IS900 targets translation initiation signals in *Mycobacterium avium* subsp. *paratuberculosis* to facilitate expression of its hed gene

Tim Doran, Mark Tizard, Douglas Millar,† Jon Ford, Nazira Sumar, Mark Loughlin and John Hermon-Taylor

Author for correspondence: Mark Tizard. Tel: +44 181 725 5580. Fax: +44 181 725 3594. e-mail: sghk400@sghms.ac.uk

**Keywords**: mycobacteria, IS900, DNA insertion elements, targeting, translation

**INTRODUCTION**

*Mycobacterium avium* subsp. *paratuberculosis* (formerly *Mycobacterium paratuberculosis*) is an atypical insertion sequence, IS900, encodes a novel gene on the complementary strand to the putative transposase, p43. This gene requires a promoter, ribosome binding site (RBS) and termination codon to be acquired upon insertion into the *M. avium* subsp. *paratuberculosis* genome and hence is designated the hed (host expression-dependent) gene of IS900. Analysis of IS900 insertion sites suggests that this element targets translation initiation signals in *M. avium* subsp. *paratuberculosis*, specifically inserting between the RBS and start codon of a putative gene sequence. This aligns the hed initiation codon adjacent to a functional RBS and possibly downstream of an active promoter, driving expression of Hed protein. We have confirmed this unique targeting process by detecting expression of hed in *M. avium* subsp. *paratuberculosis* at the level of transcription by reverse transcription-PCR. Further, two Hed-specific antibodies detected Hed translation products in Western blots of protein extracts from *M. avium* subsp. *paratuberculosis*. A recombinant form of Hed expressed and purified from *Escherichia coli* will facilitate studies of IS900 transposition and will also be assessed as a diagnostic antigen for *M. avium* subsp. *paratuberculosis* disease. Implications of IS900 insertion in *M. avium* subsp. *paratuberculosis* pathogenicity are discussed.

---

**The Mycobacterium avium subsp. paratuberculosis** (formerly *Mycobacterium paratuberculosis*) atypical insertion sequence, IS900, encodes a novel gene on the complementary strand to the putative transposase, p43. This gene requires a promoter, ribosome binding site (RBS) and termination codon to be acquired upon insertion into the *M. avium* subsp. *paratuberculosis* genome and hence is designated the hed (host expression-dependent) gene of IS900. Analysis of IS900 insertion sites suggests that this element targets translation initiation signals in *M. avium* subsp. *paratuberculosis*, specifically inserting between the RBS and start codon of a putative gene sequence. This aligns the hed initiation codon adjacent to a functional RBS and possibly downstream of an active promoter, driving expression of Hed protein. We have confirmed this unique targeting process by detecting expression of hed in *M. avium* subsp. *paratuberculosis* at the level of transcription by reverse transcription-PCR. Further, two Hed-specific antibodies detected Hed translation products in Western blots of protein extracts from *M. avium* subsp. *paratuberculosis*. A recombinant form of Hed expressed and purified from *Escherichia coli* will facilitate studies of IS900 transposition and will also be assessed as a diagnostic antigen for *M. avium* subsp. *paratuberculosis* disease. Implications of IS900 insertion in *M. avium* subsp. *paratuberculosis* pathogenicity are discussed.

---

**Keywords**: mycobacteria, IS900, DNA insertion elements, targeting, translation

---

**INTRODUCTION**

*Mycobacterium avium* subsp. *paratuberculosis* (formerly *Mycobacterium paratuberculosis*) is an obligate pathogen causing Johne's disease in ruminants and has been associated with Crohn's disease in humans. IS900 is an atypical insertion sequence (IS), exclusively present in the *M. avium* subsp. *paratuberculosis* genome. IS900 is atypical in that, unlike classical IS elements, it lacks inverted terminal repeats and flanking direct repeats and is consequently classed in a family of IS elements that includes IS901/902 of *Mycobacterium avium* subsp. *silvaticum* (Kunze et al., 1991; Moss et al., 1992), IS1110 of *M. avium* (Hernandez Perez et al., 1994) and ISI16 of *Streptomyces clavuligerus* (Leskiw et al., 1990). IS900 is 1451 bp in length and has been reported to have a single ORF possessing the translation signals necessary for the expression of a protein of 399 aa (Green et al., 1989; Collins et al., 1989). Translation analysis of the IS900 sequence has revealed expression of this single protein, which was designated p43 (Tizard et al., 1992). This protein is highly similar to the putative transposase proteins of IS901/902, IS1110 and ISI16. Although IS900 lacks inverted terminal repeats and flanking direct repeats, it does display an insertion site specificity. The consensus insertion site for IS900 was suggested to be 5' CATGN(4–6)*CNCCCTT 3' (where the asterisk denotes the site of insertion). IS900 is present in 15–20 copies in the *M. avium* subsp. *paratuberculosis* genome and always inserts in the same orientation with respect to the target sequence (Green et al., 1989).

