Appl Microbiol Biotechnol (1999) 52: 811–819

ORIGINAL PAPER

P. A. Lessard · X. M. O'Brien · N. A. Ahlgren
S. A. Ribich · A. J. Sinskey

Characterization of IS*1676* from *Rhodococcus erythropolis* SQ1

Received: 1 April 1999 / Received revision: 6 July 1999 / Accepted: 1 August 1999

Abstract To develop a transposable element-based system for mutagenesis in *Rhodococcus*, we used the *sac*B gene from *Bacillus subtilis* to isolate a novel transposable element, IS*1676*, from *R. erythropolis* SQ1. This 1693 bp insertion sequence is bounded by imperfect (10 out of 13 bp) inverted repeats and it creates 4 bp direct repeats upon insertion. Comparison of multiple insertion sites reveals a preference for the sequence 5′-(C/T)TA(A/G)-3′ in the target site. IS*1676* contains a single, large (1446 bp) open reading frame with coding potential for a protein of 482 amino acids. IS*1676* may be similar to an ancestral transposable element that gave rise to repetitive sequences identified in clinical isolates of *Mycobacterium kansasii*. Derivatives of IS*1676* should be useful for analysis of *Rhodococcus* strains, for which few other genetic tools are currently available.

Introduction

The genus *Rhodococcus* includes many bacteria with practical applications in environmental and industrial biotechnology (recently reviewed in Bell et al. 1998). Although the genus also includes species that cause disease in plants and animals, the primary interest in *Rhodococcus* focuses on the ability of several species to transform or degrade a wide variety of compounds, including aromatic compounds, herbicides and halogenated hydrocarbons. Currently, there are 12 recognized species of *Rhodococcus* (Bell et al. 1998), although classification of bacteria within this genus is continually evolving. Many of the *Rhodococcus* isolates that have been studied for their biodegradative abilities have not been classified at the species level (e.g. Allen et al. 1997; Denome et al. 1994; Shao and Behki 1995).

Buckland et al. (1999) and Chartrain et al. (1998) recently described two isolates of *Rhodococcus*, designated I24 and B264-1, that are able to degrade naphthalene or toluene. These bacteria have also been found to oxygenate indene to a variety of indandiols, although they cannot use this substrate as a sole carbon source. The ability to produce indandiol from indene may be of practical value, as indandiol is a key intermediate for the synthesis of indinavir sulfate, an HIV protease inhibitor sold under the name Crixivan (Buckland et al. 1999).

Treatment of the I24 and B264-1 strains with chemical mutagens has produced isolates with altered or impaired indene conversion properties (M. Chartrain, unpublished observations; P. Lessard and A. Sinskey, unpublished observations). However, it is difficult to isolate the genes responsible for these altered phenotypes, particularly in bacteria that are poorly characterized at the genetic level. One resource that might accelerate the identification of relevant genes would be a transposon-based mutagenesis system. Although no "transposons" (i.e. compound transposable elements that carry selectable markers) have been found in *Rhodococcus*, it should be possible to derive artificial transposons from naturally occurring insertion sequences (IS elements) or other transposable elements.

Working with different strains of *Rhodococcus*, several researchers have identified open reading frames (ORFs) that are homologous to the putative tranposases from known insertion sequences (Denome and Young 1995; Komeda et al. 1996; Grzeszik et al. 1997; Kulakova et al. 1997; Eulberg et al. 1998; Kulakov and Larkin 1998; Seibert et al. 1998). The majority of these elements have been found serendipitously by sequencing genomic clones carrying operons involved in the

P. A. Lessard · X. M. O'Brien · N. A. Ahlgren
S. A. Ribich · A. J. Sinskey (✉)
Building 68-370, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
e-mail: asinskey@mit.edu
Tel.: +1 617-253-6721
Fax: +1 617-253-8550

Present address:
S. A. Ribich
Massachusetts General Hospital, Department of Neuroscience, Room 6311, 149 13th Street, Charlestown Navy Yard, Charlestown, MA 02120, USA

degradation of various compounds. However, other than the work of Denome and Young (1995) and Komeda et al. (1996), none of the elements identified in this manner carried other features common to IS elements, such as inverted terminal repeats or target site duplications, and only Denome and Young (1995) presented evidence that one of these elements (IS*1166*) may be transpositionally acitive. The lack of structural features common to IS elements also refers to a presumed transposon that was found among the genes for chlorocatechol metabolism (Eulberg et al. 1998) for which the authors reported sequences of an "imperfect inverted repeat", which do not convincingly prove the occurrence of such a structure. In the absence of direct evidence, it would be difficult to argue that these elements are competent to transpose. Therefore, most of the above mentioned elements would not be useful for developing a transposon based mutagenesis system in *Rhodococcus*.

In contrast, Nagy et al. (1997) discovered an IS element, IS*1415*, while sequencing a gene cluster involved in the degradation of thiocarbamate herbicides. This element not only contains two ORFs homologous to transposition related sequences from IS*21*, but it also bears lengthy inverted repeats at its termini, and copies of the element were associated with 5–6 bp target site duplications in the genome of *R. erythropolis* NI86/21. These authors went on to develop two artificial transposons, Tn*5561*X1 and Tn*5561*X2, by inserting a chloramphenicol resistance marker into IS*1415*. They found these recombinant elements were able to transpose in a new host, *R. erythropolis* SQ1.

The only researchers to date who have identified transposable elements in *Rhodococcus* by their ability to transpose were also the first to identify IS elements in this genus. Jaeger et al. (1995) used the *sac*B gene from *Bacillus subtilis* to trap an insertion sequence in *R. fascians*. *sac*B encodes levan sucrase, an enzyme that hydrolyzes sucrose and polymerizes fructan polymers (Gay et al. 1983). In Gram-negative bacteria, expression of *sac*B is lethal in the presence of sucrose, making it a useful tool for trapping IS elements (Gay et al. 1985). Jaeger et al. (1995) extended this strategy to four genera of Gram-positive bacteria, including *Rhodococcus*, and identified IS-*Rf* in *R. fascians*.

