
Sequence and characteristics of IS900, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*

E.P.Green*, M.L.V.Tizard, M.T.Moss, J.Thompson, D.J.Winterbourne, J.J.McFadden[†] and J.Hermon-Taylor

Department of Surgery, St Georges Hospital Medical School, Tooting, London SW17 ORE, UK

Received August 18, 1989; Accepted October 17, 1989

EMBL accession no. X16293

ABSTRACT:

The complete sequence of an insertion element IS900 in *Mycobacterium paratuberculosis* is reported. This is the first characterised example of a mycobacterial insertion element. IS900 consists of 1451bp of which 66% is G + C. It lacks terminal inverted and direct repeats, characteristic of *Escherichia coli* insertion elements but shows a degree of target sequence specificity. A single open reading frame (ORF 1197) coding for 399 amino acids is predicted. This amino acid sequence, and to a lesser extent the nucleotide sequence, show significant homologies to IS110, an insertion element of *Streptomyces coelicolor* A3(2). It is proposed that IS900, IS110, and similar insertion elements recently identified in disease isolates of *Mycobacterium avium* are members of a phylogenetically related family. IS900 will provide highly specific markers for the precise identification of *Mycobacterium paratuberculosis*, useful in defining its relationship to animal and human diseases.

INTRODUCTION:

Insertion sequence (IS) elements are small, mobile genetic elements, containing only genes related to insertion functions; their location within structural genes causes insertional mutations [1,2]. They were first detected as the cause of strong polar mutations in the *lac* and *gal* operons of *Escherichia coli* K12 [3,4]. Subsequently many IS elements of *E.coli* K12 and closely related bacteria have been identified and characterised [2]. The following properties have been observed: 0.8–1.8kb in size; one major open reading frame (ORF) and often up to two minor ORFs; terminal inverted repeats, 8–40bp; flanking direct repeats, 2–12bp [2]. More recently it has become apparent that IS elements are widespread amongst prokaryotes, for example, IS elements have been found in *Streptomyces coelicolor* A3(2) [5,6,7,], *Corynebacterium diphtheriae* [8], *Methanobrevibacter smithii* [9], *Halobacterium halobium* [10] and *Agrobacterium tumefaciens* [11]. As more bacterial genomes are studied at a molecular level so the number of IS elements known increases. Moreover, it is now obvious that not all IS elements have terminal inverted repeats and/or flanking direct repeats [6,7,12].

IS900 is the first example of an IS element to be found in a mycobacterium. This element was identified in a clone, pMB22, derived from a genomic library prepared from a human Crohn's disease isolate of *Mycobacterium paratuberculosis* [13,14,15]. The 5.5kb *Bam*HI insert of pMB22 includes a sequence of 1–2kb which is repeated 15 to 20 times in the *M.paratuberculosis* genome (15, Green & Moss unpublished). The distribution of this repetitive element, now called IS900, is identical in all 15 strains of *M.paratuberculosis* examined so far [15]. The clone pMB22 has been characterised by restriction mapping and Southern blotting [15].

Previous DNA hybridisation studies demonstrated a sequence difference of less than

2% between *M.paratuberculosis* and serovars 2 and 5 of *M.avium* [13,14]. The use of pMB22 as a probe provides a simple means of distinguishing between these organisms, since IS900 is absent in *M.avium* serovars 2 and 5 [15]. However, related sequences have been identified in some isolates of *M. avium* and these repetitive sequences together with IS900 are believed to be members of a family of phylogenetically related IS elements [15].

In this report the nucleotide sequence and the predicted protein for IS900 are described. The homology found with a *Streptomyces* insertion element provides strong supportive evidence for IS900 being an IS element.

MATERIALS AND METHODS:

Sequencing strategy.

All of the M13 subclones represented in Figure 1 were derived from the 5.5kb BamHI insert of pMB22 using M13mp18 and/or M13mp19 [16]. DNA sequencing was performed by the chain-termination method [17] using Klenow fragment (BCL) or Sequenase (USB). For the restriction site directed sequencing the 'universal' primer (GTAAACGACGGCCAGT; Pharmacia) was used on the appropriate M13 templates. Sequence reactions with synthetic oligonucleotides were performed on M13 and plasmid templates. Double stranded sequencing was by the method of Green *et al* [18] as modified by Winship [19]. 1µg of linearised plasmid was used with 10ng of primer or 50–100ng of polymerase chain reaction (PCR) product plus 50ng of primer. Many secondary structure problems, (stops and compressions) were encountered. These were overcome by the use of dGTP analogues (7-deaza-dGTP, dITP), raised reaction temperatures and electrophoresis through 40% formamide, 7M urea, 8% acrylamide gels in 1× TBE (89mM Tris-base, 89mM boric acid, 2mM EDTA).

