Identification of *Mycobacterium tuberculosis* signal sequences that direct the export of a leaderless \(\beta\)-lactamase gene product in *Escherichia coli*


Author for correspondence: Mario R. W. Ehlers. Tel: +27 21 40 66 33 5. Fax: +27 21 47 76 69. e-mail: mehlers@physio.uct.ac.za

Proteins secreted by *Mycobacterium tuberculosis* may play a key role in virulence and may also constitute antigens that elicit the host immune response. However, the *M. tuberculosis* protein export machinery has not been characterized. A library of *M. tuberculosis* H37Rv genomic DNA fragments ligated into a signal sequence selection vector that contained a leaderless \(\beta\)-lactamase gene and an upstream Tac promoter was constructed. Transformation of *Escherichia coli* with the *M. tuberculosis* DNA library and selection on plates containing 50–100 \(\mu\)g ampicillin ml\(^{-1}\) resulted in the identification of 15 Amp\(^{+}\) clones out of a total of 14 000 transformants. Twelve of the \(\beta\)-lactamase gene fusions conferred high levels of Amp\(^{+}\) (up to 1 mg ampicillin ml\(^{-1}\)); insert sizes ranged from 350 to 3000 bp. Of ten inserts that were completely sequenced, two were identified as fragments of the genes for *M. tuberculosis* antigens 85A and 85C, which are the major secreted proteins of this pathogen. Seven of the remaining inserts were \(\geq 97\%\) identical to hypothetical ORFs in the *M. tuberculosis* genome, one of which encoded a protein with 35\% identity to a low-affinity penicillin-binding protein (PBP) from *Streptomyces clavuligerus*. Four of the seven hypothetical ORFs encoded putative exported proteins with one or more membrane interaction elements, including lipoprotein attachment sites and type I and II transmembrane (TM) segments. All of the inserts encoded typical signal sequences, with the exception of a possible type II membrane protein. It is concluded that expression of \(\beta\)-lactamase gene fusions in *E. coli* provides a useful system for the identification and analysis of *M. tuberculosis* signal-sequence-encoding genes.

**Keywords**: tuberculosis, *Mycobacterium*, signal sequences, \(\beta\)-lactamase, export

**INTRODUCTION**

*Mycobacterium tuberculosis* is an aggressive intracellular pathogen that preferentially invades mononuclear phagocytes and establishes persistent intracellular infections that are difficult to eradicate (Fenton & Vermeulen, 1996). There is evidence that these events are accompanied by the active secretion of proteins and other molecules by the pathogen. Comparative analyses by two-dimensional electrophoresis of proteins produced by intracellular versus broth-grown bacteria have revealed distinct differences, with intracellular bacteria producing proteins not evident during *in vitro* growth, some of these proteins being secreted (Lee & Horwitz, 1995). In addition to these specialized proteins, *M. tuberculosis* sheds a large array of proteins regardless of growth conditions, as revealed by extensive analyses of

**Abbreviations:** MCS, multiple cloning site; PBP, penicillin-binding protein; TM, transmembrane.

The GenBank accession numbers for the sequences reported in this paper are AF017098 (plasmid pGB14T-P), AF017099 (pGB14T-V), AF017100 (pGB14T-X), AF017101 (pGB14T-O1), AF017102 (pGB14T-O2), AF017103 (pGB14T-B) and AF017103 (pGB14T-O1).
culture filtrates (Andersen, 1994; Sonnenberg & Belisle, 1997). Many of these are true secreted proteins (Wiker et al., 1991) and are antigenic in vivo (Orme et al., 1993), and some may prove of value in the design of subunit vaccines (Andersen, 1994; Horwitz et al., 1995). However, at least two issues regarding secreted M. tuberculosis proteins remain unresolved. First, what roles, if any, do secreted proteins play in pathogenesis, particularly as virulence factors in cell invasion and intracellular survival? Second, what is the nature of the protein export machinery responsible for the secretion of these proteins?

Three types of protein export pathway have been identified in eubacteria. The best studied is the classical, sec-dependent, general secretory pathway that exports proteins with characteristic N-terminal signal peptides and that is probably the major protein export pathway in both Gram-negative and Gram-positive bacteria (Pugsley, 1993; Salmond & Reeves, 1993). This pathway is also referred to as the type II secretion system. In addition, subsets of secreted proteins are exported by type I and type III secretion systems, which are sec- and signal-sequence-independent (Salmond & Reeves, 1993).

The type III system is of particular interest because of its involvement in the export of virulence-associated proteins in diverse bacterial pathogens (Salmond & Reeves, 1993; Dreyfus et al., 1993; Silhavy, 1997). To date, none of these pathways has been characterized in mycobacteria.

