Molecular Microbiology (1991) 5(9), 2265–2272

IS*901*, a new member of a widespread class of atypical insertion sequences, is associated with pathogenicity in *Mycobacterium avium*

Z. M. Kunze,[1] S. Wall,[1] R. Appelberg,[2] M. T. Silva,[2] F. Portaels[3] and J. J. McFadden[1]*

[1]*Molecular Microbiology Group, School of Biological Sciences, University of Surrey, Guildford, Surrey GU2 5XH, UK.*

[2]*Centro de Cytologia Experimental, Universidade Do Porto, Porto, Portugal.*

[3]*Institute of Tropical Medicine, Antwerpen, Belgium.*

Summary

An insertion sequence (IS*901*), found in pathogenic strains of *Mycobacterium avium*, but absent in *M. avium* complex isolates from patients with acquired immune deficiency syndrome (AIDS), has been isolated and sequenced. This insertion element has a nucleotide sequence of 1472 bp, with one open reading frame (ORF1), which codes for a protein of 401 amino acids. The amino acid sequence, terminal ends and target site of IS*901* are similar to those of IS*900*, present in *Mycobacterium paratuberculosis*. However, the DNA sequences of these two IS elements exhibit only 60% homology, compared to a DNA homology of 98% between their respective hosts. IS*901*, like IS*900*, appears to belong to a family of related insertion elements present in actinomycetes and other bacteria. *M. avium* strains containing IS*901* were found to be more virulent in mice than closely related strains lacking IS*901*. IS*901* may be a useful tool for the study of the genetics of virulence in the *M. avium* complex and for obtaining stable integration of foreign genes into mycobacteria.

Introduction

The *Mycobacterium avium* complex contains a number of strains that are divided by restriction fragment-length polymorphism (RFLP) analysis (McFadden *et al.*, 1987b) into at least two groups of related strains: *M. avium* strains comprising serotypes 1–6, 8–10 and *Mycobacterium intracellulare* strains (serotypes 7 and 11–21). *M. avium* causes disease (mainly tuberculosis) in birds and animals and atypical tuberculosis in humans; it is also found in the environment and commonly causes opportunistic infections in immunocompromised patients, particularly patients with acquired immune deficiency syndrome (AIDS) (Snider *et al.*, 1987). Two specific pathogens are also very closely related to *M. avium*, sharing greater than 98% DNA homology (McFadden *et al.*, 1987a): *Mycobacterium paratuberculosis*, the causative agent of Johne's disease in ruminants (and implicated in Crohn's disease of humans (McFadden *et al.*, 1987a)) and *Mycobacterium lepraemurium*, the causative agent of rodent leprosy. Both these species differ from *M. avium* in having complex growth-factor requirements (the iron chelator mycobactin for *M. paratuberculosis*) and being very slow growing. However some strains of *M. avium* isolated from animals and birds may also require mycobactin for growth, particularly on initial isolation; in experimental infections these may also cause Johne's disease (Collins *et al.*, 1983). Previous studies have not detected any genetic differences between *M. avium* infecting animals and *M. avium* causing opportunistic infections, particularly in AIDS patients. However, *M. paratuberculosis* has been shown to differ from *M. avium* in that it possesses multiple copies of the insertion sequence IS*900* in its genome. This 1.45 kb insertion sequence was isolated as a repetitive DNA element from *M. paratuberculosis* (McFadden *et al.*, 1987a), and DNA sequence analysis demonstrated it to be an atypical bacterial insertion sequence that lacks terminal inverted repeats and does not generate a target-site duplication on insertion (Green *et al.*, 1989). IS*900* contains a single open reading frame (ORF) coding for a 399-amino-acid protein that shows homology to proteins encoded by various insertion sequences (ISs) isolated from *Streptomyces* (Bruton and Chater, 1987; Henderson *et al.*, 1989; 1990; Leskiw *et al.*, 1990). A recently discovered insertion sequence, IS*1000*, from *Thermus thermophilus* (Ashby and Bergquist, 1990), which also lacks terminal inverted repeats and does not produce a target-site duplication, was shown to be related to both IS*110* and IS*492*, a similar element from *Pseudomonas atlantica* (Bartlett and Silverman, 1989). Transposition of IS*900* in mycobacteria has been demonstrated recently (England *et al.*, 1991) and has been proposed as a method for obtaining stable integration of foreign genes

