel benzoylformate decarboxylase genes that can be introduced in *trans* and that may originate from a genomic, a metagenomic, or a directed-evolution library. The deletion mutant *P. putida* ATCC 12633 Δ *mdlC* was constructed by the insertion of a gentamicin resistance cassette into the functional gene *mdlC* and was cultivated on selective medium containing benzoylformate as the sole carbon source, enabling growth of the recombinant bacteria only upon expression of a functional plasmid-encoded BFD. Surprisingly, the mutant *P. putida* ATCC 12633 Δ *mdlC* was still able to grow on benzoylformate-containing medium. Therefore, this strain could not be used as a selection host; however, this result clearly indicated either the presence of so far unknown genes encoding enzymes with BFD activity or an unknown pathway for the degradation of benzoylformate.

Thus, we searched for bacterial strains having the ability to grow with benzaldehyde as the sole carbon source but lacking the enzymes catalyzing the reactions from (R)-mandelate to benzoylformate (Fig. 1). These bacteria are expected to possess a benzaldehyde dehydrogenase and to metabolize benzal-

FIG. 2. Growth of recombinant *P. putida* KT2440 displaying benzoylformate decarboxylase activity on an agar plate containing benzoylformate selective medium. *P. putida* KT2440 possessing the plasmids pBFDB (coding for BfdB from the genomic-DNA library of *P. putida* ATCC 12633), pBFDC (coding for BfdC from the genomic-DNA library of *P. putida* ATCC 12633 Δ *mdlC*), and pBFDM (coding for BfdM from a metagenomic library) expressed active BFD enzymes enabling growth by converting benzoylformate. *P. putida* strains harboring the corresponding plasmids, pBBRCm^R, pBBRGm^R, and pBBRKm^R, served as controls.

dehyde via the β -ketoacid pathway to yield the products acetyl-CoA and succinyl-CoA. A database analysis revealed that the strains *P. putida* KT2440 and *P. putida* DSM50198 should exhibit these features (21, 30). Furthermore, Tsou et al. described the presence of a β -ketoacid pathway in *P. putida* DSM50198, which lacks genes of the mandelate pathway (47). First, we confirmed that both strains were unable to grow in selective medium containing benzoylformate as the sole carbon source. As a functional test, we then cloned into the broad-host-range vector pBBR1MCS the *mdlC* genes encoding BFDs from both *P. aeruginosa* PAO1 and *P. putida* ATCC 12633. The corresponding plasmids were transferred into the putative selection host strains, *P. putida* KT2440 and *P. putida* DSM50198, and both recombinant strains were able to grow on benzoylformate-containing medium, whereas the control strains, which lack the BFD genes, did not grow at all (data not shown). These results clearly indicated that both *P. putida* strains were useful as selection hosts.

Identification of two novel BFD-encoding genes in the genome of *P. putida* ATCC 12633. Our results suggested that *P. putida* ATCC 12633 may possess at least one so far unknown gene encoding a BFD activity. We therefore constructed DNA libraries from chromosomal DNAs of *P. putida* ATCC 12633 and the deletion mutant *P. putida* ATCC 12633 Δ *mdlC* using the plasmid pBBR1MCS. The resulting genomic-DNA libraries consisted of 4,500 and 4,000 clones for the *P. putida* wild type and the *mdlC* mutant, respectively, both with an average insert size of 5 kb, thereby reaching a fourfold coverage of the *P. putida* genome. Initially, these libraries were constructed in *E. coli* and then transferred to *P. putida* KT2440 by triparental mating. We identified several clones that grew on benzoylformate selective media. The corresponding plasmids were isolated, and restriction analysis revealed two different fragment patterns. Subsequent DNA sequencing identified two novel genes, which were designated *bfdB* and *bfdC*. The two genes

TABLE 2. Plasmids encoding putative benzoylformate decarboxylases, other ORFs identified, and their closest homologues as identified from databases^a