The classical sec-dependent (type II) pathway is presumed to exist in mycobacteria because a number of secreted proteins have been found to be the products of genes that code for typical N-terminal signal sequences (Young et al., 1992; Orme et al., 1993). Moreover, recently a variety of signal-sequence-encoding genes have been identified by fusions with a leaderless alkaline phosphatase (phoA) gene and expression in a suitable mycobacterial host (Lim et al., 1995). In this study, we have extended this approach by investigating the usefulness of gene fusions with a leaderless β-lactamase gene and expression in Escherichia coli. Selection for export-competent signals by both PhoA and β-lactamase fusions depends on the principle that the enzyme must be translocated across the cytoplasmic membrane into the periplasm; PhoA is inactive in the cytoplasm, and cytoplasmic β-lactamase cannot protect cells against...
ampicillin-induced lysis (Broome-Smith et al., 1990). However, β-lactamase gene fusions have been shown to provide an informative alternative to pgb14T or pgb14TC was grown on LB-agar plates containing 200 μg/ml lactamase gene fusions was performed on LB-agar plates. pGB14 was a gift from Jan M. van Dijl (Groningen Biomolecular Sciences and Biotechnology Institute, Haren, The Netherlands); pGB14T and pGB14TC were generated as described below. pGEX2T was from Pharmacia and Signal peptide motif and cleavage site prediction, Genestream ALIGN

**Table 1. Internet servers used in the analysis of *M. tuberculosis* DNA–β-lactamase gene fusions**

<table>
<thead>
<tr>
<th>Server</th>
<th>Description</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Database searches</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCBI BLAST</td>
<td>National Center for Biotechnology Information; gapped BLAST/BLASTN/ TBLASTN searches of all non-redundant peptide and nucleotide sequence databases</td>
<td><a href="http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?form=0">http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?form=0</a></td>
</tr>
<tr>
<td>TIGR BLASTN</td>
<td>BLAST searches of Sanger and TIGR contigs</td>
<td><a href="http://ziggy.sanbi.ac.za/tb/tbsearch.htm">http://ziggy.sanbi.ac.za/tb/tbsearch.htm</a></td>
</tr>
<tr>
<td>Sanger Centre M.tb.</td>
<td>BLAST searches of Sanger Centre TB database</td>
<td><a href="http://www.sanger.ac.uk/Projects/M_tuberculosis/blast_server.shtml">http://www.sanger.ac.uk/Projects/M_tuberculosis/blast_server.shtml</a></td>
</tr>
<tr>
<td>MycDB BLAST</td>
<td>BLAST searches of Integrated Mycobacterium Database (Institut Pasteur)</td>
<td><a href="http://kiev.physchem.kth.se/MycDB/seqsearch_form.html">http://kiev.physchem.kth.se/MycDB/seqsearch_form.html</a></td>
</tr>
<tr>
<td><strong>Sequence analysis tools</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPred</td>
<td>TM domain predictions; under ExPASy tools</td>
<td><a href="http://ulrec3.unil.ch/software/TMPRED_form.html">http://ulrec3.unil.ch/software/TMPRED_form.html</a></td>
</tr>
<tr>
<td>ExPASy compute MW</td>
<td>Molecular weight predictions; under ExPASy tools</td>
<td><a href="http://expasy.hcuge.ch/ch2d/pi_tool.html">http://expasy.hcuge.ch/ch2d/pi_tool.html</a></td>
</tr>
<tr>
<td>SignalP</td>
<td>Signal peptide motif and cleavage site prediction</td>
<td><a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a></td>
</tr>
<tr>
<td>Genestream ALIGN</td>
<td>Sequence alignment tool</td>
<td><a href="http://genome.eerie.fr/bin/align-guess.cgi">http://genome.eerie.fr/bin/align-guess.cgi</a></td>
</tr>
</tbody>
</table>
et al., 1989), or modified as described previously (Da Silva Tatley & Steyn, 1993). All DNA sequencing was performed on double-stranded plasmid DNA using the Sequencing protocol (version 2.0, USB), or by automated sequencing on an ALF Express Sequenator. M. tuberculosis DNA inserts were sequenced directly in plasmids pGB14T or pGB14TC using primers Tac1B (forward) and Bla (reverse, complementary to 5′ end of TEM β-lactamase, 5′-CAGCATCCTTATCCTTTC-3′), or after subcloning EcoRI/XbaI-digested inserts into pBS and using universal end-labelled primers. Insert DNA sequences were translated in three reading frames (forward direction with respect to the β-lactamase gene) and the ORF in-frame with the β-lactamase gene (Fig. 2) was identified. Insert ORFs were then searched against sequence databases and analysed for sequence motifs using the Internet servers listed in Table 1. The primary databases were the M. tuberculosis genome sequences released by The Institute for Genomic Research (TIGR) and the Sanger Centre.