Murray et al. (1992) isolated a promoter sequence, P**AN**,
from *M. avium* subsp. *paratuberculosis*, located proximal to the 3' end of a copy of IS900. *P*$_{AN}$ could potentially drive expression of a second ORF from IS900, encoded on the complementary strand to the p43 gene. This second ORF (starting with an ATG) potentially drive expression of a second ORF from IS900 to facilitate expression of its protein product (Doran et al., 1994). This paper describes the unique process by which IS900 has recruited these transcription and translation signals in *M. avium* subsp. *paratuberculosis* and subsequently presents the detection of expression of this ORF in this bacterial host.

**METHODS**

Reverse transcription and PCR. *M. avium* subsp. *paratuberculosis* 0138 cells were grown in Dubos broth (Difco) containing 2 mg Mycobactin J 1⁻¹ (Rhone Merieux) and 20 % newborn calf serum ( Gibco). Cells were harvested by centrifugation and total RNA was isolated using the procedure described by Patel et al. (1991). Total RNA was reverse transcribed in a reaction volume of 50 μl containing 125 μM each of dATP, dCTP, dGTP and dTTP, 300 pmol of the bed-specific primer (P90, 5' GAAGGGTGTTCGGGG-CGGTCGCTTAG 3'), a buffer of 50 mM Tris/HCl, 75 mM KCl, 3 mM MgCl₂, and 10 mM DTT, and 5 units of avian myeloblastosis virus reverse transcriptase (Gibco). The RNA sample and primer in reaction buffer were heated for 15 min at 65 °C then chilled on ice before addition of the nucleotides and enzyme. The reaction was carried out at 42 °C for 60 min.

PCR reactions were performed in a Perkin Elmer Cetus DNA Thermal Cycler. The DNA polymerase (Promega) under the conditions defined by the manufacturer. Three-hundred nanomoles of each oligonucleotide (P25, 5' CCAGGGACGTCGGGTATGGC 3'; P26, 5' GGTCGG-CCTTACCAGCGGCCTG 3') was added to the reaction mix. PCR conditions were 35 cycles at 94 °C for 1 min, 50 °C for 2 min and 72 °C for 2 min. Reactions were analysed using agarose gel electrophoresis.

**Construction of the Hed expression vector.** A 1.5 kb EcoRI–BamHI fragment from the plasmid pN14 (Tizard et al., 1992) was cloned into the EcoRI–BamHI sites of pDEV19 (Tizard et al., 1992). This fragment contains the full sequence of IS900 with 27 bp of 5' flanking sequence (8 bp derived from *M. avium* subsp. *paratuberculosis* genome and 19 bp from pUC18) and 74 bp of 3' flanking sequence (66 bp from *M. avium* subsp. *paratuberculosis* and 8 bp from pUC18). This fused the *hed* coding sequence plus 11 extra codons from the 5' flanking sequence in-frame with the ATG initiation codon of lacZ on pDEV19, which is under control of the lac promoter. An oligonucleotide encoding 10 consecutive histidine codons was then inserted into the EcoRI site of this construct, again in-frame with the lacZ ATG and at the amino terminus of the *hed* sequence. The resulting plasmid was designated pTT3. Expression from pTT3 would result in a 508 aa protein beginning with the ATG from lacZ and 10 consecutive histidine residues.

**Expression and purification of recombinant H$_3$Hed protein.** Escherichia coli DH10B (Gibco) harbouring pTT3 were grown in Terrific Broth (Sigma) to an OD$_{590}$ of approximately 1.5 then induced for 16 h by addition of 0.1 mM isopropyl-β-D-thiogalactoside (NBL). Cells were harvested and a bacterial lysate prepared in TN buffer (10 mM Tris/HCl, pH 8.0, 50 mM NaCl) by treatment with 1 mg lysozyme ml⁻¹ (Sigma) for 1 h at 4 °C, followed by sonication. The lysate was then separated into soluble and insoluble fractions by centrifugation. The pellet containing inclusion bodies was dissolved in 5 M guanidinium thiocyanate in TN buffer.

