

Transposition of the IS21-Related Element IS1415 in *Rhodococcus erythropolis*

ISTVÁN NAGY, GEERT SCHOOF, JOS VANDERLEYDEN, AND RENÉ DE MOT*

F. A. Janssens Laboratory of Genetics, Catholic University of Leuven, B-3001 Heverlee, Belgium

Received 10 March 1997/Accepted 13 May 1997

Three copies of the IS21-related transposable element IS1415 were identified in *Rhodococcus erythropolis* NI86/21. Adjacent to one of the IS1415 copies, a 47-bp sequence nearly identical to the conserved 5' end of integrons was found. Accurate transposition of IS1415 carrying a chloramphenicol resistance gene (Tn5561) was demonstrated following delivery from a suicide vector to *R. erythropolis* SQ1.

The genus *Rhodococcus* is characterized by a remarkable ability to degrade or convert a broad range of organic compounds (reviewed in reference 31). *Rhodococcus rhodochrous* has already found application in the industrial production of acrylamide (32). A steadily increasing number of reports reveal the potential of these nocardioform actinomycetes for environmental biotechnology. However, progress in this area is still hampered by a lack of appropriate genetic tools, such as (broad-host) vectors and transpositional mutagenic systems (10). In particular, the existence of transposable elements in *Rhodococcus* species is poorly documented. Two recent reports describe the identification of IS elements of the IS256 family by sequence analysis of DNA regions containing biodegradative genes (5, 18), but transposition has not yet been demonstrated. In this report, we describe the identification of the *Rhodococcus* transposable element IS1415, representing the first member of the IS21 family isolated from an actinomycete. We further show that a transposon constructed from IS1415 is capable of faithful transposition following delivery from a suicide vector to *Rhodococcus* cells.

IS1415: a member of the IS21 family. The *thcRBCD* gene cluster, required for biodegradation of thiocarbamate herbicides (21) and *s*-triazine herbicides (20), is located in the region downstream of the cobalamin (vitamin B₁₂) biosynthetic operon *cobLMK* of *Rhodococcus erythropolis* NI86/21 (4). Further sequence analysis of the original LambdaGEM-12 clone (λFAJ2028) (21) indicated that the N-terminal part of a putative transposase gene was present on this *SacI* fragment. With the 900-bp *SalI-SacI* fragment (Fig. 1A) as a probe, the overlapping clone λFAJ2031 was isolated from the λEMBL3 library of the *R. erythropolis* NI86/21 genome (30). A hybridizing 6.8-kb *Bam*HI subfragment was cloned in pUC18 (pFAJ2546) and provided about 3.7 kb of additional DNA sequence. Further DNA sequencing of both strands on overlapping fragments subcloned in pUC18 was carried out with an automated sequencer (ALF; Pharmacia Biotech). Potential coding regions (ORF1–ORF4', IstA, and IstB) were identified with the GCWIND program (28). No obvious open reading frames (ORFs) were detectable between ORF1 and IstA, consistent with the lower G+C content of this region (58.8%) compared to those of the other regions of the fragment (64.7%). Database searches revealed no known homologs for ORF1, ORF2, or ORF3. The hydropathy pattern of the partial ORF4 amino

acid sequence showed several potential transmembrane helices (data not shown), consistent with a low but extended sequence similarity (21.4% identity in a 257-amino-acid overlap) with the malate permease of *Schizosaccharomyces pombe* (12). Remarkably, ORF1 (110 amino acids) and ORF3 (115 amino acids) show a high degree of amino acid sequence homology (68.2% identity) and have a highly similar hydropathy profile, indicating two potential transmembrane helices (data not shown).