Received 27 March, 1991; revised 4 June, 1991. *For correspondence. Tel. (0483) 571281, ext. 2671; Fax (0483) 300374.

```
TTCCTTAGGGTTGAAGGGGTCTGGGATTGGATGTCCTGGATGCCCAGGACGGTGGGGTGTGGCTAATGTGTTGTTGCAACGGTTGTTGCT
         10        20        30        40        50        60        70        80        90

TGAAAGGAATGGCCGCCCTGCCGTTTTGGACGGTCCTGGCCACTGATTGAGATCTGACGCGTACTCGATGACGCTGCTCTAAGGACCTGT
        100       110       120       130       140       150       160       170       180

                                  SD            M  A  E  P  D  R  V  W  V  G  I  D  V  G  K  S
TGGCGGGGTTGTCCGCGGGGACTGCGACGACAGGAGTAGCGGTATGGCCGAACCCGACCGAGTGTGGGTGGGTATCGACGTCGGTAAGTC
        190       200       210       220       230       240       250       260       270

 T  H  H  A  C  A  I  D  D  T  G  K  V  V  W  S  K  K  I  P  N  E  Q  A  A  I  E  D  L  I
CACTCATCATGCGTGCGCGATCGATGACACCGGAAAGGTGGTGTGGTCGAAGAAAATCCCGAACGAACAGGCCGCGATCGAAGACCTGAT
        280       290       300       310       320       330       340       350       360

 A  Q  G  G  R  I  A  N  H  V  V  W  A  I  D  L  T  S  R  R  R  R  L  L  I  A  V  L  L  S
CGCCCAGGGCGGCCGGATTGCTAACCACGTGGTGTGGGCGATCGATTTGACCTCGCGCCGGCGGCGGCTGCTGATCGCCGTACTGCTGAG
        370       380       390       400       410       420       430       440       450

 A  K  A  E  V  V  Y  V  P  G  R  T  V  N  T  M  S  H  A  F  R  G  E  G  K  T  D  A  K  D
CGCGAAAGCCGAGGTGGTGTATGTGCCGGGCCGCACGGTTAACACGATGAGTCATGCGTTCCGCGGCGAAGGCAAGACCGACGCCAAAGA
        460       470       480       490       500       510       520       530       540

 A  R  V  I  A  E  T  A  R  H  R  R  D  L  S  P  V  V  P  G  E  D  L  V  A  E  L  R  S  L
CGCGCGGGTAATCGCCGAAACCGCTCGGCACCGACGAGATCTGTCCCCGGTCGTACCCGGCGAAGACCTGGTTGCCGAATTGCGGTCGCT
        550       560       570       580       590       600       610       620       630

 T  A  Y  R  S  D  L  M  A  D  W  V  R  G  V  N  R  V  R  S  M  L  T  A  I  F  P  A  L  E
GACCGCATACCGGTCGGATCTGATGGCTGACTGGGTGCGAGGCGTGAACCGCGTGCGCTCGATGCTCACCGCCATCTTCCCTGCTCTGGA
        640       650       660       670       680       690       700       710       720

 A  A  F  D  Y  S  T  R  A  P  L  I  L  V  S  A  M  C  T  P  G  E  I  R  S  A  K  R  A  G
AGCTGCGTTCGACTACTCCACCCGCGCGCCGTTGATCCTGGTATCCGCTATGTGCACTCCGGGCGAAATCCGGTCGGCAAAAAGAGCTGG
        730       740       750       760       770       780       790       800       810

 V  I  K  H  L  R  K  N  R  A  W  P  N  N  I  D  T  I  A  D  K  G  L  A  A  A  A  G  Q  I
CGTGATCAAGCACCTTCGGAAAAACCGGGCATGGCCCAACAACATCGACACGATCGCCGACAAGGGCCTCGCCGCGGCAGCAGGCCAGAT
        820       830       840       850       860       870       880       890       900

 I  T  L  P  G  E  A  G  T  A  A  L  I  K  Q  L  A  A  R  L  L  D  L  D  R  Q  I  K  D  I
AATCACCCTTCCCGGCGAAGCCGGAACCGCCGCGCTCATCAAGCAACTCGCAGCACGGCTGCTGGACTTGGATCGGCAGATCAAGGACAT
        910       920       930       940       950       960       970       980       990

 D  K  Q  I  T  N  K  F  R  E  H  P  S  A  A  I  I  E  S  M  P  G  M  G  P  H  L  G  A  E
CGATAAGCAAATCACCAACAAATTTCGTGAGCATCCCAGCGCCGCCATCATCGAGTCGATGCCCGGCATGGGGCCACACCTGGGCGCTGA
       1000      1010      1020      1030      1040      1050      1060      1070      1080

 F  L  V  I  T  G  G  N  M  A  A  F  T  N  P  G  R  L  A  S  F  A  G  L  V  P  V  P  R  D
GTTCCTCGTAATCACCGGCGGCAACATGGCCGCCTTCACCAACCCCGGCCGACTGGCATCGTTCGCCGGATTGGTACCCGTCCCACGCGA
       1090      1100      1110      1120      1130      1140      1150      1160      1170

 S  G  R  I  T  G  N  L  H  R  P  K  R  Y  N  R  R  L  R  R  V  F  Y  L  A  A  L  S  S  L
TTCCGGCCGTATCACCGGCAATCTGCATCGGCCCAAGCGCTACAACCGGCGCCTGCGCCGCGTGTTCTACCTCGCCGCCCTGTCCAGCCT
       1180      1190      1200      1210      1220      1230      1240      1250      1260

 K  I  E  G  P  S  R  A  F  Y  D  R  K  R  S  E  N  H  I  H  T  Q  A  L  L  A  L  A  R  R
CAAGATCGAAGGTCCCTCGCGGGCTTTCTACGACCGCAAACGATCCGAGAACCATATCCACACCCAGGCCCTGCTTGCCCTGGCACGCCG
       1270      1280      1290      1300      1310      1320      1330      1340      1350

 H  V  D  V  L  W  A  L  L  R  D  N  R  T  W  Q  P  Q  Q  P  T  V  A  A  A  *
CCACGTCGACGTCCTGTGGGCACTGCTGCGCGACAACAGAACATGGCAACCCCAGCAACCAACCGTGGCAGCTGCCTGACGCACTCACCG
       1360      1370      1380      1390      1400      1410      1420      1430      1440

GGCGTCCTCACCGCTTGACACGCTCATTGAGA
       1450      1460      1470
```

Fig. 1. Nucleotide sequence of IS*901* and deduced amino acid sequence of ORF1. A possible Shine-Dalgarno sequence (SD) is shown to precede the predicted initiation codon (ATG). These sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Data Libraries under the accession number X59272.

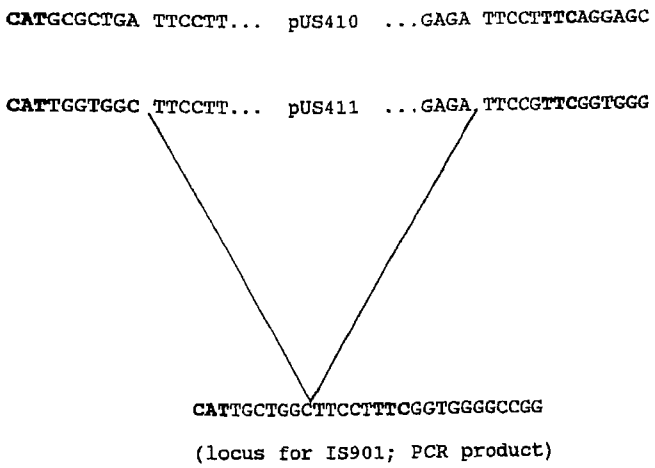

Fig. 2. The sequence of two insertion sites for IS*901* in *M. avium*, as determined from clones pUS410 and pUS411. Also shown is the sequence of a putative insertion site in an *M. avium* strain lacking the IS element.

in mycobacteria that may be useful for genetic studies and for developing mycobacteria as multivaccines (Jacobs *et al.*, 1988).

