Appl Microbiol Biotechnol (2004) 65: 168–176
DOI 10.1007/s00253-004-1589-3

Horst Priefert · Xian M. O'Brien · Philip A. Lessard · Annette F. Dexter · Ellen E. Choi · Sladjana Tomic · Geeta Nagpal · Jennie J. Cho · Melina Agosto · Lucy Yang · Sheri L. Treadway · Lance Tamashiro · Matthew Wallace · Anthony J. Sinskey

Indene bioconversion by a toluene inducible dioxygenase of *Rhodococcus* sp. I24

Received: 14 November 2003 / Revised: 29 January 2004 / Accepted: 6 February 2004 / Published online: 7 April 2004

Abstract *Rhodococcus* sp. I24 can oxygenate indene via at least three independent enzyme activities: (i) a naphthalene inducible monooxygenase (ii) a naphthalene inducible dioxygenase, and (iii) a toluene inducible dioxygenase (TID). Pulsed field gel analysis revealed that the I24 strain harbors two megaplasmids of ~340 and ~50 kb. *Rhodococcus* sp. KY1, a derivative of the I24 strain, lacks the ~340 kb element as well as the TID activity. Southern blotting and sequence analysis of an indigogenic, I24-derived cosmid suggested that an operon encoding a TID resides on the ~340 kb element. Expression of the *tid* operon was induced by toluene but not by naphthalene. In contrast, naphthalene did induce expression of the *nid* operon, encoding the naphthalene dioxygenase in I24. Cell free protein extracts of *Escherichia coli* cells expressing *tidABCD* were used in HPLC-based enzyme assays to characterize the indene bioconversion of TID in vitro. In addition to 1-indenol, indene was transformed to *cis*-indandiol with an enantiomeric excess of 45.2% of *cis*-(1*S*,2*R*)-indandiol over *cis*-(1*R*,2*S*)-indandiol, as revealed by chiral HPLC analysis. The K_m of TID for indene was 380 μM. The enzyme also dioxygenated naphthalene to *cis*-dihydronaphthalenediol with an activity of 78% compared to the formation of *cis*-indandiol from indene. The K_m of TID for naphthalene was 28 μM. TID converted only trace amounts of toluene to 1,2-dihydro-3-methylcatechol after prolonged incubation time. The results indicate the role of the *tid* operon in the bioconversion of indene to 1-indenol and *cis*-(1*S*,2*R*)-indandiol by *Rhodococcus* sp. I24.

H. Priefert · X. M. O'Brien · P. A. Lessard · A. F. Dexter · E. E. Choi · S. Tomic · G. Nagpal · J. J. Cho · S. L. Treadway · A. J. Sinskey (✉)
Department of Biology, 68-370, Massachusetts Institute of Technology,
77 Massachusetts Avenue,
Cambridge, MA 02139, USA
e-mail: asinskey@mit.edu
Tel.: +1-617-2536721
Fax: +1-617-2538550

P. A. Lessard · G. Nagpal · M. Agosto · L. Yang · L. Tamashiro · M. Wallace · A. J. Sinskey
Microbial Genetics Project Laboratory, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA

Introduction

Bacteria from the genus *Rhodococcus* have been found to metabolize a wide variety of environmental pollutants including alkanes, aromatic compounds and halogenated hydrocarbons (Bell et al. 1998). In addition to their value in bioremediation, rhodococci show promise as biocatalysts in the synthesis of chiral compounds such as derivatives of indene. These properties are due to a battery of oxygenases, enzymes that catalyze the stereoselective incorporation of oxygen into hydrocarbons.

Buckland et al. (1999), and Chartrain et al. (1998), identified a *Rhodococcus* isolate, designated I24, that could metabolize toluene and naphthalene. Although this strain could not grow on indene as a sole carbon source, the enzymes responsible for naphthalene and toluene oxygenation in I24 are also able to oxygenate indene. From indene, this strain can generate a variety of indandiols, some of which have value in the production of the chiral drug, indinavir (Buckland et al. 1999). In the I24 strain, at least three independent enzyme systems are supposed to be responsible for the conversion of indene to the different indandiols. The first of these is a naphthalene inducible dioxygenase (NID) that predominantly generates *cis*-(1*R*,2*S*)-indandiol. The second is a toluene inducible dioxygenase (TID) that is thought to predominantly generate the *cis*-(1*S*,2*R*) enantiomer. The third is a naphthalene inducible monooxygenase (NIM) that most probably generates an epoxide derivative of indene, which can then be resolved to either *trans*-(1*R*,2*R*)-indandiol or *cis*-(1*S*,2*R*)-indandiol (Chartrain et al. 1998; Stafford et al. 2002). Chartrain et al. (1998) found that these activities could also be induced to some extent by indene alone.

To examine the genetic basis of indene bioconversion, Treadway et al. (1999) screened a cosmid library derived from the I24 strain for indigogenic activity. This led to the

identification of genes that apparently encode the NID from I24 (now called the *nid* operon). When transferred to a naïve host, the *nid* operon enables the oxygenation of indene to *cis*-(1*R*,2*S*)-indandiol.

In this paper we describe the discovery of genes encoding the TID from the I24 strain. This new operon resides on a large, ~340 kb megaplasmid, and it is homologous to toluene-type dioxygenases from other organisms. Expression of this operon can be induced by toluene or indene in the I24 strain. This operon was heterologously expressed in *E. coli* and the capability of the gene products to convert indene to 1-indenol and *cis*-(1*S*,2*R*)-indandiol was confirmed by in vitro enzyme assays and chiral HPLC analysis.

Materials and methods

Strains and culture conditions

A summary of the most critical bacterial strains and plasmids used in this study is presented in Table 1. Additional strains and plasmids are described in the text. Bacteria were routinely cultured in LB medium (Sambrook et al. 1989) except where noted. Medium Rare (Yanagimachi et al. 2001) was used as the defined medium. When necessary the glucose in this medium was omitted to allow for alternative carbon sources. Where noted, antibiotics were included in media at the following concentrations: gentamicin, 10 μg/ml; streptomycin, 50 μg/ml; rifampicin, 10 μg/ml; kanamycin, 100 μg/ml.

DNA manipulation

Sequencing of cosmid DNA

Digesting cosmid 2G11 with *Not*I produced several fragments, ranging in size from 0.4 to 10 kb. Individual *Not*I fragments were ligated into pBluescript II SK+(Stratagene, La Jolla, CA). Plasmid DNA was recovered from *E. coli* using standard methods. The cloned DNA fragments were sequenced at the MIT Biopolymers Lab using the BigDye Terminator Version 2 sequencing reagents (Applied Biosystems, Foster City, CA).