Nickel ion affinity chromatography was performed on IDA chelating Sepharose FF (Pharmacia) charged with 50 mM NiSO$_4$ and equilibrated with 5 M guanidinium thiocyanate buffer. The *E. coli* extract in 5 M guanidinium thiocyanate was then applied and the column washed with 8 M urea, 60 mM imidazole in TN. Recombinant H$_3$Hed was then eluted from the column with 8 M urea, 1 M imidazole in TN buffer. Dialysis against decreasing concentrations of urea in TN resulted in a soluble H$_3$Hed sample, which was analysed by standard SDS-PAGE followed by Coomassie blue staining.

**Preparation of Hed-specific antibodies.** Computer analysis of the *hed* sequence was used to identify charged regions with turning structural characteristics (Staden-Plus, Amersham) to favour the selection of hydrophilic surface loops which may be antigenic. A region representing Arg$_{18}$-Asp$_{33}$, designated TD2 (RAPGNIKQLPSDEQGD), was synthesized using FMDC protection chemistry. An additional carboxy-terminal cysteine was added to allow coupling to solid phase. The peptide was purified by HPLC and validated by amino acid analysis and mass spectrometry. This reagent and purified H$_3$Hed were used to hyperimmunize New Zealand White rabbits. TD2 was coupled to EAH Sepharose 4B (Pharmacia) and used to affinity-purify anti-TD2 antibodies from immune sera. Likewise, H$_3$Hed was coupled to CNBr-activated Sepharose 4B (Pharmacia) and used to purify anti-H$_3$Hed antibodies from immune sera. Negative controls of pre-immune sera from both sets of rabbits were used in assays for the specific antibodies.

**Preparation of mycobacterial protein extracts.** *M. avium* subsp. *paratuberculosis* strains 0023, 0138, 0139, 0140, 0141 and 0142 were cultured in Dubos broth (Difco) containing 2 mg Mycobactin J 1⁻¹ (Rhone Merieux) and 20 % newborn calf serum (Gibco). *M. avium* subsp. *silvaticum* 0205 was grown in Dubos containing 2 mg Mycobactin J 1⁻¹ and 10 % DMA (150 mM NaCl), 5 % bovine serum albumin, 7.5 % dextrose). *M. avium* 0040 was also grown in Dubos, 10 % DMA, but without Mycobactin J. Cells were harvested by centrifugation and resuspended in 1 ml phosphate-buffered saline. Protein extracts were prepared by adding 0.5 g of 0.1 mm zirconium beads (Biospec) and bead beating for 5 x 1 min with a Biospec mini-beadbeater. Extracts were analysed by SDS-PAGE followed either by staining with Coomassie blue or by transfer to BAS nitrocellulose (Schleicher & Schuell) using a Sartoblot II semi-dry electroblotter (Sartorius) according to the manufacturer's protocol.

**Immunostaining of Western blots.** The purified antibody reagents were used at a working dilution of 1:50 for anti-H$_3$Hed and 1:10 for anti-TD2 in TBST (50 mM Tris/HCl, pH 7.6, 150 mM NaCl, 0.05 % Tween 20) using 10% dried skimmed milk as blocking agent. Second layer antibody was goat anti-rabbit Ig-horseradish peroxidase conjugate (Sigma) used at a 1:1000 dilution. Staining was revealed using ECL substrate (Amersham) combined with autoradiography according to the manufacturer's protocol.

**RESULTS AND DISCUSSION**

The *hed* gene on IS900

IS900 encodes a second ORF on the complementary strand to the p43 gene. This ORF (starting with an ATG 8 bp in from the end of the element) was initially
Expression of the IS900 *hed* gene

unreported as it requires host-encoded sequences adjacent to and outside IS900 to enable transcription and translation of a protein product. We have therefore designated this ORF as the *hed* (host expression-dependent) gene of IS900. A promoter, ribosome binding site (RBS) and termination codon are all required to be encoded by sequences that flank IS900. Intriguingly, it is the site-specific insertion event by IS900 that provides these sequences to facilitate *hed* expression. Fig. 1(a) details the complement sequence of the insertion and juncnon sites for six copies of IS900 in the *M. avium* subsp. *paratuberculosis* genome. A close analysis of each of these sequences indicates that IS900 inserts between the putative RBS and start codon of a given gene, thus aligning the *hed* initiation codon adjacent to a functional RBS and possibly downstream of an active promoter (e.g. P_AGL). If the insertion sites are aligned to derive a target consensus, this sequence is shown to comprise a RBS and start codon (ATG or GTG), correctly spaced and at either side of the insertion site (Fig. 1b). We therefore propose that IS900 specifically targets putative translation initiation sequences of host genes, so as to provide the necessary regulatory sequences for *hed* expression for each copy of IS900 on the *M. avium* subsp. *paratuberculosis* genome. A schematic representation of this process is shown in Fig. 2.

