Insertion sequence IS1137, a new IS3 family element from *Mycobacterium smegmatis*

Maria J. Garcia,¹ Christophe Guilhot,² Raju Lathigra,^{3†}
M. Carmen Menendez,¹ Pilar Domenech,¹ Carlos Moreno,³
Brigitte Gicquel² and Carlos Martin⁴

Author for correspondence: Maria J. Garcia. Tel: +34 1 3975440. Fax: +34 1 3150075.
e-mail: mjgarcia@mvax.fmed.uam.es

¹ Dept Medicina Preventiva,
Facultad de Medicina,
Universidad Autonoma,
Arzobispo Morcillo 4,
28029 Madrid, Spain

² Unit de Gntique
Mycobactrienne, Institut
Pasteur, 25 Rue Dr Roux,
75015 Paris, France

³ MRC Tuberculosis Unit,
Hammersmith Hospital,
DuCane Road, London
W12 0HS, UK

⁴ Dept Microbiologia,
Facultad de Medicina,
Universidad de Zaragoza,
Domingo Miral s/n, 50009
Zaragoza, Spain

A new insertion sequence (IS) has been isolated from *Mycobacterium smegmatis*. It is 1361 bp long and possesses characteristics of the IS3 family elements. It harbours 32 bp imperfect inverted repeats at its extremities and a 3 bp direct repeat flanks the element, possibly as the result of a transposition event. This IS, IS1137, contains three major ORFs. Two of them, ORF A and ORF B show homologies both at the amino acid sequence level and at the organization level with the ORFs encoding the transposase of the IS3 family elements. IS1137 has a narrow host range and was found only in *M. smegmatis* and *M. chitae*. The fact that IS1137 is not present in the *M. tuberculosis* complex strains makes this element a new candidate for transposon mutagenesis in mycobacteria.

Keywords: *Mycobacterium smegmatis*, insertion sequences, IS3 family

INTRODUCTION

All living organisms probably harbour mobile genetic elements. Insertion sequences (IS) are the smallest of these elements; they carry the transposition functions and components regulating this function. In addition, host factors participate in the regulation of transposition (Galas & Chandler, 1989).

The IS3 family is the most common group of ISs. About 25 % of all elements described so far belong to this family. The analysis of two elements of this family, IS150 and IS911, demonstrated that its transposases are expressed as a fusion of two ORFs by a –1 ribosomal frameshifting event (Polard *et al.*, 1991). A similar regulation was also demonstrated for IS1 and for the synthesis of retroviral integrases (Escoubas *et al.*, 1991; Fayet *et al.*, 1990).

IS3 family elements have been found both in Gram-negative and Gram-positive bacterial species from various ecological niches. Such an element, IS6110 (Thierry *et al.*, 1990) (also named IS986 (McAdam *et al.*, 1990) and IS987

(Hermans *et al.*, 1991b) has been isolated from the *Mycobacterium tuberculosis* complex. By sequence comparison, a site for frameshifting was predicted (Chandler & Fayet, 1993). Analysis of a large number of *M. tuberculosis* isolates has shown that this element transposes at a low frequency and this property has enabled the typing of *M. tuberculosis* strains by restriction fragment length polymorphism (RFLP) analysis (Van Soolingen *et al.*, 1991; Mazurek *et al.*, 1991; Otal *et al.*, 1991; Mendiola *et al.*, 1992).

The analysis of a *M. smegmatis* genomic library has allowed us to clone a new IS3 family element within the genus *Mycobacterium*. Here we describe its characteristics.

METHODS

Bacterial strains and plasmids. Tables 1 and 2 show plasmids, phages and bacterial strains used in this study.

DNA isolation and manipulation. Mycobacterial strains were grown in Lwenstein–Jensen medium (Difco) and DNA was isolated from cultures on Dubos medium containing Tween-albumin enrichment (Difco). DNA was prepared as described by Patel *et al.* (1986) with some modifications. Briefly, D-cycloserine incubations were for 48 h and reagent concentrations were: lysozyme, 4 mg ml^{–1}; proteinase K, 50 µg ml^{–1}; and SDS, 1 % (w/v). DNA was dialysed at 4 °C overnight before ethanol precipitation. DNA was quantified by ultraviolet absorbance and stored in ethanol at –20 °C. Standard tech-

† Present address: MedImmune Inc., 35 West Watkins Mill Road, Gaithersburg, MD 20878, USA.

Abbreviation: IS, insertion sequence.

The GenBank accession number for the sequence data reported in this paper is X70913.

Table 1. Plasmids and phages

Plasmid/ phage	Relevant characteristics	Source/ reference
Plasmid		
pUC18	Cloning vector	Yanisch-Perron <i>et al.</i> (1985)
pMSJ60	pUC18 construct containing a 3.2 kbp <i>KpnI</i> fragment including IS1137	This work
pMSJ61	pUC18 construct containing a 1.5 kbp <i>KpnI</i> - <i>SalI</i> fragment that includes the IS1137 end and upstream region (pMSJ60 derivative)	This work
pMSJ63	pUC18 construct containing a 700 bp <i>SalI</i> - <i>BamHI</i> fragment that includes the IS1137 end and downstream region (pMSJ60 derivative)	This work
pMSJ64	pUC18 construct containing a 580 bp <i>SalI</i> - <i>BamHI</i> fragment comprising IS1137 internal probe (pMSJ60 derivative)	This work
Phage		
M13mp18	Cloning vector	Yanisch-Perron <i>et al.</i> (1985)
M13mp19	Cloning vector	Yanisch Perron <i>et al.</i> (1985)

niques were used for plasmid isolation and restriction enzyme analysis (Sambrook *et al.*, 1989).

