Appl Microbiol Biotechnol (1999) 51: 786–793

ORIGINAL PAPER

S. L. Treadway · K. S. Yanagimachi · E. Lankenau
P. A. Lessard · G. Stephanopoulos · A. J. Sinskey

Isolation and characterization of indene bioconversion genes from *Rhodococcus* strain I24

Received: 8 December 1998 / Received revision: 26 January 1999 / Accepted: 5 February 1999

Abstract *Rhodococcus* strain I24 is able to convert indene into indandiol via the actions of at least two dioxygenase systems and a putative monooxygenase system. We have identified a cosmid clone from I24 genomic DNA that is able to confer the ability to convert indene to indandiol upon *Rhodococcus erythropolis* SQ1, a strain that normally can not convert or metabolize indene. HPLC analysis reveals that the transformed SQ1 strain produces *cis*-(1*R*,2*S*)-indandiol, suggesting that the cosmid clone encodes a naphthalene-type dioxygenase. DNA sequence analysis of a portion of this clone confirmed the presence of genes for the dioxygenase as well as genes encoding a dehydrogenase and putative aldolase. These genes will be useful for manipulating indene bioconversion in *Rhodococcus* strain I24.

Introduction

Many gram-negative and gram-positive bacteria from the genera *Pseudomonas* and *Rhodococcus* are able to metabolize aromatic hydrocarbons (Allen et al. 1997; Masai et al. 1995; Gibson and Subramanian 1984). Degradation of these compounds generally proceeds via the action of multicomponent dioxygenases (Mason and Cammack 1992) or through successive monoxygenations of the aromatic rings to produce diols (Harayama et al. 1992). The chirality of these diols make them useful for synthesizing enantiopure precursors of biologically active products (Collins 1997). In vivo, further metabolism of the diols is frequently initiated by one or more dehydrogenation steps (Fukuda et al. 1994; Furukawa et al. 1993; Patel and Gibson 1974).

Rhodococcus strain I24, is capable of metabolizing naphthalene and toluene as sole carbon sources (Chartrain et al. 1998). Presumably by using the oxygenase systems that enable metabolism of naphthalene and toluene, this strain can also oxygenate indene. Indene is converted to a variety of indandiols, two of which can be utilized as a precursor in the synthesis of the HIV protease inhibitor indinavir sulfate (Buckland et al. 1999). An indene bioconversion pathway in *Rhodococcus* strain I24 has been proposed on the basis of the analysis of indene breakdown in the presence and absence of naphthalene and toluene as inducers (Fig. 1) (Buckland et al. 1999; Chartrain et al. 1998). Initial analysis of the system suggests that there may be three different enzyme systems responsible for converting indene into the various 1,2-indandiols. The presence of multiple oxygenase enzymes in an organism capable of utilizing a common substrate is not unusual (Asturias and Timmis 1993; Kosono et al. 1997). Chartrain and colleagues (Chartrain et al. 1998) have proposed that the I24 strain possesses a naphthalene-inducible dioxygenase, a toluene-inducible dioxygenase, and a naphthalene-inducible monooxygenase, which all contribute to the oxygenation of indene. A single enzyme complex may be responsible for both the naphthalene-inducible mono- and dioxygenase activities. Earlier work with *Pseudomonas* has demonstrated that naphthalene dioxygenases can act both as dioxygenases and monooxygenases (Gibson et al. 1995). Interestingly, each of the proposed oxygenation reactions seems to produce a specific indandiol stereoisomer, suggesting discrete enzyme mechanisms. Other oxygenases have been found to act upon indene in a stereospecific manner (Gibson et al. 1995; Wackett et al. 1988).

In the proposed pathway shown in Fig. 1, each oxygenation reaction is followed by a dehydrogenation. These reactions may be carried out by a single dehydrogenase or by separate dehydrogenases specific to each

S. L. Treadway · E. Lankenau · P. A. Lessard · A. J. Sinskey (✉)
Department of Biology, Massachusetts Institute of Technology, Bldg. 68-370, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: asinskey@mit.edu
Tel.: +1-617-253-6721
Fax: +1-617-253-8550

K. S. Yanagimachi · G. Stephanopoulos
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

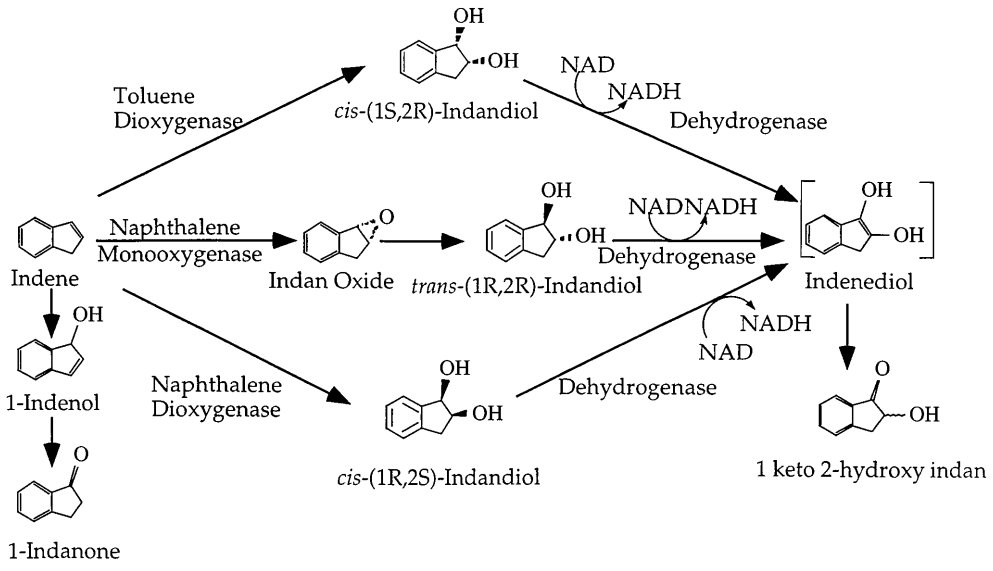


Fig. 1 Proposed indene bioconversion pathway of *Rhodococcus* I24 (modified from Chartrain et al. 1998)

pathway. We are interested in analyzing at the genetic level the oxygenation and dehydrogenation steps of this indene bioconversion in order to better understand and control the production of indandiol in this strain.