Upon screening of Southern blots of DNA from *M. avium* strains with the clone pMB22, which contains IS*900*, some strains of *M. avium*, particularly animal-derived strains, were found to produce a complex banding pattern, indicating the presence of a repetitive element related to IS*900* in their genome (McFadden *et al.*, 1990). We have designated these strains *M. avium* RFLP type A/I, to distinguish them from the *M. avium* strains that lack IS*900*-related elements, which we designate as RFLP type A strains (Hampson *et al.*, 1989, McFadden *et al.*, 1990). Interestingly, like *M. paratuberculosis*, *M. avium* RFLP type A/I strains have been isolated from Crohn's disease (J. J. McFadden, unpublished). We have now isolated and characterized the element present in *M. avium* RFLP type A/I strains which has been designated IS*901*. In addition, we have carried out *in vivo* studies to investigate the association of IS*901* with virulence in mice. The data presented indicate that IS*900* and IS*901* belong to a family of phylogenetically related atypical IS elements found in actinomycetes and other bacteria.

Results

*Isolation and sequence analysis of IS*901 *and its integration site*

A genomic library was prepared from a mycobactin-dependent goat-derived *M. avium* RFLP type A/I strain, FP8589. This library was screened with a fragment of pMB22, containing predominantly IS*900* sequence, to isolate the 2.8 kb clone pUS410. The sequence of pUS410 was determined by subcloning into M13mp18 and sequencing in both directions using the dideoxy chain-termination technique (Sanger *et al.*, 1977). DNA alignment of this sequence with IS*900* allowed for an approximation of the putative ends of the element. To identify the exact ends of IS*901*, a second clone containing IS*901* (pUS411) was isolated from the *M. avium* library and the terminal sequence determined directly by means of double-stranded plasmid DNA sequencing (Zhang *et al.*, 1988). To determine whether insertion had produced a target-site duplication, DNA primers homologous to sequences flanking IS*901* in pUS411 (for unknown reasons this approach was unsuccessful using pUS410 flanking DNA sequence) were used to amplify the unoccupied insertion site from a strain of *M. avium* RFLP type A, isolated from an AIDS patient that lacks IS*901*. This 300 bp polymerase chain reaction (PCR) product was directly sequenced (Casanova *et al.*, 1990). The complete nucleotide sequence of IS*901* and amino acid sequence of the major open reading frame are shown in Fig. 1.