Promoter fusions

The *Not*I subclone used in the sequence analysis was pEC26, which carries the fragment from nucleotide (nt) position 1 to 5788 of the cosmid 2G11 contig (numbering is relative to that in GenBank accession no. AF452376; see also Fig. 2). The *Not*I fragment from pEC26 was inserted into the *Not*I site of pAL282 to create pEC45.

pXO31 was created by transferring the *Sph*I–*Eco*RI fragment (including the *E. coli lacZ* ORF) from pAB1002 (Becker et al. 1995) into the corresponding sites in pAL245 (Stafford et al. 2002). A promoterless *gusA* vector was prepared by amplifying the *gusA* ORF from pGUS:7S 3′ (Lessard et al. 1993) using the polymerase chain reaction (PCR) using primers 3′ GUS and 5′ atgGUS (Table 2). Ligating this product into pCR2.1-TOPO (Invitrogen, Carlsbad, CA) generated pAL272b. The *gusA* ORF was then excised as an *Xho*I–*Sph*I fragment and inserted into the *Xho*I and *Sph*I sites of pXO31, displacing the entire *lacZ* ORF, creating pXO33. A suspected transcriptional terminator from the *Corynebacterium glutamicum ilvBNC* operon (Keilhauer et al. 1993) was amplified from *C. glutamicum* 21253 (American Type Culture Collection, Manassas, VA) by PCR using the primers 3′ term and 5′ term (Table 2). This product was ligated into pCR2.1-TOPO, producing pAL277a. The transcriptional terminator was then excised as a *Mun*I–*Xho*I fragment and inserted into the *Eco*RI and *Xho*I sites of pXO33, creating pAL280, the promoterless *gusA* expression cassette from which the following promoter fusions were derived (see also Fig. 2 for graphical representations of the promoter fragments).

PCR was used to amplify the region immediately upstream of *tidA* using pEC45 as a template and the primers pEC-up and pEC-down, or the primers pEC-uphalf and pEC-down (Table 2). Cloning the former into pCR2.1-TOPO generated pAY1, while cloning the latter into pCR2.1-TOPO produced pAY2. The *Pst*I–*Xho*I fragments from each of these plasmids were inserted into the *Pst*I and *Xho*I sites of pAL280, generating pAY3 and pAY4, respectively.

A portion of the *nid* operon was amplified from cosmid pR4-10 (Treadway et al. 1999) using primers AP21 and AP23 (Table 2) and ligated into pCR2.1-TOPO, producing pST120. The *nidA* upstream sequence was excised from this plasmid as a *Pst*I–*Xho*I fragment and ligated into pAL280 to produce the reporter plasmid pAL296.

Table 1 Key strains and plasmids used in this study

Name	Description[a]	Reference
Strains		
Rhodococcus sp. I24	Wild type strain, naph$^+$, tol$^+$, rif^S, strep^S	Buckland et al. (1999)
Rhodococcus sp. KY1	Derivative of *Rhodococcus* sp. I24, lacking ~340 kb megaplasmid; naph$^+$, tol$^-$	Yanagimachi et al. (2001)
Rhodococcus erythropolis SQ1	"Naïve" strain; naph$^-$, tol$^-$, rif^R, strep^R	Quan and Dabbs (1993)
Rhodococcus erythropolis XO1	Transconjugant derivative of *R. erythropolis* SQ1 carrying ~50 kb megaplasmid from *Rhodococcus* sp. I24; naph$^+$, tol$^-$, rif^R, strep^R	This study
Plasmids		
pAL280	Promoterless *gusA* gene in vector that replicates in both *E. coli* and *Rhodococcus*; gent^R	This study
pAL296	Derivative of pAL280, carrying promoter from *nid* operon	This study
pAY3	Derivative of pAL280, carrying promoter from *tid* operon	This study
pAY4	Derivative of pAL280, carrying truncated promoter from *tid* operon	This study
pSKEB38	*tidABCD* expression plasmid	This study

[a]Abbreviations: naph$^{+/-}$, ability/inability to use naphthalene as a carbon source; tol$^{+/-}$, ability/inability to use toluene as a carbon source; rif$^{R/S}$, rifampicin resistance/sensitivity; strep$^{R/S}$, resistance/sensitivity to streptomycin; gent^R, resistance to gentamicin.

Table 2 Oligonucleotide PCR primers used in this study

Name	Sequence (5′–3′)	Anneals to (in 5′–3′ orientation)[a]
3′ term	CCACCATTTTGCCAAAATCTTGTTGCCT	*C. glutamicum* genomic DNA
3′ GUS	GGCATGCTGTTGATTCATTGTTTGCCT	3′ terminus of *gusA* ORF
5′ atgGUS	CCTCGCGGAGTCCTCCATGGTCCGTCCTGTAGA	5′ terminus of *gusA* ORF
5′ term	CCAATTGCTTAAGTTTCACCCCTTTGA	*C. glutamicum* genomic DNA
AP21	TGTCCAATGCTGATGATGTG	pR4-10 nt 3039–3058
AP23	TGTTGCACGTAGAGCTCGAC	pR4-10 nt 6133–6114
IPBredA	CGTCTGTTCCTGATCGGCAAGGA	2G11 nt 5585–5601
IPBredB	GGGCCCGCATATCGTCGCCGAACCA	2G11 nt 6835–6811
pEC-down	CCTCGAGCATCAGCAACCACGAGCGACCGA	2G11 nt 3334–3312
pEC-up	CCTGCAGGGTACTTCCCATCCTGCCGTCGGA	2G11 nt 2626–2649
pEC-uphalf	CCTGCAGGATCACTATGGATGCCAGCGA	2G11 nt 2986–3006

[a]Nucleotide positions relative to contigs depicted in Fig. 2, unless otherwise indicated.

Southern analysis

Southern hybridizations of pulsed field gels were carried out using the DIG High Prime Labeling and Detection Starter Kit II for Chemiluminescent Detection with CSPD (Roche Molecular Diagnostics, Indianapolis, IN) essentially as described by the manufacturer. Gels were exposed 2 min per side on a short wave UV transilluminator prior to blotting to aid in DNA transfer. Probes were prepared from the *Not*I I24-derived fragment of cosmid 2G11 or the 1,278 bp *Mlu*I fragment containing *nidAB* from pST120.

