p43, the protein product of the atypical insertion sequence IS900, is expressed in *Mycobacterium paratuberculosis*

MARK L. V. TIZARD,* MICHAEL T. MOSS, JEREMY D. SANDERSON, BRIAN M. AUSTEN and JOHN HERMON-TAYLOR

Department of Surgery, St George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK

(Received 28 February 1992; revised 13 April 1992; accepted 24 April 1992)

The novel mycobacterial insertion sequence IS900 was analysed by coupled transcription–translation, of both strands independently, in a cell-free *E. coli* extract using an exogenous promoter. This revealed only one protein product, p43, as predicted from the nucleotide sequence. The protein was readily translated in recombinant *E. coli*, using the *tac* promoter, though it did not appear as a major product by SDS-PAGE analysis. A synthetic peptide was used to generate and affinity-purify a specific anti-p43 antibody, which clearly identified the protein in recombinant *E. coli*. p43 was relatively stable in exponential phase and stationary phase bacteria, though a 28 kDa processed form was seen to accumulate over a period of hours. Both forms appeared in the soluble fraction of the bacterial lysate. The anti-p43 antibody also identified p43, as a 28 kDa processed product, in Western blots of protein extracts from *Mycobacterium paratuberculosis*, indicating a level of expression which would be unusually high for a classical transposase. These data have important implications for the relationship between IS900 and its host.

**Introduction**

Diagnosis of *Mycobacterium paratuberculosis* infection in ruminants, causing the chronic enteritis Johne's disease, is of veterinary and economic importance. This is also of relevance to research into the cause of a similar chronic enteritis Crohn's disease in humans (Hermon-Taylor et al., 1990). Recent research into Johne's disease and Crohn's disease has utilized DNA probes based on IS900 (Moss et al., 1991; Vary et al., 1990; Sanderson et al., 1992) and ELISAs based on mixed bacterial antigen (Tanaka et al., 1990; Collins et al., 1991). The gene product of IS900, p43, represents the first sequence-defined specific antigen from *M. paratuberculosis* and as such has a potential use in the diagnosis of Johne's disease. In addition, IS900 shows some unusual features for an insertion sequence, the basis of which may be revealed by studying the functions of p43.

Insertion sequences (IS) are small units of DNA (approximately 700–2500 bp) that are usually able to transpose from one genetic locus to another, on the same replicon or between replicons. IS900 was the first IS element identified in a mycobacterium, from *M. paratuberculosis* (Green et al., 1989). Other elements were later identified in *M. tuberculosis* (Hermans et al., 1990; Thierry et al., 1990), *M. leprae* (Clark-Curtiss & Docherty, 1989) and *M. fortuitum* (Martin et al., 1990). Sequence analysis of the IS elements identified in *M. tuberculosis* and *M. fortuitum* showed that they were related to the IS3 and IS6 groups, respectively, found in Gram-negative organisms, and both possessed the classical terminal inverted and direct repeat sequences. Two other elements have recently been identified in *M. smegmatis*, IS6120 and IS1096, both of which have been completely sequenced and possess terminal inverted repeats and contain two overlapping open reading frames (Guilhot et al., 1992; Cirillo et al., 1992).

IS900 is 1451 bp in length with a single open reading frame, ORF1197, from nucleotides 236 to 1433 starting with a GTG initiation codon and proceeded at nucleotide 226 by a Shine–Dalgarno (S–D) sequence GGAGG (Green et al., 1989). IS900 is unusual in that it does not possess either of the terminal repeat structures which are characteristic of IS elements. This is significant because inverted repeat structures are known to be functionally important in transposition events (Huang et al., 1986; New et al., 1988). Specific interaction between established transposase proteins and their respective terminal inverted repeat sequences has been demonstrated for transposons Tn3 (Wishart et al., 1985), γδ (Wiater &
Grindley, 1991) and Tn10 (Morisato & Kleckner, 1984). Another unusual feature of IS900 is that it has a specific target sequence into which it inserts in a specific orientation (Green et al., 1989). Tn554, from Staphylococcus, also shows these characteristics, including the lack of terminal repeats (Murphy, 1988); it is, however, much larger and more complex, encoding three proteins (14, 43 and 74 kDa) essential for transposition. Another transposon exhibiting orientation and site-specific insertion is Tn7, from E. coli. Transposition of this element has been shown to require the formation of a nucleoprotein complex involving two molecules of DNA and at least four transposon-encoded proteins (TnsA 31 kDa, TnsB 81 kDa, TnsC 62 kDa and TnsD 59 kDa) and probably one or more host-encoded products (Bainton et al., 1991). These characteristics are found only in a few mobile elements which, with the exception of IS900 (small and encoding only one 43 kDa protein), are all large and complex. The IS900 protein (possibly in concert with host-encoded factors) appears to function in an analogous manner to the components encoded by the more complex transposons Tn7 and Tn554.