IS*901* has a nucleotide sequence of 1472 bp, with a G+C content of 62%, and contains a single long open reading frame (ORF1) from nucleotide 224 to 1426 coding for a protein of 401 amino acids and with an expected M_r of 44 kDa. The predicted ATG initiation codon is preceded by a possible Shine-Dalgarno sequence from position 212 to 216. The ends of IS*901* and flanking genomic DNA in pUS410 and pUS411, together with an unoccupied insertion site, are shown in Fig. 2. Several similarities and differences exist between the ends and target sites of IS*901*, IS*900* and the *Streptomyces* IS elements. Whereas IS*110* and IS*117* possess imperfect inverted terminal repeats (at the integration site for the minicircle), these are absent in IS*116*, IS*900* and IS*901*. Similarly, the latter three IS elements, as well as IS*117*, do not produce a target-site duplication. The consensus insertion site for IS*900* was suggested to be CATGN(4-6)*CNCCTT (where an asterisk denotes the site of insertion (Green *et al.*, 1989)). Examination of the flanking sites for IS*901* insertion leads to the suggestion that there is a consensus insertion site of CATN(7)*TTCCNTTC that is similar to the IS*900* sequence and also the integration site for IS*116* insertion (CATGGTCGG*TCTCCTGGT (Leskiw *et al.*, 1990)); this suggests that similar short sequences might be involved in the integration of these related elements, as has been proposed for other IS elements (Gringauz *et al.*, 1988; Murphy, 1988; Leskiw *et al.*, 1990). The sequence TTCCTT, present at the target site, is also present at the left end of both elements, suggesting that it may be involved in the recognition of the target site or in the recombination reaction. Minor differences in the flanking region of different IS*901* copies and the putative insertion site in *M. avium* are the result of polymorphism characteristic of the *M. avium* complex. As found

```
IS901 ORF1      :  ..MAEPDRVWVGIDVGKSTHHACAIDDTGKVVWSKKIPNEQA...AIEDLIAQGGRI.AN
IS900 ORF1      :  ....VAQP--A-V-A--AD-YCMV-N-DAQRLL-QRVA-DE-...-LLE---AVTTL.-D
IS116 ORF1197   :  .LSTRH--I-----A--GH-W-V-V-AD-ETLF-T-VI-DE-...QVLT--ETA.-E.RE
IS110 ORF1215   :  MFDT-DVG-FL-L----TA--GHGLTPA--K-LD-QL--SEPRLR-VF-KL-......AK
IS117 ORF1      :  .MWEDSLT-FC---WAERH-DV-IV----TLLAKAR-TDDV-GYNKLL--L-EH-DSS-T
IS1000 ORF1A    :  .......MTFA----S-THLDL-LVSNSP-PTRLRFPNSPEGRQALLAA-AHHNPAWV-L
IS492 ORF1      :  MNPKTNQS-N--V-T--FQLDIYIR...PHDIYFTVSND-KGIKE--KKIQQHSPGRIVI
                                                                    rtm
IS901 ORF1      :  H..VVWAIDLTSRRRRLLIAVLLSAKAEVVYV.PGRTVNTMSHAFRGEG.KTDAKDARVI
IS900 ORF1      :  GGE-T-----NAGGAA----L-IA-GQRLL-I.-----HHAAGSY----.-------AI-
IS116 ORF1197   :  E..-R--V-ISG-AST--L-L-VAHGQN----.------RM-G-YK---.----------
IS110 ORF1215   :  FGT-LVIV-QPASIGA-PLT-ARD-GCK-A-L.--LAMRRIADLYP--A.-------A--
IS117 ORF1      :  PIP-AIETSHG.....--V-A-RTGSRK-FAIN-LAAARYRDRHGVSRK.-S-PG--L-L
IS1000 ORF1A    :  EPTGAYHLP-LKLL..........-ENRLQVALVNPYHLAAFRKAK--RQ---RQ--LLL
IS492 ORF1      :  EATGRLEMPFIIAC..........-N-NLPF-VANPIHIKRFAGAI-QRA---KL--QL-

IS901 ORF1      :  AETAR.HRRDLSPVVPGEDLVAELRSLTAYRSDLMADWVRGVNRVRSMLTAIFPALEAAF
IS900 ORF1      :  -DQ--.M-H--Q-LRA-D-IAV---I--SR----V--RT-AIEPNARPAAG-LS---R--
IS116 ORF1197   :  --Q--.M---FA-LDRPPEL-TT--L--NH-A--I--R--LI--L-DL--G-C----R--
IS110 ORF1215   :  --A--TMAHT-RSLELTDEIT---SV-VGFDQ--A-EAT-TS--I-GL--QFH-S--RVL
IS117 ORF1      :  -NIL-TDMHAHR-LPADSELAQAITV-ARAQQ-AVWNRQQVA-Q---L-REYY--ALH--
IS1000 ORF1A    :  -RY-QVYHGE-WAYTLPPEALR--KA-VGY-E--AGRERAIL-QMEAAEW-GSKEVL-LL
IS492 ORF1      :  -HYGEAIQPK--Q..LKP-TMQAMSD-V-R-NQ-L.....VMQTME.......KNRLQIL

IS901 ORF1      :  DY..STR..........APLILVSAMCTPGEIRSAKRAGVIKHL..RKNRAWPNNIDTI
IS900 ORF1      :  --NK-RA..........-LI.....LL-GYQTPD-L-SAGGARVAAFLRKRKAR-A--V
IS116 ORF1197   :  --..-A..........-KGPV-..-L-EYQTPA-L-RTGV-R-TTWLG-RKVRDA--V
IS110 ORF1215   :  GPRLDHQ..........-....-TWLLERYGSPA-L-KAGRRR-VELVRPKA-RMAQRL
IS117 ORF1      :  QSKDGGLTRPDARVILTM--TPAKA-KL-LAQL-AGLKRSGRTRAFNTEIERLRGIFRSE
IS1000 ORF1A    :  QKELACV........KGLLGEVEARIQALLATLPE-EV.....................
IS492 ORF1      :  PKELAMTIKPILTAFKNQITK-ENKIVALIESCPDYQAKNCI.................

IS901 ORF1      :  ADKGLAAAAGQIITLPGEAGTAALIKQLAARLLDLDRQIKDIDKQITNKFREHPSAAIIE
IS900 ORF1      :  -ATA-Q--NA-HSIV--QQLA-TVVAR--KEVMA--TE-G-T-AM-EER--R-RH-E--L
IS116 ORF1197   :  -A-AIE--RT-QVVL--EKRATK-VCD--HQ--A--ER---N-RE-RET--TDDR-E---
IS110 ORF1215   :  I-DIFD-LDE-TVVV--TGTLDIVVPS--SS-TAVHE-RRALEA--NALLEA--LSPVLT
IS117 ORF1      :  YARQ-P-VEDAFGHQ.........LLA-LRQ-DATCLAAD-LA-AV-DA----ADSE-LL
IS1000 ORF1A    :  ..........................................................LM
IS492 ORF1      :  ..........................................................LQ

IS901 ORF1      :  SMPGMGPHLGAEFLVITGGNMAAFTNPGRLASFAGLVPVPRDSGRITGNLHRPKRYNRRL
IS900 ORF1      :  ----F-VI----F-AA---D----ASAD---GV---A---------S---K--R--D---
IS116 ORF1197   :  -------V-----VA-V.-DLSGYKDA-----H---A---------R---Y---Q------
IS110 ORF1215   :  ----V-VRT.-AV-LV-V-DGTS-PTAAH---Y---A-TTKS--TSIHGE-A-RGG--Q-
IS117 ORF1      :  -F--L--L---RV-AEI-DDRSR--DARA-K-Y--SA-IT-A---KHFVGR-FVK.-N--
IS1000 ORF1A    :  AL--V--QVA-AV-ALLPPELWGRAKRAASYA..--I-EREE--KSVERSRLSKKGPPL-
IS492 ORF1      :  --K-I-KIAS-SIISNLPELGYMTNKQASALV..-VA-MN-E---YK-L.RKIQGGRHQV

IS901 ORF1      :  RRVFYLAALSSLKIEGPSRAFYDRKRSENHIHTQALLALARRHVDVLWALLR.DNRTWQ.
IS900 ORF1      :  L-AC--S--V-IRTDPS--TY-----T-GKR----V------RLN----M--.-HAVYH.
IS116 ORF1197   :  -WL--MS-QTAMMRP----DY-LK--G-GLL------S----R------M--.-K-LFT.
IS110 ORF1215   :  K-AMF-S-FACMNADPA--TY---Q-ARGKT------R---QRIS--F-M--.-GTFYES
IS117 ORF1      :  MNAGF-W-FAA-QASPGANAH-R-R-EHGDW-AA-QRH-LN-FLGQ-HHC-QTRQHFDEQ
IS1000 ORF1A    :  --KL-MG--VAVRHDPEM----H-LLSRGKRKK---V-V-HKLLRRMMGR--EYYA-QLD
IS492 ORF1      :  -T-L-M-MM-AIQSNPVFKET-Q-LVAAGKPKKV-II-CIRKM-VI-NSM--DGVMWEAP

IS901 ORF1      :  ..PQQPTVAAA
IS900 ORF1      :  ..-ATT-A--.
IS116 ORF1197   :  ..-AP-VTQTA
IS110 ORF1215   :  RM-AGVEL--.
IS117 ORF1      :  RAFAPLLQ--A
IS1000 ORF1A    :  QGVA.......
IS492 ORF1      :  KTTN.......
```

Fig. 3. CLUSTAL alignment of IS*901* ORF1, IS*900* ORF1197, IS*110* ORF1215, IS*116* ORF1197, IS*117* ORF1, IS*1000* ORF1A and IS*492* ORF1. Sequence identity with IS*901* is shown (–); (·) denotes gaps and deletions in each sequence. Sequence homology with a reverse transcriptase motif is highlighted (rtm).