Two ORFs showed significant homology with similarly organized ORFs (IstA and IstB) present in insertion sequences belonging to the IS21 family (reviewed in reference 13). IS21-related elements have been identified in *Pseudomonas aeruginosa* (IS1326), *Pseudomonas fluorescens* (IS1162), *Burkholderia cepacia* (IS408), *Bacteroides fragilis*, *Yersinia pestis* (IS100), *Bacillus thuringiensis* (IS232), and *Bacillus stearothermophilus* (IS5376) (1, 13). The highest sequence similarities were found with the gene products of the prototype element IS21 from *P. aeruginosa*: 28.5% identity in a 382-amino-acid overlap for IstA and 31.8% identity in a 239-amino-acid overlap for IstB (alignment not shown). Our database search identified additional putative members of this family revealed by recently released genomic sequencing data for *Mycobacterium tuberculosis* (GenBank accession no. Z77165 and Z83858) and *Rhizobium* sp. strain NGR234 (11). The *R. erythropolis* element was assigned IS1415 by the Plasmid Reference Center (Stanford University School of Medicine, Stanford, Calif.).

Motifs characteristic of IstA and IstB proteins (13) were also present in the corresponding IS1415 sequences (Fig. 1B). IstA contained a putative N-terminal DNA binding region and the DDE triad which is found in the catalytic domains of many bacterial transposases and integrases of retroviruses and retrotransposons (24). Two possible translational starts separated by six codons were found in the IS1415 *istA* gene. In IS21, two functionally different proteins are encoded by the *istA* gene depending on which of two starts, separated by seven codons, is used (27). The shorter protein (cointegrator) was found to possess an increased cointegration activity compared to the full-length IstA transposase (13). The possible N-terminal extension of the *Rhodococcus* IstA (MKSSREI) resembles the equivalent region of IstA in IS21 (MLSREDFY). *Rhodococcus* IstB and its homologs in the IS21 family (13, 29) have two well-conserved regions separated by about 50 residues (Fig. 1B). A similar ATP-binding motif is found in the DnaA and DnaC proteins (19). Apparently, IstB functions as a helper protein for faithful IstA-catalyzed transposition of IS21 (13). Clustered genes for a DDE-type of transposase and a nucleoside triphosphate-binding protein have been identified in a

* Corresponding author. Mailing address: F. A. Janssens Laboratory of Genetics, Catholic University of Leuven, Willem de Croylaan 42, B-3001 Heverlee, Belgium. Phone: 32-16-32 23 98. Fax: 32-16-32 29 90. E-mail: rene.demot@agr.kuleuven.ac.be.


```

*****
N186/21 I 5'-agcagtcagtggtggcgggctcagcagtgAGTGGTGTTC...GACAAGTGGCGcgtatgacagatgatgacgtgtggtccg-3'
II 5'-tcttcagccaatcccagatcgcttgcgggGCTGGTGTTC...GACAGCTGGtagatcgaggaatattcttggcgagctat-3'
III 5'-tcacccgatcgcgtcgtcgggacggtGTGGTGTTC...GACAGTGGCGgaggaacatttcacgctcgtacagtgccg-3'
SQ1 IV 5'-tgaccgctcttgaagtcgacgggctctcgaAATCGTGTTC...GACAATCGgaatcgattgctccgaaggacccaatcgc-3'
V 5'-aattcctcgttcgagggctccgatcgcgaaCATCGTGTTC...GACATCGCaggggttcgcacctttctcgtgtaacag-3'
*****

```

FIG. 2. Sequences of the DNA flanking the three cloned *IS1415* copies in *R. erythropolis* N186/21 (I, pFAJ2546; II, pFAJ2557; and III, pFAJ2589) and insertion sites of Tn5561X1 (IV) and Tn5561X2 (V) in *R. erythropolis* SQ1 mutants FAJ2043 and FAJ2044, respectively. For the transposable elements (delineated by *), only the boundary sequences (italicized) are shown. The 5- to 6-bp duplicated sequences (direct repeats) are in bold.

duplications (13). The *IS1415* terminal sequences bear no homology with the ends of a partially sequenced putative insertion sequence of *Rhodococcus fascians* (16).

IS1415 is not related to other putative IS elements that have recently been identified by sequence analysis of biodegradation gene clusters in *Rhodococcus* species and that belong to the IS256 family (23). These elements, *IS1166* and *IS1295*, were identified in the region downstream of the sulfur oxidation genes *soxABC* in *Rhodococcus* sp. strain IGTS8 (5). A related element, *IS1164*, is located adjacent to the nitrile hydratase structural genes of *R. rhodochrous* J1 (18). Yet another putative member of this family is clustered with the hydrogenase genes of *Rhodococcus opacus* (GenBank accession no. U70364).