In this paper, we describe a genetic analysis of the *Rhodococcus* I24 indene bioconversion pathway. Using a functional screen for dioxygenase activity, we identify genes encoding the large and small subunits of a naphthalene-type dioxygenase, a dehydrogenase, and a putative aldolase. We demonstrate that the dioxygenase is capable of oxygenating indene. We discuss the implications of the findings on the proposed model for indene bioconversion in this strain and the relationship of these genes to previously identified dioxygenase genes.

Materials and methods

Reagents

All strains and plasmids are listed in Table 1. All chemicals were reagent-grade and purchased from Sigma (St. Louis, Mo.) or Aldrich Chemical Co. (Milwaukee, Wis.) unless otherwise noted. All media components were purchased from Difco (Detroit, Mich.).

Pulse-field gel analysis

Pulse-field gel analysis used a modified protocol from Lai and Birren (1990). Genomic DNA from *Rhodococcus* I24 was prepared from cultures in Luria-Bertani (LB) medium grown to an A_{600} of 0.8–1.0 and treated with 0.2 mg/ml chloramphenicol and 0.01% isoniazid for 2 h. The cells and DNA were then treated as described (Lai and Birren 1990) except that the agitation step during the 24-h lysis was omitted. InCert agarose (FMC, Philadelphia, Pa.) was used to embed the cells. Gel slices were equilibrated in TEN buffer (10 mM TRIS pH 8.0, 1 mM EDTA, 50 mM NaCl) overnight at room temperature. Individual gel slices were then treated with 10 mg Pefabloc SC (Boehringer Mannheim, Indianapolis, Ind.) in 1 ml TE (10 mM TRIS pH 8.0, 1 mM EDTA) for 2 h at 37 °C, and equilibrated in 45 ml TEN buffer for 1 h. Gel slices were digested with *Ase*I and *Ssp*I (New England Biolabs, Beverly, Mass.) in their appropriate buffers, incubated at 4 °C overnight and then for 6–8 h at 37 °C.

The digested samples were analyzed in 1% agarose (Gibco BRL, Grand Island, N.Y.) 0.5× TBE (0.045 M TRIS-Borate, 0.001 M EDTA) gels run in a BioRad Chef-DR II pulse-field gel apparatus at 6 V/cm at 14 °C for 16–19 h at a ramp time of 15–75 s for *Ase*I and 1–40 s for *Ssp*I. For 20 to 350-kb a ramp time of 1–25 s was used. A

Table 1 Strains and plasmids

Strain/plasmid	Description	Reference
Strains		
Rhodococcus I24	Converts indene to indandiol, orange colonies	Buckland et al. 1999
Rhodococcus SQ1	Easily transformable isolate of *R. erythroplis*	Quan and Dabbs 1993
Escherichia coli JM109	*endA1*, *recA1*, *gyrA96*, *thi*, *hsdR17*, (r^-, m^+), *relA1*, *supE44*, λ^-, Δ(*lac-proAB*), [F′, *traID36*, *proAB*, *lacI*^{q}*ZΔM15*]	Yanisch-Peron et al. 1985
Plasmids		
SuperCOS	Amp^R, Kan^R, colE1 ori, COS sites	Strategene
pEP2	Kan^R, NG2 ori	Zhang et al. 1994
pRhodoCOS	Kan^R, NG2 ori, COS sites	This study
pR4	Kan^R, NG2 ori, I24 genomic cosmid clone	This study
pR4-10	Kan^R, NG2 ori, *Sau3a* pR4 subclone	This study

ramp time of 1–40 s resolved 20 to 450-kb fragments. Fragments of 250–850 kb were resolved at 15–75 s and 450 to 750-kb fragments by 30–60 s ramp time. The lambda DNA pulse-field gel ladder, used as a molecular mass marker, was obtained from New England Biolabs (Beverly, Mass.).

DNA manipulation and plasmid construction

Restriction enzymes were purchased from New England Biolabs (Beverly, Mass.) and used according to the manufacturer's recommendations. Plasmid DNA was prepared, using the Wizard Maxiprep Kit, the Wizard Miniprep Kit from Promega (Madison, Wis.), or a boiling lysis miniprep (Sambrook et al. 1989). To construct the cosmid vector pRhodoCOS, pEP2 (Zhang et al. 1994) was partially digested with *Hin*CII, then digested with *Bam*HI and treated with shrimp alkaline phosphatase (Boehringer Mannheim, Indianapolis, Ind.). The 3.1-kb *Hin*CII-*Bam*HI fragment of pEP2 was ligated to the 4.2-kb *Bgl*II-*Sca*I fragment of SuperCOS (Stratagene, La Jolla, Calif.). *E. coli* JM109 were transformed by electroporation (Sambrook et al. 1989) and plated on LB plates containing 100 μg/ml kanamycin. The pR4-10 subclone was made by a *Sau*3AI partial digest of pR4 DNA.