An engineered mycobacterial site-specific recombination system has recently been used to mediate stable integration of foreign genes into M. bovis BCG, exploiting the adjuvant activity of this organism to create novel vaccine strains (Stover et al., 1991). The demonstration that IS900 can promote integration of an antibiotic resistance marker in mycobacteria highlights the potential for development of this system (England et al., 1991). A defined recombination system, with the unique characteristics exhibited by IS900, would be another useful tool for the molecular genetic manipulation of mycobacteria. Definition of the mechanism controlling transposition of IS900 is essential for construction of vectors analogous to mini-Tn10 (Petit et al., 1990), which can mediate controllable, terminal integrations. To investigate this mechanism we have studied translation of the single open reading frame in vitro and in vivo, establishing a source of recombinant protein for purification and functional analysis.

Methods

Mycobacteria. Four M. paratuberculosis strains were isolated from cattle, a deer and a goat suffering chronic enteritis (kindly supplied by Dr W. Donachie, Morehead Institute, Edinburgh, UK). M. avium subsp. silvaticum strains were isolated from deer suffering chronic enteritis. M. avium strains were isolated from an AIDS patient suffering intestinal mycobacteriosis. The latter were cultured in Middlebrook 7H9 medium supplemented with 0-02% glycerol and ADC (Difco) at 30 °C with shaking. The former were cultured under the same conditions in media supplemented with 1 μg mycobactin J (Rhone Merieux) ml⁻¹.

Recombinant DNA methods. All plasmids were maintained in E. coli XL-1 (Stratagene). Recombinants were selected and maintained by growth on Luria–Bertani medium agar with 100 μg ampicillin ml⁻¹, at 37 °C. Plasmid DNA for restriction analysis and in vitro transcription-translation was isolated by alkaline lysis coupled with purification by the Qiagen method (Qiagen GmbH).

Conventional DNA manipulations were performed as described by Sambrook et al. (1989). DNA fragments were purified from agarose gels for use in ligation using Gene Clean (Bio 101 Inc.). All restriction enzymes and DNA modification enzymes were used according to the manufacturer's specifications (NBL).

Construction of plasmids. A 1·5 kbp NruI fragment from the plasmid pMB22 containing IS900 (McFadden et al., 1987) was cloned into the Smal site of pUC18. This fragment contains the full sequence of IS900 with 66 bp of 5' flanking sequence and 8 bp of 3' flanking sequence derived from the M. paratuberculosis genome (Green et al., 1989). This generated two orientations of the fragment: pN12 with the coding strand for ORF1197 transcribed from the lac promoter, and pN14 with the antisense strand for ORF1197 transcribed from the lac promoter (Fig. 1).

In other constructs, coding sequences were placed under control of the lac promoter in the vector pDEV19 (a gift from Dr Rayment, Wellcome Research Laboratories, Beckenham, UK), which carries ampicillin resistance, lacZ with an EcoR1–BamH1 cloning site at the amino-terminus and lacP. The 5' non-coding region of IS900 was deleted with BstS1 nuclease prior to cloning. Two clones were selected: pDI13 representing a transcriptional fusion of ORF1197 to Pₘ₉₅, and pD19 representing a translational fission of the first 10 codons of lacZ to codon 8 of ORF1197.

Fig. 1. Circular maps of the plasmids pN12 and pN14, which contain IS900, showing the orientations of the open reading frame ORF1197 with respect to the lac promoter in the vector pUC18.
coli extract (Amersham): 0.5-1.0 mg of template was used and translation followed by incorporation of [35S]methionine (ICN) according to the manufacturer's protocol. Proteins were analysed by SDS-PAGE (Laemmli, 1970) on a 10% (w/v, acrylamide) gel, which was dried and subject to autoradiography.