By DNA homology searches, a particular sequence motif was identified about 500 bp from the left border of the originally identified copy of *IS1415* (Fig. 1). The 47-bp sequence showed strong homology with the 5' end of several integrons (14, 25) and Tn5053 (17), which are typically found in gram-negative bacteria (Fig. 3). The first 25 nucleotides constitute the "Brown repeat," which was originally identified as an inverted repeat occurring at the boundaries of an 11.2-kb insertion (In2) in Tn21 (2), but clearly the homology extends beyond this sequence. The *Rhodococcus* 47-bp sequence contains an internal imperfect direct repeat (19/21 match) (Fig. 3). Such repeats also occur in the integron sequences (16/18 match) and in Tn5053 (17/22 match). However, none of the essential components of an integron, namely, an *int* gene (encoding a site-specific recombinase of the λ integrase family, called integrase) and the adjacent *attI* site (to be recognized by this integrase) at which incoming cassettes are integrated (15, 26), were found adjacent to the conserved 47-bp sequence. The assumption that this sequence may constitute one end of a transposon (like that in Tn5053, which is bracketed by Brown repeats) is not supported by sequence data, which have not (yet) revealed a putative remote end with an inverted Brown repeat. In total, we have determined about 22 kb in the 3' direction of this motif, covering a cluster of cobalamin biosynthetic genes (4), thiocarbamate degradation genes (21), and 20S proteasome structural genes (22, 30).

A transposon derived from *IS1415* transposes in *Rhodococcus*. A system for insertional mutagenesis in *R. fascians* based on illegitimate integration of nonreplicative vectors has been described (6), but a more convenient system based on transposition is not yet available. We therefore decided to construct

```

      ↓
Tn5053 5'-TGTCAGTTTCAGAAAGCCGACAGCAGCAGABATAATCAAAAGTCGACCCGACGGgtctt-3'
      5'-TGTCAGTTTCAGAAAGCCGACTGCACT-GTTCAGAAAGTCGACTGCACTaatca-3'
In's   5'-TGTCGTTTTCAGAAAGCCGCTGCACTGAACGTCAGAAAGCCGACTGCACTatagc-3'
Consensus  TGTCRBTTTCAGAAAGRCRCRCWCACK--RYDTCARAAGYSGACVGCACK

```

FIG. 3. Alignment of the 47-bp element (in capitals) located near *IS1415* (middle sequence) with the 5' end of integrons (In's) and with the left border of Tn5053. Nucleotides that make up an imperfect direct repeat within each of these sequences are underlined (the end of the first repeated sequence is marked by the arrow). Gaps introduced for optimal alignment are indicated by dashes. In the deduced consensus sequence, perfectly conserved nucleotides (in bold) that are part of the sequence repeat are underlined.

a transposon by inserting an antibiotic resistance marker into *IS1415*. To leave the *istA* promoter region unaffected and because of the apparent translational coupling of *istA* and *istB* (Fig. 1B), the region between *istB* and the right inverted repeat was modified to facilitate the insertion of suitable marker genes.