Expression of *tidABCD* in *E. coli* and in vitro analysis of enzymatic function of the corresponding gene products

tidABCD expression

Sequence analysis of cosmid 2G11 revealed the incompleteness of *tidD* on this cosmid. We amplified the remaining portion of *tidD* by PCR of a ~1.3 kb fragment from I24 genomic DNA using primers IPBredA and IPBredB (Table 2). Ligation of this product into pCR2.1-TOPO produced pEC47. An *Spe*I–*Nde*I fragment carrying the majority of the *tid* operon from pEC45 was ligated into the *Spe*I and *Nde*I sites of pEC47 to reconstitute the operon, creating pSTS1. The entire *tid* operon was then moved as an *Sph*I fragment into the *Sph*I site of pAL282 (Stafford et al. 2002), creating pSTS2. From pSTS2 a 3,863 bp *Bgl*II/*Eco*RI fragment comprising *tidABCD* was isolated and ligated to *Bam*HI-*Eco*RI-digested pBluescript SK$^-$ (Stratagene). In the resulting hybrid plasmid, pSKEB38, *tidABCD* was located downstream of and codirectionally with the *lacZ* promoter of the vector. *E. coli* XL1-Blue was transformed with pSKEB38 and the corresponding recombinant strain was used for *tidABCD* expression. Cells were grown at 37°C in 50 ml LB medium containing 100 μg/ml ampicillin and 12.5 μg/ml tetracycline, and OD_{600} was monitored using a Spectronic 20 GENESYS spectrophotometer (Spectronic Instruments, Leeds, UK). When the cultures had reached an OD_{600} of 0.6, expression of *tidABCD* was induced by the addition of IPTG to a final concentration of 1 mM. Cultures were further incubated at 30°C for 5 h. Cells were harvested by centrifugation at 4000*g* for 20 min, washed with 0.85% NaCl, and finally resuspended in 3 ml 100 mM HEPES, pH 7.6, containing 0.4 mM DTT, 10% (v/v) glycerol and complete protease inhibitor mix without EDTA (Roche, Indianapolis, IN).

Enzyme assay

Cells were disrupted by twofold French press passage at 96 MPa. Cell free protein extracts were obtained by centrifugation at 38,720*g* and 4°C for 20 min. The supernatants of these centrifugations were directly applied in HPLC-based enzyme assays. Protein extract (400 μl, about 80 mg protein/ml) was mixed with 50 μl 100 mM $NADH+H^+$ in a 2 ml cryogenic vial (Corning, Acton, MA), and the reaction was started by the immediate addition of 50 μl 50 mM Indene (or at variable concentrations from 5 to 400 mM in DMSO). Vials were incubated horizontally in a orbital shaker at 200 rpm at 30°C. Samples of 75 μl were taken after 0, 20, 40, 60, and 90 min of incubation and mixed vigorously with 75 μl of acetonitrile/isopropyl alcohol (1:1; v/v) prior to HPLC analysis. Indene was substituted by naphthalene (from 0.5 to 20 mM in DMSO) or toluene (from 5 to 1,000 mM in DMSO) in the enzyme assay.

HPLC analyses

To separate indene, naphthalene, or toluene-derived metabolites, samples were mixed with an equal volume of a 1:1 (v/v) mixture of acetonitrile and isopropyl alcohol. After 20 min centrifugation, the supernatant was filtered through a 0.2 μm PVDF 13 mm syringe filter (Alltech, Deerfield, IL) and 20 μl of the filtrate were analyzed with a Hewlett Packard Series 1050 HPLC (Agilent, Palo Alto, CA) equipped with a Zorbax RX-C8 column (4.6 mm×25 cm) (Agilent) as described previously (Chartrain et al. 1998). The chirality of the *cis*-indandiol was determined using a normal phase Chiralpak AD column (Chiral Technologies, Inc., Exton, PA) on the aforementioned HPLC system, as described previously (Chartrain et al. 1998).

Rhodococcus transformation

Rhodococcus sp. I24 was transformed via electroporation essentially as described by Stafford et al. (2002) to transform the KY1 strain, except that the I24 cells were cultured in MB 1.5% (w/v) glycine medium instead of tryptic soy broth. MB 1.5% glycine medium contains, per liter, 5 g yeast extract (Difco, Detroit MI), 15 g Bacto-tryptone (Difco), 5 g Bacto-soytone (Difco), 5 g NaCl, and 15 g glycine.

β-Glucuronidase (GUS) assays for responsiveness to aromatic compounds

Induction and sampling

Rhodococcus sp. I24 or strains harboring the promoter: *gusA* fusion plasmids were grown in 100 ml Medium Rare, supplemented with 10 μg/ml gentamicin (for plasmid bearing strains), at 30°C for four days. Saturated cultures were diluted in 250 ml fresh medium to an OD_{600} of 1.5–2.0 and allowed to grow at 30°C overnight to an OD_{600} of 2.5–3.5. Each culture was split into five 45 ml cultures and induced with either 2 mM indene, 2 mM toluene, 2 mM salicylate, a small volume of naphthalene flakes (sufficient to saturate the

culture) or no inducer. OD_{600} was monitored and 1.5 ml culture samples were taken over the course of three days. The samples were quickly spun down, the supernatant aspirated and the pellet stored immediately at −80°C until assayed for GUS activity.

Cell permeabilization and GUS activity assay

Cell pellets were thawed on ice and resuspended in 324 µl B-PER Bacterial Protein Extraction Reagent (Pierce, Rockford, Ill.), 50 µl Protease Inhibitor Cocktail for Bacterial Cell Extracts (Sigma, St. Louis, Mo.), 1 µl 34 mg/ml chloramphenicol and 6 µl freshly prepared 10 mg/ml lysozyme. Samples were vortexed vigorously for 1 min and incubated on ice for 5 min to permeabilize cells. Two-hundred microliters of this permeabilized cell slurry was used in GUS assays essentially as described by Wilson et al. (1992) except that at each time point, the assay mixture was briefly centrifuged, a 100 µl aliquot transferred to 800 µl 0.4 M Na_2CO_3 and the remainder gently resuspended for further incubation. The rate of GUS activity is reported as nanomoles product per minute per OD_{600} unit.