Southern blots and hybridizations. Two micrograms of genomic mycobacterial DNA was digested, the fragments were separated by electrophoresis on 0.75% agarose gels and transferred to a nylon membrane. Southern blots and hybridizations were performed as described elsewhere (Zhang *et al.*, 1992). Hybridization conditions were 30% formamide at 55 °C ($T_m - 18$ °C). After hybridization, filters were washed in $2 \times$ SSC (sodium saline citrate; $1 \times$ SSC is 0.15 M NaCl, 0.015 M tri-sodium citrate) + 0.1% SDS for 15 min at room temperature and then in $1 \times$ SSC + 0.1% SDS for 15 min at 55 °C. A 580 bp *BamHI*-*SalI* fragment from pMSJ64 (see Table 1) was used as an IS1137 internal probe for hybridizations. A 804 bp fragment from 16S rRNA gene of *M. fortuitum* ATCC 6841 was used as a mycobacterial probe as described elsewhere (Domenech *et al.*, 1994). The IS1137 and 16S rRNA probes were labelled with [α -³²P]dCTP (Amersham) using the Prime-a-Gene system (Promega) according to the manufacturers' instructions.

DNA sequencing. Double-stranded plasmid DNAs (pMSJ60, pMSJ61, pMSJ63 and pMSJ64, see Table 1) were sequenced using both universal primers (Pharmacia) and synthetic oligonucleotide primers (as indicated in Fig. 2). Oligonucleotide primers for sequencing were synthesized at the Unité de Génétique Micobactérienne, Institut Pasteur, Paris, France. The Sequenase kit (USB) was used according to the manufacturer's protocol.

The 1.5 kbp *KpnI*-*SalI* fragment from pMSJ61 (see Table 1) was cloned by standard shotgun methodology (Sambrook *et al.*,

Table 2. Bacterial strains

Bacterial strain*	Relevant characteristics	Source/ reference†
<i>Escherichia coli</i> TG1	Standard host strain used for cloning	Sambrook <i>et al.</i> (1989)
<i>M. abscessus</i> ATCC 19977	Type strain	A. Manrique
<i>M. aurum</i> L1		Hermans <i>et al.</i> (1991a)
<i>M. aurum</i> 11	Supplied as DNA	
<i>M. agri</i> CIPT 1320001	Type strain	CIPT
<i>M. bovis</i> BCG-Pasteur	Supplied as DNA	G. Torrea
<i>M. chelonae</i> ATCC 35752	Type strain	A. Manrique
<i>M. chitae</i> CIPT 1160001	Type strain	CIPT
<i>M. fortuitum</i> ATCC 6841	Type strain	A. Manrique
<i>M. fortuitum</i> ATCC 13756		A. Manrique
<i>M. fortuitum</i> FC1	Supplied as DNA	
<i>M. gadium</i> ATCC 27726	Type strain	Med. Prev. Collection
<i>M. gilvum</i> Stanford 132		Med. Prev. Collection
<i>M. kansasii</i> 890493	Supplied as DNA	G. Torrea
<i>M. peregrinum</i> ATCC 14467	Type strain	A. Manrique
<i>M. scrofulaceum</i> ATCC 19981	Supplied as DNA	G. Torrea
<i>M. smegmatis</i> strain		Med. Prev. Collection
<i>M. smegmatis</i> ATCC 607	Type strain	A. Manrique
<i>M. smegmatis</i> CIPT 1330001		CIPT
<i>M. smegmatis</i> CIPT 1330010	Type strain	CIPT
<i>M. smegmatis</i> mc ² -6	Supplied as DNA	
<i>M. smegmatis</i> mc ² -155	Supplied as DNA	
<i>M. tuberculosis</i> H37 Rv	Supplied as DNA	
<i>M. tuberculosis</i> H37 Ra	Supplied as DNA	
<i>M. ulcerans</i> HB 3436	Supplied as DNA	G. Torrea
<i>M. vaccae</i> SN-902		Med. Prev. Collection

* ATCC, American Type Culture Collection, Rockville, MD, USA; CIPT, Collection Institut Pasteur Tuberculose, Paris, France.

† G. Torrea, Unité de Génétique Micobactérienne, Institut Pasteur, Paris, France; A. Manrique, CNMVIS, Majadahonda, Madrid, Spain; Med. Prev. Collection, Dept Medicina Preventiva, Facultad de Medicina, Universidad Autonoma, Madrid, Spain.

1989) in M13mp18 and mp19 (Pharmacia). Universal primers and T7 DNA polymerase (Pharmacia) were used for sequence determination according to the manufacturer's protocol.

Sequence computer analysis. M13 sequences were analysed using the Staden-Plus program. Comparisons between IS1137 ORFs and other IS3 family elements were undertaken using a FASTA program in three different databases (GenBank, PIR and SwissProt). Amino acid sequences from closely related ISs were selected and multiple alignments were done using GCG package programs.