**Western blot analyses of fusion proteins.** pGB14T(H37Rv)-transformed, ampicillin-resistant (Amp') E. coli were grown in LB broth containing 100 µg ampicillin ml⁻¹ to OD₆₀₀ 0.5, induced with 0.3 mM IPTG for 2 h and centrifuged at 4000 r.p.m. The supernatants were concentrated 16-fold in a Centricon (30 kDa cut-off; Amicon) and saved as the culture filtrate fraction. The cells were divided into three fractions. (a) The pellet was resuspended in buffer A (20 mM Tris/HCl, pH 7.4, 200 mM NaCl, 1 mM EDTA), frozen at -20°C, thawed and sonicated (six 15 s pulses) and centrifuged at 8500 r.p.m., after which the supernatant was saved as the soluble cell fraction. (b) The pellet from (a) was resuspended in buffer A and saved as the insoluble cell fraction. (c) The cells were resuspended in 30 mM Tris/HCl, pH 8.0, containing 20% (w/v) sucrose, adjusted to 1 mM EDTA, incubated with shaking for 10 min at room temperature, centrifuged at 8500 r.p.m. for 10 min and the pellet resuspended in ice-cold 5 mM MgSO₄; after incubation on ice with shaking for 10 min, the suspension was centrifuged as before and the supernatant was saved as the periplasmic fraction.

The various cell fractions were mixed with SDS sample buffer and resolved by SDS-PAGE on a 10% (w/v) polyacrylamide gel. Electrophoresed proteins were transblotted onto Hybond-C nitrocellulose filters (Amersham) under standard conditions (see Hoppe et al., 1997). The filters were incubated in blocking buffer (50 mM Tris/HCl, pH 7.4, 200 mM NaCl, 5% w/v, skimmed milk powder, 0.1% Tween 20) for 1 h at room temperature, and then incubated with rabbit anti-ampicillinase IgG (CP Laboratories; 1:1000 in blocking buffer) for 1 h at room temperature. The blots were washed in blocking buffer without skimmed milk powder (TBS/Tween), incubated with peroxidase-conjugated anti-rabbit IgG (ECL Western blotting system, Amersham; 1:1500 in TBS/Tween) for 1 h at room temperature, washed again in TBS/Tween, and then treated with ECL detection reagents and exposed to photographic film, as recommended by the manufacturer.

**RESULTS**

**Construction of plasmids pGB14T and pGB14TC**

pGB14 has proven useful in identifying signal-sequence-encoding genes from the Gram-positive bacteria Bacillus subtilis and Lactococcus lactis by selection in an E. coli background (Smith et al., 1987; Perez-Martinez et al., 1992). However, this vector may not function well with M. tuberculosis inserts because the leaderless β-lactamase gene in pGB14 also lacks a promoter and mycobacterial promoters are, in general, poorly recognized by the E. coli transcriptional apparatus (Das Gupta et al., 1993). We therefore introduced the efficient, IPTG-inducible E. coli Tac promoter, to generate pGB14T. Selection for pGB14-transformed E. coli depends on erythromycin resistance, which is a poor marker in Gram-negative organisms; we found that even at 200 µg erythromycin ml⁻¹, selection for true Em' proved difficult due to a high rate of spontaneous low-level resistance. Hence we constructed pGB14TC as an alternative vector, in which the Em' gene was interrupted by a Cm' cassette. As expected, growth on chloramphenicol facilitated selection for E. coli (pGB14) transformants.

**Construction in E. coli of a library of β-lactamase fusions with M. tuberculosis genomic DNA fragments**

Genomic DNA from M. tuberculosis H37Rv digested with Sau3A, Hae3 or RsaI was ligated into the MCS of pGB14T and pGB14TC between the Tac promoter and the leaderless β-lactamase gene. The highest transformation efficiency of E. coli with the ligation products was achieved with electroporation (10-100-fold greater than transformation of CaCl₂-competent cells). However, even with electroporation, transformations were only in the order of 2 × 10⁶ c.f.u. per µg pGB14T(H37Rv) or pGB14TC(H37Rv) DNA. Transformation with the pGB14T or pGB14TC vectors alone gave similar results, compared to approximately 10⁸ c.f.u. per µg for pBS, indicating that pGB14-based vectors are inefficient transformation substrates.

Generally, E. coli transformed with pGB14T(H37Rv) or pGB14TC(H37Rv) were selected directly on plates containing 50-100 µg ampicillin ml⁻¹, and an aliquot of each batch of transformants was grown on plates containing 200 µg erythromycin ml⁻¹ or 30 µg chloramphenicol ml⁻¹, respectively, to assess transformation efficiencies. All Amp' clones were tested on erythromycin or chloramphenicol plates, as appropriate, to select for true transformants. From a total of 6000 Em' clones and 8000 Cm' clones, 10 Amp'/Em' and 5 Amp'/Cm' transformants, respectively, were isolated.