A termination codon for *hed* is not generated by the target site and thus will vary in position for each genomic location of an IS900 element. Consequently, expression of *hed* could result in translation products either in- or out-of-frame with the ORF of the putative *M. avium* subsp. *paratuberculosis* gene that has been targeted by IS900 (Fig. 2). Of the target sequences listed in Fig. 1, the *hed* sequence would be out-of-frame with the downstream initiation codons in two cases (pMB22, pMBJ3) and in-frame in the remaining four (pMB15, pMB55, pMBJ2, pDL2). It is known that the out-of-frame fusion on pMB22 terminates almost immediately with a TGA sequence, where the T and G residues are supplied by the initiation codon of the targeted gene (Doran et al., 1994).

Once IS900 has targeted and disrupted the translation initiation signals of a given gene, it is possible that it then provides this gene with a replacement RBS (AAGGA) at the 3' terminus of the element (Fig. 1a). If a polycistronic

---

**Fig. 1.** IS900 insertion sites in *M. avium* subsp. *paratuberculosis*. Sequence data from Green et al. (1989) and Moss et al. (1992). (a) Insertion and junction sites for six copies of IS900. The IS900 sequence is italicized and underlined, with 5' and 3' ends of the element indicated with asterisks. Putative target-encoded and IS900-encoded RBSs and initiation codons are in bold. (b) Alignment of insertion sites to derive a target consensus sequence. Putative *M. avium* subsp. *paratuberculosis* gene RBSs and initiation codons are in bold. *E. coli* and mycobacteria translation initiation consensus sequences taken from Dale & Patki, (1990). The asterisk denotes the site of IS900 insertion.

**Fig. 2.** IS900 targets translation initiation signals in *M. avium* subsp. *paratuberculosis*. IS900 is represented by a shaded box and a putative target gene by a shaded arrow. The target gene RBS and potential replacement RBS on IS900 are shown as black ovals. The *hed* and target gene initiation codons are shown as black diamonds. The arrows (a), (b) and (c) represent the possible translation outcomes resulting from IS900 insertion between the RBS and start codon of a target gene. The locations of primers P90, P25 and P26 within IS900 are shown.
mRNA is transcribed from an upstream promoter, then perhaps translation of the disrupted gene initiates from this putative RBS (Fig. 2). This provides the novel scenario of an IS element actively ‘hijacking’ promoter and RBS signals for its own expression needs and then restoring function to the disrupted gene, assuring minimal insertional damage to the host.

Other elements of the IS900 family also appear to specifically target translation initiation sequences in their respective hosts (Doran et al., 1994; Hernandez Perez et al., 1994) although only IS901/902 and IS116 encode a hed homologue. Within IS1110 there is an in-frame stop codon that would terminate translation of this ORF after only 12 aa (Hernandez Perez et al., 1994). Unlike IS900, IS901/902 and IS116 encode a termination codon for their hed homologues, which are 477 and 413 aa in size, respectively. Alignment of these respective homologues with the IS900 Hed sequence revealed overall sequence identities of 35 and 38%, respectively.

Given that the hed gene is structurally capable of encoding a protein, there are several lines of evidence that suggest it is indeed expressed in M. avium subsp. paratuberculosis. Firstly, the potential coding region is very large (> 1444 bp), with codon usage revealing a bias complying with other mycobacterial genes (Dale & Patki, 1990). Secondly, a DNA fragment containing the M. avium subsp. paratuberculosis promoter P
\(_A\)N and part of the hed gene was fused to lacZ and expressed in Mycobacterium bovis BCG (Murray et al., 1992). This indicates that in M. avium subsp. paratuberculosis, at least one hed sequence has procured an upstream promoter capable of driving expression of a polypeptide encoded by this ORF. Finally, in vitro transcription–translation studies of an IS900 element cloned into a pUC18 vector detected expression of Hed (M. L. V. Tizard, unpublished). The DNA fragment encoding IS900 was derived from plasmid pMB22 (Green et al., 1989) and contains sufficient flanking sequence to provide the putative RBS and termination codon for its hed gene. The predicted size of Hed encoded on this plasmid is approximately 55 kDa and the in vitro expression of a protein this size was detected.