Preparation of *Rhodococcus* genomic DNA and library

Genomic DNA was prepared from a 200-ml LB culture of *Rhodococcus* grown to saturation at 30 °C. Cells were harvested by centrifugation for 10 min at 4000 *g*, the supernatant was removed and the pellet was frozen at −20 °C for 30–60 min. The pellet was resuspended in 5 ml TE with 20 mg lysozyme and 200 μg mutanolysin, and incubated for 1 h at 37 °C with shaking. A 1-ml sample of 0.5 M EDTA, 1 ml 10% sodium dodecyl sulfate, and 1 ml 5 M NaCl were added, and incubated on ice for 10 min. The cell suspension was treated with 2 mg proteinase K for 1 h at 37 °C. This was followed by the addition of 3.77 g sodium perchlorate; the mixture was incubated at room temperature with gentle agitation for 30 min. Following phenol and chloroform extraction, two volumes of ice-cold 100% ethanol were added to the aqueous phase. The DNA was spooled onto a glass rod and resuspended in 5 ml TE. Next, 50 μl RNase (Boehringer Mannheim, Indianapolis) was added and the sample was incubated at 37 °C for 30 min. The solution was again extracted with phenol and chloroform, and the aqueous phase was precipitated with two volumes of cold ethanol and 0.1 vol of 3 M sodium acetate. The DNA was spooled on a glass rod and washed in 70% ethanol for 3 min. The DNA was air-dried for 5–10 minutes and resuspended in TE. The genomic DNA was quantified and subjected to partial digest with *Sau*3AI.

The cosmid library of *Rhodococcus* I24 was made using pRhodoCOS, following the Stratagene (La Jolla, Calif.) protocol provided with the SuperCOS vector kit and Gigapack III XL packaging extract. JM109 *E. coli* was used as the host strain for the cosmids (Yanisch-Perron et al. 1985).

Screen for dioxygenase activity in vivo

A total of 286 cosmid clones in JM109 *E. coli* cells were grown on LB plates containing 100 μg/ml kanamycin at room temperature for 3–4 days and screened for blue color development. In *Rhodococcus* SQ1, blue color development was assayed by first growing colonies at 30 °C. After the colonies had reached 2–3 mm in diameter, filter-paper containing 600 μl 3% indole (w/v in dimethylformamide; Fisher Scientific Fair Lawn, N.J.) was placed in the lid of the petri dish. Plates were kept at room temperature for 24–48 h.

Transformation of *Rhodococcus* SQ1

Competent *Rhodococcus* SQ1 cells were prepared from a 100-ml culture in MB medium (5 g/l yeast extract, 15 g/l Bacto-tryptone, 5 g/l Bacto-soytone, 5 g/l NaCl), 1.5% glycine, 1.8% sucrose and 0.01% isoniazid grown at 30 °C to an A_{600} of approximately 1.6–1.8 (Hewlett Packard diode-array spectrophotometer 8452A). The culture was treated with 1 μl of 100 mg/ml ampicillin and incubated at 30 °C for 1 h. The cells were harvested by centrifugation at 4000 *g* for 10 min and washed twice with 30 ml cold EPB1 medium (20 mM HEPES pH 7.2, 5% glycerol). The washed cells were resuspended in 2 ml cold EPB2 medium (5 mM HEPES pH 7.2, 15% glycerol). Competent cells were stored at −80 °C.

Electroporation using a BioRad gene pulser set to 2.50 kV, 400 Ω, 25 F was used to transform 70 μl competent cells with about 1 μg DNA. Immediately after electroporation, 400 μl recovery broth [80 g/l brain heart infusion mixed with an equal volume of solution 2 (80 g/l sorbitol, 20 g/l sucrose)] was added and cells were incubated at 30 °C for 1 h. The transformation was plated on LB plates with 200 μg/ml kanamycin and grown at 30 °C for 3–5 days.

Indene bioconversion

Cultures containing 25 ml LB, 5 ml silicone oil, 150 μl indene, and 150 μg/ml kanamycin (if appropriate) were grown at 30 °C for 4 days. Media samples (1 ml) were extracted with 6 ml HPLC-grade isopropanol (Mallinckrodt, Paris, Ky.) and 3 ml MilliQ water. Samples were vortexed and cleared of cells in a 5-min centrifugation at 4000 *g*. Samples were filtered through a 0.22-μm polvinylidene difluoride syringe filter (Alltech, Deerfield, Iu.), and 20 μl filtrate was injected onto a Zorbax Rx-C8 (4.6 × 250 mm) HPLC column on a Rainin Dynamax HPLC System. The HPLC protocol was followed as previously described (Chartrain et al. 1998). Indene, silicone oil, and the HPLC standards *cis*-indandiol, *trans*-indandiol, 1-indenol, 1-indanone, and ketohydroxyindan were gifts from M. Chartrain (Merck Research Laboratories, Rahway, N.J.). HPLC-grade acetonitrile was purchased from Mallinckrodt (Paris, Ky.). The chirality of the *cis*-indandiol was determined by chiral HPLC analysis as previously described (Chartrain et al. 1998).

DNA sequencing and analysis

Sequencing of pR4 was performed at the MIT Biopolymers Facility using an ABI cycle sequencer by primer walking. Primers were made by Gibco/BRL Life Technologies (Grand Island, N.Y.). pR4-10 was sequenced by primer walking by Lark Technologies Inc. (Houston, Tex.) using Big Dye terminator cycle sequencing reactions (PE-ABD) analyzed on an ABI 373A-S or an ABI 377 automated sequencer.

All sequence data were analyzed using GenBank, EMBL, Swissprot databases, the BLASTN, BLASTP, and BLASTX (Altschul et al. 1990; Gish and State 1993) programs via the National Center for Biotechnology Information server, and the PROSITE protein database (Bairoch 1992) via the ExPASy web site (http://expasy.hcuge.ch/sprot/prosite.html). Phylogenetic analysis was carried using the Lasergene program (DNA star, Madison, Wis.) via a CLUSTAL alignment using a PAM250 residue-weight table.

GenBank accession number

The sequence was lodged in the GenBank database under the accession number AF121905.