Bacterial conjugation

Five milliliter cultures of donor and recipient were grown to saturation at 30°C in LB. Five-hundred microliters of each strain were combined in sterile microcentrifuge tubes, mixed by inverting and centrifuged briefly. One milliliter samples of donor strain alone and recipient strain alone were used as controls. Cell pellet mixtures were resuspended in a minimal volume of supernatant and dripped onto Millipore (Bedford MA) HA filters (25 mm, 0.45 µm pore size) that had been placed in the center of LB plates, presterilized by saturating with 100 µL of ethanol and dried in air. Donor and recipient were coincubated on filter plates overnight at room temperature. Cells from filters were resuspended in 500 µl 2X M9 salts (Sambrook et al. 1989) and spread onto Medium Rare 2% (w/v) agar supplemented with 10 µg/ml rifampicin and 150 µg/ml streptomycin with 10% (w/v) naphthalene paraffin underlays as a sole carbon source. Plates were wrapped in foil and incubated at 30°C for seven days. Putative transconjugants were verified using Gram staining, qualitative microscopic examination of morphology, and repeated growth passaging on selective media.

Pulsed field gel electrophoresis (PFGE)

Rhodococcus strains were grown in 50 ml MB supplemented with 3.5% (w/v) glycine at 30°C to an OD_{600} of 0.8–1.2. Isoniazid and chloramphenicol were added to a final concentration of 0.01% (w/v) and 0.2 mg/ml, respectively, and cultures were incubated 2 h at 30°C with shaking. Cells were harvested by centrifugation in 50 ml conical tubes at 4,000*g* for 10 min at 4°C and pellets frozen at −80°C for 45 min. Cell pellets were thawed on ice and resuspended in 1% (v/v) Triton X-100 in TE. Tubes were rocked gently at 37°C for 2 h. Cells were then pelleted at 4,000*g* for 10 min, washed twice in 10 ml Wash buffer (200 mM NaCl, 10 mM Tris, pH 8, 100 mM EDTA) and recentrifuged. Pellets were gently resuspended in 1 ml Wash buffer, mixed with an equal volume of molten 2% (w/v) InCert agarose (Bio-Rad Laboratories, Richmond, Calif.) in 1X TBE (0.49 M Tris, 0.49 M boric acid, 0.001 M EDTA; pH 8), which had been pre-cooled to 50°C. The resulting mixture was pipetted into plug molds (Bio-Rad). After incubating at 4°C for 15 min, the solidified plugs were pushed out of the molds into 15 ml conical tubes containing 5 ml Bacterial Lysis solution (10 mM Tris; pH 8, 50 mM NaCl, 100 mM EDTA, 0.2% (w/v) sodium deoxycholate, 0.5% (w/v) lauryl sarcosine, 5 mg/ml lysozyme, 0.04 mg/ml mutanolysin). After incubation at 37°C for 24 h with gentle rocking, the plugs were transferred to fresh 15 ml conical tubes containing 15 ml Bacterial Digestion buffer (0.01 M Tris; pH 9, 0.5 M EDTA, 1% (w/v) lauryl sarcosine, 1 mg/ml proteinase K) and incubated at 50°C for 24–48 h. For PFGE, the plugs were inserted into the wells of agarose gels containing 1% (w/v) PFG certified agarose (Bio-Rad) in 1X TBE. Electrophoresis was performed at 14°C in a CHEF DRII PFGE apparatus (Bio-Rad) at 6 V/cm in 1X TBE for a total run time of 19 or 20 h. Specific pulse switch time conditions for individual experiments are provided in the figure legends. A bacteriophage λ pulsed field marker and Mid-Range I pulsed field marker (New England Biolabs, Beverly, MA) were used as size markers.

Results

Conjugal transfer of naph$^+$ phenotype

To investigate whether a subset of the degradative functions of the I24 strain are localized on a transmissible element, mating experiments were carried out using I24 as a donor strain and the SQ1 strain as the recipient. The SQ1 strain is unable to utilize naphthalene or toluene as sole carbon sources. Additionally, the SQ1 strain is resistant to streptomycin and rifampicin. One of the resulting transconjugants, designated "XO1," had gained the ability to utilize naphthalene but not toluene as a sole carbon source. Finding that the transconjugant had gained the naph$^+$ phenotype but not the tol$^+$ phenotype implies that the genes encoding these two properties reside in at least two distinct physical loci in the I24 genome. Furthermore, these observations are consistent with the hypothesis that the naphthalene functions are carried on a transmissible extrachromosomal element, as has been seen in other systems (Herrick et al. 1997; Stuart-Keil et al. 1998).

Indigogenic screen of cosmid library

Individual cosmid clones from the I24 library were screened as described previously (Treadway et al. 1999). Restriction enzyme digests and Southern analysis were used to identify indigogenic cosmids that were not contiguous with the *nidAB*-containing cosmids described in the earlier study (Treadway et al. 1999) to ensure that distinct genes were responsible. Southern analysis also showed that one of these cosmids (2G11) contained DNA that was apparently lost during the genesis of the KY1 strain (data not shown).

PFGE analysis of indene oxygenating strains

As shown in Fig. 1, PFGE revealed two large extra-chromosomal elements in *Rhodococcus* sp. I24, with estimated sizes of ~340 and ~50 kb. In comparison, the KY1 strain, a derivative of the I24 strain that has lost the ability to grow on toluene as well as the TID activity (Stafford et al. 2002), has also lost the ~340 kb megaplasmid while retaining the ~50 kb megaplasmid (Fig. 1). The naïve SQ1 strain, which cannot metabolize naphthalene or toluene, possesses a megaplasmid of

approximately ~340 kb. When the XO1 strain was generated, the transfer of naphthalene metabolism from I24 was apparently accompanied by the transfer of the ~50 kb element.

We used Southern blotting of pulsed field gels to determine the relationship between cosmid 2G11 and the megaplasmids. As shown in Fig. 1, probe derived from cosmid 2G11 hybridized strongly to the ~340 kb element, indicating that this clone was derived from the larger of the megaplasmids. In contrast, probe derived from *nidAB* (Fig. 1) hybridized to the ~50 kb element. From these results, we conclude that the genes contained within cosmid 2G11 were lost along with the ~340 kb plasmid during the genesis of the KY1 strain, and the *nid* genes were transferred during the genesis of the XO1 strain along with the ~50 kb plasmid. We propose that genes responsible for naphthalene utilization are genetically and physically linked to genes found in cosmid pR4, while genes involved in toluene metabolism are linked to genes found in cosmid 2G11.