Plasmid name (insert size [kb])	ORF	Closest database similarity		
		Protein (accession no.)	Organism	Identity (%)
pBFDB (6.234)	BkdA1	Ketoacid dehydrogenase E1 α -subunit (AAA 65614)	<i>Pseudomonas putida</i>	100
	BkdR	Transcription regulator (AAA 65613)	<i>Pseudomonas putida</i>	100
	Orf3	α/β Hydrolase (ZP_00898945)	<i>Pseudomonas putida</i>	95
	Orf4	Glutamin synthetase (ZP_00900555)	<i>Pseudomonas putida</i>	95
	BenC	Benzoate dioxygenase (AAF 63450)	<i>Pseudomonas putida</i>	85
	BfdB	Benzoylformate decarboxylase (NP_253588)	<i>Pseudomonas aeruginosa</i>	83
pBFDC (2.952)	Orf1'	Putative LysR regulator (AAL 27559)	<i>Pseudomonas putida</i>	100
	BfdC	Benzoylformate decarboxylase (NP_253588)	<i>Pseudomonas aeruginosa</i>	83
	Orf3'	Probable MFS ^b transporter (NP-253587)	<i>Pseudomonas aeruginosa</i>	82
pBFDM (6.586)	Orf1''	Short-chain dehydrogenase (ZP_00243984)	<i>Rubrivivax gelatinosus</i>	65
	Orf2''	Transcriptional regulatory protein (ZP-01022169)	<i>Polaromonas naphthalenivorans</i>	60
	BfdM	Benzoylformate decarboxylase (ZP-01022170)	<i>Polaromonas naphthalenivorans</i>	72
	Orf4''	Hypothetical protein (ZP_01022174)	<i>Polaromonas naphthalenivorans</i>	60
	Orf5''	Sigma e factor sigma 24 (ZP_01022480)	<i>Polaromonas naphthalenivorans</i>	60

^a The diagram shows the organization of *P. putida* genomic (pBFDB and pBFDC) and environmental (pBFDM) DNAs present in plasmids that conferred benzoylformate decarboxylase activity on *P. putida*. ORFs encoding benzoylformate decarboxylase activity are shown as black arrows; partial arrows represent incomplete ORFs.

^b MFS, major facilitator superfamily.

share identical DNA sequences, except for those nucleotides encoding the C-terminal nine amino acids. The clones expressing these genes were able to grow on benzoylformate selective media, whereas the respective control strain showed no growth at all (Fig. 2). The two BFD-encoding genes are located in different regions of the *P. putida* ATCC 12633 genome, as deduced from the sequences of the respective flanking regions (Table 2). The genes up- and downstream of *bfdC* resemble the genomic region surrounding *mdlC* in *P. aeruginosa*, where a putative benzoate transporter of the major facilitator superfamily and a putative transcriptional regulator are encoded in the same orientation. Thus, the *bfdC* gene seems to be part of a gene cluster encoding proteins for the degradation of aromatic compounds, as suggested for *mdlC* of *P. aeruginosa* (44). On the other hand, *bfdB* seems more likely to have originated from a gene transfer. To our knowledge, the region surrounding this gene does not show any homology to known regions adjacent to other BFD genes identified in published bacterial genomes. The enzymes BfdB and BfdC exhibit 83% identity to the benzoylformate decarboxylase from *P. aeruginosa* and 63% to the previously described enzyme *mdlC* from *P. putida* ATCC 12633. These results demonstrate that (i) *P. putida*

KT2440 is an appropriate selection host strain and (ii) *P. putida* ATCC 12633 possesses two additional hitherto-unknown BFD-encoding genes, *bfdB* and *bfdC*, one of which was most likely duplicated during evolution.