From plasmid pFAJ2547 (pUC18 with the 4.9-kb *KpnI* fragment of pFAJ2546 carrying *IS1415*), the region downstream of the *EcoRV* site in *istB* was removed (by digestion with *EcoRV* and *EcoRI*) and replaced by a PCR product (digested with the same enzymes) which introduced unique restriction sites immediately downstream of *istB* (*HpaI*, *StuI* and *BglII*). The forward PCR primer covered the *EcoRV* site of *istB* (5'-GGTCCTTGAGCGATATCATCC-3'; *EcoRV* site in italics), and the reverse primer covered the TAA stop codon of *istB* (5'-AGG AATTCAGATCTAGGCCTGTAAACGGGGTTAGCTGAAGG GCTTCACC-3'; the introduced restriction sites are italicized, and the complementary sequence of the stop codon is underlined). In the resulting plasmid (pFAJ2548), the region downstream of *istB* (providing the right inverted repeat) was reintroduced as a PCR product digested with *BglII* and *EcoRI* (forward primer, 5'-GTCAGATCTCAGAAAGCGGGTGGG GAAAAA-3'; reverse primer, 5'-AGGAATTCCCGGGTACC GTGCAGTGTGCATCCTCC-3' [introduced restriction sites are italicized]). This pUC18 derivative (pFAJ2549) contained *IS1415* with unique *BglII*, *StuI*, and *HpaI* sites positioned between the 3' end of the *istB* gene and the right inverted repeat region (Fig. 1B). The DNA region of pDA71 conferring chloramphenicol resistance (3) was finally inserted into the *StuI* site as a 1.8-kb *StuI-BbrPI* fragment in both orientations. This generated the transposon delivery constructs pFAJ2571, containing Tn5561X1, and pFAJ2572, containing Tn5561X2 (numbers assigned by the Plasmid Reference Center). Sequence analysis of the DNA fragment conferring chloramphenicol resistance (unpublished data) revealed a gene highly homologous to the *cmr* gene of *R. fascians* which encodes a transmembrane protein most likely mediating active efflux of the antibiotic (8). In Tn5561X1 the *cmr* homolog is in the same orientation as *istA* and *istB*.

To be useful as a transpositional mutagenesis system, the target cells should not contain the incoming transposable element. By Southern hybridization, *IS1415* was not detectable in other *R. erythropolis* strains (SQ1, DSM 743, DSM 1069 and DSM 43066) or in several other *Rhodococcus* species (*Rhodococcus coprophilus* N774, *R. fascians* D188, *Rhodococcus globerulus* R58, *R. rhodochrous* N54 and ATCC 12674, and *Rhodococcus ruber* N361). The same result was obtained with species of related genera, *Dietzia maris* N1015 and *Gordona rubropertincta* ATCC 25593.

To investigate whether Tn5561 transposition can occur in *Rhodococcus*, plasmid pFAJ2571 was introduced into *R. erythropolis* SQ1 by electroporation (7). Electroporated cells were plated on Luria-Bertani medium and incubated overnight at 30°C to allow phenotypic expression of chloramphenicol resistance. Then, chloramphenicol was added (250 μ l from a 4 mg/ml stock solution) beneath the agar to obtain a concentra-

tion of 40 $\mu\text{g/ml}$ after diffusion. Assuming a typical transformation efficiency by electroporation of about 10^4 per μg of DNA, chloramphenicol-resistant colonies were found at an estimated frequency of about 10^{-4} . Similar results were obtained with pFAJ2572. For ten randomly selected putative transposon mutants, colony hybridization with a pUC18 probe and colony PCR with primers specific for each transposon (previously described forward primer covering the *istB* *EcoRV* site and reverse primer inside the *cmr* homolog as indicated below) were used for verification. All putative mutants lacked the vector but had retained the *cmr* homolog, as expected for genuine transposon mutants. Southern hybridization indicated that independent and apparently random single transposition events had occurred (data not shown). Two mutants (FAJ2043, containing Tn5561X1, and FAJ2044, containing Tn5561X2) were further characterized by isolation of the disrupted DNA regions. From size-selected pUC18 libraries of these mutants, pFAJ2590 (with a 10-kb *PstI* fragment of mutant strain FAJ2043) and pFAJ2591 (with a 12-kb *PstI* fragment of mutant strain FAJ2044) were selected by colony hybridization. Partial sequence analysis of the *IS1415*-flanking DNA was performed with the *istA*-specific sequencing primer (see above) and outward sequencing primers corresponding to the region downstream of the *cmr*-like gene in Tn5561X1 (5'-CGTCGG CACAGATGCATTGGC-3') or to the promoter region of the latter gene in Tn5561X2 (5'-GATCGACCGTCCTGGGTAT TC-3'), respectively. These data showed unequivocally that independent transposition of *IS1415* had taken place, resulting in a 5-bp duplication in mutant FAJ2043 and a 6-bp duplication in mutant FAJ2044, consistent with the analysis of insertion sites in strain NI86/21 (Fig. 2). A comparison of these target sites revealed no obvious sequence preference, which is a prerequisite for the development of this system for random mutagenesis. The relatively high proportion of G and C probably merely reflects the overall base composition of *Rhodococcus* genes. Homology searches for the interrupted DNA regions revealed no known homologous genes.