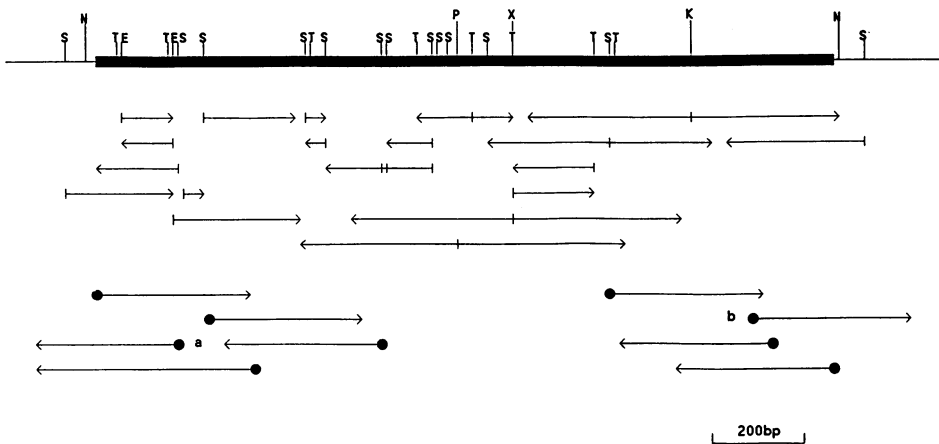


Figure 1 The Strategy used to sequence IS900 in pMB22. IS900 (—) and flanking region (—) sequence was sequenced using site directed sub clones in M13 (—) [derived from pMB22 using *Sau3A* (S), *TaqI* (T), *EcoRI** (E), *PvuII* (P), *XhoI* (X) and *KpnI* (K)] and specific primer extension (●) on plasmid and M13 clones. The primers labelled a and b were used to determine the end points of ISMy1 in pMB22 and other cloned copies. The positions of the two *NruI* (N) sites are also shown.

PCR

Putative insertion loci were isolated by PCR amplification [20] of *M. avium* DNAs (see results). The amplification conditions were as follows: 60s denaturation at 93°C; 60s annealing at 58°C; 180s extension at 72°C in a reaction buffer of 67mM Tris-Cl pH 8.8, 16.6mM ammonium sulphate, 6.7mM magnesium chloride, 10mM β-mercaptoethanol, 170μg BSA, 0.5mM dNTPs, 500ng of each primer, 10ng of target DNA and, 2.5 units Taq polymerase (Cetus). One tenth of the reaction mixture was assayed on a gel of 1% Nusieve agarose (ICN Biomedicals) 1% agarose Type1 (Sigma). The 295bp fragment obtained was purified from the remaining reaction mixture using the Gene-Clean™ (Bio 101) procedure and one fifth was sequenced directly using the amplification primers separately.

Computer analysis of nucleotide and predicted amino acid sequences.

Sequence data were stored, assembled and analysed using computer software developed by Staden [21–27] on a PC AT. Codon usage analysis of Staden and McLachlan [28] was used to determine the coding strand and frame of ORF1197. Data base searches were conducted using the Wisconsin GCG package running on the SEQNET node at SERC Daresbury [29]. The program CLUSTAL [30] was used for comparison and alignment of ORF1197 with the ORF1215 of IS110 in *Streptomyces coelicolor* A3(2).

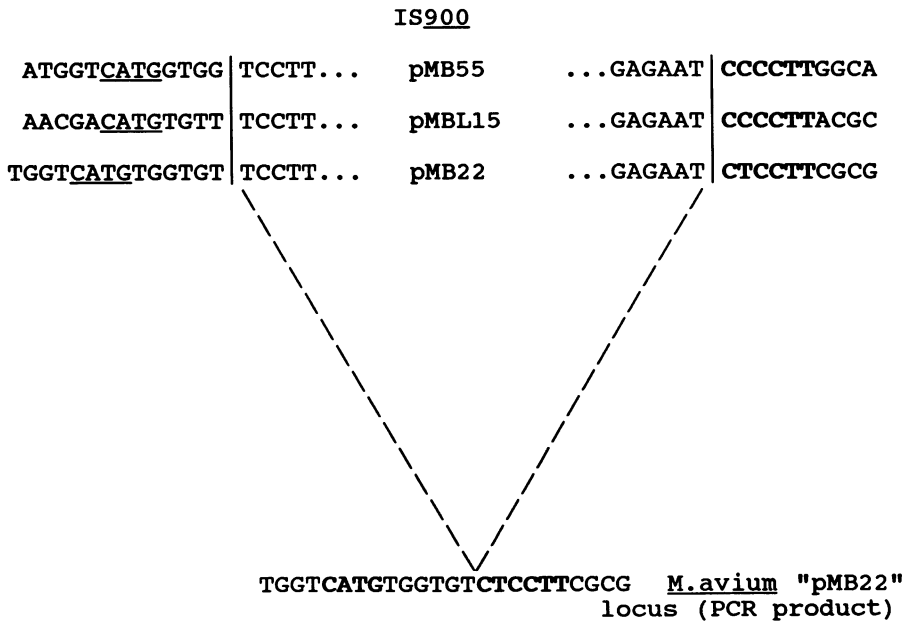


Figure 2 The sequence of three insertion sites for IS900 in *M. paratuberculosis*, isolated in clones pMB55, pMBL15 and pMB22, is shown above. Below is the sequence of a potential insertion site in *M. avium*, within the pMB22 equivalent locus. Possible conserved nucleotides are underlined.