Our goal is to develop a transposon-based mutagenesis system to study the genes involved in indene conversion in *Rhodococcus* sp. strains I24 and B264-1. To avoid the problems associated with homologous recombination between elements or autorepression of transposable elements (Ohtsubo and Sekine 1996), we have chosen to isolate a transposable element from *R. erythropolis* SQ1, an easily transformed derivative of ATCC strain 4277-1 (Quan and Dabbs 1993).

Here we describe the use of the *sac*B gene from *B. subtilis* to isolate an insertion sequence, IS*1676*, from *R. erythropolis* SQ1. This element is approximately 1.7 kb in length, including a single large open reading frame. It has imperfect, inverted repeats at its termini and it creates a 4 bp duplication of the target site upon insertion. This element is not homologous to any of the other elements described in *Rhodococcus* and represents a new family of insertion sequences. IS*1676* should prove useful for genetic analysis of *Rhodococcus* strains such as I24 and B264-1.

Materials and methods

Reagents

Fine chemicals were purchased from Sigma (St. Louis, Mo.). Reagents for culture media were purchased from Difco Laboratories (Detroit, Mich). LB medium was prepared as described in Sambrook et al. (1989). LB plates contained 2% Bactoagar. When necessary, kanamycin (150 μg/ml) or sucrose (5%) were included in liquid or solid media. Unless otherwise stated, oligonucleotide primers were purchased from Gibco/BRL Life Technologies (Grand Island, N.Y.).

Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are presented in Table 1. *R. erythropolis* SQ1 colonies were tested via microscopic examination and a sugar acidification assay to rule out contamination. The sugar acidification assay involved mixing a small amount of filter-sterilized phenol red solution with minimal medium (Chartrain et al. 1998) containing one of six different sugars (glucose, sucrose, maltose, lactose, galactose, or mannose) as primary carbon source. Samples from candidate *R. erythropolis* SQ1 colonies were streaked onto these media, as were samples of the reference strain (control). Following 3–5 days incubation at 30 °C, media containing glucose and sucrose alone take on a yellow hue in the presence of *R. erythropolis* SQ1, while other media remained red.

Minipreparation of plasmid DNA

Plasmid DNA was isolated from *Escherichia coli* strains as described in Sambrook et al. (1989). Plasmid DNA was isolated from *R. erythropolis* SQ1 by a method modified from that of Vogt Singer and Finnerty (1988). *R. erythropolis* SQ1 culture (1.5 ml) was centrifuged at 13 000 rpm in a microcentrifuge for 60 s. The pellet was then resuspended in 400 μl TENS (50 mM Tris · HCl, pH 8; 10 mM EDTA, pH 8; 50 mM NaCl; 20% sucrose) containing 5 mg/ml freshly prepared lysozyme and incubated at 37 °C for 30–120 min. Each sample was then supplemented with 185 μl 10% sodium dodecyl sulfate and 30 μl 1 M Tris · NaOH, pH 12.6, and incubated at 55 °C for 30–120 min. Following addition of 300 μl potassium acetate (3 M potassium, 5 M acetate), each sample was incubated on ice for 5 min, then centrifuged at 4 °C at top speed in an Eppendorf microcentrifuge for 20 min. Supernatants were transferred to fresh microcentrifuge tubes and precipitated with 450 μl isopropanol. Following 20 min microcentrifugation at 4 °C, pellets were washed once with cold 70% ethanol, dried, and resuspended in 50 μl TE (10 mM Tris · HCl, pH 8; 1 mM EDTA, pH 8).

DNA manipulations

Restriction enzyme digestions, ligations, and other routine plasmid manipulations were carried out using enzymes from New England Biolabs (Beverly, Mass.) according to the manufacturer's recommendations. During construction, plasmids were maintained in *E. coli* strains XL1-Blue or JM109 (Table 1). The polymerase chain reaction (PCR; Mullis and Faloona 1987) was carried out with reagents from the Boehringer Mannheim (Indianapolis, Ind.) PCR Core Kit using 100 μl reactions in an MJ Research DNA Engine

Table 1 Strains and plasmids. (*ampR* ampicillin resistance, *cmR* chloramphenicol resistance, *kanR* kanamycin resistance, *NalR* confers nalidixic acid resistance, *strR* confers streptomycin resistance, *sucS* confers sucrose sensitivity)

Bacterial strains and plasmids	Genotype or relevant characteristics	Source or reference
Rhodococcus erythropolis SQ1	Highly transformable derivative of ATCC strain number 4277-1	(Quan and Dabbs 1993)
Escherichia coli		
JM109	F′ *traD36 lacI*^{q} Δ(*lacZ*)*M15 proA*$^+$ *B*$^+$/*e14*$^-$(*McrA*$^-$) Δ(*lac-proAB*) *thi gyrA96*(NalR) *endA1 hsdR17* ($r_k^- m_k^+$) *relA1 supE44 recA1*	NE Biolabs
TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX*74 *recA*1 *deoR araD*139 Δ(*ara-leu*)7697 *galU galK rpsL* (strR) *endA*1 *nupG*	Invitrogen
XL1-Blue	F′::Tn*10 proA*$^+$ *B*$^+$ *lacI*^{q} Δ(*lacZ*)*M15*/*recA1 endA1 gyrA96* (NalR) *thi hsdR17* ($r_k^- m_k^+$) *supE44 relA1 lac*	NE Biolabs
Plasmids		
pAL233	derivative of pPWS2 carrying IS*1676*	This report
pCR2.1	Vector for cloning PCR products; ampR, kanR	Invitrogen
pEP2	Broad host range plasmid derived from pNG2 of *Corynebacterium diphtheriae*; replicates in *E. coli* and *Rhodococcus*; kanR	(Zhang et al. 1994)
pPWS1	*lacZsac*:B PCR product in pCR2.1; sucS	This report
pPWS2	*lacZsac*:B from pPWS2 in pEP2; sucS, kanR	This report
pUCD4121	pTZ18RC::*sac*B, cmR sucS	(Kamoun et al. 1992)
pXS26	derivative of pPWS2 carrying IS*1676*	This report

Peltier Thermal Cycler. Amplification conditions included an initial denaturation step at 94 °C for 3 min, 30 cycles of 1 min at 94 °C, 2 min at 43 °C, and 3 min at 72 °C, followed by an additional 10 min at 72 °C. Genomic DNA was prepared from *Rhodococcus* and *Corynebacterium* strains as described previously (Treadway et al. 1999). Sequencing of plasmid DNA was carried out at the MIT Biopolymers Facility using an ABI Cycle Sequencer and the Big-Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, Calif.). DNA sequence comparisons with entries from the GenEMBL non-redundant database were conducting using the BLAST program (Altschul et al. 1997).