**Analysis of M. tuberculosis DNA–β-lactamase fusions conferring Amp' in E. coli**

Twelve of the Amp' clones were tested for their level of Amp' and all were resistant to up to 1 mg ampicillin ml⁻¹ on LB plates. Insert sizes ranged from 350 to 3000 bp. Ten inserts from Amp'/Em' and Amp'/Cm' clones were sequenced and the results are summarized in Table 2. To facilitate analysis, insert DNA sequences were translated and the ORF in-frame with the β-lactamase gene was identified. Comparison of the deduced amino acid sequences with known sequences in databases enabled complete identification in all cases (Fig. 2). pGB14T-L4 and pGB14TC-W were found to be identical to the M. tuberculosis antigen 85A and 85C genes, respectively (Fig. 2a, b). pGB14T-L4 showed 100% identity over 180 amino acid residues with...
Ag85A; pGB14TC-W showed 100% identity over 40 residues with Ag85C. Both inserts included the entire signal sequence, translational start site and upstream ribosome-binding site.

A group of four inserts was found to be homologous to mycobacterial genes that comprise hypothetical ORFs encoding one or more potential membrane-interaction domains (in addition to the signal peptides) (Fig. 3).

The first insert, pGB14T-P, was 98% identical (176 out of 180 amino acid residues) to the product of an unannotated, hypothetical ORF (MTcY493.19) on M. tuberculosis cosmid Y493 (positions 26636–27264, EMBL accession number Z95844), which we have designated Vmtbl (Fig. 2d). Vmtbl was a 202 amino acid preprotein that contained an extended apolar, glycine- and proline-rich central domain (residues 60–156) that was not typical of a TM domain but may participate in hydrophobic interactions (Fig. 3b).

The second insert, pGB14TC-V, was 98% identical (83 out of 84 residues) to the protein encoded by an annotated, hypothetical ORF (MTcY04D9.16~ on cosmid SCY04D9, positions 147 out of 147 residues) to a hypothetical M. tuberculosis protein (MTcY04D9.16~ on cosmid SCY04D9, EMBL accession number Z84725) (Xmlepl in Fig. 2e). Further, Xmlepl was 44% identical (65 out of 147 residues) to a hypothetical, 15.9 kDa M. tuberculosis protein (SWISS-PROT accession number P54880) (Xmlepl in Fig. 2e). Two of the homologues, Xmlepl and Xmleb, contained putative TM domains near their C termini (Fig. 3c).

The fourth insert in this group, pGB14T-O2, was 100% identical to the product of an unannotated, hypothetical ORF in the M. tuberculosis genome, referred to here as Xmtbl (Fig. 2e). Xmtbl was, in turn, 75% identical (104 out of 137 residues) to a hypothetical, 15.9 kDa M. tuberculosis protein (MTcY270.05; EMBL accession numbers Z80233 and Z95388). Both Pmtbl and Pmlepl contained potential membrane lipoprotein lipid attachment sites (as predicted by GCG motifs; see also Pugsley, 1993); Pmtbl contained a potential TM domain near its C terminus (Fig. 3a).

The third insert, pGB14T-O1, was 98% identical (83 out of 84 residues) to the product of an annotated, hypothetical ORF (MTcY04D9.16~ on cosmid SCY04D9, positions 26636–27264, EMBL accession number Z95844), which we have designated Vmtbl (Fig. 2d). Vmtbl was a 202 amino acid preprotein that contained an extended apolar, glycine- and proline-rich central domain (residues 60–156) that was not typical of a TM domain but may participate in hydrophobic interactions (Fig. 3b).

The third insert, pGB14T-O1, was 98% identical (83 out of 84 residues) to the product of an annotated, hypothetical ORF (MTcY04D9.16~ on cosmid SCY04D9, positions 26636–27264, EMBL accession number Z95844), which we have designated Vmtbl (Fig. 2d). Vmtbl was a 202 amino acid preprotein that contained an extended apolar, glycine- and proline-rich central domain (residues 60–156) that was not typical of a TM domain but may participate in hydrophobic interactions (Fig. 3b).