RESULTS

Sequence determination and insertion site homologies.

The position of IS900 in pMB22 was assigned to a 1.5kb *Nru*I fragment (Figure 1) by Southern blot analysis of the insert with radiolabelled *M. paratuberculosis* genomic DNA. With multiple copies of the element in the *M. paratuberculosis* genome, restriction fragments carrying part or all of this sequence are strongly highlighted by this method. The strategy used to sequence IS900 is shown in Figure 1. The region represented has been sequenced 2–4 times in both directions. It was not possible to identify the precise end points of IS900 from this sequence alone, as no significant inverted or direct repeats were found by computer analysis [23]. This problem was overcome by isolating further copies of IS900 from our *M. paratuberculosis* genomic library [13,14]. The library was screened with the 1.5kb *Nru*I fragment of pMB22 (Figure 1). The restriction enzyme site profiles of positive clones were compared to that of pMB22 and Southern blot hybridisations were used to confirm the presence of IS900. Two unique isolates were identified: clones pMB55 and pMB115. The nucleotide sequence of these two clones, directed from primers a and b (Figure 1), was determined. These data were compared to that of pMB22 and points of divergence were located. Further confirmation of the termini was obtained by examining the pMB22 equivalent locus in two strains of *M. avium* which did not contain IS900 related elements.

```

TCCTTACCTTTCTTGAAGGGTGTTCGGGGCCGTCGCTTAGGCTTCAATTGCCCAGGGAC
    10          20          30          40          50          60

GTCGGGTATGGCTTTCATGTGGTTGCTGTGTTGGATGGCCGAAGGAGATTGGCCGCCCGC
    70          80          90         100         110         120

GGTCCC CGCAGACTCGACCGCTAATTGAGAGATGCGATTGGATCGCTGTGTAAGGACAC
   130         140         150         160         170         180

GTCGGCGTGGTTCGCTGCTGGGTTGATCTGGACAATGACGGTTACGGAGGTGGTTGTGGC
   190         200         210         220         230         240
                        SD                               V A

Q P V W A G V D A G K A D H Y C M V I N
ACAACCTGTCTGGGCGGGCGTGGACGCCGGTAAGGCCGACCATTAATGCTGGTTATTTAA
   250         260         270         280         290         300

D D A Q R L L S Q R V A N D E A A L L E
CGACGACGCGCAGCGATTGCTCTCGCAGCGGGTGGCCAACGACGAGGCCGCGCTGCTGGA
   310         320         330         340         350         360

L I A A V T T L A D G G E V T W A I D L
GTTGATTGGCGGGTACGACGTTGGCCGATGGAGGCGAGGTACCGTGGGCGATCGACCT
   370         380         390         400         410         420

N A G G A A L L I A L L I A A G Q R L L
CAACGCCGGCGGGCGCCGCTGCTGATCGCCTTGCTCATCGCTGCCGGGACGCGGCTGCT
   430         440         450         460         470         480

Y I P G R T V H H A A G S Y R G E G K T
TTATATTCCCGGGCGCACGGTCCATCACGCCCGGGTAGTTACCGCGGCGAAGGCAAGAC
   490         500         510         520         530         540

D A K D A A I I A D Q A R M R H D L Q P
CGACGCCAAAGACGCTGCGATCATCGCCGATCAGGCCCGGATGCGCCACGACTTGCAGCC
   550         560         570         580         590         600
    
```

L R A G D D I A V E L R I L T S R R S D
 TCTGCGCGCCGGCGATGACATCGCAGTCGAGCTGCGCATCCTGACCAGCCGACGTTCCGA
 610 620 630 640 650 660

L V A D R T R A I E P N A R P A A G I L
 TCTGGTGGCTGATCGGACCCGGGCGATCGAACCGAATGCGCGCCAGCTGCTGGAATACT
 670 680 690 700 710 720

S A L E R A F D Y N K S R A A L I L L T
 TTCGGCGCTGGAACGCGCCTTCGACTACAACAAGAGCCGTGCCGCGCTGATCCTGCTTAC
 730 740 750 760 770 780

G Y Q T P D A L R S A G G A R V A A F L
 TGGTACCAAACCTCCGACGCGCTGCGCAGCGCCGGTGGCGCTCGAGTAGCCGCGTTCTT
 790 800 810 820 830 840

R K R K A R N A D T V A A T A L Q A A N
 GCGTAAACGCAAGGCCCGCAACGCCGATACCGTCGCAGCCACCGCGCTGCAGGCCGCTAA
 850 860 870 880 890 900