Construction of the transposon trap, pPWS2

The *sac*B gene and the associated *lac*Z promoter were amplified from pUCD4121 (Kamoun et al. 1992) via PCR with the forward and reverse primers purchased from Promega (Madison Wisconsin, Table 2). The resulting ca. 2 kb PCR product was ligated into pCR2.1 using the TOPO-TA cloning kit (Invitrogen, Carlsbad, Calif.), generating the plasmid pPWS1. A 2 kb *Nsi*I-*Bam*HI fragment from pPWS1 was ligated into the *Pst*I and *Bam*HI sites of pEP2 (Zhang et al. 1994) to create pPWS2 (Fig. 1A).

Transformation of bacterial cells

Plasmids were introduced into *E. coli* strains via electroporation using the Gene Pulser apparatus (BIO-RAD, Richmond, Calif.) and the manufacturer's recommended protocol. *R. erythropolis* SQ1 cells were also transformed via electroporation as described previously (Treadway et al. 1999) except that LB was used as the recovery medium following electroporation.

Southern blot analysis (Southern 1979)

Genomic DNA from *Rhodococcus* and *Corynebacterium* was digested overnight with various restriction enzymes, then separated on 0.6% agarose gels. DNA was transferred to charged nylon membranes (Boehringer-Mannheim) via alkaline transfer (Reed and Mann 1985). The probe was prepared with a 1.5 kb *Bst*BI-*Eag*I fragment from IS*1676* (see also Fig. 2) using the Boehringer-Mannheim DIG High Prime DNA Labeling and Detection Starter Kit II. Hybridization and detection of the Southern blot were carried out using reagents as described in the kit. The final washes of the filter following hybridization were carried out in 0.5 × SSC, 0.1% SDS at 65 °C.

Inverse PCR

Inverse PCR (Ochman et al. 1989) involved digesting ca. 4 μg genomic DNA from *R. erythropolis* SQ1 in 40 μl reaction volumes with one or two restriction enzymes as follows: *Mse*I; *Apa*I; *Bss*HII and *Mlu*I; *Mse*I and *Bfa*I; or *Pst*I. Subsequently, heating to 80 °C for 20 min inactivated the restriction enzymes. 20 μl ligations were prepared using either 2 μl or 17.5 μl of the digested DNA mixture and incubated overnight at 16 °C. 2 μl from each ligation reaction were used for PCR amplification with primers SQ1-IS5 and

Table 2 Primers. For reference, the *lac*Z promoter and *sac*B gene lie between nucleotide positions 200 and 1772 in pUCD4121; the *sac*B gene in pPWS2 lies between positions 40 and 1461. Positions in IS*1676* are numbered relative to the first nucleotide in the left inverted repeat (see also Fig. 2)

Primer name	Sequence	Anneals to (relative nucleotide positions):
Forward	5′-ACCGTATTACCGCCTTTG-3′	pUCD4121 (2130–2113)
Reverse	5′-TAAGTTGGGTAACGCCAG-3″	pUCD4121 (103–121)
pEP2-L	5′-GCTTCAAAGCATGACTTCCT-3′	pPWS2 (5024–5044)
sacBKpn	5′-GCTCTCGGTATGATTGAGCT-3′	pPWS2 (528–509)
SQ1-IS1	5′-CGTACCAACAAGATTCGTCGGT-3′	IS*1676* (1542–1521)
SQ1-IS2	5′-CCAAACTCTCCATCGACACCA-3′	IS*1676* (148–168)
SQ1-IS3	5′-CCAAGGTTTGCAGGCTTCCGA-3′	IS*1676* (1173–1153)
SQ1-IS4	5′-CCATCCCTTGTTGGCGTTCGT-3′	IS*1676* (644–664)
SQ1-IS5	5′-GGTTGCAATGTCTGCGAGGCA-3′	IS*1676* (365–345)
SQ1-IS6	5′-GCAACCTACCTTTGTACGGT-3′	IS*1676* (1240–1259)