The fourth insert in this group, pGB14T-O2, was 100% identical to the product of an unannotated, hypothetical ORF in the M. tuberculosis genome, referred to here as Xmtbl (Fig. 2e). Xmtbl was, in turn, 75% identical (104 out of 137 residues) to a hypothetical, 15.9 kDa M. tuberculosis protein (MTcY270.05; EMBL accession numbers Z80233 and Z95388). Both Pmtbl and Pmlepl contained potential membrane lipoprotein lipid attachment sites (as predicted by GCG motifs; see also Pugsley, 1993); Pmtbl contained a potential TM domain near its C terminus (Fig. 3a).
**Fig. 2.** Sequence alignments of predicted proteins encoded by export-competent *M. tuberculosis* gene fusions. In each case (a–h), *M. tuberculosis* insert sequences identified with the pGB14T and pGB14TC selection vectors are shown above their respective homologues retrieved from mycobacterial databases (Table 1). Sequences are in the single-letter amino acid code. Only one line of sequence per insert is shown, truncated as necessary (by the number of residues indicated), ending in the underlined sequence (G)DPLESTAQAC (spacer sequence at the fusion junction) followed by *ppet*... (start of mature β-tubiamin in lower-case italics). Hydrophobic sequences that represent potential signal peptides or TM domains are in bold; arrows indicate probable signal peptide cleavage sites in the native proteins, as predicted by SignalP (Table 1). Sequence nomenclature is our own. Sources are as follows (in parentheses are accession numbers; gb, GenBank; emb, EMBL; sp, SWISS-PROT): (a) Ag85A (gb M27016); (b) Ag85C (sp P31953); (c) Pmtb1, unannotated ORF, *M. leprae* cosmid B1529, 19336-20019 (gb L78824); (d) Pmepl, unannotated ORF, *M. leprae* cosmid B983, 9760-10797 (gb L78828); (e) Pmtb1, unannotated ORF, *M. tuberculosis* (gb MTCY493.19 on cosmid Y493, 26656-27264 (emb 295844)); (f) Pmepl, unannotated ORF on *M. tuberculosis* contig gmt551, 5197-4167 (TIGR); (g) Bmtb1, unannotated ORF, *M. leprae* cosmid 8124, complement 24139-24603 (gb U15183); Bmtb2, ORF labelled *cyoC* on *M. tuberculosis* cosmid 31159-31602 (gb Z77162); Bmtb3, ORF labelled *mip* on *M. tuberculosis* cosmid 1771-15199 (emb Z84725); (h) Omtb1, unannotated ORF on *M. leprae* contig gmt374, 1342-1040 (TIGR).
5197–4167) (O2mtbl in Fig. 2f). O2mtbl did not contain a typical cleavable signal sequence. A stretch of 21 hydrophobic and apolar residues that could serve as an export signal was preceded by at least 37 N-terminal residues, including three negatively charged residues and a cluster of five arginines (Fig. 2f), which is very uncharacteristic of cleavable signal sequences (Pugsley, 1993). O2mtbl may be a type II TM protein with an uncleaved signal peptide acting as a true TM segment (Fig. 3d). Apart from this exception, all Amp−-conferring β-lactamase gene fusions that were sequenced coded for readily identifiable signal sequences, defined as N-terminal extension sequences that direct export across the cytoplasmic membrane and characterized by shared physico-chemical features (Pugsley, 1993; Izard & Kendall, 1994) (Fig. 2). Thus, Pmtb1, Vmtb1, Xmtb1, O2mtbl and homologues represent a group of hypothetical exported mycobacterial proteins that contain putative TM, hydrophobic, or membrane-interaction domains.

As was found for pGB14T-P and pGB14TC-X, pGB14T-B was homologous to several mycobacterial genes. pGB14T-B was 100% identical (50 out of 50 residues) to a hypothetical M. tuberculosis protein (MTCY19H5.32c on cosmid SCY19H5, EMBL accession number Z97182) (Bmtb1 in Fig. 2g), which was 65% identical (168 out of 257 residues) to a second hypothetical M. tuberculosis protein (MTC128.10 on cosmid 128, EMBL accession number Z97050) (Bmtbl2 in Fig. 2g). Bmtb2, in turn, was 75% identical (260 out of 346 residues) to the product of an unannotated, hypothetical ORF in the M. leprae genome (cosmid B983, positions 9760–10797, GenBank accession number L78828; Bmlep1 in Fig. 2g). The functions of each of these putative exported proteins is unknown. Of note is that Bmtb2 contained four potential N-linked glycosylation sites (NXS/T) typical of eukaryotic exported proteins.

The insert in pGB14T-O1 was cloned independently on two occasions, as pGB14T-O1 and -B3, perhaps because it encoded an extremely favourable signal peptide (Fig. 2h), or because multiple copies of the gene were present in the genome. pGB14T-O1 was 100% identical (44 out of 44 residues) to the product of an unannotated, hypothetical ORF in M. tuberculosis contig gmt374 (positions 1324–1040) in the TIGR genome (O1mtbl in Fig. 2h). Lastly, the insert in pGB14T-Q5 (at 3000 bp the largest) was found to be 100% identical (193 out of 193 residues) to the E. coli cytochrome C-type biogenesis protein CCMH precursor, which has an excellent N-terminal signal peptide (not shown). This is presumably an E. coli chromosomal DNA contaminant that was copurified and codigested with the pGB14T plasmid preparation prior to M. tuberculosis DNA insertion.