**Detection of hed mRNA in M. avium subsp. paratuberculosis**

To confirm that hed is indeed expressed, we studied its expression within M. avium subsp. paratuberculosis at the level of transcription. Total RNA was extracted from M. avium subsp. paratuberculosis, cDNA was reverse transcribed from a primer within IS900 and then used as a template for PCR. Analysis of the products of this reaction on an agarose gel identified a specific hed fragment of the correct 228 bp size (Fig. 3). Controls were performed to demonstrate that the PCR fragment was amplified from an mRNA starting template and not contaminating DNA. RNase H was added to one reaction tube prior to reverse transcription and PCR, or alternatively, reverse transcriptase was not included in one reaction tube. As predicted, analysis of the PCR products on an agarose gel failed to detect an amplified fragment (Fig. 3). These results suggest that the hed PCR fragment was amplified from a cDNA template, reverse transcribed from mRNA. PCR detection of a mRNA specific to hed indicates that this gene is transcribed in M. avium subsp. paratuberculosis.

**Detection of Hed in M. avium subsp. paratuberculosis**

Further confirmation of hed expression was obtained by using two Hed specific polyclonal antibodies to probe Western blots of six M. avium subsp. paratuberculosis protein extracts (Fig. 4). The first, anti-H,\(_A\) Hed, was raised against recombinant Hed expressed and purified from E. coli; the second, anti-TD2, was raised against a Hed-specific synthetic peptide. Both polyclonal antibodies reacted specifically with several proteins in M. avium subsp. paratuberculosis, in particular a protein of approximately 55 kDa, but also several larger proteins that are putative Hed translational fusion products. As previously mentioned, a termination codon for hed is not encoded on IS900 or within its target sequence and thus will vary in position for each copy of IS900. Hence, hed expression could result in either an in-frame or out-of-frame protein fusion with the M. avium subsp. paratuberculosis gene upstream of which IS900 has inserted. If an out-of-frame hed fusion was to terminate promptly (e.g. pMB22), the predicted translation product would correlate approximately in size to the predominant 55 kDa protein in each of the M. avium subsp. paratuberculosis extracts. We propose that this major protein band represents at least one, and possibly more of these Hed translation products. Alternatively, this may represent one copy of IS900 which has inserted downstream of a stronger promoter, over-expressing just one Hed species.

Since IS900 is present in 15–20 copies in M. avium subsp. paratuberculosis, theoretically this number of Hed
fusión proteinas pueden ser expresadas. Sin embargo, aproximadamente 2-5 proteínas dentro de cada M. avium subsp. paratuberculosis strain pueden reaccionar específicamente con la antígeno Hed. Como se mencionó anteriormente, el major 55 kDa protein band puede constituir más de lo que una terminación prompte Hed translation product. It is also possible that of all the promotors procured by IS900 for Hed expression, only promoters driving transcription of these proteins are active during the in vitro growth of M. avium subsp. paratuberculosis. Alternatively, other Hed fusions may be expressed but in amounts too low to be observed at the level of detection for these Western blots.

We have expressed and purified a recombinant form of Hed in E. coli. It is hoped that this material, in conjunction with recombinant p43, will facilitate an investigation of the molecular mechanism of IS900 transposition. p43 is the predicted transposase (Tizard et al., 1992) and as IS900 actively targets translation initiation signals for Hed expression, we believe that Hed will also play an important role in IS900 transposition.

Fig. 4. Detection of Hed by Western blotting and immunostaining with anti-Hed (a) or anti-TD2 (b) of M. avium subsp. paratuberculosis protein extracts. Lanes: 1, positive control, E. coli pPT3 expressing recombinant Hed; 2-7, M. avium subsp. paratuberculosis strains 0023, 0138, 0139, 0140, 0141 and 0142, respectively; 8, M. avium subsp. silvaticum 0205; 9, M. avium 0040.

ACKNOWLEDGEMENTS

This work was supported by grants from Action Research, The Dinwoodie Trust and The Wellcome Trust, to whom we express our appreciation. We are also grateful to Professor Brian Austen for producing the peptide TD2.

REFERENCES


of a repetitive sequence specific to Mycobacterium paratuberculosis. FEMS Microbiol Lett 60, 175–178.


Received 27 June 1996; revised 10 September 1996; accepted 25 September 1996.