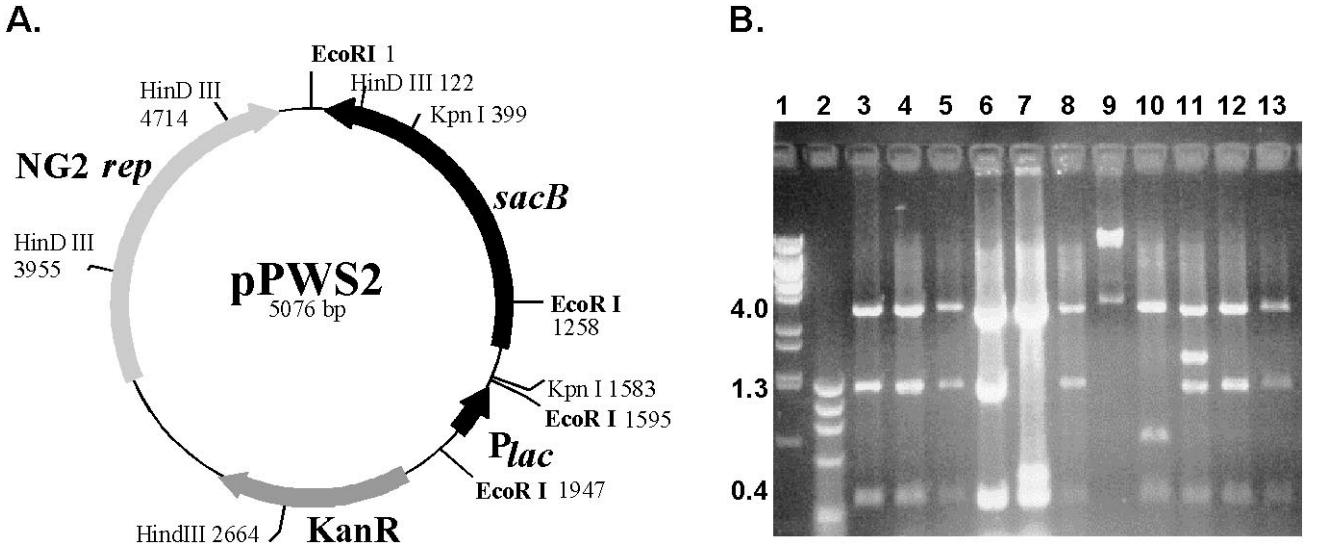


Fig. 1 **A** Diagram of pPWS2 containing the *sac*B gene from *B. subtilis*. Abbreviations: *NG2 rep* ORF required for replication from pNG2 (Zhang et al. 1994); P_{lac} lac promoter; *KanR* kanamycin resistance marker. **B** Restriction enzyme analysis of plasmids from several sucrose-resistant candidates carrying derivatives of pPWS2. Relative sizes (in kb) of major bands are indicated to the left of the image. Plasmid DNA was digested with *Eco*RI and separated on a 1% agarose gel. *Lanes 1* and *2* DNA molecular weight markers; *lane 3* pPWS2 (control); *lanes 4–13* pPWS2 derivatives from sucrose-resistant candidates. Plasmids in *lanes 7* and *10* have suffered deletions affecting the 1.3 kb *Eco*RI fragment from the *sac*B gene. Further testing of the plasmid in *lane 9* suggested that this candidate either had suffered some gross rearrangement or was not related to pPWS2. The plasmid in *lane 11* carried an insertion of approximately 1.7 kb and was labeled pAL233

SQ1-IS6 (Table 2), except for the amplification of *Pst*I-digested DNA, in which the primers SQ1-IS3 and SQ1-IS6 were used. PCR products were then ligated into pCR2.1 using the TOPO-TA cloning kit (Invitrogen) and propagated in the TOP10 strain of *E. coli* (Table 1) DNA sequencing of the cloned PCR products was carried out as described above, using forward and reverse sequencing primers (Invitrogen).

Nucleotide sequence accession number

The complete nucleotide sequence of IS*1676* described in this report has been deposited in GenBank under the accession number AF126281.

Fig. 2 Schematic diagram of IS*1676*. Positions of the left (*LIR*) and right (*RIR*) inverted repeats, the large central open reading frame (*ORF1*), and restriction sites mentioned in the text are indicated. The relative positions and 5′–3′ orientations of primers SQ1-IS1 through SQ1-IS6 that were used for sequencing and inverse PCR are indicated beneath the element. Primers pEP2-L and sacBKpn anneal outside this element in pAL233 and are not shown

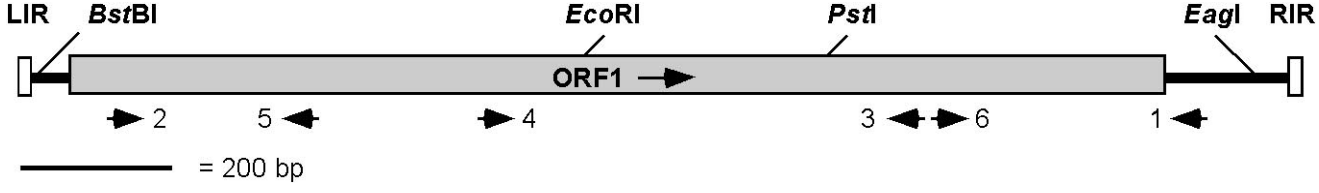
Results

Analysis of sucrose resistant clones

To prepare a plasmid for trapping a transposable element in *R. erythropolis* SQ1, we inserted the *sac*B gene from *B. subtilis* into the plasmid pEP2, creating the plasmid pPWS2 as described in Materials and methods. Derived from pNG2, a plasmid from *Corynebacterium diphtheriae*, pEP2 can replicate in *Corynebacterium* species, *Mycobacterium* species and *E. coli* (Zhang et al. 1994), making it a versatile plasmid vector. We have found that pEP2 will also replicate in *Rhodococcus* strains, including *R. erythropolis* SQ1, *Rhodococcus* sp. strain I24, and *Rhodococcus* sp. strain B264-1 (data not shown).

The plasmid pPWS2 was introduced into *R. erythropolis* SQ1 via electroporation. Restriction enzyme digests showed that plasmids recovered from kanamycin resistant colonies were identical to pPWS2 (not shown). We tested colonies for sensitivity to sucrose by streaking samples onto LB plates containing kanamycin and sucrose. Candidates that had demonstrated sensitivity to sucrose were then inoculated into 5–10 ml aliquots of LB with kanamycin and cultured at 30 °C for 2–3 days. 100 µl aliquots from these cultures were then spread onto LB plates containing kanamycin and sucrose. Sucrose-resistant colonies were recovered after 3–5 days.

Sucrose-resistant candidates were then inoculated into 3 ml LB with kanamycin and incubated for 2–3 days at 30 °C and plasmid recovered by minipreparа-



tion. Restriction enzyme analysis of these plasmids revealed that the majority recovered in this manner either contained deletions within the *sac*B element or showed no noticeable change in structure, presumably indicating the presence of small deletions, insertions, or point mutations that disrupted the function of *sac*B (Fig. 1B). Another possibility is that other mutations in the genome may have affected localization of the levan sucrase or substrate transport, resulting in sucrose resistance (Jaeger et al. 1995). Fewer than 1% of the sucrose-resistant bacteria carried plasmids with insertions of any appreciable size. Plasmids from two such candidates were pAL233 (Fig. 1B) and pXS26.

Sequence analysis of the insertion element

Restriction enzyme analysis of pAL233 demonstrated that the inserted element lay between the *Hin*DIII and *Kpn*I sites near the 3′ end of the sacB gene (not shown). This region of pAL233 was sequenced, first using the primers pEP2-L and sacBKpn, followed by the primers SQ1-IS1, SQ1-IS2, SQ1-IS3 and SQ1-IS4 (Table 2). A diagram of the inserted element is presented in Fig. 2.