Results

Genome size analysis and genomic cosmid library construction

We determined the approximate size of the *Rhodococcus* I24 genome by pulse-field gel analysis. Figure 2 shows a

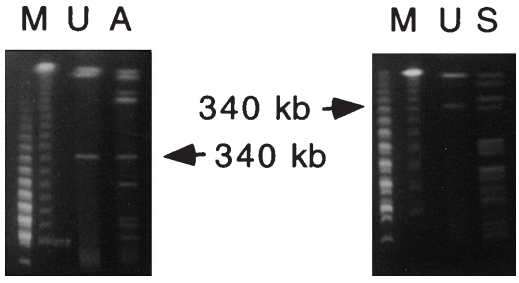


Fig. 2 Pulsed-field gel analysis of *Rhodococcus* I24. The size of the I24 genome was analyzed by pulse-field gel analysis with two enzymes, *Ssp*I (*S*) and *Ase*I (*A*). *Lane 1* lambda pulse-field gel marker (*M*), *Lane 2* uncut (*U*) I24 genomic DNA. The I24 genome contains a large plasmid of 340 kb

representative gel from this analysis. By varying the ramp times we were able to increase resolution over several size ranges to get a more accurate representation of the genomic content (Table 2). We estimate that the genome size is approximately 3 Mb, including a large plasmid of roughly 340 kb. This plasmid could be either linear or circular, and in single or multiple copies. We have named this plasmid pI24.

To prepare a cosmid library, we first constructed pRhodoCOS, which contains two COS elements and an origin of replication that allows the plasmid to replicate in a broad range of hosts, including *Escherichia coli* and *Rhodococcus*. Using this new vector, we constructed a genomic library of *Rhodococcus* I24 containing 286 cosmid clones with inserts of approximately 30–40 kb. On the basis of our calculations of genome size, the library provides approximately threefold coverage of the *Rhodococcus* I24 genome.

Table 2 Pulse-field gel analysis of *Rhodococcus* I24

	Size of bands (kb)	
	Ase I	*Ssp* I
	800	680
	380[a]	630[b]
	340	340
	210[a]	240
	195	130
	175	100
	130[a]	60
	110	
	95	
	90	
	65	
	60	
	35	
	25	
	20	
	15	
Total	2.75 Mb	2.8 Mb
Total with multiples	3.1–3.2 Mb	3.4[c] Mb

[a] Could be doublets due to intensity of ethidium bromide staining
[b] At least a doublet. Used in the initial total
[c] Total based on 630-kb band as a triplet

Identification of ring-hydroxylating dioxygenase activity

A functional screen was used to identify cosmid clones encoding ring-hydroxylating dioxygenases. Other researchers have found that certain dioxygenases will convert indole to indoxyl, which spontaneously dimerizes to form indigo in *E. coli* (Ensley et al. 1983; Hart et al. 1992). We screened 286 transformants of *E. coli* JM109, each carrying a single cosmid from the library, for blue color development. Of the 286 strains, 9 turned blue, suggesting the presence of ring-hydroxylating dioxygenase activity. Restriction-enzyme analysis of cosmid DNA isolated from these 9 strains suggests that these 9 clones contain overlapping portions of the genome (data not shown). The smallest clone, pR4, had an insert of 20–30 kb and was chosen for further study.

The indigo formation assay can also be carried out in *Rhodococcus* provided that the host strain does not have any inherent detectable ring-hydroxylating dioxygenase activity and an external source of indole is provided. A strain that does not change color in the presence of indole is *R. erythropolis* SQ1 (Kesseler et al. 1996; Quan and Dabbs 1993). We transformed the cosmid pR4 into *R. erythropolis* SQ1 and found that *R. erythropolis* SQ1 harboring the plasmid pR4 will turn blue when exposed to indole. This result strongly suggests that pR4 encodes a dioxygenase activity, and is consistent with the results obtained in *E. coli*.

Indene bioconversion analysis of pR4

To characterize the indigogenic activity further, we tested whether pR4 permitted *R. erythropolis* SQ1 to oxygenate indene. In addition to being unable to convert indole to indigo, wild-type *R. erythropolis* SQ1 is unable to oxygenate indene. Transformation of the pR4 cosmid into *R. erythropolis* SQ1 allows the strain to oxygenate indene (Fig. 3). A series of bioconversion products were identified from *R. erythropolis* SQ1(pR4) including *cis*-indandiol, ketohydroxyindan, 1-indenol, and 1-indanone. We did not detect any production of *trans*-indandiol. The product profile of *R. erythropolis* SQ1(pR4) represents a subset of the *Rhodococcus* I24 indene bioconversion product profile and suggests that pR4 encodes only some of the genes responsible for indene bioconversion in *Rhodococcus* I24.

In *Rhodococcus* I24, production of the two the *cis*-indandiol enantiomers is differentially induced by naphthalene and toluene (Chartrain et al. 1998). To characterize further the dioxygenase enzyme activity encoded by pR4, the chirality of the *cis*-indandiol produced by *R. erythropolis* SQ1(pR4) was examined by chiral HPLC analysis. Only *cis*-(1*R*,2*S*)-indandiol, not *cis*-(1*S*,2*R*)-indandiol, was produced (data not shown). On the basis of these results, we conclude that pR4 carries a gene or genes that encode a specific dioxygenase activity and a possible dehydrogenase activity.