RESULTS AND DISCUSSION

Isolation of IS1137

A *M. smegmatis* genomic DNA library was constructed in pUC18. Initially our aim was to characterize the *M. smegmatis* genetic region homologous to the *M. tuberculosis* DNA fragment containing the *sod* gene encoding superoxide dismutase. By using Southern hybridization, we screened the *M. smegmatis* library with a 1.1 kbp fragment

Fig. 1. Restriction map of pMSJ60 showing the IS1137 location and orientation; flanking inverted repeats are shown as boxes. Horizontal arrows indicate fragments sequenced using synthetic oligonucleotide primers (see Methods). Fragments corresponding to pMSJ61, 63 and 64 (see Table 1) are highlighted. No sites were found for *EcoRI*.

containing the *M. tuberculosis* *sod* gene as a probe (Zhang *et al.*, 1991). Among the recombinant plasmids which hybridized with the 1.1 kbp fragment, one (pMSJ60; Table 1) contained a 3.2 kbp segment. Unexpectedly, the nucleotide sequence of this insert, which has been determined, did not reveal significant homology with the probe (the restriction map of the pMSJ60 insert and the sequencing strategy are shown in Fig. 1). A search in the EMBL database revealed homologies with ISs of the IS3 family.

A similar situation has already been encountered with the cloning of a repetitive sequence from *M. tuberculosis* using the pAL5000 plasmid as a probe (Zainuddin & Dale, 1989). Although this plasmid does not contain IS6110, cross-hybridization was observed with a clone containing this IS which is present in multiple copies in the *M. tuberculosis* genome. We interpreted these findings as resulting from the high G+C content of mycobacterial species which could be responsible for non-specific hybridization.

Genomic DNAs of *M. smegmatis*, *M. fortuitum*, *M. gadium* and *M. tuberculosis* were cleaved by restriction enzymes and subjected to Southern blot analysis using the 1.5 kbp fragment internal to the 3.2 kbp fragment of pMSJ60 (pMSJ61; Table 1). As shown in Fig. 2, several hybridizing bands were seen with DNA from *M. smegmatis*, but none were detected with DNAs from the other mycobacterial species included in this study. Therefore the repetitive sequence present in pMSJ61 seems to be specific to *M. smegmatis* and is present in multiple copies in the genome of this bacterium.

Characteristics of IS1137

The complete nucleotide sequence of the fragment exhibiting similarities with IS3 family elements is shown in Fig. 3. As expected, it has characteristics of IS3 family

Fig. 2. Ethidium bromide staining (a) and hybridization of mycobacterial DNAs using the pMSJ61 fragment as probe (b). Southern blots of digested mycobacterial DNAs probed with the 1.5 kbp *SalI*-*KpnI* fragment from pMSJ61 (see Table 1). Sizes are shown in kbp. L, Ladder DRigest II (Pharmacia); A, *M. fortuitum* ATCC 6841; B, *M. gadium* ATCC 27726; C, *M. smegmatis* laboratory strain; D, *M. tuberculosis* H37Rv; 1, *BamHI*; 2, *EcoRI*; 3, *KpnI*.

```

1  GACTGATCCG CCCCGGCGTG TCTGGAGAGT TCTTCCGTTG GGAAGGATGG GCACGTGGGA 60
61  TCGAAGTCAT CGAGCCGGTA TCCAGACGAG CTGAAGGTGC GGGCGGTGCA GATGGTGGCC 120
121  GATCTGCGCA CGAGACAGTT TCGGAGTGGG AGGCGATGGG CCGGGTTGCT GACCTGCTGG 180
181  GCGTCGGTAC TGCCGAAACG GTGCGCAAAT GGGTCCGCCA GGCTGAGATC GACGCCGGCT 240
241  CTCGGGCGGG GCAGACCAGC GAGGAATCCG AGGTCTGCGC CAAGCTGCGC CGGGAGAATG 300
301  CCGAACTCAA GCGGGCCAAC GCGATCTTGA AGGCGGCCTC GGTTTTCTTC GCCGCCGAGC 360
361  TCGACCGGCC CTCTCAGTAG TCGTGGAGTT CATCAGCGCC CACCAGCACA TGCGGGTGGG 420
421  CGTTGATGGT CTAAGTGGG GTGTCGAGTC GATGTGCGCT GTGCTTTCCG AGTACGGCGT 480
481  CACGATCGCC CCGTCGACGT ATTACGCCCA CCGCGCCCGT CAGGCCCTCG AAAGCCGACT 540
541  GGGCCGATGC GCAGGTGATC GATGCGATCT TGGCAGCTTC GCCAATCGCA GAGTCTTGTA 600
601  CCGAGTCCCTG GCGCTCGTA AGACATGGAT TGTATTGCGC ACCAATGGTA TCGACGTATC 660
661  GCGCTGTGTC GTGAACGGG TCATGCGGGA GATGGGTGGG CGGGGTGCCT GCAAGCGCCG 720
721  GCGGGTGC GC ACCACCGTC CCGATCCCGG CAGCGACGGC AGCCCCGGAT CCGTCCGGCG 780
781  TAATTTGCTC GCCGGTGC GC CTGACCGATT GTGGGTGGCC GATTTACCTT ACTGCGGGAC 840
841  TCGTGCCGGC TGGGCCTACA CGGCGTTCGT CACCGATGTC TATGCCCGCA AGATCGTGGG 900
901  TTGGAAGGTG GCTACCGAGA TGACCCAGAA GCTGGTGACC GATGCGATCG ACCACGCTAT 960
961  CGACACCAGA AAGCGTTCGG GTGCAGCATC TTTGGATTCA CTTATCCACC ACAGCGACGC 1020
ORF C ***
1021 GGGCTCGCAA TACACGCGGT CGCGTTCACC GAACGCTCG CCGCTGAGGG GATCCTGCCT 1080
1081 TCAGTACGGA TCCGTCGGGG ATAGTTTCGA TAATGCTCTG GCCGAATCGG TGAACAGCAG 1140
1141 CTACAAGACC GAACTCATCG ACCGCCAGCC GCTCTATCCC GGTGCCACCG AACTGGCGCT 1200
1201 GGGGACCGCC GAATGGGTGG CCTTCTACAA CCGTCAGCGA CCAACGGCT ACTGCCAGAC 1260
ORF B ***
1261 CTGACCCCCG ATCGGCCGAA GCGCTCTACT ACCATCGCCA ACGGCACCC TATGCCGAGG 1320
1321 AGGCACTCAG ATAAGAGAAC CTCCCGACAT GCCGGGGCGG ATCAGAGC 1367