The element captured in pAL233 bears many of the characteristics of classical IS (Galas and Chandler 1989). The element is 1693 bp in length. It is bounded by imperfect (in this case, 10 of 13 base pairs match), inverted repeats. Insertion of the element created a 4 bp direct repeat "footprint" in the target sequence (see below), and the element contains a single, large (1446 bp) ORF, predicted to encode a protein of 482 amino acids. Predictions of protein structure (Rost and Sander 1993, 1994) and properties (Wilkins et al. 1998) suggest that the encoded protein has a pI of 9.99 and one or more helix-loop-helix domains. Based on these criteria this element has been designated IS*1676* by the Plasmid Reference Center (Stanford University School of Medicine).

Occurrence of IS*1676* in *Rhodococcus* and *Corynebacterium*

Southern blot analyses demonstrated that 4–7 copies of IS*1676* occur in the genome of *R. erythropolis* SQ1 (Fig. 3). Digestion of genomic DNA from *R. erythropolis* SQ1 with *Mse*I exposed four prominent bands, some of which, by virtue of their size, might comprise DNA fragments that carry two copies of IS*1676*. Digestion of the same DNA with *Bfa*I produced three prominent hybridization products as well as four weaker signals. It is possible that the weaker hybridization may indicate IS*1676*-like elements of lower homology or incomplete digestion of the genomic DNA with *Bfa*I. Interestingly, the smallest of the hybridizing bands in the *Bfa*I-digested is approximately 1.7 kb in length, corresponding precisely in size to IS1676. This suggests that at least one copy of IS*1676* is flanked immediately by

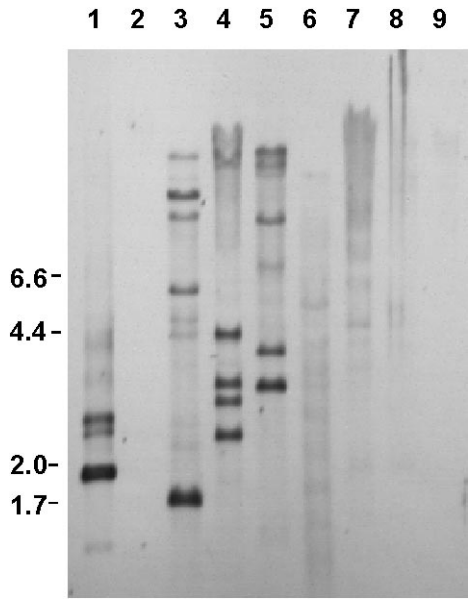


Fig. 3 Occurrence of IS*1676* in *Rhodococcus* and *Corynebacterium*. Genomic DNA from three strains of *Rhodococcus* and two strains of *Corynebacterium* was digested and examined for homology to IS*1676* via Southern analysis. *Lane 1* positive control plasmid carrying IS*1676*; *lane 2* DNA molecular weight markers (did not hybridize to the IS*1676* probe); *lanes 3–5* DNA from *R. erythropolis* SQ1 digested with *Bfa*1, *Mse*I, and *Apa*I, respectively; *Apa*I-digested DNA from *Rhodococcus* sp. strain I24 (*lane 6*), *Rhodococcus* sp. strain B264-1 (*lane 7*), *C. lactofermentum* ATCC 21799 (*lane 8*), and *C. glutamicum* E12 (*lane 9*). Sizes of DNA fragments (in kb) are indicated to the left of the figure

*Bfa*I sites. *Apa*I digestion similarly revealed four strongly hybridizing bands and two or three weaker signals.

Rhodococcus sp. strains I24, B264-1 and *R. erythropolis* SQ1 are three distantly related rhodococci (Chartrain et al. 1998), and *Corynebacterium* represents a related genus among the actinomycetes. The IS*1676* probe from *R. erythropolis* SQ1 did not hybridize appreciably to genomic DNA from the two other *Rhodococcus* strains or the two *Corynebacterium* strains tested, suggesting that no homologues of the element exist in these bacteria. Lower stringency washes or prolonged exposure of the filters revealed very diffuse, weak hybridization of the probe to these DNAs (not shown).

Target site specificity

Upon insertion into pPWS2, IS*1676* duplicated the 4 bp sequence 5′-TTAA-3′. To determine whether any homology exists among target sites for IS*1676* insertion, we compared the sequences flanking the two trapped elements in pAL233 and pXS26 with the sequences flanking several of the elements residing in the *R. erythropolis* SQ1 genome.

Terminal portions of the IS*1676* homologues were recovered from *R. erythropolis* SQ1 genomic DNA

along with adjacent sequences via inverse PCR. By comparing the flanking sequences recovered with each element, we determined that we had recovered at least five distinct IS1676 homologues in this manner (Fig. 4). This number is consistent with the data from the Southern hybridizations, which predicted 4–7 copies of the element. One element (genomic copy 6) was recovered from an inverse PCR reaction in which *Bfa*I (which cleaves the sequence 5′-CTAG-3′) had been used to digest the genomic DNA. Apparently *Bfa*I cleaved immediately in the 4 bp flanking the element, allowing the product to recircularize upon ligation. Thus, the base pairs adjacent to the left inverted repeat are identical to the base pairs comprising the right inverted repeat in this specimen. This result is consistent with the data from the Southern hybridization (Fig. 3), which suggested that at least one copy of IS*1676* is flanked by *Bfa*I sites in the genome. We found no sequence discrepancies among the recovered elements in the ca. 240 bp sequenced from inside the left terminus and the ca. 400 bp sequenced from inside the right terminus of each element (not shown). These observations indicate a strong conservation among the resident copies of IS*1676* in *R. erythropolis* SQ1.

Comparing the sequences flanking the elements (Fig. 4), we found no evidence for extended target sequence homology. G + C content in the regions adjacent to the inverted repeats ranges from 35% (for the element trapped in pAL233) to 62% (for the element represented by genomic copy 5). Within the 4 bp duplicated upon insertion, there is a clear preference for the degenerate sequence 5′-(C/T)TA(A/G)-3′. This may indicate some sequence specific interaction on the part of the transposase and its substrate DNA.