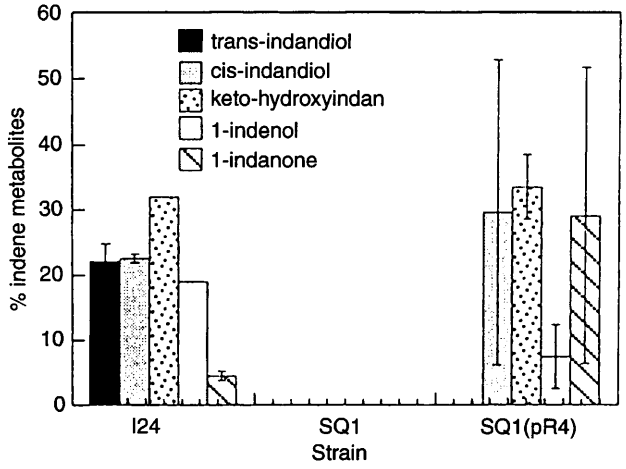


Fig. 3 Indene bioconversion analysis of *Rhodococcus* SQ1(pR4). The presence of indene metabolites was measured by HPLC analysis

Sequence analysis of pR4

To identify which regions of pR4 contained the genes involved in the partial bioconversion of indene we constructed subclones of pR4. The indigo formation assay was repeated to screen for clones that retained ring-hydroxylating dioxygenase activity. Positive clones were analyzed by restriction-enzyme digestion. Subclone pR4-10, which retained indigo-forming activity and had an insert approximately 7.0 kb in length, was chosen for sequence analysis. Subclone pR4-10, when transformed into *R. erythropolis* SQ1, is also capable of converting indene to indandiol (data not shown).

The entire insert of pR4-10 (7 kb) and approximately 4 kb of pR4 were subjected to DNA sequencing. Four open reading frames (ORF) in the sequenced DNA have significant homology to genes found in dioxygenase operons (Fig. 4). The largest open reading frame is 1407 bp and shows homology to genes encoding the large subunit of a variety of dioxygenases. We have named this gene *nidA* for naphthalene-inducible dioxygenase system. Analysis of the translated open reading frame shows a potential Rieske-type iron-sulfur-binding center that exactly matches the published consensus sequence ($CXHRGX_8GNX_5CXYHG$) (Mason and Cammack 1992). Furthermore, there are four histidines and three tyrosines near the middle of the polypeptide encoded by pR4, any of which could contribute to a potential iron-binding site.

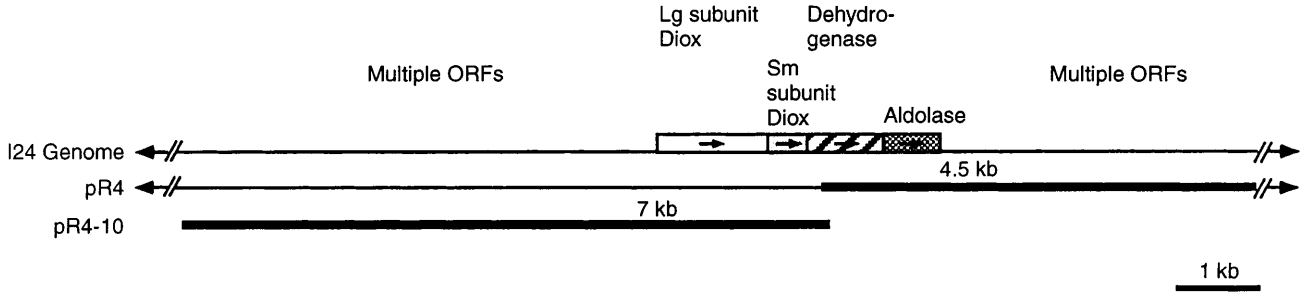
Next to *nidA* is a 519-bp ORF with homology to dioxygenase small subunit genes. We have named this ORF *nidB*. The small subunit of the dioxygenase is followed by another ORF of approximately 1.0 kb with high homology to *cis*-dihydrodiol dehydrogenases, which we have named *nidC*. Analysis of the derived protein sequence shows that NidC has the characteristic C-terminal YXXXK motif thought to be the active site of these dehydrogenases, and a potential NAD^+ binding site (consensus site GXXXGXG) in the N terminus of the polypeptide (Jornvall et al. 1995). Phylogenetic analysis based on aligned sequences demonstrates that the *nidA*, *nidB* and *nidC* genes diverged early from their nearest neighbors (Fig. 5).

A fourth ORF, located downstream of the *nidC* dehydrogenase, is 756 bp in length and exhibits homology to aldolase genes. Phylogenetic analysis suggests that this gene is not very similar to its most homologous neighbors (Fig. 5D). We have named the putative aldolase *nidD*. The regions upstream and downstream of these four genes (*nidABCD*) contain open reading frames that exhibit no significant homology to anything in the GenEMBL non-redundant databases. All four of the open reading frames identified by homology appear to have ribosome-binding sites associated with them, based on the canonical ribosome-binding sites from *E. coli* (Shine and Dalgarno 1975).

Discussion

Using a functional assay in *E. coli* and DNA sequencing, we identified an 11.0-kb fragment of *Rhodococcus* I24 DNA encoding four genes (*nidABCD*) that includes a dioxygenase and a dehydrogenase capable of breaking down indene to *cis*-(1*R*,2*S*)-indandiol, 1-indenol, 1-indanone, and ketohydroxyindan. Nine overlapping cosmid clones were identified that carry these genes. On the basis of our genomic analysis of the strain, we believe

Fig. 4 Sequence analysis and genome structure. The structure of the region encoding the dioxygenase was determined by sequencing pR4 and pR4-10. The portion of pR4 sequenced is denoted by the *thick line*. In addition to the large and small subunits of the dioxygenase, the dehydrogenase and putative aldolase, multiple open reading frames (*ORF*) were identified both upstream and downstream of the denoted gene cluster. None of these ORF showed significant homology to known genes in BLAST searches of the GenEMBL databases