```

Fig. 3. Nucleotide sequence of IS1137. Three base-pair direct repeats are in bold letters; imperfect terminal inverted repeats are underlined; ###, potential initiation codon; ***, termination codons; << and >>, ORF A limits. A potential ribosome binding site is overlined and labelled RBS; potential ribosomal frameshifting motifs and loop structures are indicated by +++. Explanations of ORF B and ORF C are given in the text.

elements (Fayet *et al.*, 1990). It is 1361 bp long with two 32 bp imperfect inverted repeats at its extremities. A 3 bp direct repeat is found flanking this element that might correspond to the duplication of the target sequence during the transposition event. We have named this element IS1137.

Analysis of the sequence has revealed a G+C content of 63.4% which is typical of mycobacteria. Three main ORFs (Fig. 3) with a codon usage characteristic of mycobacteria were found.

(1) ORF A of 75 amino acids possesses a putative start codon at nucleotide 156 and a sequence similar to a consensus Shine-Dalgarno (SD) motif at an expected position (6 nucleotides before the start codon).

(2) ORF B of 295 amino acids.

(3) A third ORF of 196 amino acids was found on the strand complementary to that encoding ORF A and ORF B.

Alignment of the amino acid sequence deduced from

Fig. 4. Amino acid sequence alignments of two main ORFs in several IS3 family elements. Consensus sequences indicate a block of more than three identical amino acids. Amino acid positions identical to those in IS1137 sequences are indicated by periods, and insertions that maximize alignments are denoted by dashes. (a) ORF A alignment. The highly conserved sequence is boxed. (b) ORF B alignment. The highly conserved region related to retroviral integrase is boxed; * indicates main conserved motifs in this region (Fayet *et al.*, 1990).

(a)

```

1                               60
IS11a  MTYSRNRYDQ DFKKNVHLS FNSS----- -KPVKII.SE ...PESALY. W----RKLYT
IS600a  MSRKTRYSK EFKAQAVRTV PENQ----- -LSISEG.SR .SLPEGLTGQ WVTAARKGLG
IS3411a MTKNTRFSP EVRQRAVRMV LESQSEYDSQ WATICSI.PK I.-C.R..L. V....H.R.T
IS629a  MTKNTRFSP EVRQRAVRMV LESQGEYDSQ WATICSI.PK I.-C.P..L. VR...H.R.T
IS51a   MKKLPKYSP EVRERAIRMV FEHLPEYESQ WATISAI.PK I.-C.P..L. L....S.RNS
IS6110a                               V. ....V..
IS1137a                               MGRVADL LGVGTAETVR KWRVQAEIDA

Consensus  ---K--R-S- E----AVR-V -E----- ---I--IA-- -G--T-ET-R -WVRQ-E-D-

61                               110
IS11a  EDGKQTPFAS LEAEN.A.K. ....ALERD M..K.AAY.. --S.QK*
IS600a  TPGSRT-VA. LESEILQ..K ALN.ARLERD ...K.TAY.. QES.KNTR*
IS3411a .GGDG.L.TA .RQR.KELE. ...R..R.S.D ..RQ..AY.. K..F..LWKK
IS629a  .GGDG.L.TA .RQR.KE.E. ...R..R.S.D ..RQ..AY.. K..F..LWKK
IS51a   .Q.D-.LSTV .R.RVKL.E. ..RQ.RQ..E ..RK..AY.. Q.....FKR
IS6110a .A.-P.T.T. .AEIKR... D....R.... ..T..A... -.....AR*
IS1137a  GSR-AGQTSE ESEVLRKLRR ENAELKRANA ILKAASVFFA -AELDRSQ*

Consensus  G-G--G-TT- E----K-L-R EN-ELRR-ND ILK-ASAYFA -AELDR-W--