**Immunodetection of β-lactamase fusions**

Cell fractions of Amp+ E. coli transformed with the clones described in Fig. 2 were probed by Western blotting using an anti-TEM β-lactamase antibody.
in size to those seen in the periplasmic and culture filtrate fractions. Hence we were unable to detect the unprocessed pre-forms, except as occasional faint traces, including a 32 kDa immunoreactive band in the pBS fraction that presumably corresponded to the unprocessed form (faintly visible in Fig. 4b). We conclude that under these experimental conditions, signal peptide cleavage was so efficient that the vast majority of detectable fusion proteins was processed.

**DISCUSSION**

The mycobacterial protein export machinery has not been characterized. Although they are formally considered as members of the Gram-positive family, mycobacteria have unusual cell walls that contain unique features. Most notable is the presence of an arabinogalactan–mycolic acid superpolymer that is covalently linked, and external, to the peptidoglycan. The mycolic acids of this superpolymer have been proposed to constitute the inner leaflet of a pseudo outer membrane; the outer leaflet is provided by various non-covalently associated phospholipids and glycolipids (Brennan & Draper, 1994). Thus, the mycobacterial cell wall bears a superficial resemblance to the lipid-rich Gram-negative envelope, although the mycobacterial envelope is markedly waxy and much less fluid (Liu et al., 1996). It can be anticipated, therefore, that some of the special problems of protein export in Gram-negative organisms are also encountered in mycobacteria, particularly with regard to targeting to various cell wall compartments, including complete secretion into the extracellular milieu (Pugsley, 1993). These considerations notwithstanding, it seems likely that the first step in protein translocation, export across the cytoplasmic membrane, which is common to the general secretory pathways of both Gram-positive and Gram-negative organisms, is also shared by mycobacteria. Indeed, many of the genes encoding secreted mycobacterial proteins encode typical N-terminal signal sequences (Young et al., 1992), indicating that mycobacteria probably possess a standard Sec-dependent export pathway. A search of annotated genes in the *M. tuberculosis* genome (Sanger Centre) revealed homologues of *secA*, *secB*, *secD*, *secE* and *secF* [EMBL accession numbers and loci Z95121 (MTCY20B11.15c), Z79701 (MTCY277.40), Z77724 (MTCY227.14), Z92772 (MTCY20H10.19) and Z77724 (MTCY227.15), respectively]. The *Mycobacterium bovis* BCG secY has been reported (GenBank accession number U77912) and we found the *M. tuberculosis* homologue on TIGR contig gmt7540 (positions 12060–13382).

In efforts to gain further insights into protein secretion in *M. tuberculosis* we have explored the usefulness of a signal sequence selection vector for the identification and characterization of mycobacterial signal-sequence-encoding genes in an *E. coli* host. We have exploited a vector relying on β-lactamase gene fusions, which has proved useful in identifying *B. subtilis* and *L. lactis* signal sequence genes in transformed *E. coli* (Smith et al., 1987; Perez-Martinez et al., 1992). After modifying the vector by inserting a Tac promoter upstream of the

---

**Fig. 4.** Western blots of Amp′, pGB14T(H37Rv)-transformed *E. coli* fractions probed with an anti-β-lactamase antibody. *E. coli* cell fractions were prepared as described in Methods. (a) Periplasmic (peri) and culture filtrate fractions, respectively, of *E. coli* transformed with pGB14T-P (P peri, P filtrate), -O2 (O2 peri, O2 filtrate), -O1 (O1 peri, O1 filtrate) and pBS (pBS peri, pBS filtrate). (b) Insoluble fractions of *E. coli* transformed with the indicated vectors; no band was seen in the pGB14T-01 lane on this blot. The arrow on the right indicates the position of an immunoreactive band in the pGB14T-01 lane that was seen clearly on the original blot. Molecular mass marker positions (in kDa) are indicated.
leaderless β-lactamase gene, we were able to isolate a variety of signal sequence genes from an \textit{M. tuberculosis} genomic DNA library. The ability of the selection vector to successfully identify \textit{bona fide} secreted proteins was evident from clones pGB14T-L4 and -W that contained fragments of the Ag85A and Ag85C genes, respectively, which together with Ag85B constitute the major secreted proteins of the \textit{M. tuberculosis} complex (Content et al., 1991). Ag85 proteins account for more than 30% (by weight) of the total extracellular proteins in \textit{M. tuberculosis} culture filtrates (Horwitz et al., 1995).