Preparation of bacterial protein extracts. Recombinant E. coli were grown to an OD_{550} of 0.3 then induced by addition of isopropyl β-D-galactopyranoside (NBL) to 120 mg ml^{-1} final concentration. Cultures were sampled at various intervals. Cells were harvested and lysed with lysozyme (Sigma) then treated with DNaseI (Sigma). Protein extracts were analysed by SDS-PAGE followed either by staining with Coomassie brilliant blue or by transfer to BAS nitrocellulose (Schleicher and Schuell) using a SartoblotII semi-dry electroblotter (Sartorius) according to the manufacturer’s protocol.

Mycobacterial protein extracts were prepared by boiling the harvested cell pellet in SDS loading buffer (1:5%, w/v, SDS, 0.18 M-Tris/HCl, pH 6.5, 15% v/v, glycerol, 10 mM-DTT, 0.05% bromophenol blue) followed by sonication for 1 min to shear released DNA. Extracts were analysed as described above.

Preparation of p43-specific anti-peptide antibody. Computer analysis of ORF1197 was used to identify charged regions with turning structural characteristics using Staden-Plus (Amersham) to favour the selection of the putative transcriptional control region of IS900. The putative transcriptional control region of IS900 (Fig. 3). The full insert from pN12 was cloned into the pUC18 vector to give pDI12, designated CP-1 (YRGEGKTDAKDAHAD) was synthesized using FMOC 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 was used to hyperimmunize two New Zealand White rabbits. CP-1 was coupled to Sepharose CL-4B (Pharmacia). The immobilized peptide was used to affinity-purify anti-peptide antibodies from immune sera with ELISA titres greater than 1:3000. Pre-immune sera from both rabbits were negative when assayed by ELISA, and failed to recognize any specific bands in Western blots of protein extracted from recombinant E. coli or M. paratuberculosis.

Immunostaining of Western blots. The purified antibody reagent was used at a working dilution of 1:2000 in TBST (50 mM-Tris/HCl, pH 7.6, 150 mM-NaCl, 0.05% Tween 20), using 10% (w/v) dried skimmed milk as blocking agent. The second layer antibody was goat(anti-rabbit IgG)-horseradish peroxidase conjugate (Sigma) used at a 1:1000 dilution. Staining was revealed with hydrogen peroxide and 3-amino-9-ethylcarbazole (Sigma), or using ECL substrate (Amersham) combined with autoradiography according to the manufacturer’s protocol.

Results

Transcription–translation of both strands of IS900

Coupled in vitro transcription–translation of the plasmid template pN12 (Fig. 2, lane 2) suggested that the predicted S-D sequence and open reading frame ORF1197 (Green et al., 1989) were functional in an E. coli 30S cell free extract. Migration of the protein product, designated p43, indicated an apparent molecular mass of 44.5 kDa. No specific proteins could be identified in transcription–translation of the opposite strand, from the plasmid template pN14 (Fig. 2, lane 3). A long open reading frame runs antiparallel to ORF1197; however, it lacks S-D sequences adjacent to initiation codons. The molecular mass of pUC18-vector-encoded products (Fig. 2, lane 1) does not coincide with the predicted molecular mass of potential products of the antiparallel open reading frame.

Regulation of in vitro expression of p43

The putative transcriptional control region of IS900 upstream of ORF1197 was deleted using Bal31 nuclease. pDI13 represented the S-D sequence and codons 1–399 (nucleotides 200–1451) inserted downstream and out of frame with the vector-encoded S-D sequence and ATG initiation codon (Fig. 3). pDI9 represented a fusion of the S-D sequence and codons 1–399 out of frame with the vector-encoded S-D sequence and ATG initiation codon (Fig. 3). The full insert from pN12 was cloned into pDEV19 to give pDI12. These four constructs were analysed for comparative levels of expression of p43 and β-lactamase, as an internal control (Fig. 4). The results of scanning densitometry on autoradiographs from a number of experiments (Table 1) show that the tac
promoter gives a significantly higher level of expression of p43 than the lac promoter. Comparing expression of p43 between pDI12 and pDI13 shows that the presence or absence of the region of IS900 (nucleotides 1–199) upstream of the predicted ribosome binding site (RBS) has no significant effect. However, replacement of the p43 translation signals, the RBS and the first seven codons including the GTG initiation codon, in pDI9 with optimized E. coli translation signals increased the level of expression threefold.