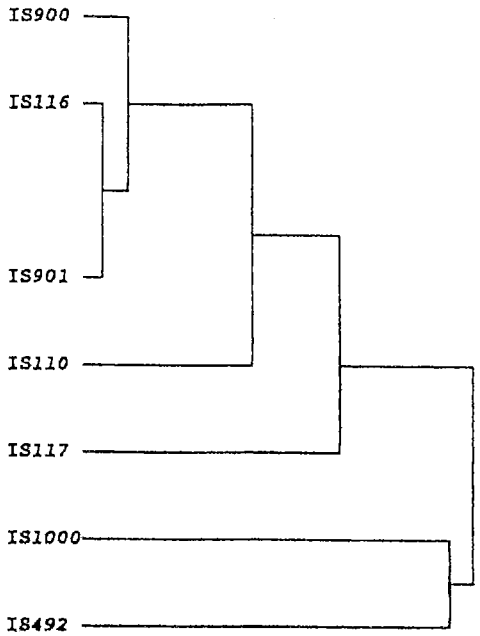

Fig. 4. Phylogenetic relationship between IS*901* and other IS elements derived from PHYLIP analysis.

for IS*900*, both copies of IS*901* insert in the same orientation with respect to the target site.

The sequence homologies between IS*901* and IS*900* (Green *et al.*, 1989), IS*110* (Bruton and Chater, 1987), IS*117* (Henderson *et al.*, 1989; 1990), IS*116* (Leskiw *et al.*, 1990), IS*1000* (Ashby and Bergquist, 1990) and IS*492* (Bartlett and Silverman, 1989) were determined by sequence alignment (Table 1). CLUSTAL (Higgins and Sharp, 1988) was used to align the amino acid sequences of the major ORFs from each of these sequences. As can be seen (Fig. 3), homologous regions are present in each of these proteins, particularly the conserved motif K-D--DA at position 111–117, which, as has been noted, is also found in reverse transcriptase enzymes (Johnson *et al.*, 1986; Xiong and Eickbush, 1988). Additional conserved motifs are also present. Presumably these motifs are conserved functional domains. The ORFs exhibit greatest overall homology at the carboxy terminus and it is interesting that the ORFs of both mycobacterial ISs terminate with three alanine residues. Both of the non-actinomycete sequences appeared to have an approximately 80-amino-acid deletion (or conversely an insertion in the actinomycete elements), with respect to the actinomycete sequences, at position 219–298.

Table 1. DNA sequence homologies and amino acid sequence homologies of the ORFs (in brackets) between IS*901* and IS*900*, IS*100*, IS*116*, IS*117*, IS*1000* and IS*492*.

	IS901
IS*900*	60% (49%)
IS*110*	51% (32%)
IS*116*	58% (50%)
IS*117*[a]	49% (27%)
IS*1000*	34% (23%)
IS*492*	34% (18%)

a. DNA and amino acid homologies between ORF1 of IS*901* and ORF1 of IS*117*.

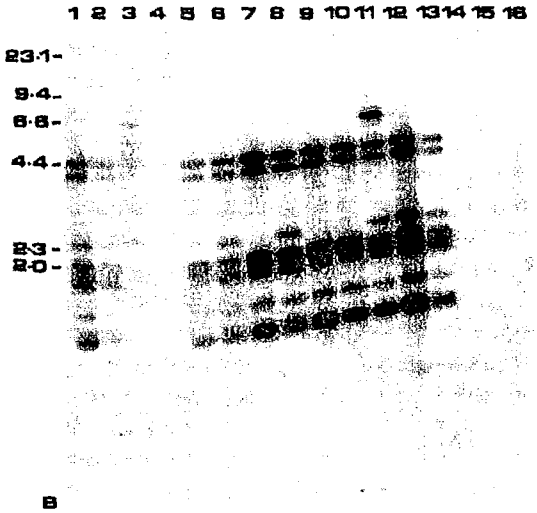

Fig. 5. Southern blots of *M. avium* DNA digested with *Pvu*II, probed with pUS410. Lanes 1 and 8, deer isolates; lanes 2–4, and 9, bovine isolates; lanes 5–7, badger isolates; lanes 10 and 11, wood pigeon isolates; lanes 12 and 13, bird isolates; lanes 14 and 15, AIDS patients isolates; lane 16, *M. paratuberculosis* control. DNA size markers are shown along the side.