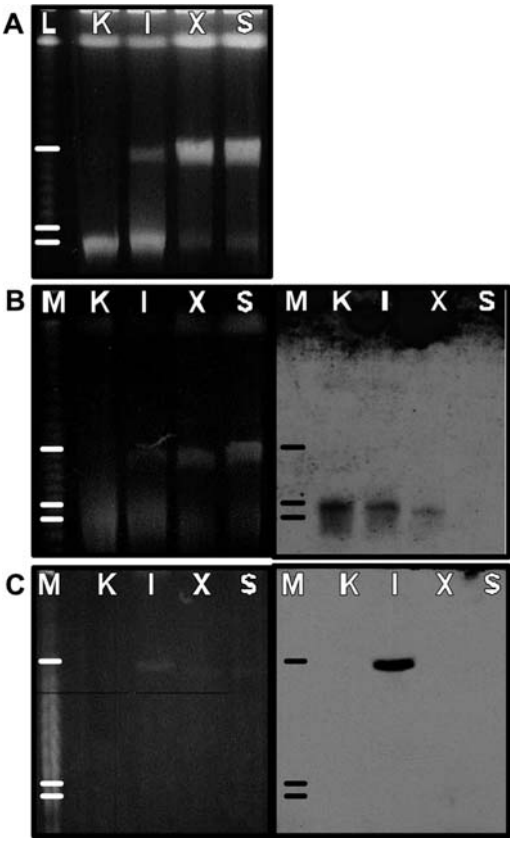

Fig. 1 PFGE analysis of *Rhodococcus* strains. In each image, lanes correspond to: M, midrange I, or L, lambda PFG molecular weight marker; K, DNA from the KY1 strain; I, DNA from the I24 strain; X, DNA from the XO1 transconjugant; and S, DNA from the SQ1 naïve strain. **a** PFGE with a switch time 30–60 s, 20 h, 6 V/cm, 14°C; **b** PFGE with a switch time 45–75 s, 19 h, 6 V/cm, 14°C (*left*) and Southern blot of this gel hybridized with *nidAB* probe; **c** PFGE with a switch time 35 s, 19 h, 6 V/cm, 14°C (*left*) and Southern blot of this gel hybridized with cosmid 2G11 probe. Bars in molecular weight marker lanes in each panel indicate positions, from top to bottom, of 388, 97, and 48.5 kb

Sequence analysis of cosmids

Portions of cosmid 2G11 were sequenced (see "Materials and Methods"). From this, we were able to identify contigs that encompassed open reading frames (ORFs) that may encode the oxygenase activities that were detected in the indigogenic screen. These ORFs are depicted schematically in Fig. 2. Analysis of a ~5.7 kb contig from cosmid 2G11 revealed three ORFs with strong homology to a number of toluene-type dioxygenases (Zylstra and Gibson 1989), including the isopropyl benzene 2,3-dioxygenase from *R. erythropolis* (Kesseler et al. 1996). At the very end of the contig was an incomplete ORF that bore homology to the fourth ORF of the *ipbA* locus from *R. erythropolis*. This sequence ended at a *Sau3AI* site and was followed by cosmid vector sequence (Treadway et al. 1999), indicating that the clone extended to the terminus of the original cosmid clone. To recover the remaining portion of the final ORF, we consulted with Integrated Genomics, Inc. (Chicago, IL), which had been sequencing the genome of the I24 strain as this work was ongoing. Based on their recommendations we recovered the remaining portion of the final ORF by PCR of a ~1.3 kb fragment from I24 genomic DNA using primers IPBredA and IPBredB (Table 2). Together, the overlapping PCR-derived and cosmid-derived sequences constitute a contig of approximately 6.8 kb (Fig. 2). Based on sequence homology (see Table 3) and other results (see below), we have named these four ORFs *tidA* through *tidD*, denoting their possible roles encoding a TID.

Induction of gene expression

As implied by the naming of the ORFs, we inferred that the contig from cosmid 2G11 bore genes that might correspond to the toluene inducible activity observed by Chartrain et al. (1998). To test whether genes from this contig are in fact toluene-responsive, we isolated the region immediately upstream of *tidA* (see Fig. 2) and ligated it into the plasmid pAL280 upstream of the promoterless *gusA* gene, creating the plasmid pAY3. Both pAY3 and pAL280 were introduced into *Rhodococcus* sp. I24 by electroporation. These strains were then cultured in the presence or absence of candidate inducer molecules. As shown in Fig. 3a, the presumed promoter of the *tid* operon directs approximately ten-fold higher expression of the *gusA* gene when cells are grown in the presence of toluene. Indene shows a similar effect on *gusA* expression from this promoter. In contrast, naphthalene was not able to induce expression from this promoter, indicating the specificity of this response (Fig. 3). A shorter segment of the region upstream of the *tid* operon did not confer toluene responsiveness to the *gusA* gene (pAY4; Table 4). Nor did this fragment confer indene-responsiveness (Table 4). These results indicate that a *cis*-acting, toluene-(and indene-) responsive element lies upstream of the *tid* operon, and that the toluene response requires sequences that are found in pAY3 but missing from the shorter

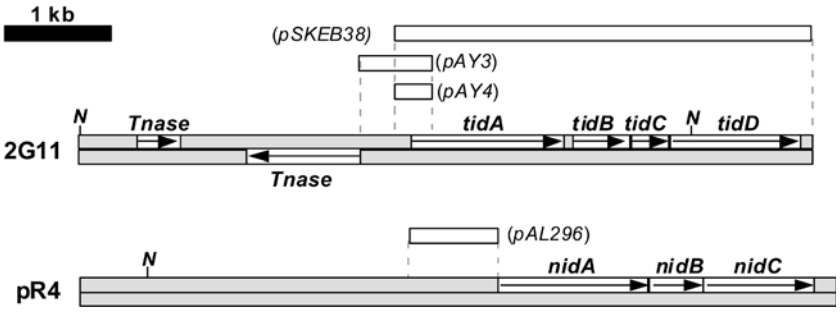

Fig. 2 Diagram of contigs derived from cosmids 2G11 and pR4. The relative positions, sizes and orientations of major ORFs are indicated as *arrows* in white boxes (see also Table 3) Hatched *gray bars* denote fragments that are carried in the indicated plasmids. **N**, *Not*I sites. Information for pR4 is adapted from Treadway et al. (1999)

element in pAY4. From this, it is reasonable to infer that expression of *tidA* (and perhaps the entire *tid* operon) increases in the presence of toluene. This observation resembles the increase in toluene dioxygenase activity that Chartrain et al. (1998) observed when *Rhodococcus* sp. I24 was cultured in the presence of toluene. Also consistent with our observations, Chartrain et al. (1998) reported that indene is similarly able to induce toluene dioxygenase activity in *Rhodococcus* sp. I24.