With the stronger tac promoter, expression yields an additional band which migrates slightly faster than p43, corresponding to a lower molecular mass with a difference of $<1.0$ kDa. This product represents approximately 50% of the higher molecular mass product and is produced by both pDI12 and pDI13. The primary product of pDI9 is of slightly higher molecular mass than p43 (due to the amino-terminal modification), there is a second product with an equivalent 1.0 kDa molecular mass difference (approximately 20% of the primary product) and a third product with an apparent molecular mass slightly lower still (approximately 30% of the primary product). There are no potential signal sequences at the amino-terminus of p43 which might explain these observations.

**Over-expression of p43 in vivo**

SDS-PAGE analysis of IPTG-induced bacterial lysates of hosts carrying pDI13 (Fig. 5a) shows a protein with a molecular mass of approximately 45 kDa accumulating during a time-course of up to 2 h with no product present in induced lysates from hosts containing the vector (pDEV19) only. Coomassie staining of this gel was carried out after a partial electro-transfer as the band of interest was obscured by a host band of similar molecular mass, but which electro-transferred at a higher rate (data not shown).

**Anti-peptide antibody binds p43**

Western blot analysis of the recombinant E. coli cell lysates (Fig. 5b) shows staining p43 with the monospecific antibody. This result confirmed the reading frame and identity of p43, even though the protein migrates with an apparent molecular mass of 45 kDa. This disparity is almost certainly due to the high basic character of p43 (predicted pI 10.4), which may retard its migration during SDS-PAGE and clearly reduces its rate of electro-transfer to nitrocellulose.

Of the p43 expressed in E. coli, 80–90% was found in the soluble fraction of the bacterial lysate (data not shown). p43 continued to accumulate to a maximum up to 5 h after induction of a mid exponential phase culture,
Fig. 5. Detection of recombinant p43 expression. Protein extracts from recombinant E. coli carrying PDI13 or pDEV19 were analysed by SDS-PAGE. Samples were taken at 0, 30, 60 and 120 min after induction of expression. (a) Coomassie-stained gel after partial electrotransfer; (b) immunostain with antipeptide antibody after transfer to nitrocellulose; the HRP-conjugate second-layer antibody was revealed by reaction with AEC.
and the majority was intact in cultures 22 h after induction. A specific cleavage product was observed at approximately 28 kDa, which first appeared between 1 and 2 h post-induction and was recognized by the antibody. This accumulated to a maximum (at 5 h) of approximately 28 kDa, which first appeared between 1 and 2 h post-induction and was recognized by the antibody. This product co-migrated with the major cleavage product of the recombinant molecule.

**p43 in M. paratuberculosis**

The anti-CP-1 antibody reacted specifically with a 28 kDa protein present in protein extracts of four strains of *M. paratuberculosis* on Western blot (Fig. 6, lanes 1–4). This product co-migrated with the major cleavage product of the recombinant *p43* from the positive control *E. coli* (lane 11). In the extract of *M. paratuberculosis* in lane 4, a faint but sharp band was observed at 45 kDa (particularly on longer exposures); this was interpreted with caution since there was a background smear in this region in the IS900-free *M. avium* (Fig. 6, lanes 7 and 8), though no distinct bands. The *M. avium* and *M. avium* subsp. *silvaticum* extracts did not show clearly positively staining bands at either 45 kDa or 28 kDa.

**Discussion**

Coupled transcription–translation of both strands of IS900 (Fig. 2) showed only one identifiable product, *p43*, corresponding to the predicted molecular mass of the single open reading frame. When the level of transcription was increased by coupling to the *tac* promoter, an additional band was resolved with a slightly lower apparent molecular mass (Fig. 4). This was observed in all experiments and suggested a post-translational modification of the polypeptide in vitro. Though the doublet was not resolved in Fig. 5 or Fig. 6, it was frequently observed in recombinant *E. coli* in cultures induced for 2 h or longer (data not shown). The nature of the two components of the doublet is currently under investigation. The reaction of the specific antibody, generated against a synthetic peptide, with *p43* on Western blots (Fig. 5b) confirmed the correct interpretation of the sequence and assignment of the open reading frame ORF1197 (Green et al., 1989). However, the affinity-purified polyclonal antibody did not recognize the native form of *p43* produced in vitro or in vivo. This suggests that the region may be blocked by secondary structure or buried in a pocket or groove inaccessible to an antibody molecule.