A Q H S I V P G Q Q L A A T V V A R L A
 CGCCAAACAGCATCGTCCCGGCCAACAACCTGGCGGCCACTGTGGTGGCCCGCCTGGC
 910 920 930 940 950 960

K E V M A L D T E I G D T D A M I E E R
 CAAGGAGGTGATGGCCCTCGACACCGAAATCGGCGACCCGACGCGATGATCGAGGAGCG
 970 980 990 1000 1010 1020

F R R H R H A E I I L S M P G F G V I L
 ATTTCCGCCGCCACCGCCACGCCGAAATCATCCTGAGCATGCCCGGATTCCGGCGTCATCCT
 1030 1040 1050 1060 1070 1080

G A E F L A A T G G D M A A F A S A D R
 GGGCGCTGAGTTCCTCGCCGCCACCGCGGGACATGGCCGATTGCCCTCCGCCGACCG
 1090 1100 1110 1120 1130 1140

L A G V A G L A P V P R D S G R I S G N
 CCTCGCCGGCGTCGCCGGCTGGCGCCGGTACCACGAGATTCCGGCCGCATCAGCGGAAA
 1150 1160 1170 1180 1190 1200

L K R P R R Y D R R L L R A C Y L S A L
 CCTCAAACGCCCCGACGCTACGACCGGCGCCTGCTGCGCGCCTGTACCTGTCCGCCTT
 1210 1220 1230 1240 1250 1260

V S I R T D P S S R T Y Y D R K R T E G
 GGTCAGCATCCGCACCGACCCCTCCTCGCGCACCTACTACGCCGAAAACGCACCGAAGG
 1270 1280 1290 1300 1310 1320

K R H T Q A V L A L A R R R L N V L W A
 AAAACGCCACACCCAAGCCGTCCTCGCCCTGGCCCGCCGCGCCTCAACGTCCTGTGGGC
 1330 1340 1350 1360 1370 1380

M L R D H A V Y H P A T T T A A A
 CATGTCGCGACACGCTGTCTACCACCCGCAACCACTACCGCGCGGCTTGACAACG
 1390 1400 1410 1420 1430 1440

TCATTGAGAAT
 1450

Figure 3 The nucleotide sequence of IS900 and the amino acid sequence of ORF1197 also showing a possible Shine-Dalgarno sequence preceding the predicted initiation codon (GTG).

From sequence flanking IS900 in pMB22 (data not shown), a pair of 20mer primers were selected one located 40bp 5' of IS900 and the other located 255bp 3'. Polymerase chain reaction (PCR) amplification of *M. avium* genomic DNA, with this pair of primers, allowed isolation of an IS element free locus. The sequence of this 295bp product was determined directly [19,20]. Figure 2 shows three sites of insertion from *M. paratuberculosis* and the unoccupied locus from *M. avium*. From these data it was possible to confirm that IS900 does not have terminal inverted repeats or flanking direct repeats. However, these results suggest a consensus sequence CATG within 10 bases flanking the 5' terminus of IS900 and a consensus sequence CNCCTT flanking the 3' terminus. Thus, CATGN(4-6)CNCCTT may represent a recognition sequence for IS900 insertion and in the three cases examined the element has the same orientation with respect to this sequence.

Analysis of the nucleotide sequence of IS900.

The nucleotide sequence of IS900 is shown in Figure 3. The overall G + C content is 66%, as is the G + C composition of the flanking region. There is a long open reading frame (ORF 1197) from nucleotide 236 to 1432 corresponding to a protein of 399 amino acids (Mr. 43 kDa). The predicted GTG initiation codon is preceded by a Shine-Dalgarno sequence from nucleotide 226 to 229. Codon usage compared to mycobacterial genes retrieved from the GenBank and EMBL Databases (releases 59.0 and 18.0 respectively) confirmed that this was the likely coding frame. Table 1 shows the codon usage of ORF1197. *In vitro* translation has identified a 43kDa product from the element (Tizard et al, manuscript in preparation).

The first 236 nucleotides of IS900 contain two positions which show a resemblance to the -35 and -10 regions of the consensus for *E. coli* promoters (Figure 3). Some form of promoter would be expected in this region for autonomous regulation, though the significance of the sequences indicated has not yet been determined.

Table 1 Codon usage for predicted protein ORF1197 of IS900.