Using aligned sequences from the terminal 163 amino acids of each of the proteins (excluding the above-mentioned deletion), a phylogenetic 'tree' was constructed by the method of maximum parsimony using the program PROTPARS (part of the PHYLIP package of phylogeny inference programs; Felsenstein, 1990). The most robust and parsimonious tree (Fig. 4) indicated that the coding region of IS*901* is most closely related to IS*116*, only slightly less closely related to IS*900*, and more distantly related to IS*110*, IS*117*, IS*1000* and IS*492*, in that order. The DNA homology between IS*901* and IS*900* (60%), is much less than that estimated between *M. avium* and *M. paratuberculosis* (>98% (McFadden *et al.*, 1987a,b)), indicating that two separate infection events involving two distinct insertion sequences gave rise to *M. paratuberculosis* and *M. avium* RFLP type A/I rather than divergence of a single insertion sequence after separation of *M. paratuberculosis* and *M. avium* RFLP type A/I. Both genera (*Mycobacterium* and *Streptomyces)* belong to the family of actinomycetes and it is therefore not surprising

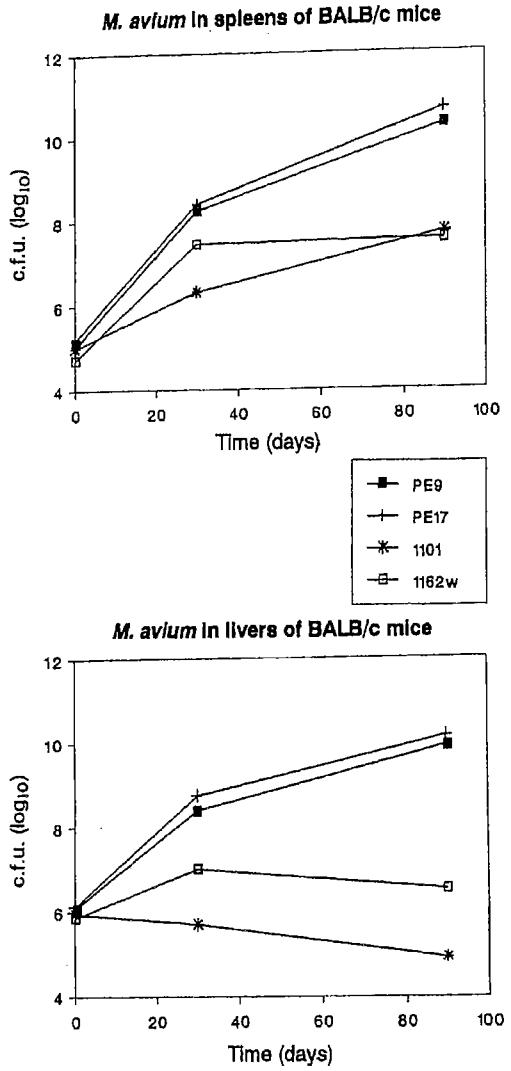

Fig. 6. Proliferation of *M. avium* type A and type A/I strains in liver and spleen of infected BALB/c mice.

that these elements show the greatest relatedness; however, the analysis clearly demonstrates that this class of element is not confined to the actinomycetes.The genus *Thermus* is only distantly related to other eubacteria (Hartmann *et al.*, 1989) and therefore this class of atypical insertion sequences may be far more widely distributed than is presently realized.

*Distribution of IS*901 *in the* M. avium *complex*

To identify strains containing IS*901*, DNA from 202 *M. avium* strains was probed with pUS410 containing IS*901*. Figure 5 shows a selection of these strains screened with pUS410. IS*901* was found in 48/55 *M. avium* strains isolated from birds and other animals. However, IS*901* was not found in any of the 127 *M. avium* strains isolated from AIDS patients, or in any of the 20 environmental isolates screened. All strains containing IS*901* gave very similar banding patterns with the IS*901* probe, indicating that, like IS*900* in *M. paratuberculosis* strains (McFadden *et al.*, 1987b), IS*901* is relatively stable in the genome of *M. avium* RFLP type A/I strains. This is consistent with IS*901* being a site-specific insertion sequence, although some differences in banding patterns were found (Fig. 5). Thus IS*901* segregates between two groups of *M. avium*: *M. avium* RFLP type A, lacking IS*901*, which is commonly found in AIDS patients and also in environmental samples (Z. M. Kunze *et al.*, submitted), and *M. avium* RFLP type A/I, which contains IS*901* and appears to be a more specific animal pathogen.

*Association of IS*901 *with virulence in mice*

The above data indicated that *M. avium* strains containing IS*901* differed in host range from those lacking IS*901*. However, it was not clear whether this was due to a difference in virulence. In order to clarify this, the proliferation of two *M. avium* type A and two *M. avium* type A/I strains was studied in BALB/c mice, which are innately susceptible to infection by these mycobacteria (Appelberg and Sarmento, 1990). As can be seen from Fig. 6, the two *M. avium* type A strains (strains 1101 and 1162w), obtained from AIDS patients, were essentially avirulent in this model, exhibiting minimal proliferation in either the spleen or the liver of mice throughout the three-month period of infection studied (proliferation was arrested after the first month of infection). In contrast, the two animal isolates (PE9 and PE17), both IS*901*-containing strains, proliferated extensively in both organs. Although the number of strains examined is small the difference in virulence for these very closely related *M. avium* strains that differ in possession of IS*901* is striking. It is perhaps surprising that in this system the strains isolated from AIDS patients are essentially avirulent, indicating that other factors (e.g. environmental prevalence) rather than virulence (as measured in BALB/c mice) must be responsible for the prevalence of these strains (and the complete absence of type A/I strainsl) in AIDS.

Discussion

The evolution of *M. avium* type A/I (from a type A strain) clearly involved the acquisition of multiple copies of IS*901*. Genetic changes must also have taken place which have modified the pathogenicity and host-range of type A/I strains relative to type A strains. There is, of course, no direct evidence to link the difference in pathogenicity to acquisition of IS*901*; however, it is notable that the evolution of *M. paratuberculosis* from *M. avium* type A has also involved acquisition of a similar insertion sequence — IS*900* (Green *et al.*, 1989). Clearly, it would be of value to directly determine the phenotypic effect of infection of *M. avium* type A with IS*900* or IS*901*. Transposition of IS*900* into *M. smegmatis* has been

achieved recently (England *et al.*, 1991); however, unfortunately we have so far failed to transform *M. avium* with DNA. Nevertheless, IS*901* may be a useful tool for studying the genetics of virulence of mycobacteria; like IS*900* (England *et al.*, 1991), it may be used to obtain stable integration of foreign genes in mycobacteria. DNA sequence data from IS*901* could also be used to develop a *M. avium*–RFLP type A/I-specific PCR, similar to that already developed for *M. paratuberculosis* (Vary *et al.*, 1990), which could be used for rapid diagnosis of disease caused by this pathogen.