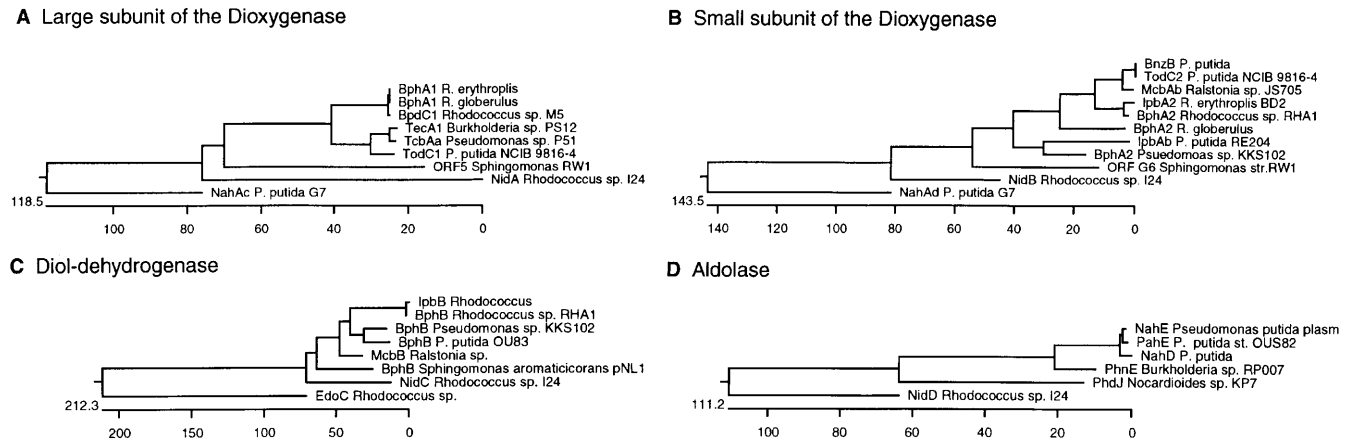


Fig. 5A–D Phylogenetic analysis of *nidABCD* and related genes. Phylogenetic analysis of aligned DNA-derived protein sequences with the highest homology to the *nid* genes was conducted. The alignment was performed using the CLUSTAL method with a PAM250 residue-weight table. The *x*-axis of this unbalanced tree represents the number of substitutuion events, thus describing the divergence between sequences. **A** The large subunit of the dioxygenase is analyzed. The GenBank accession numbers for the indicated genes are (from *top* to *bottom*) D88020, X80041, U27591, U15298, J04996, AJ223219, AF121905, 484406. **B** Analysis of the small subunit of the dioxygenase. GenBank accession numbers are (from *top* to *bottom*) P08085, J04996, 3184045, U24277, D32142, S51758, 2822266, Q52439, AJ223219, AF121905, 484406. **C** The dehydrogenase in analyzed. GenBank accession numbers for the genes indicated are (from *top* to *bottom*) AJ006127, D32142, D17319, Y07655, AJ006307, AF07931, AF121905, AJ006126. **D** Analysis of the aldolase. GenBank accession numbers are (from *top* to *bottom*) U09057, AB004059, AF010471, AF061751, AB000735, AF121905

that these genes could be located on the large plasmid found in the strain. This plasmid could be present in multiple copies, thereby accounting for the disproportionate number of clones identified. To address these issues we are conducting a further analysis of this plasmid and its association with these genes.

Bioconversion analysis suggests that the genes *nidAB* correspond to the subunits of the naphthalene-inducible dioxygenase proposed for *Rhodococcus* I24 (Chartrain et al. 1998) since pR4 only produces the *cis*-(1*R*,2*S*)-indandiol when transformed into the naïve host *R. erythropolis* SQ1 (Fig. 1). This finding strengthens the hypothesis that multiple dioxygenases produce the different indandiols in *Rhodococcus* I24 in a stereospecific manner (Buckland et al. 1999; Chartrain et al. 1998). The lack of *cis*-(1*S*,2*R*)-indandiol and *trans*-(1*R*,2*R*)-indandiol production by the naïve *R. erythropolis* SQ1 containing pR4 also supports this notion. The absence of these enantiomers suggests that they are produced by separate oxygenases. We believe these enzymes to be a toluene-inducible dioxygenase and a naphthalene-inducible monooxygenase respectively (Chartrain et al. 1998).

Other evidence that is consistent with pR4 encoding a functional dioxygenase is the production of indenol and indanone. It has previously been shown that indenol and indanone can be produced by both naphthalene and toluene dioxygenases when indene is a substrate (Gibson 1995; Wackett et al. 1988). The formation of indenol and indanone in *Rhodococcus* I24 can now be attributed, at least in part, to the *nidAB* dioxygenase. It may be possible that the other oxygenases in the system can contribute to the formation of these products as well. We are currently investigating substrate utilization, gene deletion and complementation studies with the *nidAB* dioxygenase genes so that we may modify this bioconversion pathway.

Sequence analysis strongly suggests that the *nidABCD* genes belong to a family of genes capable of degrading aromatic compounds. Comparison of the *nidABCD* dioxygenase system with other dioxygenase systems in their overall gene organization suggests that the *nid* gene organization and genes in *Rhodococcus* I24 are different from other dioxygenase operons. Some examples of the gene organization of operons involved in aromatic compound degradation are given in Fig. 6. By sequence homology, these operons are some of the most closely related to the *nidABCD* genes. A general pattern can be seen in the gene organizations. The *nid* genes appear to diverge from this general pattern. When the homologous sequences are aligned and subjected to phylogenetic analysis the *nid* genes diverge early from their most homologous relatives (Fig. 5). This is consistent with the divergence seen in the organization of the ORF.