F TTT	1.	S TCT	0.	Y TAT	1.	C TGT	0.
F TTC	5.	S TCC	4.	Y TAC	9.	C TGC	2.
L TTA	0.	S TCA	0.	* TAA	0.	* TGA	0.
L TTG	8.	S TCG	4.	* TAG	0.	W TGG	3.
L CTT	3.	P CCT	2.	H CAT	2.	R CGT	3.
L CTC	9.	P CCC	7.	H CAC	8.	R CGC	28.
L CTA	0.	P CCA	2.	Q CAA	6.	R CGA	7.
L CTG	24.	P CCG	2.	Q CAG	6.	R CGG	6.
I ATT	3.	T ACT	4.	N AAT	1.	S AGT	1.
I ATC	17.	T ACC	14.	N AAC	8.	S AGC	7.
I ATA	1.	T ACA	0.	K AAA	5.	R AGA	0.
M ATG	7.	T ACG	4.	K AAG	5.	R AGG	0.
V GTT	1.	A GCT	10.	D GAT	7.	G GGT	3.
V GTC	11.	A GCC	41.	D GAC	21.	G GGC	17.
V GTA	2.	A GCA	5.	E GAA	6.	G GGA	5.
V GTG	9.	A GCG	21.	E GAG	8.	G GGG	3.

Sequence homologies between IS900 and an insertion sequence from Streptomyces coelicolor A3(2), IS110.

Comparisons of the IS900 sequence with the EMBL and GenBank data bases revealed an overall 52% homology (after maximum alignment) with IS110 of *S.coelicolor* A3(2) [5,6,7]. The 3' ends of IS900 and IS110 show the greatest degree of homology. No significant homologies were found with IS elements from *E.coli* or any other bacterial species.

DIAGON comparison of the amino acid sequences encoded by IS900 and IS110 (respectively ORF1197 and ORF1215) produced a single broken diagonal, whereas a comparison of the DNA sequences was inconclusive (Figure 4). CLUSTAL alignment of the amino acid sequences showed that 36% of the residues are identical and 12% are conserved. These figures reflect a strong conservation of particular motifs between these ORFs, which are highlighted in Figure 5. As with the DNAs of IS110 and IS900, the similarities are highest between the 3' ends of the predicted proteins. This is further emphasised by hydrophobicity and charge plots for the two proteins (figure 5). The 3' ends have an almost identical charge and hydrophobicity distribution.

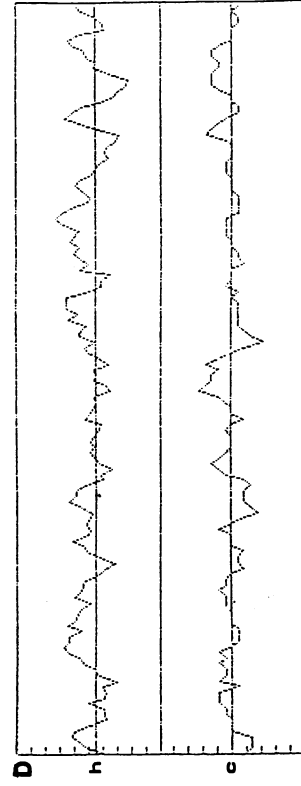
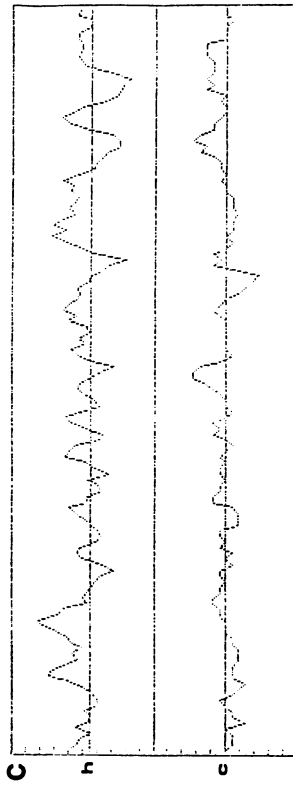
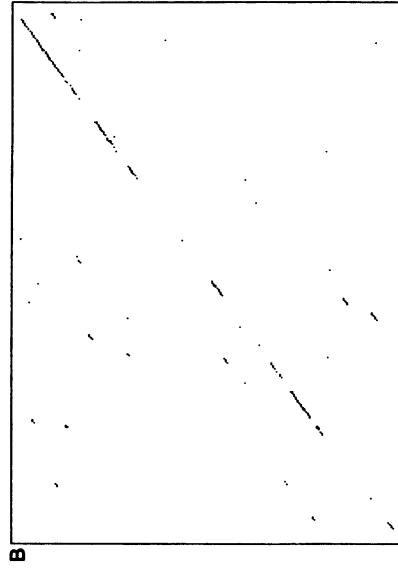
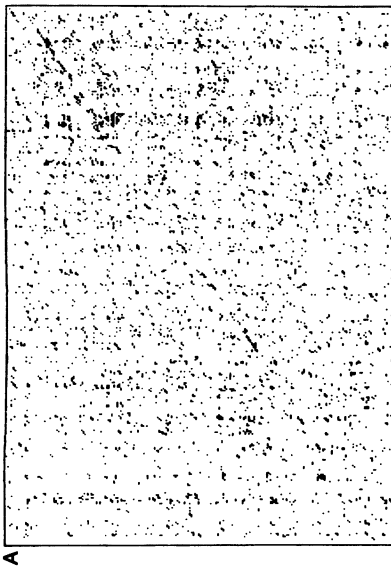
DISCUSSION

Analysis of the nucleotide sequence of IS900 shows that it has a size of 1451bp and a major ORF of 1197 nucleotides coding for a predicted protein of 399 amino acids. The sequence of three separate copies of IS900 is identical (results not shown). Southern blot analysis using IS900 derived probes has shown that the genome of *M.paratuberculosis* includes 15 to 20 dispersed copies of IS900. These features together with the homologies found with the IS110 of *S.coelicolor* A3(2) support the proposal that IS900 is an IS element [15].