All but one of the Amp\textsuperscript{R} β-lactamase gene fusions that were sequenced were found to encode clearly recognizable signal sequences; the exception (pGB14T-O2) appeared to encode a sequence more typical of type II TM proteins with uncleaved signal peptides. All of the gene fusions tested conferred wild-type levels of Amp\textsuperscript{R} (up to 1 mg ampicillin ml\textsuperscript{-1} on agar plates), indicating that the \textit{M. tuberculosis} genome contains a number of genes encoding signal sequences that satisfy all the requirements of the \textit{E. coli} Sec-dependent export machinery. This conclusion is supported by previous data demonstrating that the secreted \textit{M. bovis} protein MBP70 was efficiently secreted into the periplasm of \textit{E. coli} transformed with the entire MBP70 gene, including the mycobacterial ribosome-binding site, start codon and signal peptide (Hewinson & Russell, 1993). The MBP70 signal peptide also directed the export of the 19 kDa \textit{M. bovis} antigen in \textit{E. coli} (Hewinson et al., 1996a). The nature of the selection system used in our experiments will be biased towards the selection of \textit{E. coli}-compatible signal sequences and provides no information on the presence of atypical signal sequences and/or export pathways in \textit{M. tuberculosis}. Nevertheless, these results argue strongly that a subset of proteins secreted by \textit{M. tuberculosis} is exported by a Sec-dependent pathway which has characteristics similar to the \textit{E. coli} machinery. This is consistent with the notion that despite differences in detail, the sec-dependent pathways of Gram-positive and Gram-negative bacteria are broadly similar (Pugsley, 1993; Salmond & Reeves, 1993; Bolhuis et al., 1996).

All of the signal-peptide-encoding DNA fragments sequenced could be matched with database sequences. Nine of ten gave perfect matches with sequences in the \textit{M. tuberculosis} genome (TIGR and Sanger Center databases), whereas one (pGB14T-Q5) contained an \textit{E. coli} chromosomal DNA contaminant. Strikingly, apart from the Ag85 proteins that were recently shown to be mycolyltransferases (Belisle et al., 1997), only one of the other putative exported proteins shown in Fig. 2 could be assigned a function, despite extensive BLAST searches against all of the major databases. Pmtb1 was an unannotated hypothetical ORF that after a gapped BLAST search was found to be 35\% identical to a low-affinity PBP (PcbR) from \textit{S. clavuligerus}, a protein that belongs to the high-molecular-mass group B PBPs (Paradkar et al., 1996). Low-affinity PBPs of this type may play an important role in β-lactamase resistance in actinomycetes (Paradkar et al., 1996). Further analysis revealed that Pmtb1 was 31\% identical to two annotated \textit{M. tuberculosis} low-affinity PBPs, MTCY10H4.16c and MTCY270.05, which were 23\% identical to each other. In contrast to the latter two, Pmtb1, and its \textit{M. leprae} homologue Pmlepl, contained a signal peptide with features typical of a lipoprotein lipid attachment site (Fig. 2; see Pugsley, 1993). High-molecular-mass PBPs typically contain N-terminal hydrophobic domains that probably function as uncleaved signal peptides and membrane anchors, as has been inferred for PcbR (Paradkar et al., 1996); similar sequences were present in the putative \textit{M. tuberculosis} PBPs MTCY10H4.16c and MTCY270.05. Interestingly, Pmtb1 was only 6 kb removed from the gene encoding MPT83 (on contig gnt18 in the TIGR \textit{M. tuberculosis} genome), which is a known lipoprotein that is expressed by members of the \textit{M. tuberculosis} complex and which is homologous to, and in turn only 2-4 kb upstream of, the gene for the secreted \textit{M. tuberculosis} protein MPT70 (Hewinson et al., 1996b; Vosloo et al., 1997). The functions of MPT83 and MPT70 are unknown, but expression of MPT83 may be elevated during an \textit{in vivo} \textit{M. tuberculosis} infection (Hewinson et al., 1996b), and MBP70 is an immunodominant \textit{M. bovis} antigen (Hewinson & Russell, 1993).