Fig. 4 Alignment of insertion sites for IS*1676* homologues. *LIR* and *RIR* sequences are indicated in upper case, and the intervening IS*1676* sequences are indicated by *dots*. Flanking sequences are indicated in lower case, with the duplicated 4-bp elements *underlined*. Genomic copies *1* through *4* were each recovered multiple times from separate inverse PCR reactions involving different restriction enzymes. The right-flanking sequence from genomic copy *4* is identical to that from genomic copy *2*, suggesting this product may be a chimera of two different elements in the genome. Genomic copy *7* provides information only for the right flanking sequence as this product was obtained using *Pst*I to digest the genomic DNA. Unlike the other enzymes used in this strategy, *Pst*I cuts within IS*1676*, making it possible to recover only "one half" of an element through inverse PCR

IS*1676* homology to other transposable elements

A BLAST search (Altschul et al. 1997) of the GenEMBL non-redundant database with the IS*1676* sequence retrieved only one other sequence of significant homology. The right half of IS*1676* bears extensive homology to an insertion sequence-like element from *Mycobacterium kansasii* (Yang et al. 1993). Approximately 66% of 960 bases from IS*1676* can be aligned with bases from the *M. kansasii* element (Fig. 5). Sequences from the helix-loop-helix region of the predicted protein encoded by IS*1676* showed very weak homology to a similar ORF from an IS element found in *R. opacus* 1CP (Seibert et al. 1998). In this span, 18 of 56 amino acids are identical (26 of 56 are similar) between the two putative proteins. However, the BLAST search used to identify this homology (tblastn Altschul et al. 1997) assigned a poor probability ($P = 0.13$) that this homology was significant. No homology was found for IS elements recovered from other *Rhodococcus* strains, including IS*1166* and IS*1295* (Denome and Young 1995), IS*1415* (Nagy et al. 1997), IS*2112* (Kulakov and Larkin 1998), IS*1164* (Komeda et al. 1996), and the transposase-like ORFs from *R. opacus* MR11 (Grzeszik et al. 1997), *R. opacus* 1CP (Eulberg et al. 1998), and the terminal sequences reported for IS-*Rf* (Jaeger et al. 1995).

Discussion

Bacterial insertion sequences are the smallest transposable elements capable of independent mobilization (Galas and Chandler 1989). Ranging in size from 800 to 2500 bp, the majority of IS elements are bracketed by imperfect, inverted repeats of about 10–40 bp, and they create short, direct repeats of the target DNA upon insertion (Ohtsubo and Sekine 1996). IS elements also carry one or more ORFs that are required for transposition (at least in the cases that have been tested); thus the encoded proteins are frequently termed "transposases." IS*1676* is a 1693 bp element with imperfect, inverted 13 bp repeats at its termini. It creates 4-bp target site duplications, and carries a single, large ORF, which is predicted to encode a 482 amino acid polypeptide.

The predicted protein is rather basic, and includes at least one helix-loop-helix domain, two features that are common among transposases from characterized IS elements (Galas and Chandler 1989). IS elements from the IS*4*/IS*231* family, which have been isolated from

Source	LIR		RIR
IS*1676*	----------CCTCGCGCTTTCA	.....	TGAAAGATTGAGG----------
pAL233	caaaactggccttgtgttaaCCTCGCGCTTTCA	.....	TGAAAGATTGAGGttaaaaatggatcttgatcc
pXS26	aaaacgcacggctgagttagCCTCGCGCTTTCA	.....	TGAAAGATTGAGGttagcaaacggcgctctcgg
genomic1	acactcgaatgaagccctaaCCTCGCGCTTTCA	.....	TGAAAGATTGAGGctaacgcaacccgcagacat
genomic2	ctggttccgttattagttagCCTCGCGCTTTCA	.....	TGAAAGATTGAGGttaggggcccgccaaaagt
genomic3	gttacccgctggtagtttagCCTCGCGCTTTCA	.....	TGAAAGATTGAGGttagcgtcatcgacgaaagg
genomic4	ctgattgcccacctagttagCCTCGCGCTTTCA	.....	TGAAAGATTGAGGttaggggcccgccaaaagt
genomic5	cgaatgctgccgggctctagCCTCGCGCTTTCA	.....	TGAAAGATTGAGGctagacggagagaggaggtg
genomic6	tcatgaaagattgaggctagCCTCGCGCTTTCA	.....	
genomic7		.....	TGAAAGATTGAGGctagcgatccgactatttcc