The degree of homology between IS900 and IS110 at the nucleotide level clearly shows that they are distinct. However, the presence of conserved amino acid sequence motifs suggests a degree of protein structure and function conservation. Homologies between IS900 and IS elements in disease isolates of *M.avium* have been shown by Southern blot analysis: probes derived from the 3' end of IS900 were found to cross hybridise whereas those derived from the 5' end did not [15]. This is comparable with the location of homologies found between IS900 and IS110. These observations indicate that IS900, IS110 and the related *M.avium* elements may belong to a family of phylogenetically related IS elements. This family of IS elements may also include further sequences related to IS110, which have been detected by Southern blot hybridisations in a range of *Streptomyces* sp. [5].

IS900 and IS110 have G + C contents comparable to that of their hosts (66% and 71% respectively). A similar degree of conservation of G + C content between host and IS element DNA is expected for the *M.avium* elements, as the genomes of *M.avium* and *M.paratuberculosis* are closely related [14]. Mycobacteria and Streptomyces belong to the family Actinomycetes and both genera characteristically have DNA with a high G + C content (62–73%). Thus, it is interesting that the IS elements of these organisms may be related. The conserved amino acid sequence motifs could be domains involved in functions such as binding to G + C rich DNA, excision, integration and/or duplication of the elements.

The sequences adjacent to the termini of IS900 do not contain any direct repeats which could have been produced by a duplication of host sequences during integration, as occurs




```

ORF1215      : MFDTEDVGVFLGLDVGKTAHHGHGLTPAGKKVLDKQLPNSEPRLRVFDKLA--AKFG
ORF1197      : ----VAQPVWAGVDAGKADHYCMVINDDAQRLLSQRVANDEAALELELIAAVTTLADGG
                *  * . * * * *      .      . * * * *      .      * *
ORF1215      : LVIVDQPASIGALPLTVARDAGCKVAYLPLGLAMRRIADLYPGEAKTDAKDAAVIADAA
ORF1197      : TWAILDLNAGGAALLIALLIAAGQRLLYIPGRTVHHAAGSYRGEKTDAKDAAIADQA
                . * * * * .      . * * . * * * * . . . * * * * * * * * * * * * *
ORF1215      : MAHTLRSLLELTDEITAELSVLVGFDQDLAAEATRTSNRIRGLLTQFHPSLERVLGPRL
ORF1197      : MRHDLQLRAGDDIAVELRIILTSRRSDLVADRTRAIEPNARPAAGILSALERAFDYNK
                * * * * * * * * * * * * * * * * * * * * * * * * * * * *
ORF1215      : QAVTWLLERYGSPAALRKAGRRRLVELVRPKAPRMAQRLLIDDFDALDEQTVVVPGTG
ORF1197      : AALI-LLTGYQTPDALRSAGGARVA AFLRKRKARNADTVAATALQAANAQHSIVPGQQ
                * . * * * * * * * * * * * . . . * * . .      . * * * * * *
ORF1215      : DIVVPSLASSLTAVHEQRRALEAQINALLEAHPLSPVLTSMPGVGVRT-AAVLLVTVG
ORF1197      : ATVVARLAKVEMALDTEIGDTDAMI EERFRHRHAEIILSMPGFGVILGAEFLAATGG
                * * * * . * . .      . * * . *      . . * * * * * * * * * *
ORF1215      : TSFPTAAHLASYAGLAPTTKSSGTSIHGEHAPRGGNRQLKRAMFLSAFACMNADPASR
ORF1197      : AAFASADRLAGVAGLAPVPRDSGRISGNLKRPRRYDRLLRACYLSALVSIRTPSSR
                * * * . * * * * * * . * *      * * . * * * * * * * * . * * * *
ORF1215      : YDRQRARGKTHQTQALLRLRQRISVLFAMLRDGTTFYESRMPAGVELAA
ORF1197      : YDRKRTEGKRHTQAVLALARRRLNVLWAMLRDHAVYH---PATTTAAA
                * * * * * * * * * * * * * * * * * * * * * * * * * * *
    
```

Figure 5 CLUSTAL (30) alignment of ORF1197 with ORF1215 of the *S.coelicolor* insertion element IS110, showing identical (*) and conserved (.) residues.

with classical *E.coli* K12 IS elements [2]. IS110 frequently integrates into a poly C site in phage C31 and in the process may produce a duplication of this target sequence [6]. This is quite different from the proposed target site for IS900, where homologies detected between three insertion sites in *M.paratuberculosis* and a potential site in *M. avium* suggests a specific recognition sequence of CATG(N)₄₋₆CNCCTT. There are a number of other distinct differences between IS900 and IS110: IS900 is approximately 100bp smaller than IS110; IS900 has no form of terminal inverted repeats, and IS900 has no equivalent of the short open reading frame (ORF330) of IS110 [6].