```

(b)

```

1                               75
IS3b    MKY.FIEKHQAEPFSA..R..R- ----V.R.GW.TWCQ.RT---IST RQOF.QHCDVVLAFTRSKQRY..
IS904b                               MHR---.PSK QQVE.EI.SEKIKAVFHEHKGRY..
IS600b                               M.Q.FG- ----VSR.G..NWQHEP-----SD RKQSDER.KLEIKVAHIRTRETY.T
IS3411b MMPLL.K.REQY..GPL.SE.H- ----.....HCQQQRHHPDKRSA R.QRDDW.KKQIQ.VYDENHK.Y.V
IS629b  MMPLL.K.REQY..GPL.SE.H- ----.....HCQQQRHHPDKRSA R.QRDDW.KKQIQ.VYDENHK.Y.V
IS51b   MKTFI.ENRQSYG..PI.K..P- ----.....RYA.Q..DP.QQSA R.QRDQE..EHIQ.VWDEHFQ.Y.V
IS6110b FGH.E.P...R-.....I.TQ.T. L..P.....D-----HINREPS. RELRDGE.KEHIS.VHAAN.G.Y..
IS1137b QHMRVGVGDLK--WGVESMCAVLSE YGVTIAPSTYYAHRAQALSRIGR CAGDRCDLGSFANR--RVLYRVLGA

Consensus  --M----D--R---GV---C-VL-- ----IAPSTYY---Q-----S- RAQRD--LK--I-RV--E---VYGA

76                               150
IS3b    PRLTDE..AQ.YPFNVKT.AASL.R Q..L.AKAS.-KFSPVSYRAH.LP-V .ENLLEQD.Y.SG.NQK.AG.I..L
IS904b  VRITK..HNT..MTNTRK.GKL.HL ..LYAKGS.YKYKHYNRKAGSL-- R.NLINQI.K.T..NKV.LG.M..I
IS600b  .RLQTE.AE...I.G.DRLA.LRK. LRL.CKQ..-KFRA..NSNHNLP-V A.NLLNQ.T.APT..NQV....L..V
IS3411b .V.RQ.LRE..R.A..T.A.L.AV ..LA.-VL.GKK...ISRKAVA-- AGHR.N.Q...ER..Q.....V
IS629b  .V.RQ.LRE..R.A..T.A.L.AV ..LA.-VL.GKK...ISRKAVA-- AGDR.N.Q...ER..Q.....V
IS51b   .V.RQ.LRD.LQ.A..T.Q.L.G. L.LQAWYAASPSR.RSATK..RAR- RIMT.DSLWLSARTPYGYRTLP.SV
IS6110b .V.LT.NRE..E.A..T..L.TK L.LS.-TT.GKAR...I...ATAR- PADL.Q.R.GPP..N.....L..V
IS1137b RKTWIVLRTNGIDVSRVVERVMRE MGWRGACKR-RRVRTTVADPGSDG- SPGSVRRNFVAGAPDRLWVADFTYC

Consensus  RKVW--L-R-GI-VARCTVARLM-- MGL-G---R-K--RTT-----A-- ---LVN--F-A-APNQLWVAD-TYV

151                               225
IS3b    R.PE..L.L.V.I.LWS.AVI..SM SPR..AQ.AC..LQM.LWR...PR- ----NV.V.T.R.G..CSADYQAQL
IS904b  P.KE.TL.L.VNI..FS.....SM SSR.QD...R.CFLQ.CGKEHPQP- ----G..V.T.Q....S..YQSTL
IS600b  A.QE..L.L.GIK...TCE..RYAM GER..KE.TGK.LFM.LRSQRPPA- ----G.....R....CAYDYRVIQ
IS3411b S.WR.FV.V.FII..F.GY....R. SSS.ETTF.L..LEQ.LW..RP.. ----TV....K....VSLAYTQRL
IS629b  S.WQ.FV.V..II..F.GY....R. SSS.ETTF.L..LEQ.LWA.RP.. ----TV....K....VSLAYTQRL
IS51b   PG.VLFTWPSSLSLPLDSSS..RCP VQRVPILYSMH.LEQALYA.RP--- VKQGG.....R.V..VSI.YTERL
IS6110b S.W..F..V.....A...R.L..R. .ST.ATSM.L...EQ..W..QQEGV LD.KDV...T.R....SI.FSERL
IS1137b GTRAGWAYTAFVTDVYARKIVGWKV ATEMQKLVTDALDHAIDTRKRSGA ASLDSLHSDAGSQYTRSRFPNVW

Consensus  -T--GF-Y-AF-IDV-AR-IVGW-V SSRM---LV-DALEQALW-RRP-G- ----LIHSDRGSOY-S-RY--RL
* *

226                               300
IS3b    KRHNLRGMSAK.CCY...CV..FF HSL.V.C.HGEH-FISREIMRATVF NYIECD...W.RHSWCGGLSPEQFE
IS904b  RQV.AQSSMRK.NPY...MM..FY KTL.R...NDAH-FETRA.ATQEIF KYIETY..TK.MHSGLDYKSPKDF-
IS600b  EQS.LKTSMSRK.NCY...PM..FW GTL.N.SLSHYR-FNNRD.AISVIR .YIEI.....RHSRLGNISPAAPR
IS3411b KEA.LLAST..T...Y...M...I. GL..A.V.H.-KSWKNRA.VE.A.L T..D-W..NR.LLERLGHTPPAEAE
IS629b  KEA.LLAST..T...Y...M...I. GL..AKV.H.-KSWKNRA.VE.A.L T..D-W..NR.LLERLGHTPPAEAE
IS51b   VEA.IEPSV.....Y.....TI. GL..A.VIH.-RSW.TRGAVE.E.L K..D-WF.HR.LLEPIGHIIPPVEAE
IS6110b AEA.IQPSV.A..S.Y.....TI. GL.....KPGKPWRSIEDVE.A.. R..D-WF.HR.LYQ..GDVPPVELE
IS1137b PLRGSLCQYGSVGSFDNALAESVN SSKYTELIDRQPLYPGATELALGTA EWVA-FYNRQRPNGYCQT*.....

Consensus  -EAGL--S-GS-GDSYDNA AESIN GLYK-E-IHR--W--R-EVEL-T- -WVD-WYN-RRL--LG--PP-E-E
*

```

Fig. 4(a, b). For legend see facing page.