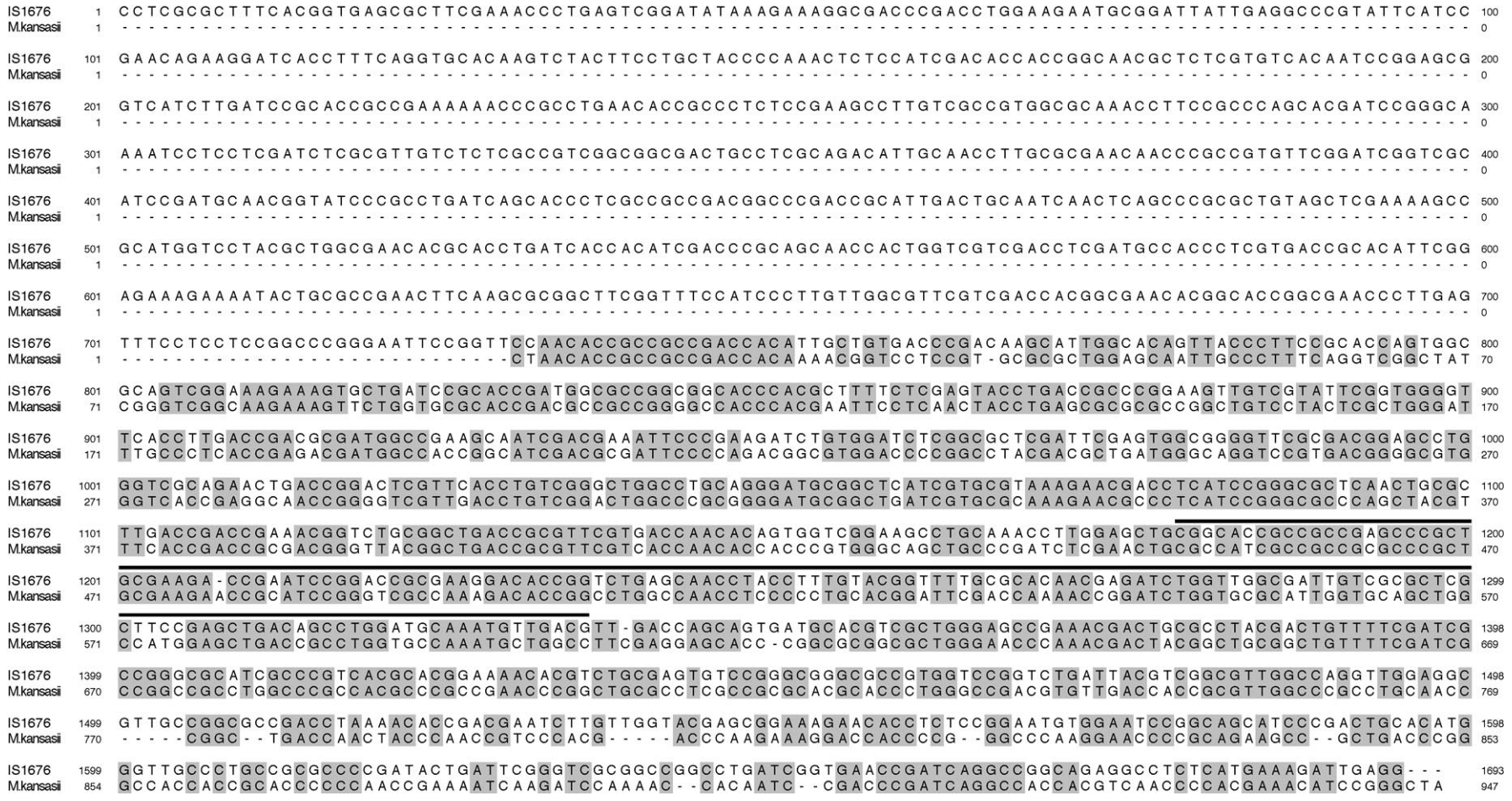


Fig. 5 Homology between IS*1676* and an insertion sequence-like element from *M. kansasii*. Identical base pairs are boxed in *gray*. Heavy overlining indicates the region encoding a putative helix-loop-helix domain. The sequence for the element from *M. kansasii* is derived from GenBank accession number L11041

both Gram-positive and Gram-negative hosts, are very similar to IS*1676* in size (Murphy 1989). These elements also characteristically possess a single, predominant ORF. However, the similarities between IS*1676* and members of the IS*4*/IS*231* family do not extend to the sequence level. Sequence comparisons at both the nucleotide and amino acid level revealed that IS*1676* has no significant homology to any other element found in *Rhodococcus*. Beyond the genus *Rhodococcus*, the only homology found to IS*1676* was that from a repetitive element from *M. kansasii* (Yang et al. 1993).

Yang et al. (1993) identified the IS-like element (which they labeled IS*1652*) while examining repetitive elements in a particular subspecies of *M. kansasii*. They found 1–10 copies of the element distributed among other isolates of this subspecies and sequenced four of these elements from a single isolate. IS*1652* contains ten short ORFs, the largest of which are on the order of 400 bp in length. The elements are not bounded by inverted repeats; however, they appear to be associated with 3 bp 5′-TAG-3′ repeats at their termini. Although they have not observed transposition directly, Yang et al. (1993) discovered evidence of a cointegrate form of the element, suggesting a replicative transposition event had occurred at some time during the evolution of the subspecies.

While the homology between IS*1676* and IS*1652* is striking, the differences between these two elements is perhaps more interesting. The numerous gaps in the homology between these elements account for several disruptions in the reading frame(s) of IS*1652* relative to the large ORF of IS*1676* (Fig. 5). Similarly, homology between these elements extends into the right inverted repeat of IS*1676*. Here again, the mismatches between the two elements (and the complete absence of the left inverted repeat) make it impossible to recognize the inverted repeat element in the *M. kansasii* sequence alone.

The above observations suggest that IS*1652* from *M. kansasii* may in fact be a degenerate form of an ancestral element homologous to IS*1676*. Even the 5′-TAG-3′direct repeats reported by Yang et al. (1993) resemble degenerate versions of the 5′-(C/T)TA(A/G)-3′ direct repeats induced by IS*1676* transposition (Fig. 4). Given that Yang et al. (1993) were unable to find homologues of IS*1652* in any mycobacterial species other than *M. kansasii*, it is more likely that both *M. kansasii* and *R. erythropolis* SQ1 obtained the ancestral version of this element through horizontal gene transfer, rather than from a common ancestor, despite the close phylogenetic relationship between these genera (Pascual et al. 1995). It is also possible, that a species of *Rhodococcus* (perhaps even *R. erythropolis*) was the donor that introduced the ancestral element into *M. kansasii*. Picardeau et al. (1997) also proposed horizontal transfer between *Rhodococcus* and *Mycobacterium* when they discovered an IS element (IS*1512*) in *M. gordonae* that shared homology with IS*1164* from *R. rhodochrous* J1 (Komeda et al. 1996) and the putative IS from *R. opacus* MR11 (Grzeszik et al. 1997).

Regardless of its ancestry, the fact that IS*1676* was isolated following its insertion into pPWS2 demonstrates that this element is transpositionally active. Prospects for converting this element into a useful mutagenic tool are encouraging. The element can apparently transpose into regions of either low or high G + C content. The latter is especially relevant to research on *Rhodococcus* sp. strains I24 and B264-1, both of which have an estimated G + C content above 50% (unpublished observations). The apparent requirement for a 5′-(C/T)TA(A/G)-3′ motif at the target site imposes no great limitation on this system considering that this degenerate motif should occur once per 64 bp in a genome of 50% G + C. Also, there appear to be no homologues of IS*1676* in either *Rhodococcus* sp. strain I24 or strain B264-1. Thus we can avoid the problem of homologous recombination with an endogenous copy of the element when an IS*1676* derivative is introduced into the strain. Also, this limits the problem of autorepression that may occur among multiple copies of an IS element. Future work on IS*1676* will focus on modifying the element to carry selectable markers and studying its transposition in other host species.