Whereas the upstream region from the *tid* operon was not responsive to naphthalene, we sought to assess whether expression of the *nid* operon identified by Treadway et al. (1999) was responsive to naphthalene (as its name implied). To test this, we placed the putative promoter from *nidA* (see Fig. 2) into pAL280, creating pAL296. As shown in Fig. 3b, GUS activity is induced by naphthalene when this plasmid is carried in the I24 strain. Additionally, we also found that the *nidA* promoter was not induced in the presence of salicylate at the concentration tested (not shown).

Indene conversion by the *tidABCD* gene products

To obtain physiological evidence that the genes *tidABCD* might correspond to the toluene inducible indene bioconversion activity observed by Chartrain et al. (1998), we introduced into *E. coli* cells pSKEB38, which bears the *tidABCD* ORFs under control of the *lac* promoter. Protein extracts derived from IPTG-induced *E. coli* XL1-Blue (pSKEB38) exhibited indene bioconversion capability (Fig. 4a). Indene was transformed to 1-indenol and *cis*-indandiol. The activities typically obtained in expression experiments were 85 mU/g protein, with respect to the production of 1-indenol, and 12.5 mU/g protein, with respect to the production of *cis*-indandiol (one unit is defined as the enzyme activity that catalyzes the formation of 1 μmol product per minute). The *cis*-indandiol produced in the enzyme assays was isolated by semi-preparative HPLC and analyzed by chiral-HPLC. The *cis*-indandiol that was produced exhibited an enantiomeric excess (e.e.) of 45.2% of *cis*-(1*S*,2*R*)-indandiol over *cis*-(1*R*,2*S*)-indandiol. The K_m value of TID for indene was 380 μM, as determined by Lineweaver–Burk plots of the activities determined for 1-indenol and *cis*-indandiol formation from variable indene concentrations (Fig. 4b). The enzyme also dioxygenated naphthalene to *cis*-dihydronaphthalenediol with an activity of 9.75 mU/g protein, which corresponded to 78% compared to the production of *cis*-indandiol from indene. The K_m value of TID for naphthalene was 28 μM. Applying toluene as a substrate in the enzyme assay, only trace amounts of 1,2-dihydro-3-methylcatechol were detected after a prolonged incubation time of 20 h (not shown). In control experiments omitting the substrate or co-substrate, or using extracts of *E. coli* XL1-Blue harboring only the vector pBluescript SK$^-$, no conversion products were obtained.

Due to the very weak expression of *tidABCD* in *E. coli*, we were not successful in obtaining reasonable toluene conversion rates. This problem has also been reported for the expression of the isopropyl benzene 2,3-dioxygenase and biphenyl dioxygenase genes of *R. erythropolis* BD2 (Kesseler et al. 1996) and *R. globerulus* (McKay et al.

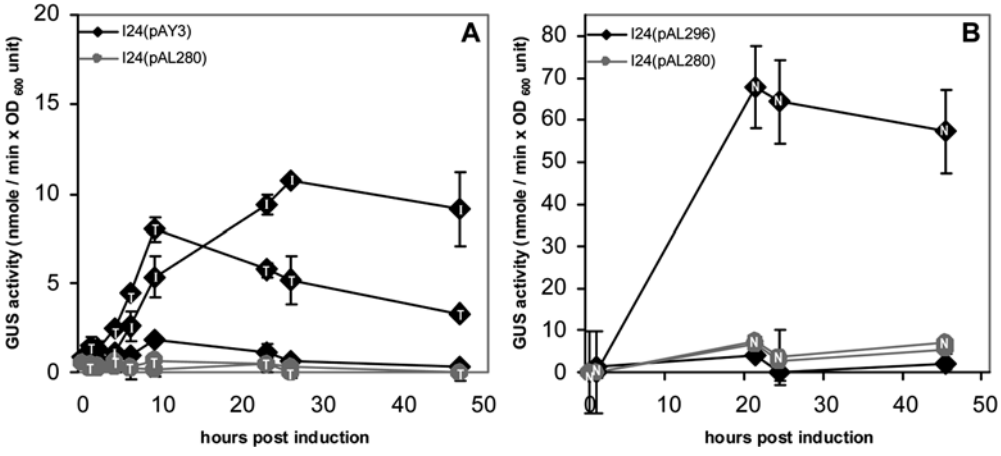

Fig. 3 Induction of gene expression by aromatic hydrocarbons. **a** *tidA* promoter analysis; **b** *nidA* promoter analysis. GUS activity was measured in lysates from cells grown in the presence of toluene (2 mM, denoted as "T"), indene (2 mM, denoted as "I"); or naphthalene (saturated, ~0.3 mM, denoted as "N"); unlabelled symbols are from uninduced cultures. Note that expression of GUS activity from the negative control construct (pAL280) did not respond to any of the inducers tested. Data are averages of 3–5 replicate cultures. See also Table 4

Table 3 Homology-based functional assignments for ORFs from cosmid 2G11

ORF name	GenBank accession	Nucleotide position (5′–3′)	Accession of homolog[a]	Function of homolog	Reference for homolog
Tnase	AF452375	589–939	T29424	Probable transposase from *Streptomyces coelicolor*	Direct GenBank submission
Tnase	AF452375	2637–1591	AAC29482	Probable transposase from IS*1121* of *Clavibacter michiganense*	Mogen and Oleson (1987)
tidA	AF452375	3332–4540	AAB08025	Isopropyl benzene 2,3-dioxygenase, large subunit	Kesseler et al. (1996)
tidB	AF452375	4622–5173	AAB08026	Isopropyl benzene 2,3-dioxygenase, small subunit	Kesseler et al. (1996)
tidC	AF452375	5182–5496	AAB08027	Isopropyl benzene 2,3-dioxygenase, ferredoxin component	Kesseler et al. (1996)
tidD	AF452375	5501–6739	AAB08028	Isopropyl benzene 2,3-dioxygenase, reductase component	Kesseler et al. (1996)

[a]These accession numbers refer to protein database entries; homologies were identified using the BLASTP program (Altschul et al. 1997) on the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/).