Experimental procedures

Cloning of pUS410 in M13mp18

Standard genetic manipulations were carried out as described by Maniatis *et al.* (1982). Plasmid DNA of pUS410 was prepared using the cleared lysate method. Insert DNA of pUS410 was subcloned into M13mp18 (Amersham). The orientation of insert DNA was determined by means of restriction digest mapping. Screening of the *M. avium* plasmid library was performed using the multiprime DNA labelling system (Amersham).

Sequencing of M13,410; pUS411 and PCR products

The insert of recombinant clone M13,410 was sequenced completely in both directions by means of dideoxy chain-termination sequencing utilizing the Sequenase 2 Kit (US Biochemicals). Two clones, oriented in either direction, were sequenced using the 'universal' primer on the M13 template. Subsequently, DNA sequencing primers were synthesized using an automated DNA oligonucleotide synthesizer (Applied Biosystems).

Primers derived from the putative ends of IS*901*, contained in pUS410 and identified by sequence comparison of M13,410 and IS*900*, were used to sequence the termini and flanking regions of IS*901* contained in pUS411 by means of double-stranded plasmid sequencing employing the Sequenase 2 Kit.

In turn, primers derived from the flanking region of IS*901* contained in pUS411 were employed to amplify a putative insertion site in an *M. avium* strain lacking IS*901*. Samples of undigested chromosomal DNA (0.1 ng) from this *M. avium* strain were chosen as target DNA. Briefly, PCR (Biometra) conditions included a 30-cycle reaction with initial denaturation at 94°C for 1 min, annealing at 60°C for 30 s, and extension at 72°C for 150 s; the final cycle consisted of one annealing at 60°C for 2 min followed by extension at 72°C for 5 min. PCR primers were used at a concentration of 100 ng μl^{-1} each. The expected 300 bp product obtained was sequenced directly (Casanova *et al.*, 1990) using the Sequenase 2 Kit. Modifications of this procedure included denaturation of 100 ng of template DNA in the presence of 25 ng of primer at 99°C for 10 min, annealing of primer for 30 s at room temperature, and extension at room temperature for 45 s followed by the termination step at 37°C for 15 min. Assembly and editing of sequence data were performed using the Staden-Plus Software for Molecular Biology (Amersham). DNA and amino acid homology studies were carried out using the MicroGenie Sequence Analysis Program (Beckman).

Screening of M. avium *strains with pUS410*

A total of 202 *M. avium* isolates from birds and other animals, the environment and AIDS patients, as well as a control strain of *M. paratuberculosis*, were screened with pUS410, containing IS*901*. DNA from these samples was prepared using a previously described method (Visuvanathan *et al.*, 1989). Total DNA was digested with restriction endonuclease *Pvu*II and Southern-blotted (Southern, 1975). Hybridization, using Hybond-N filters (Amersham), was carried out as decribed by Hampson *et al.* (1989).

In vivo *study of the proliferation of* M. avium *isolates in BALB/c mice*

Mycobacterial inocula were prepared as described previously (Appelberg and Sarmento, 1990). Briefly, smooth-transparent colony-forming bacteria from the different isolates, two animal-derived isolates (PE9 and PE17) and two AIDS patient-derived isolates (1101 and 1162w), were grown until mid-log phase in Middlebrook 7H9 medium (Difco) supplemented with 0.04% Tween-80 (Sigma). The bacteria were pelleted by centrifugation, resuspended in saline with 0.04% Tween-80, and sonicated briefly. Aliquots were frozen at –70°C until use. Female BALB/c mice were obtained from the Gulbekian Institute (Oerias) and were infected at 8–10 weeks of age with approximately 10^6 c.f.u. of the different *M. avium* isolates by inoculating the bacteria suspended in 0.2 ml of saline/Tween-80 through one lateral vein of the tail. After 6 h, three mice were killed and the numbers of c.f.u. in the spleens and livers were evaluated as described below. This reflected the numbers of bacteria initially implanted in these organs (at time zero). At 30 and 90 days of infection, the spleens and livers of infected mice (three animals per time point) were aseptically harvested and homogenized in a 0.04% Tween-80 solution. The homogenates were serially diluted and plated in duplicate on Middlebrook 7H10 agar plates. After incubation for 2 weeks at 37°C, the colony-forming units were counted. Results in Fig. 6 are shown as the mean of the $\log_{10}$ c.f.u. in the organs of the infected animals and the standard deviation of this geometric mean.

Notes added in proof

Since submitting this manuscript we have discovered that an insertion sequence from *Mycobacterium avium* has been independently characterized by Moss, M.T., Tizard, M.L.V., Sanderson, J.D., Malik, Z., and Hermon-Taylor, J. (submitted). This shows 98% DNA homology to IS*901*. In order to simplify nomenclature the sequence has been designated IS*902*.

Acknowledgements

We are grateful to Ann Nolan at the Central Veterinary Laboratory, Weybridge, UK, and Tobin Hellyer at St Mary's Hospital, London for provision of animal-derived *M. avium* strains. This investigation was supported by the Science and Technology for Development Programme of the European Community, Contract No. TS2-0080-UK, and by a grant from the Crohn's & Colitis Foundation of America, Inc. This work was carried out in

partial fulfilment of the requirements for the degree of Ph.D. (University of Surrey, UK) by Z.K.

References

Appelberg, R., and Sarmento, A.M. (1990) The role of macrophage function(s) in the control of *Mycobacterium avium* infection in mice. *Clin Exp Immunol* **80:** 324–331.

Ashby, M.K., and Bergquist, P.L. (1990) Cloning and sequence of IS*1000*, a putative insertion sequence from *Thermus thermophilus* HB8. *Plasmid* **24:** 1–11.

Bartlett, D.H., and Silverman, M. (1989) Nucleotide sequence of IS*492*, a novel insertion sequence causing variation in extracellular polysaccharide production in the marine bacterium *Pseudomonas atlantica*. *J Bacteriol* **171:** 1763–1766.