Acknowledgements This work was supported financially and intellectually by Merck Research Laboratories. We would also like to thank Priya Bhargava and Wilson Leong for technical assistance, Dr. C.I. Kado for the gift of pUCD4121, Dr. A.J. Pittard for providing pEP2, Dr. M. Schlömann and Dr. D. Berg for helpful discussions during the course of this work, and Dr. E. Lederberg for registering IS*1676* in the Plasmid Reference Center.

References

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

Altschul SF, Madden TL, Schaeffer 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

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

Buckland B, Drew S, Connors N, Chartrain M, Lee C, Salmon P, Gbewonyo K, Gailliot P, Singvi 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*-(1S,2R)-indandiol and *trans*-(1R,2R)-indandiol by *Rhodococcus* species. J Ferment Bioeng 86: 550–558

Denome SA, Young KD (1995) Identification and activity of two insertion sequence elements in *Rhodococcus* sp. strain IGTS8. Gene 161: 33–38

Denome SA, Oldfield C, Nash LJ, Young KD (1994) Characterization of the desulfurization genes from *Rhodococcus* sp. strain IGTS8. J Bacteriol 176: 6707–6716

Eulberg D, Kourbatova EM, Golovleva LA, Schloemann M (1998) Evolutionary relationship between chlorocatechol catabolic enzymes from *Rhodococcus opacus* 1CP and their counterparts in proteobacteria: sequence divergence and functional convergence. J Bacteriol 180: 1082–1094

Galas DJ, Chandler M (1989) Bacterial insertion sequences. In: Berg E, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington DC, pp 109–162

Gay P, Le Coq D, Steinmetz M, Ferrari E, Hoch JA (1983) Cloning structural gene *sac*B, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli*. J Bacteriol 153: 1424–1431

Gay P, Le Coq D, Steinmetz M, Berkelman T, Kado CI (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J Bacteriol 164: 918–921

Grzeszik C, Lubbers M, Reh M, Schlegel HG (1997) Genes encoding the NAD-reducing hydrogenase of *Rhodococcus opacus* MR11. Microbiology 143: 1271–1286

Jaeger W, Schaefer A, Kalinowski J, Puehler A (1995) Isolation of insertion elements from gram-positive *Brevibacterium*, *Corynebacterium* and *Rhodococcus* strains using the *Bacillus subtilis sac*B gene as a positive selection marker. FEMS Microbiol Lett 126: 1–6

Kamoun S, Tola E, Kamdar H, Kado CI (1992) Rapid generation of directed and unmarked deletions in *Xanthomonas*. Mol Microbiol 6: 809–816

Komeda H, Kobayashi M, Shimizu S (1996) Characterization of the gene cluster of high-molecular-mass nitrile hydratase (H-NHase) induced by its reaction product in *Rhodococcus rhodochrous* J1. Proc Natl Acad Sci USA 93: 4267–4272

Kulakov LA, Larkin MJ (1998) Sequencing and functional analysis of the haloalkane dehalogenase gene region. Direct GenBank Accession number AF060871

Kulakova AN, Larkin MJ, Kulakov LA (1997) The plasmid-located haloalkane dehalogenase gene from *Rhodococcus rhodochrous* NCIMB 13064. Microbiology 143: 109–115

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155: 335–350

Murphy E (1989) Transposable elements in Gram-positive bacteria. In: Berg E, Howe MM (eds) Mobile DNA. American Society for Microbiology, Washington DC, pp 269–288

Nagy I, Schoofs G, Vanderleyden J, De Mot R (1997) Transposition of the IS21-related element IS*1415* in *Rhodococcus erythropolis*. J Bacteriol 179: 4635–4638

Ochman H, Ajioka JW, Garza D, Hartl D (1989) Inverse polymerase chain reaction. In: Erlich HA (ed) PCR technology: principles and applications for DNA amplifications. Stockton Press, New York, pp 105–111

Ohtsubo E, Sekine Y (1996) Bacterial insertion sequences. In: Saedler H, Gierl A (eds) Transposable elements. Springer, Berlin Heidelberg New York, pp 1–26

Pascual C, Lawson PA, Farrow JA, Gimenez MN, Collins MD (1995) Phylogenetic analysis of the genus *Corynebacterium* based on 16S rRNA gene sequences. Int J Syst Bacteriol 45: 724–728

Picardeau M, Bull TJ, Vincent V (1997) Identification and characterization of IS-like elements in *Mycobacterium gordonae*. FEMS Microbiol Lett 154: 95–102

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

Reed KC, Mann DA (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res 13: 7207–7221

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 232: 584–599

Rost B, Sander C (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19: 55–72

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Seibert V, Kourbatova EM, Golovleva LA, Schlomann M (1998) Characterization of the maleylacetate reductase MacA of *Rhodococcus opacus* 1CP and evidence for the presence of an isofunctional enzyme. J Bacteriol 180: 3503–3508

Shao ZQ, Behki R (1995) Cloning of the genes for degradation of the herbicides EPTC (*S*-ethyl dipropylthiocarbamate) and atrazine from *Rhodococcus* sp. strain TE1. Appl Environ Microbiol 61: 2061–2065

Southern EM (1979) Measurement of DNA length by gel electrophoresis. Anal Biochem 100: 319–323

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

Vogt Singer ME, Finnerty WR (1988) Construction of an *Escherichia coli-Rhodococcus* shuttle vector and plasmid transformation in *Rhodococcus* spp. J Bacteriol 170: 638–645

Wilkins MR, Gasteiger E, Bairoch A, Sanchez J-C, Williams KL, Appel RD, Hochstrasser DF (1998) Protein identification and analysis tools in the ExPASy server. In: Link AJ (ed) 2-D Proteome analysis protocols. Humana Press, New Jersey

Yang M, Ross BC, Dwyer B (1993) Identification of an insertion sequence-like element in a subspecies of *Mycobacterium kansasii*. J Clin Microbiol 31: 2074–2079

Zhang Y, Praszkier J, Hodgson A, Pittard AJ (1994) Molecular analysis and characterization of a broad-host-range plasmid, pEP2. J Bacteriol 176: 5718–5728