It is clear that IS900 cannot transpose by any of the mechanisms proposed for the classical IS elements of *E.coli* [2,32, 33,34,35,36,]. There are no terminal inverted repeats to form stem loop structures and no flanking direct repeats. These features are not essential for transposition: IS492, which causes a reversible inactivation of extracellular polysaccharide production in *Pseudomonas atlantica*, does not have terminal inverted repeats but does generate a 5bp direct repeat [37]; a recently identified IS element of *Streptomyces clavuligerus* transposes into plasmids during propagation in this organism, yet does not have terminal inverted repeats and does not produce a duplication of target sequence [7]; the Staphylococcus transposon Tn554 also does not have terminal inverted repeats and flanking direct repeats [12,38].

The source of IS900 is not known but perhaps it was originally carried on a mycobacteriophage. Many IS elements are known to move between phage genomes,

Figure 4 Comparison of IS900 with IS110. A) DIAGON comparison of IS900 DNA (vertical) with IS110 DNA (horizontal). B) DIAGON comparison of IS900 predicted protein ORF1197 with ORF1215 of IS110. C) Hydrophobicity (h) and charge (c) plots for ORF1197 of IS900. D) Hydrophobicity (h) and charge (c) plots for ORF1215 of IS110.

plasmids and chromosomes [2]. IS110 is known to be an occasional passenger on the temperate phage C31. Strains of *S.coelicolor* A3(2) usually carry several copies of IS110 with infrequent changes in number and location. Thus, IS110 has been seen to transpose, whereas at present there is no such evidence for this with IS900. All 15 strains of *M.paratuberculosis* examined so far appear to have identical insertions of IS900 [15]. The possibility that the presence of IS900 in *M.paratuberculosis* provides the organism with a selective advantage such as altered pathogenicity has already been discussed [15]. It is known that IS50 of *E.coli* transposes most efficiently when it first enters a cell and thereafter inhibition of further transposition is seen [39]. Perhaps an extreme case of this type of phenomenon is responsible for the lack of mobility of IS900 in *M.paratuberculosis*. Alternatively all targets for IS900 in the *M.paratuberculosis* genome may be occupied.

Repetitive elements, which may be IS elements have now been identified in *M.tuberculosis* [40] and *M.leprae* [41]. The distribution and location of the *M.leprae* element was shown to be identical in four unique isolates of the organism, whereas, probes for the *M.tuberculosis* repetitive element(s) show that there is some variation in distribution between organisms of the *M.tuberculosis* complex. IS900 probes do not hybridise to *M.tuberculosis* DNA and no cross-hybridization was seen between the *M.leprae* element and *M.tuberculosis* [41]. The possibility that they may show amino acid homologies has yet to be excluded. A lack of homology at the nucleic acid level may reflect an inability to transpose between related species. Unlike some *E.coli* IS elements (e.g IS1) which are found in other species, IS900 appears to be confined to *M.paratuberculosis* (Moss *et al* manuscript in preparation).

M.paratuberculosis is an obligate pathogen causing chronic granulomatous intestinal infections of economic importance in farm animals. There is accumulating evidence which links this pathogen with a similar chronic intestinal disease in humans, Crohn's disease [15,42]. The uniqueness of sequences within the 5' portion of IS900 will provide a source of probes and primers for the specific and sensitive detection of *M.paratuberculosis*.

ACKNOWLEDGMENTS

We would like to thank K. Chater and colleagues at the John Innes Institute, Norwich for their helpful discussion and introduction to the program CLUSTAL. We would also like to thank P. M. Green (Guy's Hospital) for providing computer facilities and assistance and to Dr. B. Guicquel-Sanzev, Institut Pasteur for advice. We are grateful to the SERC for the use of their computing facilities at Daresbury. This work was supported by grants from Action Research for the Crippled Child and the Wellcome Trust to whom we express our appreciation.

*To whom correspondence should be addressed

[†]Present address: Department of Microbiology, University of Surrey, Guildford, Surrey GU2 5XH, UK

REFERENCES

1. Campbell, A., Berg, D. E., Botstein, D., Lederberg, E. M., Novick, R. P., Starlinger, P. and Szybalski, W. (1979) *Plasmid*, **2**, 466–473
2. Iida, S., Meyer, J. and Arber, W. (1983) In J. A. Shapiro (ed), *Mobile Genetic Elements*, pp 159–221, Academic Press, New York.
3. Malamy, M. H. (1966) *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 189–195.
4. Saedler, H. and Starlinger, P. (1967) *Mol. Gen. Genet.*, **100**, 178–179.
5. Chater, K. F., Bruton, C. J., Foster, S. G. and Tobek, I. (1985) *Mol. Gen. Genet.*, **200**, 235–239.
6. Bruton, C. J. and Chater, K. F. (1987) *Nucleic Acids Res.*, **15**, 7053–7065.