Construction and screening of a metagenomic library. The successful identification of novel biocatalysts exhibiting carboxylase activities from environmental DNA would provide additional support for a more general applicability of the growth selection system we have constructed. Therefore, we isolated metagenomic DNA directly from soil and constructed plasmid-based libraries in *P. putida* which were subsequently screened for BFD activity. The randomly digested metagenomic DNA was cloned into pBBR1MCS, transferred into *E. coli*, and finally propagated in *P. putida* KT2440. The metagenome library consisted of about 14,000 clones, with an average insert size of 2 to 10 kb. Plating of *P. putida* KT2440 on benzoylformate selective medium revealed one clone which was able to grow after 2 days of incubation (Fig. 2). The respective plasmid contained an insert with a size of 7 kb. DNA sequence determination revealed an open reading frame (ORF), which was designated *bfdM*. The deduced protein exhibits 72% identity to

FIG. 3. Sequence alignment of MdlC (BfdA), BfdB, and BfdC from *P. putida* ATCC 12633 (Pp), BFD from *Pseudomonas aeruginosa* (Pa), and BfdM isolated from the metagenome. Highly conserved regions are highlighted in black, and similar amino acid residues are shown in gray. Amino acids forming the catalytic site are indicated by an asterisk. Residues tagged with a dot are involved in substrate binding (34, 40), and the boxed residues are involved in thiamine diphosphate cofactor and metal binding.

the amino acid sequence of a putative benzoylformate decarboxylase from *Polaromonas naphthalenivorans* CJ2T (Table 2). We further identified additional DNA regions upstream and downstream of *bfdM* which also showed similarity to genes from this gram-negative bacterium, including a putative transcriptional regulatory protein and a putative sigma factor. Recently, *P. naphthalenivorans* CJ2T was isolated from a coal tar-contaminated freshwater sediment, and it was demonstrated that this strain could grow on the polycyclic aromatic hydrocarbon naphthalene as the sole carbon and energy source (20). These findings suggest that *bfdM* may originate from an organism that shares the ability to grow on aromatic compounds.

Overexpression and characterization of the novel BFDs. The primary structures of the three novel BFDs we have identified were analyzed with respect to conserved regions, including active and cofactor binding sites. Figure 3 shows a sequence alignment of the novel enzymes with BFDs from *P. putida* KT244 and *P. aeruginosa* (MdlC), which shows the highest identity to BfdB and BfdC. All amino acid side chains forming the active site, the substrate binding site, and

the thiamine diphosphate binding site are highly conserved in the new enzymes. This also holds for the metagenome-derived enzyme BfdM, indicating that the reaction mechanism of benzoylformate decarboxylases is well conserved.

All three enzymes were detected by an activity screen toward the substrate benzoylformate upon expression in the homologous host *P. putida* KT2440; however, in all cases, the amounts of active enzyme produced were very low. The overexpression of the three putative *bfd* genes in *E. coli* BL21(DE3) under the control of the T7 promoter did not result in the formation of enzymatically active enzymes, presumably because inclusion bodies were formed. Therefore, we coexpressed the chaperones GroEL, GroES, and trigger factor from plasmid pG-TF2, trying to prevent the formation of inclusion bodies as described previously (31). While BfdB and BfdM remained inactive, BfdC now displayed an enzymatic activity of 22 U/ml with benzoylformate as the substrate. Interestingly, BfdC and BfdB are nearly identical, except for the number and type of the carboxy-terminal amino acids, which therefore seem to determine the functional expression of benzoylformate decarboxylases (Fig. 3).

FIG. 4. Growth selection on benzaldehyde to identify industrially important biocatalysts from genomic- or metagenomic-DNA libraries. Accessible enzymes include (i) alcohol dehydrogenase, (ii) hydroxynitrile lyase, (iii) benzoylformate decarboxylase, and (iv) benzaldehyde lyase.

Conclusions. We have described here an efficient bacterial selection system which is based on a *P. putida* strain growing with benzaldehyde as the sole carbon and energy source. The selective capacity of this system was demonstrated by the successful identification of three novel BFDs, using benzoylformate as a model substrate. Apart from identifying BFDs, this system is more generally applicable, because benzaldehyde is not only part of the mandelate pathway but is also formed during phenylalanine degradation (25), in the veratryl alcohol pathway (19), from toluene degradation (32) and by cyanogenesis in plants (5). Therefore, the system we have constructed can be used to identify a broad range of different enzymes, including the industrially important enzymes benzaldehyde lyase, benzoylformate decarboxylase (33), hydroxynitrile lyase (5), and alcohol dehydrogenase, which all produce benzaldehyde by conversion of benzoin, benzoylformate, mandelonitrile, or benzyl alcohol, respectively (Fig. 4).