Bruton, C.J., and Chater, K.F. (1987) Nucleotide sequence of IS*110*, an insertion sequence of *Streptomyces coelicolor* A3(2). *Nucl Acids Res* **15:** 7053–7065.

Casanova, J.L., Pannetier, C., Jaulin, C., and Kourilsky, P. (1990) Optimal conditions for directly sequencing double-stranded PCR products with sequenase. *Nucl Acids Res* **18:** 4028.

Collins, P., Mathews, P.R.J., McDiarmid, A. and Brown, A. (1983) The pathogenicity of *Mycobacterium avium* and related mycobacteria for experimental animals. *J Med Microbiol* **16:** 27–35.

England, P.M., Wall, S., and McFadden, J.J. (1991) IS*900*-promoted stable integration of a foreign gene into mycobacteria. *Mol Microbiol* **5:** 2047–2052.

Felsenstein, J. (1990) *PHYLIP Manual.* Version 3.3. University Herbarium, University of California, Berkeley, California, USA.

Green, E.P., Tizard, M.L.V., Moss, M.T., Thompson, J., Winterbourne, D.J., McFadden, J.J., and Hermon-Taylor, J. (1989) Sequence and characteristics of IS*900*, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. *Nucl Acids Res* **17:** 9063–9073.

Gringauz, E., Orle, K.A., Waddell, C.S., and Craig, N.L. (1988) Recognition of *Escherichia coli att*Tn*7* by transposon Tn*7*: lack of specific sequence requirements at the point of Tn*7* insertion. *J Bacteriol* **170:** 2832–2840.

Hampson, S., Portaels, F., Thompson, J., Green, E.P., Hermon-Taylor, J., and McFadden, J.J. (1989) DNA probes demonstrate a highly conserved strain of *Mycobacterium avium* infecting AIDS patients. *Lancet* **i:** 65–68.

Hartmann, R.K., Wolters, J., Kroger, B., Schultze, S., Specht, T., and Erdmann, V.A. (1989) Does *Thermus* represent another deep eubacterial branching? *Syst Appl Bacteriol* **11:** 243–249.

Henderson, D.J., Lydiate, D.J., and Hopwood, D.A. (1989) Structural and functional analysis of the mini-circle, a transposable element of *Streptomyces coelicolor* A3(2). *Mol Microbiol* **3:** 1307–1318.

Henderson, D.J., Brolle, D.F., Kieser, T., Melton, R.E., and Hopwood, D.A. (1990) Transposition of IS*117* (the *Streptomyces coelicolor* A3(2) mini-circle) to and from a cloned target site and into secondary chromosomal sites. *Mol Gen Genet* **224:** 65–71.

Higgins, D.G., and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73:** 237–244.

Jacobs, W.R., Snapper, S.B., and Bloom, B.R. (1988) Beyond BCG: developing recombinant BCG multivaccine vehicle. In *Molecular Biology and Infectious Diseases.* Schwartz, M. (ed.). New York: Elsevier, pp. 207–212.

Johnson, M.S., McClure, M.A., Feng, D.F., Gray, J., and Doolittle, R.F. (1986) Computer analysis of retroviral *pol* genes: assignment of enzymatic function to specific sequences and homologies with non-viral enzymes. *Proc Natl Acad Sci USA* **83:** 7648–7652.

Leskiw, B.K., Mevarech, M., Barritt, L.S., Jensen, S.E., Henderson, D.J., Hopwood, D.A., Bruton, C.J., and Chater, K.F. (1990) Discovery of an insertion sequence, IS*116*, from *Streptomyces clavigerus* and its relatedness to other transposable elements from actinomycetes. *J Gen Microbiol* **136:** 1251–1258.

McFadden, J.J., Butcher, P.D., Chiodini, R., and Hermon-Taylor, J. (1987a) Crohn's disease-isolated mycobacteria are identical to *Mycobacterium paratuberculosis*, as determined by DNA probes that distinguish between Mycobacterial species. *J Clin Microbiol* **25:** 796–801.

McFadden, J.J., Butcher, P.D., Thompson, J., Chiodini, R., and Hermon-Taylor, J. (1987b) The use of DNA probes identifying restriction fragment-length polymorphisms to examine the *Mycobacterium avium* complex. *Mol Microbiol* **1:** 283–291.

McFadden, J.J., Kunze, Z., and Seechurn, P. (1990) DNA probes for detection and identification. In *Molecular Biology of Mycobacteria.* McFadden, J.J. (ed.). London: Academic Press, pp. 139–172.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Murphy, E. (1988) Transposable elements in *Staphylococcus.* In *Transposition.* Kingsman, A.J., Chater, K.F., and Kingsman, S.M., (eds). Cambridge: Cambridge University Press, pp. 59–89, .

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74:** 5463–5467.

Snider, D.E., Hopewell, P.C., Mills, J., and Reichman, L.B. (1987) Mycobacterioses and the acquired immunodeficiency syndrome. *Am Rev Respir Dis* **136:** 492–496.

Southern, E.M. (1975) Detection of specific sequences amongst DNA fragments separated by gel electrophoresis. *J Mol Biol* **908:** 503–517.

Vary, C., Andersen, P., Green, E., Hermon-Taylor, J., and McFadden, J.J. (1990) Use of DNA probes and the polymerase chain reaction to detect *Mycobacterium paratuberculosis* in Johne's disease. *J Clin Microbiol* **28:** 933–937.

Visuvanathan, S., Moss, M.T., Stanford, J.L., Hermon-Taylor, J., and McFadden, J.J. (1989) Simple enzymic method for isolation of DNA from diverse bacteria. *J Microbiol Meth* **10:** 59–64.

Xiong, Y., and Eickbush, T.H. (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol Biol Evol* **5:** 675–690.

Zhang, H., Scholl, R., Browse, J., and Somerville, C. (1988) Double stranded DNA sequencing as a choice for DNA sequencing. *Nucl Acids Res* **16:** 1220.