Fig. 5. Distribution of IS1137 among various mycobacteria. Sizes are shown in kbp. (a) Southern blot of *Pst*I-digested *M. smegmatis* DNA probed with the pMSJ64 insert (see Table 1): 1, ATCC 607; 2, laboratory strain; 3, mc²-6; 4, mc²-155. (b) Southern blots of *Pst*I-digested mycobacterial DNA probed with the pMSJ64 insert and 16S rRNA gene amplified fragment (see Methods): 1, *M. peregrinum* ATCC 14467; 2, *M. fortuitum* ATCC 6841; 3, *M. chelonae* ATCC 35752; 4, *M. abscessus* ATCC 19917; 5, *M. agri* CIPT 1320001; 6, *M. chitae* CIPT 1160001; 7, *M. smegmatis* CIPT 1330001; 8, *M. smegmatis* CIPT1330010; 9, *M. aurum* L1; 10, *M. gadium* ATCC 27726; 11, *M. gilvum* Stanford 132; 12, *M. vaccae* SN-902.

IS1137 ORF A and ORF B with apparently homologous ORFs from other IS3 family elements shows significant similarities (Fig. 4a and b). Even though both ORF A and ORF B from IS1137 show greater homology with the corresponding ORFs of the IS3 family element from *M. tuberculosis* (IS6110), significant homologies are also observed with the following elements: IS3 (Timmerman & Tu, 1985); ISL1 (Shimizu-Kadota *et al.*, 1985); IS51 (Yamada *et al.*, 1986); IS600 (Matsutani *et al.*, 1987); IS3411 (Ishiguro & Sato, 1988); IS629 (Matsutani & Ohtsubo, 1990) and IS904 (Rauch *et al.*, 1990).

Members of the IS3 family always exhibit two ORFs on the same DNA strand. For two of these members, IS911 and IS150, it has been shown that ORF B might be translated independently or with ORF A by a -1 translational frameshift to give a fusion polypeptide. This polypeptide seems to be the transposase, while the product of the first ORF (ORF A) could modulate the nature of the final product of transposition (intramolecular circle formation in the absence of the product of ORF A and intermolecular transposition in the presence of that product) (Polard *et al.*, 1991). It is also hypothesized that the product of the first ORF (ORF A) has a DNA binding capacity and could repress transcription by binding to the inverted repeats (Prre *et al.*, 1990).

In IS1137 a signal for frameshifting could be the TTTT sequence located at nucleotide 343 (Fig. 3). Inverted repeats (overlined in Fig. 3) can lead to a short stem and loop structure downstream from the frameshifting signal and, similar to other frameshifting events already described, may play a role in translation regulation of ORF A/ORF B (Chandler & Fayet, 1993). As was described for IS911, the frameshifting signal might allow a slippage of the ribosome along two codons possessing identical first and second nucleotides in 0 and in -1 frames (Polard *et*

al., 1991). The first potential initiation codon for ORF B is at nucleotide 383 (GTG) but it has no adjacent SD sequence. Since Polard *et al.* (1992), studying IS911, were not able to detect any role for ORF B, we hypothesize that ORF B might not be translated independently in IS1137 but, more probably, as a fused polypeptide with ORF A after a -1 translational frameshift occurring at nucleotide 343. A -1 frameshift was also described for the synthesis of the retroviral integrases, which also share amino acid sequence similarities with the transposases of IS3 family elements. The similarities between the IS1137 ORF B sequence, the putative transposases of different IS3 family elements and the retroviral integrases are indicated in Fig. 4(b).

As in many IS sequences, an ORF is found on the strand complementary to that encoding the transposase, ORF C. The first initiation codon for this ORF is at nucleotide 955 (GTG) but, again, it has no adjacent SD sequence.

Although the role of these ORF Cs is not yet evident, they could play a role in the regulation of transposition.

IS1137 distribution in mycobacterial species

Genomic Southern blot analysis using the 1.5 kbp fragment (pMSJ61; Table 1) containing the 5' end of the IS1137 sequence as a probe, has shown that this IS was present in multiple copies in the *M. smegmatis* genome but absent in three other mycobacterial species (Fig. 2).

Using the 580 bp fragment of pMSJ64 (see Table 1) which corresponds to an internal probe of IS1137, an analysis of a greater number of mycobacterial species and strains was performed (Table 2). Results of this study are presented in Fig. 5. All *M. smegmatis* strains tested show positive hybridization signals (Fig. 5a and b). Two strains, mc²6

and its high-efficiency transformation mutant mc²155, showed the same hybridization pattern (Fig. 5a); a similar finding was described by Cirillo *et al.* (1991) for IS1096, another IS isolated from *M. smegmatis*.

Except for a faint single signal in *M. chitae*, no hybridization was detected with all other mycobacterial species tested including three members of the *M. tuberculosis* complex (ten fast growers and five slow growers; Fig. 5b and data not shown). As a control, a segment of the 16S rRNA gene, which is a conserved sequence (see Methods), was used as a probe for rehybridization of the same filters. Signals of similar intensity were detected for all genomic DNAs tested (Fig. 5b and data not shown).