Furthermore, as most enzymes also catalyze reverse reactions, this method can presumably be employed to identify enzymes that use benzaldehyde as the substrate. This compound serves as a precursor to synthesize important chiral building blocks for organic chemistry, e.g., mandelonitrile or ephedrine. If relaxed substrate specificities and putative moonlighting activities of enzymes are taken into account, a rather broad synthetic-substrate spectrum should be accessible by the biocatalysts newly discovered with this selection method. Furthermore, the method is directly linked to the activity of the respective enzyme, thereby constituting a powerful tool to (i) identify new enzymes from very large libraries and (ii) provide experimental proof of enzymatic activities of putative enzyme-encoding ORFs identified within numerous genome and metagenome sequencing projects.

ACKNOWLEDGMENTS

This work was supported by the Federal Ministry of Education and Research, Berlin, Germany, in the framework of the Competence Network Göttingen entitled "Genome Research on Bacteria for the

Analysis of Biodiversity and Its Further Use for the Development of New Production Processes."

We thank Gerhard Gottschalk, Göttingen Genomics Laboratory, Göttingen, Germany, for continuous support.

REFERENCES

- Baker, K., C. Bleczinski, H. Lin, G. Salzar-Jimenez, D. Sengupta, S. Kranse, and V. W. Cornish. 2002. Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc. Natl. Acad. Sci. USA* **99**: 16537–16542.
- Barrowman, M. M., and C. A. Fewson. 1985. Phenylglyoxylate decarboxylase and phenylpyruvate decarboxylase from *Acinetobacter calcoaceticus*. *Curr. Microbiol.* **12**:235–239.
- Barrowman, M. M., W. Harnett, A. J. Scott, C. A. Fewson, and J. R. Kusel. 1986. Immunological comparison of microbial TPP-dependent nonoxidative alpha-keto acid decarboxylases. *FEMS Microbiol. Lett.* **34**:57–60.
- Buskirk, A. R., A. Landrigan, and D. R. Liu. 2004. Engineering a ligand-dependent RNA transcriptional activator. *Chem. Biol.* **11**:1157–1163.
- Drepper, T., S. Gross, A. F. Yakunin, P. C. Hallenbeck, B. Masepohl, and W. Klipp. 2003. Role of GlnB and GlnK in ammonium control of both nitrogenase systems in the phototrophic bacterium *Rhodospirillum rubrum*. *Microbiology* **149**:2203–2212.
- Fechter, M. H., and H. Griengl. 2004. Hydroxynitrile lyases: biological sources and application as biocatalysts. *Food Technol. Biotechnol.* **42**:287–294.
- Figurski, D. H., and D. R. Helinski. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. *Proc. Natl. Acad. Sci. USA* **76**:1648–1652.
- Gamper, M., D. Hilvert, and P. Kast. 2000. Probing the role of the C-terminus of *Bacillus subtilis* chorismate mutase by a novel random protein-termination strategy. *Biochemistry* **39**:14087–14094.