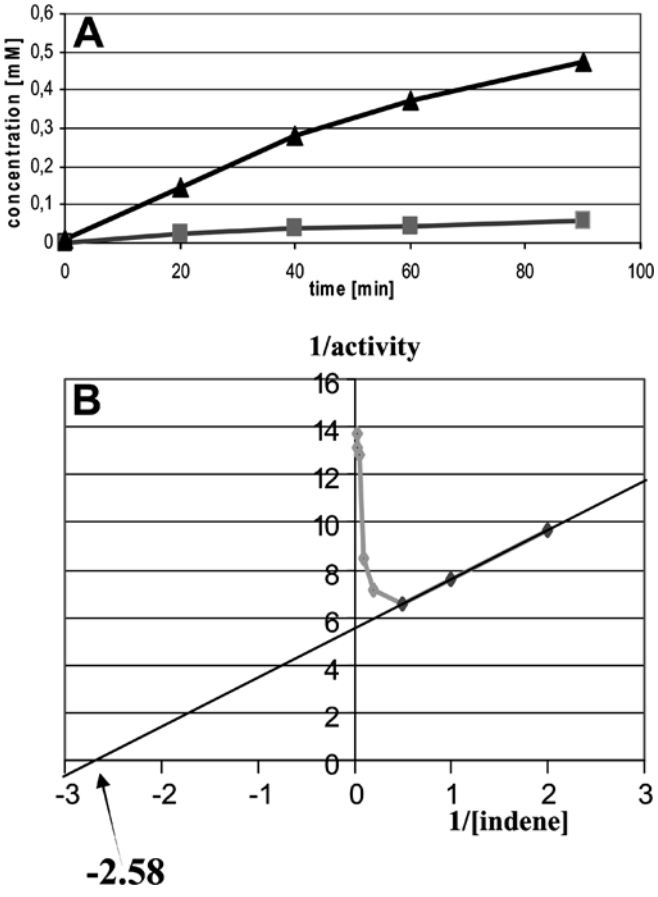

Fig. 4 Enzyme activity of the *tidABCD* gene products. The genes were expressed in *E. coli* XL1-Blue(pSKEB38) as described in "Materials and Methods". **a** time course of 1-indenol(▲) and *cis*-indandiol(▪) formation from 5 mM indene by cell free protein extracts (80 mg protein ml^{-1}). **b** Lineweaver–Burk plot of the activities determined for 1-indenol formation from 0.5, 1, 2, 5, 10, 20, 30, and 40 mM indene averaged from two independent activity determinations, respectively

1997), respectively, which are the most closely related genes found by sequence similarities. In these cases the weak expression was attributed to inefficient translation or rapid gene product degradation (McKay et al. 1997), or to modified protein assembly and inadequate post-translational modifications (Kesseler et al. 1996).

Discussion

Chartrain et al. (1998) found that the I24 strain of *Rhodococcus* can oxygenate indene via at least three independent pathways: a NID, a TID, and a NIM. Treadway et al. (1999) cloned a series of genes from the I24 strain, which they called the *nid* operon because it apparently encodes enzymes for the NID pathway. Four lines of evidence suggest that we have now identified an operon that contributes to the TID system. The supporting data can be summarized as (i) genetic and phenotypic evidence; (ii) sequence and homology evidence; (iii) gene expression evidence; and (iv) enzymatic evidence.

Table 4 β-Glucuronidase activity (nmole/min×OD_{600} unit) in transformed strains approximately 24 h after induction

	No inducer	Indene 2 mM	Toluene 2 mM	Naphthalene saturated	Salicylate 2 mM
I24 (pAL280)	0.5±0.28	0.7±0.04	0.5±0.39	0.8±0.26	1.3±0.08
I24 (pAY3)	1.1±0.48	9.4±0.56	5.8±0.38	0.4±0.02	0.9±0.06
I24 (pAY4)	1.3±0.14	1.5±0.42	1.0±0.27	1.0±0.06	*nd*

Data are averages of 2–5 replicate cultures.
nd, not determined.

Genetically, the ability to degrade toluene and to convert indene to *cis*-(1*S*,2*R*)-indandiol (which reflect activity of the TID system; (Chartrain et al. 1998)) is found only in the strains that carry the *tid* genes. Southern analysis and growth studies show that, whereas the I24 strain possesses the *tid* operon and can carry out these functions, a derivative of the I24 strain, the KY1 strain (Yanagimachi et al. 2001), has lost both the *tid* genes and the ability to grow on toluene. Unlike the I24 strain, growth of the KY1 strain in the presence of toluene does not induce the formation of *cis*-(1*S*,2*R*)-indandiol from indene. PFGE analysis shows that the *tid* genes reside on a ~340 kb extrachromosomal element in the I24 strain, and that this element was lost from the KY1 strain. In contrast, the same studies revealed that the *nid* operon resides on a smaller, ~50 kb extrachromosomal element. Both the I24 and KY1 strains have this element, and both have the NID activity. A transconjugant strain, XO1, that received the ~50 kb element from the I24 strain simultaneously gained the ability to grow on naphthalene and the ability to convert indene into *cis*-(1*R*,2*S*)-indandiol (indicative of NID activity). However, the transconjugant had not inherited the ~340 kb element nor the TID system. The localization of these biodegradative phenotypes to large extrachromosomal elements is often seen in rhodococci and pseudomonads (Larkin et al. 1998; Poelarends et al. 2000; Stuart-Keil et al. 1998). Horizontal transfer of such elements may be the basis of the biocatalytic diversity seen among these genera.

At the DNA sequence level, the *tid* operon shows sequence homology and gene organization that are consistent with its suspected function. The four ORFs of the *tid* operon, *tidA-D*, have extensive homology to genes encoding other toluene-type dioxygenases (Zylstra and Gibson 1989; Kesseler et al. 1996). Furthermore, the cistrons for each, the large and small subunits of the iron-sulphur protein, the ferredoxin and the reductase components of the dioxygenase, are arranged in an order common to the toluene-type dioxygenases (e.g., Kesseler et al. (1996)).

The third line of evidence that suggests that the *tid* and *nid* operons correspond to the enzyme activities reported by Chartrain et al. (1998) is that expression of genes within these operons precisely reflects the induction profiles for each of the oxygenase systems described in that earlier study. Using regions from each of the two operons that were likely to include promoters and regulatory elements, we found that transcription from the *tid* promoter is induced by toluene and indene but not by naphthalene, just as Chartrain et al. (1998) had shown that enzymatic activity of this dioxygenase was induced by toluene or indene, but not by naphthalene. In contrast, we found that transcription of the *nid* operon is naphthalene inducible, again reflecting the results of Chartrain et al. (1998) an issue that had not been addressed directly by Treadway et al. (1999) when they identified the *nid* operon.