These results indicate that, like many other IS sequences, IS1137 has a narrow host range.

Besides IS6110 described in the *M. tuberculosis* complex, other ISs have been found in the genus *Mycobacterium*: IS900 in *M. paratuberculosis* (Collins *et al.*, 1989); IS6100 in *M. fortuitum* (Martin *et al.*, 1990); IS1081 in *M. bovis* (Collins & Stephens, 1991); IS901 which is only found in several *M. avium* serovars (Kunze *et al.*, 1991) and its isoform IS902 isolated from *M. avium* subsp. *silvaticum* (Moss *et al.*, 1992).

Two insertion elements were described in *M. smegmatis*: IS1096 (Cirillo *et al.*, 1991) and IS6120 (Guilhot *et al.*, 1991). Analysis of IS6120 has also shown a narrow host range, it is absent in *M. bovis* BCG, *M. tuberculosis* and *M. fortuitum*; however, hybridization was detected with *M. aurum* DNA (our unpublished results), a fast-growing *Mycobacterium*. This situation is similar to that observed here with IS1137 which is only present in *M. smegmatis* and *M. chitae*, although the hybridization signal with the latter species is not very strong. Therefore, an alternative hypothesis could be that *M. chitae* carries another IS3 family IS that exhibits cross-hybridization with IS1137.

Analysis of the occurrence and distribution of repeat elements in mycobacteria provides information that could be useful for taxonomic definition of species, and when present at different locations, could allow typing of the strains within a species. This is extremely useful for tracing the transmission of mycobacterial strains in various ecological niches.

The characteristics of IS1137 make this element a potential tool for gene manipulation in mycobacteria, i.e. it is absent in a wide range of different mycobacterial species, particularly in species belonging to the *M. tuberculosis* complex. Therefore, it could be used for further studies of transposition in these mycobacteria of clinical interest.

ACKNOWLEDGEMENTS

We would like to thank Dr Y. Zhang who kindly provided the superoxide dismutase gene from *M. tuberculosis*, G. Torrea who kindly provided DNAs from several slow-growing mycobacteria, Dr V. Vincent Lévy-Frébault and A. Manrique who kindly provided CIPT and ATCC mycobacterial strains, respectively. We would like to thank J. Gavigan for the review of the manuscript. This work was supported by grants from the

Comision Interministerial de Ciencia y Tecnologia (SAL91-1034 and BIO 742-93; Spain) and from the EC (BIO2-CT92-0520).

REFERENCES

- Chandler, M. & Fayet, O. (1993). Translation frameshifting in the control of transposition in bacteria. *Mol Microbiol* **7**, 497–503.
- Cirillo, J. D., Barletta, R. G., Bloom, B. R. & Jacobs, W. R. (1991). A novel transposon trap for Mycobacteria: isolation and characterization of IS1096. *J Bacteriol* **173**, 7772–7780.
- Collins, D. M. & Stephens, D. M. (1991). Identification of an insertion sequence, IS1081, in *Mycobacterium bovis*. *FEMS Microbiol Lett* **83**, 11–16.
- Collins, D. M., Gabric, D. M. & De Lisle, G. W. (1989). Identification of a repetitive DNA specific to *Mycobacterium paratuberculosis*. *FEMS Microbiol Lett* **60**, 175–178.
- Domenech, P., Menendez, C. & Garcia, M. J. (1994). Restriction fragment length polymorphisms of 16S rRNA genes in the differentiation of fast-growing mycobacterial species. *FEMS Microbiol Lett* **116**, 19–24.
- Escoubas, J. M., Prere, M. F., Fayet, O., Savignol, I., Galas, D., Zerbido, D. & Chandler, M. (1991). Translational control of transposition activity of the bacterial insertion sequence IS1. *EMBO J* **10**, 705–712.
- Fayet, O., Ramond, P., Prere, M. F. & Chandler, M. (1990). Functional similarities between retroviruses and IS3 family of bacterial insertion sequences? *Mol Microbiol* **4**, 1771–1777.
- Galas, D. J. & Chandler, M. (1989). Bacterial insertion sequences. In *Mobile DNA*, pp. 109–162. Edited by D. E. Berg & M. M. Howe. Washington DC: American Society for Microbiology.
- Guilhot, C., Gicquel, B., Davies, J. & Martin, C. (1991). Isolation and analysis of IS6120, a new insertion sequence from *Mycobacterium smegmatis*. *Mol Microbiol* **6**, 107–113.
- Hermans, J., Martin, C., Huijberts, G. N. M., Goosen, T. & De Bont, J. A. M. (1991a). Transformation of *Mycobacterium aurum* and *Mycobacterium smegmatis* with the broad host-range Gram-negative cosmid vector pJRD215. *Mol Microbiol* **5**, 1561–1566.
- Hermans, P. W. M., Van Soolingen, D., Bik, E. M., De Haas, P. E. W., Dale, J. W. & Van Embden, J. D. A. (1991b). Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect Immun* **59**, 2695–2705.
- Ishiguro, N. & Sato, G. (1988). Nucleotide sequence of insertion sequence IS3411, which flanks the citrate utilization determinant of transposon Tn3411. *J Bacteriol* **170**, 1902–1906.
- Kunze, Z. M., Wall, S., Appelberg, R., Silva, M. T., Portaels, F. & McFadden, J. J. (1991). IS901, a new member of a widespread class of atypical insertion sequences, is associated with pathogenicity in *Mycobacterium avium*. *Mol Microbiol* **5**, 2265–2272.
- McAdam, R. A., Hermans, P. W. M., Van Soolingen, D., Zainudin, Z. F., Catty, D., Van Embden, J. D. A. & Dale, J. W. (1990). Characterization of a *Mycobacterium tuberculosis* insertion sequence belonging to the IS3 family. *Mol Microbiol* **4**, 1607–1613.
- Martin, C., Timm, J., Rauzier, J., Gomez-Lus, R., Davies, J., Gicquel, B. (1990). Transposition of an antibiotic resistance element in mycobacteria. *Nature* **345**, 739–743.
- Matsutani, M. & Ohtsubo, E. (1990). Complete sequence of IS629. *Nucleic Acids Res* **18**, 1899.
- Matsutani, S., Ohtsubo, H., Maeda, Y. & Ohtsubo, E. (1987). Isolation and characterization of IS elements repeated in the bacterial chromosome. *J Mol Biol* **196**, 445–455.