One way the *nid* genes are different from the general pattern is in the lack of an identifiable reductase, ferredoxin, or catechol dioxygenase within the 11 kb of DNA sequenced. The ferredoxin and the reductase components are required for electron transfer to the dioxygenase components in systems that are similar to *nidAB* (Mason and Cammack 1992). Therefore it is reasonable to suspect that these two components are needed for the dioxygenase system from pR4. This suggests that there may be as yet unidentified genes for the ferredoxin and the reductase on pR4, or that these components are unnecessary for this dioxygenase system. Another possibility is that the *nidAB* dioxygenase subunits can borrow the ferredoxin and reductase components from elsewhere in the cell. It has been

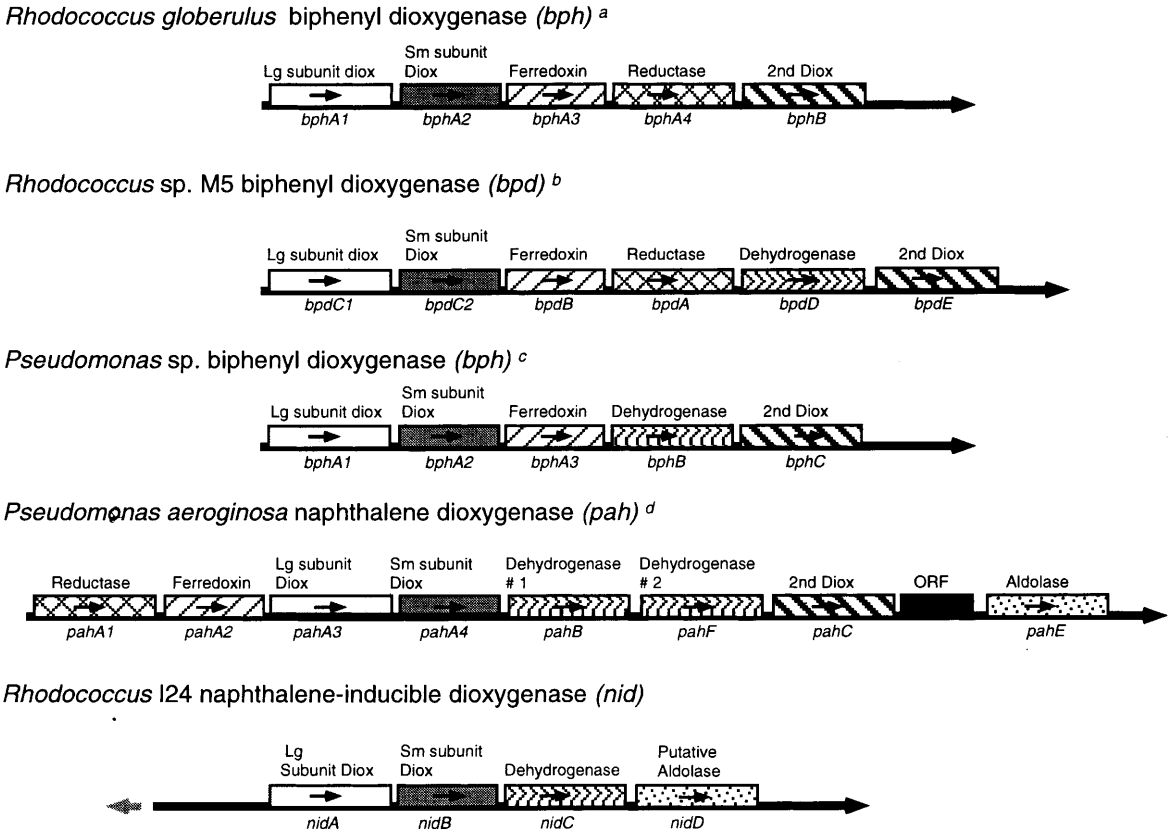


Fig. 6 Gene organization of dioxygenase systems. Four different dioxygenase systems that are similar to the *Rhodococcus* I24 *nidABCD* system by homology are compared to each other by their gene organization. [a] Asturias et al. 1995; [b] Wang et al. 1995; [c] Fukuda et al. 1994; [d] Takizawa N, Iida T, Yamauchi K, Satoh S, Wang Y, Fukuda M, and Kiyohara H, direct submission to Genbank, accession number: D84146

previously demonstrated that the dioxygenase subunits from one system or hybrid dioxygenase subunits, can borrow the ferredoxin and reductase components of another system (Beil et al. 1998; Furukawa et al. 1993). In addition, it has been demonstrated in *Pseudomonas* that the reductase component can be located elsewhere in the operon, separated from the terminal dioxygenase subunits by a number of intervening genes (Kikuchi et al. 1994) or scattered in the genome (Armengaud et al. 1998). However, if the latter were the case in *Rhodococcus* I24, a functional dioxygenase might not be identified in our screen. We plan to investigate the unidentified ORFs in the I24 *nid* region further in functional studies.

Many strains of *Rhodococcus* are capable of breaking down aromatic compounds. *Rhodococcus* I24 is unique in that it employs at least three different types of oxygenases capable of acting upon indene. This bioconversion is stereospecific and inducible. The gene structure for the *Rhodococcus* I24 naphthalene-inducible dioxygenase is unlike any dioxygenase gene cluster identified to date. We are currently investigating other genes on this cosmid clone in order to better understand this dioxygenase system and indene bioconversion in *Rhodococcus* I24. In our study of bacteria capable of stereospecific bioconversion, we hope to develop efficient and controllable methods of producing important chiral compounds and to gain a better understanding of how this organism and others like it metabolize aromatic substrates.

Acknowledgements This work was supported by a grant from Merck Research Laboratories. S. L. T. and K. S. Y. were supported in part by NIH Biotechnology Training Grant 2T32 GM08334-10. We would like to thank Merck Research Laboratories for the strains, sequencing, and chiral analysis. We also thank members of the Sinskey Lab for their discussion, comments and support. We thank A. Chapman, M. Hillis, and I. Kim for all of their assistance. The experiments presented in this publication comply with the current laws of the United States of America.