- Griffiths, J. S., M. Cheriyan, J. B. Corbell, L. Pocivavsek, C. A. Fierke, and E. J. Toone. 2004. A bacterial selection for the directed evolution of pyruvate aldolases. *Bioorg. Med. Chem.* **12**:4067–4074.
- Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**:557–580.
- Harwood, C. S., and R. E. Parales. 1996. The beta-ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* **50**:553–590.
- Hasson, M. S., A. Muscate, M. J. McLeish, L. S. Polovnikova, J. A. Gerlt, G. L. Kenyon, G. A. Petsko, and D. Ringe. 1998. The crystal structure of benzoylformate decarboxylase at 1.6 Å resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes. *Biochemistry* **37**:9918–9930.
- Hegeman, G. D. 1970. Methods Enzymol. **17**:674–678.
- Hegeman, G. D. 1966. Synthesis of the enzymes of the mandelate pathway by *Pseudomonas putida*. III. Isolation and properties of constitutive mutants. *J. Bacteriol.* **91**:1161–1167.
- Hegeman, G. D. 1966. Synthesis of the enzymes of the mandelate pathway by *Pseudomonas putida*. I. Synthesis of enzymes by the wild type. *J. Bacteriol.* **91**:1140–1154.
- Hegeman, G. D. 1966. Synthesis of the enzymes of the mandelate pathway by *Pseudomonas putida*. II. Isolation and properties of blocked mutants. *J. Bacteriol.* **91**:1155–1160.
- Holloway, B. W., V. Krishnapillai, and A. F. Morgan. 1979. Chromosomal genetics of *Pseudomonas*. *Microbiol. Rev.* **43**:73–102.
- Iding, H., T. Dünwald, L. Greiner, A. Liese, M. Müller, P. Siegert, J. Grötzinger, A. S. Demir, and M. Pohl. 2000. Benzoylformate decarboxylase from *Pseudomonas putida* as stable catalyst for the synthesis of chiral 2-hydroxy ketones. *Chem. Eur. J.* **6**:1483–1495.
- Jaeger, K. E. 2004. Protein technologies and commercial enzymes: white is the hype—biocatalysts on the move. *Curr. Opin. Biotechnol.* **15**:269–271.
- Jensen, K. A., K. M. Evans, T. K. Kirk, and K. E. Hammel. 1994. Biosynthetic pathway for veratryl alcohol in the ligninolytic fungus *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.* **60**:709–714.
- Jeon, C. O., W. Park, W. C. Ghiorse, and E. L. Madsen. 2004. *Polaromonas naphthalenivorans* sp. nov., a naphthalene-degrading bacterium from naphthalene-contaminated sediment. *Int. J. Syst. Evol. Microbiol.* **54**:93–97.
- Jimenez, J. I., B. Minambres, J. L. Garcia, and E. Diaz. 2002. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**:824–841.
- Konarzyska-Bessler, M., and K. E. Jaeger. 2006. Select the best: novel biocatalysts for industrial applications. *Trends Biotechnol.* **24**:248–250.
- Kovach, M. E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop II, and K. M. Peterson. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**:175–176.
- Kovach, M. E., R. W. Phillips, P. H. Elzer, R. M. Roop II, and K. M. Peterson. 1994. pBBR1MCS: a broad-host-range cloning vector. *Bio-Techniques* **16**:800–802.
- Lapadatescu, C., C. Ginies, J. L. Le Quere, and P. Bonnamme. 2000. Novel