7. Chater, K. F., Henderson, D. J., Bibb, M. J. and Hopwood, D. A. (1988) In A. J. Kingsman, K. F. Chater and S. M. Kingsman (eds), *Transposition*, pp 7–42, Cambridge University Press, Cambridge.
8. Rappuoli, R., Perugini, M. and Ratti, G. (1987) *J. Bacteriol.*, **169**, 308–312.
9. Hamilton, P. T. and Reeve, J. N. (1985). *Mol. Gen. Genet.*, **200**, 47–49.
10. Das Sarma, S., Raj Bandary, U. L. and Khorana, H. G. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2201–2205.
11. Machida, Y., Sakurai, M., Kijokawa, S., Ubasawa, A., Suzuki, Y. and Ikeda, J. E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7495–7499.
12. Murphy, E. (1989) In Berg, D. E. and Howe, M. M. (eds), *Mobile DNA*, pp 269–288. American Society for Microbiology, Washington.
13. McFadden, J. J., Butcher, P. D., Chiodini, R. and Hermon-Taylor, J. (1987) *J. Clinical Microbiol.*, **25**, 796–801.
14. McFadden, J. J., Butcher, P. D., Thompson, J., Chiodini, R. and Hermon-Taylor, J. (1987) *Molec. Microbiol.* **1**, 283–291.
15. McFadden, J. J., Green, E. P., Thompson, J., Moss, M., Tizard, M., Hampson, S., Portaels, F. and Hermon-Taylor, J. (1989). *Infection and Immunity*, submitted.
16. Messing, J. (1983) *Methods in Enzymol.*, **101**, 20–78.
17. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
18. Green, P. M., Bentley, D. R., Mibashan, R. S., Nilson, I. M. and Gianelli, F. (1989) *EMBO. J.*, **8**, 1067–1072.
19. Winship, P. (1989) *Nucleic Acids Res.*
20. Saiki, R. K., Gelfund, G. H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) *Science*, **239**, 487–494.
21. Staden, R. (1982) *Nucleic Acids Res.* **8**, 8673–8694.
22. Staden, R. (1982) *Nucleic Acids Res.* **10**, 4731–4751.
23. Staden, R. (1984) *Nucleic Acids Res.* **12**, 505–519.
24. Staden, R. (1984) *Nucleic Acids Res.* **12**, 521–538.
25. Staden, R. (1984) *Nucleic Acids Res.* **12**, 551–567.
26. Staden, R. (1986) *Nucleic Acids Res.* **14**, 217–231.
27. Staden, R. (1987) In Bishop, M. J. and Rawlings, C. J. (eds), *Nucleic Acid and Protein Sequence Analysis*. IRL Press, Oxford.
28. Staden, R. and Mclachlan, A. D. (1982) *Nucleic Acids Res.* **10**, 141–156.
29. Devereux, J., Haerberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
30. Higgins, D. G. and Sharp, P. M. (1988) *Gene*, **73**, 237–244.
31. Grindley, N. D. F., and Sherratt, D. J. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 1257–1261.
32. Shapiro J. A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 1933–1937.
33. Shapiro, J. A., (1980) In Stuttard, C and Rozee, K. R. (eds) *Plasmids and Transposons—Environmental Effects and Maintenance Mechanisms*, pp. 229–244, Academic Press, New York.
34. Berg, D. E. (1977) In Bukhari, I. A., Shapiro, J. A. and Adhya, S. L. (eds) *DNA Insertion Elements, Plasmids and Episomes*. pp. 205–212 Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
35. Berg, D. E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 792–796.
36. Bukhari, A. I. (1981) *Trends in Biochem. Sci.*, **6**, 56–60.
37. Bartlett, D. H. and Silverman, M. (1989) **171**, 1763–1766.
38. Murphey, E. (1988) In Kingsman, A. J., Chater, K. F. and Kingsman, S. M., *Transposition*, Cambridge University Press, Cambridge, pp. 59–89.
39. Berg, D. E., Kazic, T, Phadnis, S. H., Dodson, K. W. and Lodge, J. K. (1988) In Kingsman, A. J., Chater, K. F. and Kingsman, S. M., *Transposition*, Cambridge University Press, Cambridge, pp.107–129.
40. Eisenach, K. D., Crawford, J. T. and Bates, J. H. (1988) **26**, 2240–2245.
41. Clark-Curtiss J. E. and Docherty, M. A. (1989) *J. Infect. Dis.* **159**, 7–15
42. Chiodini, R. (1989) *Clin. Microbiol. Rev.* **2**, 90–117.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.