Nucleotide sequence accession number. The GenBank accession no. of the DNA sequence is AF002247.

R. De Mot is a senior research associate with the Fund for Scientific Research (Flanders, Belgium) and acknowledges the financial support provided by this organization.

REFERENCES

- Brown, H. J., H. W. Stokes, and R. M. Hall. 1996. The integrons In0, In2, and In5 are defective transposon derivatives. *J. Bacteriol.* **178**:4429–4437.
- Brown, N. L., T. K. Misra, J. N. Winnie, S. Schmidt, M. Seiff, and S. Silver. 1986. The nucleotide sequence of the mercuric resistance operons of plasmid R100 and transposon Tn501: further evidence for *mer* genes which enhance the activity of the mercuric ion detoxification system. *Mol. Gen. Genet.* **202**:143–151.
- Dabbs, E., B. Gowan, S. Quan, and S. J. Andersen. 1995. Development of improved *Rhodococcus* plasmid vectors and their use in cloning genes of potential commercial and medical importance, p. 129–135. In *Proceedings of the 9th International Symposium on Biology of Actinomycetes*, Moscow, Russia.
- De Mot, R., I. Nagy, G. Schoofs, and J. Vanderleyden. 1994. Sequences of the cobalamin biosynthetic genes *cobK*, *cobL*, and *cobM* from *Rhodococcus* sp. NI86/21. *Gene* **143**:91–93.
- Denome, S. A., and K. D. Young. 1995. Identification and activity of two insertion sequence elements in *Rhodococcus* sp. strain IGTS8. *Gene* **161**:33–38.
- Desomer, J., M. Crespi, and M. Van Montagu. 1991. Illegitimate integration of non-replicative vectors in the genome of *Rhodococcus fascians* upon electrotransformation as an insertional mutagenesis system. *Mol. Microbiol.* **5**:2115–2124.
- Desomer, J., P. Dhaese, and M. Van Montagu. 1990. Transformation of *Rhodococcus fascians* by high-voltage electroporation and development of *R. fascians* cloning vectors. *Appl. Environ. Microbiol.* **56**:2818–2825.
- Desomer, J., D. Vereecke, M. Crespi, and M. Van Montagu. 1992. The plasmid-encoded chloramphenicol-resistance protein of *Rhodococcus fascians* is homologous to the transmembrane tetracycline efflux proteins. *Mol. Microbiol.* **6**:2377–2385.
- Dodd, I. B., and J. B. Egan. 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res.* **18**:5019–5026.
- Finnerty, W. R. 1992. The biology and genetics of the genus *Rhodococcus*. *Annu. Rev. Microbiol.* **46**:193–218.
- Freiberg, C., X. Perret, W. J. Broughton, and A. Rosenthal. 1996. Sequencing of the 500-kb GC-rich symbiotic replicon of *Rhizobium* NGR234 using dye terminators and a thermostable "sequenase": a beginning. *Genome Res.* **6**:590–600.
- Grobler, J., F. Bauer, R. E. Subden, and H. J. Van Vuuren. 1995. The *maeI* gene of *Schizosaccharomyces pombe* encodes a permease for malate and other C4 dicarboxylic acids. *Yeast* **11**:1485–1491.
- Haas, D., B. Berger, S. Schmid, T. Seitz, and C. Reimmann. 1996. Insertion sequence IS21: related insertion sequence elements, transpositional mechanisms, and application to linker insertion mutagenesis, p. 238–249. In T. Nakazawa, K. Furukawa, D. Haas, and S. Silver (ed.), *Molecular biology of pseudomonads*. ASM Press, Washington, D.C.
- Hall, R. M., H. J. Brown, D. E. Brookes, and H. W. Stokes. 1994. Integrons found in different locations have identical 5' ends but variable 3' ends. *J. Bacteriol.* **176**:6286–6294.
- Hall, R. M., and C. M. Collis. 1995. Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Mol. Microbiol.* **15**:593–600.