Further interesting observations that emerged from analyses of putative proteins encoded by the ORFs identified in our screen (Fig. 2) include the following. Of a set of four homologues, Xmlepl and Xmtb3 contained a possible TM domain near their C termini, making them potential type I TM proteins. Together with O2mtb1, a potential type II TM protein, and Vmtb1, a protein containing an extended glycine- and proline-rich apolar domain, Pmtb1, Pmlepl, Xmlepl and Xmtb3 constitute a group of exported proteins that are potentially associated with the cell membrane or other compartments of the lipid-rich mycobacterial cell wall (Fig. 3). Finally, also of interest was the presence of four potential N-linked glycosylation sites in Bmtb2. The existence of O-linked glycoproteins in mycobacteria is now well-established (Dobos et al., 1996; Herrmann et al., 1996) and there is evidence for N-linked carbohydrates in the secreted 19 kDa antigen of \textit{M. tuberculosis} (M. Ward, W. Blackstock, M.-P. Gares, D. B. Young & C. Abou-Zeid, personal communication). The significance of this form of glycosylation in mycobacteria remains to be determined. It is noteworthy that the major outer-membrane protein of \textit{Chlamydia trachomatis} contains N-linked high-mannose type oligosaccharides that are required for binding and infection of HeLa cells \textit{in vitro} (Kuo et al., 1996). We have shown recently that surface-exposed glycoconjugates mediate direct and opsonin-enhanced binding of \textit{M. tuberculosis} to various mammalian cells \textit{in vitro} (Hoppe et al., 1997; Cywes et al., 1997). Thus the possible involvement of mycobacterial glycoproteins in such processes warrants further investigation.

It has been shown that alkaline phosphatase (\textit{pboA}) gene fusions are useful in studying protein export in mycobacteria (Timm et al., 1994) and that this system
allows identification of *M. tuberculosis* genes encoding secreted proteins in an *M. smegmatis* host (Lim et al., 1995; Jackson et al., 1997). PhoA reporter constructs gave a yield of 0.09% Amp\(^\beta\) transformants from an *M. tuberculosis* genomic DNA library (Lim et al., 1995), compared to our rate of 0.11%. Enhanced vector design enabling inserts to be cloned in all three reading frames with respect to *phoA* and optimization of mycobacterial genomic DNA digestions reportedly improved the yield of export-competent inserts to 1.5% of transformants; a complete analysis of these inserts has still to be presented (Mdluli et al., 1995).

Compared to *phoA* gene fusions, the spectrum of signal sequence genes identified with β-lactamase fusions may be different because selection is based on antibiotic resistance with the latter versus a colour reaction with the former (Smith et al., 1987). The advantages of β-lactamase fusions in contrast to PhoA fusions as membrane protein topology probes have been reviewed (Broome-Smith et al., 1990). Particularly notable for the purposes of selection for export-competent sequences is the fact that β-lactamase fusions allow direct selection whereas PhoA fusions require screening of all colonies. PhoA fusions also have the disadvantage that some cyttoplasmic fusions display weak enzymic activity, essentially giving false positives (Broome-Smith et al., 1990). In our protocol we potentially selected against fusions that were poorly translocated by direct selection where precursor forms of β-lactamase have reduced activity (Roggenkamp et al., 1985; Perez-Martinez et al., 1992), we may have selected for fusions that were processed efficiently by the *E. coli* signal peptidase to generate mature, soluble proteins (Fig. 4). Moreover, by using conditions for inducing the strong Tac promoter, we probably selected against the recovery of β-lactamase fusions with multiple membrane-spanning, polytopic membrane proteins, which are toxic when over-expressed in *E. coli*. These considerations probably account for the exclusive identification of *M. tuberculosis* gene fragments encoding typical N-terminal signal sequences in our screen. Interestingly, none of the reported *M. tuberculosis* signal-sequence-encoding genes isolated by *phoA* gene fusions (Lim et al., 1995; Mdluli et al., 1995; Jackson et al., 1997) appears to match those identified by β-lactamase gene fusions described here.

With the total sequencing of the *M. tuberculosis* genome virtually complete (Williams, 1998), the utility of both β-lactamase and *phoA* gene fusions lies not so much in identifying mycobacterial signal-sequence-encoding genes per se, but rather in testing the export competence of putative signal sequences in various bacterial hosts to probe for similarities between mycobacterial, Gram-negative and Gram-positive export pathways. The preliminary characterization of export-competent β-lactamase and *phoA* gene fusions reported here and elsewhere (Lim et al., 1995; Mdluli et al., 1995; Jackson et al., 1997) are first attempts at throwing some light on this important question. In view of the strong likelihood that exported proteins play a key role in mycobacterial virulence and in potentially eliciting a protective host immune response (Lee & Horwitz, 1995; Orme et al., 1993; Andersen, 1994; Horwitz et al., 1995), an understanding of the structure, function and regulation of the *M. tuberculosis* protein export machinery may prove decisive in the design of novel preventive and therapeutic strategies for tuberculosis.

**ACKNOWLEDGEMENTS**

This work was supported by funds provided under the Glaxo Wellcome Action TB initiative and by the University of Cape Town. We thank Jan Maarten van Dijl for the vector pGB14 and John T. Belisle for the *M. tuberculosis* H37Rv genomic DNA. We are indebted to Di James for the automated DNA sequencing.

**REFERENCES**


Received 12 August 1997; revised 2 February 1998; accepted 17 February 1998.