References

Allen CCR, Boyd DR, Larkin MJ, Reid KA, Sharma ND, Wilson K (1997) Metabolism of napthalene, 1-napthol, indene, and indole by *Rhodococcus* sp. strain NCIMB 12038. Appl Environ Microbiol 63: 151–155

Altschul SF, Gish W, Miller W, Myers EW, Lipman AJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Armengaud J, Happe B, Timmis KN (1998) Genetic analysis of dioxin dioxygenase of *Sphingomonas* sp. Strain RW1: catabolic genes dispersed on the genome. J Bacteriol 180: 3954–66

Asturias JA, Timmis KN (1993) Three different 2,3-dihydroxybiphenyl-1,2-dioxygenase genes in the gram-positive polychlorobiphenyl-degrading bacterium *Rhodococcus globerulus* P6. J Bacteriol 175: 4631–4640

Asturias JA, Diaz E, Timmis KN (1995) The evolutionary relationship of biphenyl dioxygenase from gram-positive *Rhodo-*

coccus globerulus P6 to multicomponent dioxygenases from gram-negative bacteria. Gene 156: 11–18

Bairoch A (1992) PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res 20: 2013–2018

Beil S, Mason JR, Timmis KN, Pieper DH (1998) Identification of chlorobenzene dioxygenase sequence elements involved in dechlorination of 1,2,4,5-tetrachlorobenzene [in process citation]. J Bacteriol 180: 5520–5528

Buckland B, Drew S, Connors N, Chartrain M, Lee C, Salmon P, Gbewonyo K, Gailliot P, Singhvi R, Olewinski R, Sun W, Reddy J, Zhang J, Zhou W, Jackey B, Goklen K, Junker B, Greasham R (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

Collins AN, Sheldrake GN, Crosby J (1997) Chirality in industry. II. Developments in manufacturing and applications of optically active compounds. Wiley, New York

Ensley BD, Ratzkin BJ, Osslund TD, Simon MJ, Wackett LP, Gibson DT (1983) Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. Science 222: 167–9

Fukuda M, Yasukochi Y, Kikuchi Y, Nagata Y, Kimbara K, Horiuchi H, Takagi M, Yano K (1994) Identification of the *bphA* and *bphB* genes of *Pseudomonas* sp. strains KKS102 involved in degradation of biphenyl and polychlorinated biphenyls. Biochem Biophys Res Commun 202: 850–6

Furukawa K, Hirose J, Suyama A, Zaiki T, Hayashida S (1993) Gene components responsible for discrete substrate specificity in the metabolism of biphenyl (*bph* operon) and toluene (*tod* operon). J Bacteriol 175: 5224–32

Gibson DT, Subramanian V (1984) Microbial degradation of aromatic hydrocarbons. In: Gibson DT (ed) Microbial degradation of organic compounds, vol 13. Macel Dekker, Inc., New York, pp 181–252

Gibson DT, Resnick SM, Lee K, Brand JM, Torok DS, Wackett LP, Schocken MJ, Haigler BE (1995) Desaturation, dioxygenation, and monooxygenation reactions catalyzed by napthalene dioxygenase from *Pseudomonas* sp. strain 9816-4. J Bacteriol 177: 2615–2621

Gish W, State DJ (1993) Identification of protein coding regions by database similarity search. Nat Genet 3: 266–272

Harayama S, Kok M, Neidle EL (1992) Functional and evolutionary relationships among diverse oxygenases. Annu Rev Microbiol 46: 565–601

Hart S, Koch KR, Woods DR (1992) Identification of indigo-related pigments produced by *Escherichia coli* containing a cloned *Rhodococcus* gene. J Gen Microbiol 138: 211–6

Jornvall H, Persson B, Krook M, Atrian S, Gonzalez-Duarte R, Jeffery J, Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). Biochemistry 34: 6003–13

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

Kikuchi Y, Nagata Y, Hinata M, Kimbara K, Fukuda M, Yano K, Takagi M (1994) Identification of the *bphA4* gene encoding ferredoxin reductase involved in biphenyl and polychlorinated biphenyl degradation in *Pseudomonas* sp. strain KKS102. J Bacteriol 176: 1689–94

Kosono S, Maeda M, Fuji F, Arai H, Kudo T (1997) Three of the seven *bphC* genes of *Rhodococcus erythropolis* TA421, isolated from a termite ecosystem, are located on an indigenous plasmid associated with biphenyl degradation. Appl Environ Microbiol 63: 3282–3285

Lai E, Birren B (1990) Electrophoresis of large DNA molecules: theory and applications. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Masai E, Yamada A, Healy JM, Hatta T, Kimbara K, Fukuda M, Yano K (1995) Characterization of biphenyl catabolic genes of gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1. Appl Environ Microbiol 61: 2079–2085

Mason JR, Cammack R (1992) The electron-transport proteins of hydroxylating bacterial dioxygenases. Annu Rev Microbiol 46: 277–305

Patel TR, Gibson DT (1974) Purification and propeties of (+)-cis-naphthalene dihydrodiol dehydrogenase of Pseudomonas putida. J Bacteriol 119: 879–88

Quan S, Dabbs ER (1993) Nocardioform arsenic resistance plasmid characterization and improved *Rhodococcus* cloning vectors. Plasmid 29: 74–9

Sambrook J, Fritisch E, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd edition edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Shine J, Dalgarno L (1975) Determinant of cistron specificity in bacterial ribosomes. Nature 254: 34–38

Wackett LP, Kwart LD, Gibson DT (1988) Benzylic monooxygenation catalyzed by toluene dioxygenase from *Pseudomonas putida*. Biochemistry 27: 1360–7

Wang Y, Garnon J, Labbe D, Bergeron H, Lau PC (1994) Sequence and expression of the *bpdCICZBADE* genes involved in the initial steps of biphenyl/chlorobiphenyl degradation by *Rhodococcus* sp MS

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103–19

Zhang Y, Praszkier J, Hodgson A, Pittard AJ (1994) Molecular Analysis and Characterization of a Broad-Host-Range Plasmid, pEP2. J Bacteriology 176: 5718–5728

Zylstra GJ, McCombie WR, Gibson DT, Finette BA (1988) Toluene degradation by *Pseudomonas putida* F1: genetic organization of the tod operon. Appl Environ Microbiol 54: 1498–503