- Mazurek, G. H., Cave, M. D., Eisenach, K. D., Wallace, R. J., Bates, J. H. & Crawford, J. T. (1991). Chromosomal DNA fingerprint patterns produced with IS6110 as strain-specific markers for epidemiologic study of tuberculosis. *J Clin Microbiol* **29**, 2030–2033.
- Mendiola, M. V., Martin, C., Ota, I. & Gicquel, B. (1992). Analysis of the regions responsible for IS6110 RFLP in a single *Mycobacterium tuberculosis* strain. *Res Microbiol* **143**, 767–772.
- Moss, M. T., Malik, Z. P., Tizard, M. L. V., Green, E. P., Sanderson, J. D. & Hermon-Taylor, J. (1992). IS902 an insertion element of the chronic-enteritis-causing *Mycobacterium avium* subsp. *silvaticum*. *J Gen Microbiol* **38**, 139–145.
- Ota, I., Martin, C., Vincent Lévy-Frébault, V., Thierry, D. & Gicquel, B. (1991). Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. *J Clin Microbiol* **29**, 1252–1254.
- Patel, R., Vach, J. T. & Mounts, P. (1986). Isolation and restriction endonuclease analysis of mycobacterial DNA. *J Gen Microbiol* **132**, 541–551.
- Polard, P., Prere, M. F., Chandler, M. & Fayet, O. (1991). Programmed translational frameshifting and initiation codon in gene expression of bacterial insertion sequence IS911. *J. Mol Biol* **222**, 465–477.
- Polard, P., Prere, M. F., Fayet, O. & Chandler, M. (1992). Transposase-induced excision and circularization of bacterial insertion sequence IS911. *EMBO J* **11**, 5079–5090.
- Prere, M. F., Chandler, M. & Fayet, O. (1990). Transposition in *Shigella dysenteriae*: isolation and analysis of IS911, a new member of the IS3 group of insertion sequences. *J Bacteriol* **172**, 4090–4099.
- Rauch, P. J., Beerthuyzen, M. M. & De Vos, W. M. (1990). Nucleotide sequence of IS904 from *Lactococcus lactis* subsp. *lactis* strain NIZOR5. *Nucleic Acids Res* **18**, 4253–4254.
- Sambrook J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Shimizu-Kadota, M., Kiwaki, M., Hirokawa, H. & Tsuchida, N. (1985). ISL1: a new transposable element in *Lactobacillus casei*. *Mol Gen Genet* **200**, 193–198.
- Thierry, D., Cave, M. D., Eisenach, K. D., Crawford, J. T., Bates, J. H., Gicquel, B. & Guesdon, J. L. (1990). IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Res* **18**, 188.
- Timmerman, K. P. & Tu, C.-P. D. (1985). Complete sequence of IS3. *Nucleic Acids Res* **13**, 2127–2139.
- Van Soolingen, D., Hermans, P. W. M., De Haas, P. E. W., Soll, D. R. & Van Embden, J. D. A. (1991). Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of Tuberculosis. *J Clin Microbiol* **29**, 2578–2586.
- Yamada, T., Lee, P.-D. & Kosuge, T. (1986). Insertion sequence elements of *Pseudomonas savastanoi*: nucleotide sequence and homology with *Agrobacterium tumefaciens* transfer DNA. *Proc Natl Acad Sci USA* **83**, 8263–8267.
- Yanisch-Perron, C. J., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.
- Zainuddin, Z. F. & Dale, J. W. (1989). Polymorphic repetitive DNA sequences in *Mycobacterium tuberculosis* detected with a gene probe from a *Mycobacterium fortuitum* plasmid. *J Gen Microbiol* **135**, 2347–2355.
- Zhang, Y., Lathigra, R., Garbe, T., Catty, D. & Young, D. (1991). Analysis of superoxide dismutase, the 23 kilodalton antigen of *Mycobacterium tuberculosis*. *Mol Microbiol* **5**, 381–391.
- Zhang, Y., Garcia, M. J., Lathigra, R., Allen, B., Moreno, C., Van Embden, J. D. A. & Young, D. (1992). Alterations in the superoxide dismutase gene of an isoniazid-resistant strain of *Mycobacterium tuberculosis*. *Infect Immun* **60**, 2160–2165.

Received 22 March 1994; revised 31 May 1994; accepted 20 June 1994.