- Jäger, W., A. Schäfer, J. Kalinowski, and A. Pühler. 1995. Isolation of insertion elements from Gram-positive *Brevibacterium*, *Corynebacterium* and *Rhodococcus* strains using the *Bacillus subtilis* *sacB* gene as a positive selection marker. *FEMS Microbiol. Lett.* **126**:1–6.
- Kholodii, G. Y., S. Z. Mindlin, I. A. Bass, O. V. Yurieva, S. V. Minakhina, and V. G. Nikiforov. 1995. Four genes, two ends, and a *res* region are involved in transposition of Tn5053: a paradigm for a novel family of transposons carrying either a *mer* operon or an integron. *Mol. Microbiol.* **17**:1189–1200.
- Komeda, H., M. Kobayashi, and S. Shimizu. 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**:4268–4272.
- Koonin, E. V. 1992. DnaC protein contains a modified ATP-binding motif and belongs to a novel family of ATPases including also DnaA. *Nucleic Acids Res.* **20**:1997.
- Nagy, I., F. Compennolle, K. Ghys, J. Vanderleyden, and R. De Mot. 1995. A single cytochrome P-450 system is involved in degradation of the herbicides EPTC (*S*-ethyl dipropylthiocarbamate) and atrazine by *Rhodococcus* sp. strain NI86/21. *Appl. Environ. Microbiol.* **61**:2056–2060.
- Nagy, I., G. Schoofs, F. Compennolle, P. Proost, J. Vanderleyden, and R. De Mot. 1995. Degradation of the thiocarbamate herbicide EPTC (*S*-ethyl dipropylcarbamothioate) and biosurfacting by *Rhodococcus* sp. NI86/21 involve an inducible cytochrome P-450 system and aldehyde dehydrogenase. *J. Bacteriol.* **177**:676–687.
- Nagy, I., G. Schoofs, J. Vanderleyden, and R. De Mot. Further sequence analysis of the DNA regions with the *Rhodococcus* 20S proteasome structural genes reveals extensive homology with *Mycobacterium leprae*. *DNA Seq.*, in press.
- Picardeau, M., A. Varnerot, J. Rauzier, B. Gicquel, and V. Vincent. 1996. *Mycobacterium xenopi* IS1395, a novel insertion sequence expanding the IS256 family. *Microbiology* **142**:2453–2461.
- Polard, P., and M. Chandler. 1995. Bacterial transposases and retroviral integrases. *Mol. Microbiol.* **15**:13–23.
- Rådström, P., O. Sköld, G. Swedberg, J. Flensburg, P. H. Roy, and L. Sundström. 1994. Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. *J. Bacteriol.* **176**:3257–3268.
- Recchia, G. D., and R. M. Hall. 1995. Gene cassettes: a new class of mobile elements. *Microbiology* **141**:3015–3027.
- Reimmann, C., R. Moore, S. Little, A. Savioz, N. S. Willetts, and D. Haas. 1989. Genetic structure, function and regulation of the transposable element IS21. *Mol. Gen. Genet.* **215**:416–424.
- Shields, D. C., D. G. Higgins, and P. M. Sharp. 1992. GCWIND: a micro-computer program for identifying open reading frames according to codon positional G+C content. *Comput. Appl. Biosci.* **8**:521–523.
- Solinas, F., A. M. Marconi, M. Ruzzi, and E. Zennaro. 1995. Characterization and sequence of a novel insertion sequence, IS1162, from *Pseudomonas fluorescens*. *Gene* **155**:77–82.
- Tamura, T., I. Nagy, A. Lupas, F. Lottspeich, Z. Cejka, G. Schoofs, K. Tanaka, R. De Mot, and W. Baumeister. 1995. The first characterization of a eubacterial proteasome: the 20S complex of *Rhodococcus*. *Curr. Biol.* **5**:766–774.
- Warhurst, A. M., and C. A. Fewson. 1994. Biotransformations catalyzed by the genus *Rhodococcus*. *Crit. Rev. Biotechnol.* **14**:29–73.
- Yamada, H., and M. Kobayashi. 1996. Nitrile hydratase and its application to industrial production of acrylamide. *Biosci. Biotechnol. Biochem.* **60**:1391–1400.