As the only product of an insertion element, *p43* will almost certainly be responsible for transposition of IS900, possibly with the involvement of host cofactors. Detection, by Western blotting, of an immunoreactive 28 kDa protein in all disease isolates of *M. paratuberculosis* suggests that the IS900 transposase may be equivalent to the specific cleavage product seen in recombinant *E. coli* (Fig. 6), indicates an unusually high level of expression for a putative transposase. The identity of these two 28 kDa polypeptides is under investigation. Failure to detect the full-length product is probably due to the slow growth rate of *M. paratuberculosis* and the consequent need to culture for 4–6 weeks in order to obtain sufficient inoculum for protein analysis. The transposase encoded by IS10 has an experimentally determined abundance of 0.15 polypeptides per cell per generation (Raleigh & Kleckner, 1986).
The transposase of Tn3 was only detectable by SDS-PAGE in mutants showing a 104-fold increase in translation (Casadaban et al., 1982). As a p43 cleavage product can be detected by Western blotting, it must be expressed at a level several orders of magnitude higher than other transposases. Despite this, transposition of IS900 has not been observed within a strain of M. paratuberculosis.

Restriction patterns from various strains of M. paratuberculosis isolated in Europe and North America are very similar with only minor variations, which may be accounted for by point mutations affecting individual restriction sites. Some strains have an additional copy of the IS900 element (Levy-Frebault et al., 1989; Whipple et al., 1990). The target-site specificity of IS900 for the sequence CATG(N)CNCCCTT may account for the apparent immobility of this element (Green et al., 1989). Assuming a random distribution of bases with 66% G+C content, this sequence would be expected to occur 15–35 times (with a confidence limit of 95%) in the M. paratuberculosis genome, reported as 3.9 Mbp (McFadden et al., 1987b). The observed copy number of approximately 18–20 (Green et al., 1989) falls in this range. It is conceivable that all the available sites are occupied, as would be expected if the transposase were constitutively expressed. The extra copy observed in some strains of M. paratuberculosis (Whipple et al., 1990) may be due to a fresh target-site being created through point mutation and then immediately filled. Alternatively, the 28 kDa form of p43 may be inactive and the full-length active form may be processed before it can mediate a transposition.

Whether expression of p43 in M. paratuberculosis derives from a promoter within IS900 has not been determined. The results of in vitro transcription–translation indicate that there are no functional E. coli transcription termination signals in the 5’ non-coding region (Fig. 4). If there are also no mycobacterial transcription termination signals, then one or more copies of IS900 inserted in an active operon in M. paratuberculosis would result in expression regulated by the host, which may well be constitutive. It is known, however, that there are signals in mycobacteria which may not function in a species as distant as E. coli (Kieser et al., 1986).

Purification of p43 will provide material to assess its potential as a diagnostic antigen in Johne’s disease and for the investigation of potential immune responses in Crohn’s disease. In addition, it will facilitate investigation of the mechanism of transposition of IS900. This may provide a useful tool for the genetic manipulation of M. bovis BCG, which has potential as a recombinant vaccine carrier (Stover et al., 1991; Aldovini & Young, 1991).

We would like to thank Dr Liz Green for her contribution to the early part of this work. We would also like to thank Dr Fred Rayment (Wellcome Research Laboratories, Beckenham, UK) for supplying the vector pDEV19 and Drs Carolyn Hyde and Ahmed Jehanli for producing the peptide CP-1 and the immobilized peptide. We are grateful to Dr Hazel Griffiths for critical reading of this manuscript. This work was supported by grants from the Wellcome Trust, Action Research and the Ileostomy Association of Great Britain and Ireland, to whom we express our appreciation.

References


restriction-fragment-length polymorphisms to examine the Mycobacterium avium complex. Molecular Microbiology 1, 283–291.