- scheme for biosynthesis of aryl metabolites from L-phenylalanine in the fungus *Bjerkandera adusta*. *Appl. Environ. Microbiol.* **66**:1517–1522.
26. Lingen, B., J. Grötzinger, D. Kolter, M. R. Kula, and M. Pohl. 2002. Improving the carbonylase activity of benzoylformate decarboxylase from *Pseudomonas putida* by a combination of directed evolution and site-directed mutagenesis. *Protein Eng.* **15**:585–593.
 27. Lingen, B., D. Kolter-Jung, P. Duenkelmann, R. Feldmann, J. Grötzinger, M. Pohl, and M. Müller. 2003. Alteration of the substrate specificity of benzoylformate decarboxylase from *Pseudomonas putida* by directed evolution. *Chembiochemistry* **4**:721–726.
 28. MacBeath, G., P. Kast, and D. Hilvert. 1998. Redesigning enzyme topology by directed evolution. *Science* **279**:1958–1961.
 29. Mohn, W. W., J. Garmendia, T. C. Galvao, and V. de Lorenzo. 2006. Surveying biotransformations with a la carte genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into *lacZ*-based phenotypes. *Environ. Microbiol.* **8**:546–555.
 30. Nelson, K. E., C. Weinel, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A. Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanan, R. T. DeBoy, S. Daugherty, J. Kolonay, R. Madupu, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. Chris Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzes, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Wedler, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. A. Eisen, K. N. Timmis, A. Düsterhöft, B. Tümmeler, and C. M. Fraser. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**:799–808.
 31. Nishihara, K., M. Kanemori, H. Yanagi, and T. Yura. 2000. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl. Environ. Microbiol.* **66**:884–889.
 32. Panke, S., J. M. Sanchez-Romero, and V. de Lorenzo. 1998. Engineering of quasi-natural *Pseudomonas putida* strains for toluene metabolism through an *ortho*-cleavage degradation pathway. *Appl. Environ. Microbiol.* **64**:748–751.
 33. Pohl, M., B. Lingen, and M. Müller. 2002. Thiamin-diphosphate-dependent enzymes: new aspects of asymmetric C-C bond formation. *Chem. Eur. J.* **8**:5288–5295.
 34. Polovnikova, E. S., M. J. McLeish, E. A. Sergienko, J. T. Burgner, N. L. Anderson, A. K. Bera, F. Jordan, G. L. Kenyon, and M. S. Hasson. 2003. Structural and kinetic analysis of catalysis by a thiamin diphosphate-dependent enzyme, benzoylformate decarboxylase. *Biochemistry* **42**:1820–1830.
 35. Regenhardt, D., H. Heuer, S. Heim, D. U. Fernandez, C. Strompl, E. R. Moore, and K. N. Timmis. 2002. Pedigree and taxonomic credentials of *Pseudomonas putida* strain KT2440. *Environ. Microbiol.* **4**:912–915.
 36. Robertson, D. E., J. A. Chaplin, G. DeSantis, M. Podar, M. Madden, E. Chi, T. Richardson, A. Milan, M. Miller, D. P. Weiner, K. Wong, J. McQuaid, B. Farwell, L. A. Preston, X. Tan, M. A. Snead, M. Keller, E. Mathur, P. L. Kretz, M. J. Burk, and J. M. Short. 2004. Exploring nitrilase sequence space for enantioselective catalysis. *Appl. Environ. Microbiol.* **70**:2429–2436.
 37. Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**:2541–2547.
 38. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 39. Schwimmer, L. J., P. Rohatgi, B. Azizi, K. L. Seley, and D. F. Doyle. 2004. Creation and discovery of ligand-receptor pairs for transcriptional control with small molecules. *Proc. Natl. Acad. Sci. USA* **101**:14707–14712.
 40. Siegert, P., M. J. McLeish, M. Baumann, H. Iding, M. M. Kneen, G. L. Kenyon, and M. Pohl. 2005. Exchanging the substrate specificities of pyruvate decarboxylase from *Zymomonas mobilis* and benzoylformate decarboxylase from *Pseudomonas putida*. *Protein Eng. Des. Sel.* **18**:345–357.
 41. Simon, R., U. Priefer, and A. Pühler. 1983. A broad host range mobilization system for *in vivo* genetic-engineering—transposon mutagenesis in gram-negative bacteria. *Biotechnology* **1**:784–791.
 42. Stanier, R. Y. 1947. Acetic acid production from ethanol by fluorescent pseudomonads. *J. Bacteriol.* **54**:191–194.
 43. Stanier, R. Y. 1947. Simultaneous adaptation: a new technique for the study of metabolic pathways. *J. Bacteriol.* **54**:339–348.
 44. Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrook-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* **406**:959–964.
 45. Studier, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219**:37–47.
 46. Studier, F. W., and B. A. Moffatt. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. *J. Mol. Biol.* **189**:113–130.
 47. Tsou, A. Y., S. C. Ransom, J. A. Gerlt, D. D. Buechter, P. C. Babbitt, and G. L. Kenyon. 1990. Mandelate pathway of *Pseudomonas putida*: sequence relationships involving mandelate racemase, (S)-mandelate dehydrogenase, and benzoylformate decarboxylase and expression of benzoylformate decarboxylase in *Escherichia coli*. *Biochemistry* **29**:9856–9862.
 48. van Sint Fiet, S., J. B. van Beilen, and B. Witholt. 2006. Selection of biocatalysts for chemical synthesis. *Proc. Natl. Acad. Sci. USA* **103**:1693–1698.
 49. Zhou, J., M. A. Bruns, and J. M. Tiedje. 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**:316–322.