Finally, the physiological functions of the *tidABCD* gene products were proven by heterologous expression in *E. coli*. Indene was converted by TID to 1-indenol and *cis*-indandiol, as also reported for the toluene dioxygenase (TDO) of *P. putida* (Wackett et al. 1988). The enantiomeric excess of the *cis*-(1*S*,2*R*)-indandiol produced by TID of *Rhodococcus* sp. I24 was 45.2%, higher than the e.e. of 32% observed with the TDO of *P. putida* F39/D (Wackett et al. 1988).

Based on the evidence presented, we believe we have identified the *tid* operon corresponding to one of the three dominant indene-oxygenating systems first described by Chartrain et al. (1998), and that the *tid* system is responsible for conversion of indene to *cis*-(1*S*,2*R*)-indandiol and 1-indenol in the I24 strain. Continued study of the operons involved in indene bioconversion will lead to new insights in the evolution and regulation of these genes. The enzymes encoded by these operons will also serve to develop biocatalysts for the production of chiral molecules.

Acknowledgements The authors would like to acknowledge D. Stafford, M. Fonstein and Integrated Genomics for helpful contributions during this research, as well as the support of B. Buckland and S. Drew. This work was supported by a grant from the Cambridge-MIT Institute and Merck Research Laboratories. GN, JJC, MA, LY, LT and MW received additional support from the Howard Hughes Medical Institute grant #71195-501306. ST was supported in part by NIH Biotechnology Training Grant 2T32GM08334-10. The experiments reported in this publication comply with the current laws of the United States of America.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Becker A, Schmidt M, Jager W, Puhler A (1995) New gentamicin-resistance and *lacZ* promoter-probe cassettes suitable for insertion mutagenesis and generation of transcriptional fusions. Gene 162:37–39

Bell KS, Philp JC, Aw DW, Christofi N (1998) The genus *Rhodococcus*. J Appl Microbiol 85:195–210

Buckland BC, Drew SW, Connors NC, Chartrain MM, Lee C, Salmon PM, Gbewonyo K, Zhou W, Gailliot P, Singhvi R, Olewinski RC Jr, Sun WJ, Reddy J, Zhang J, Jackey BA, Taylor C, Goklen KE, Junker B, Greasham RL (1999) Microbial conversion of indene to indandiol: a key intermediate in the synthesis of CRIXIVAN. Metab Eng 1:63–74

Chartrain M, Jackey B, Taylor C, Sandford V, Gbewonyo K, Lister L, DiMichele L, Hirsch C, Heimbuch B, Maxwell C, Pascoe D, Buckland B, Greasham R (1998) Bioconversion of indene to *cis*-(1*S*,2*R*)-indandiol and *trans*-(1*R*,2*R*)-indandiol by *Rhodococcus* species. J Ferment Bioeng 86:550–558

Herrick JB, Stuart-Keil KG, Ghiorse WC, Madsen EL (1997) Natural horizontal transfer of a naphthalene dioxygenase gene between bacteria native to a coal tar-contaminated field site. Appl Environ Microbiol 63:2330–2337

Keilhauer C, Eggeling L, Sahm H (1993) Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. J Bacteriol 175:5595–5603

Kesseler M, Dabbs ER, Averhoff B, Gottschalk G (1996) Studies on the isopropylbenzene 2,3-dioxygenase and the 3-isopropylcatechol 2,3-dioxygenase genes encoded by the linear plasmid of *Rhodococcus erythropolis* BD2. Microbiology 142:3241–3251

Larkin MJ, De Mot R, Kulakov LA, Nagy I (1998) Applied aspects of *Rhodococcus* genetics. Antonie Van Leeuwenhoek 74:133–153

Lessard PA, Allen RD, Fujiwara T, Beachy RN (1993) Upstream regulatory sequences from two beta-conglycinin genes. Plant Mol Biol 22:873–885

McKay DB, Seeger M, Zielinski M, Hofer B, Timmis K (1997) Heterologous expression of biphenyl dioxygenase-encoding genes from a Gram-positive broad-spectrum polychlorinated biphenyl degrader and characterization of chlorobiphenyl oxidation by the gene products. J Bacteriol 179:1924–1930

Mogen B, Oleson A (1987) Homology of pCS1 plasmid sequences with chromosomal DNA in *Clavibacter michiganense* subsp. *sepedonicum*: evidence for the presence of a repeated sequence and plasmid integration. Appl Environ Microbiol 53:2476–2481

Poelarends GJ, Kulakov LA, Larkin MJ, van Hylckama Vlieg JE, Janssen DB (2000) Roles of horizontal gene transfer and gene integration in evolution of 1,3-dichloropropene- and 1,2-dibromoethane-degradative pathways. J Bacteriol 182:2191–2199

Quan S, Dabbs ER (1993) Nocardioform arsenic resistance plasmid characterization and improved *Rhodococcus* cloning vectors. Plasmid 29:74–79

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Stafford DE, Yanagimachi KS, Lessard PA, Rijhwani SK, Sinskey AJ, Stephanopoulos G (2002) Optimizing bioconversion pathways through systems analysis and metabolic engineering. Proc Natl Acad Sci USA 99:1801–1806

Stuart-Keil KG, Hohnstock AM, Drees KP, Herrick JB, Madsen EL (1998) Plasmids responsible for horizontal transfer of naphthalene catabolism genes between bacteria at a coal tar-contaminated site are homologous to pDTG1 from *Pseudomonas putida* NCIB 9816-4. Appl Environ Microbiol 64:3633–3640

Treadway SL, Yanagimachi KS, Lankenau E, Lessard PA, Stephanopoulos G, Sinskey AJ (1999) Isolation and characterization of indene bioconversion genes from *Rhodococcus* strain I24. Appl Microbiol Biotechnol 51:786–793

Wackett LP, Kwart LD, Gibson DT (1988) Benzylic monooxygenation catalyzed by toluene dioxygenase from *Pseudomonas putida*. Biochemistry 27:1360–1367

Wilson KJ, Jefferson RA, Hughes SG (1992) The *Escherichia coli gus* operon: induction and expression of the *gus* operon in *E. coli* and the occurrence and use of GUS in other bacteria. In: Gallagher SR (ed) GUS protocols: using the GUS gene as a reporter of gene expression. Academic, San Diego

Yanagimachi KS, Stafford DE, Dexter AF, Sinskey AJ, Drew S, Stephanopoulos G (2001) Application of radiolabeled tracers to biocatalytic flux analysis. Eur J Biochem 268:4950–4960

Zylstra GJ, Gibson DT (1989) Toluene degradation by *Pseudomonas putida* F1. Nucleotide sequence of the *todC1C2BADE* genes and their expression in *Escherichia coli*. J Biol Chem 264:14940–14946