Lipoproteins of *Mycobacterium tuberculosis:* an abundant and functionally diverse class of cell envelope components

Iain C. Sutcliffe[1*] and Dean J. Harrington[2]

[1] Division of Biomedical Sciences, Northumbria University, UK

[2] Department of Biomedical Sciences, University of Bradford, UK

[*] Corresponding author:

Iain Sutcliffe, Division of Biomedical Sciences, School of Applied Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK

Telephone: + 44 191 227 3176

Fax: + 44 191 227 3176

E-mail: iain.sutcliffe@ northumbria.ac.uk

Keywords: Bioinformatics; Genome; Lipoprotein; *Mycobacterium tuberculosis*; Periplasm; Virulence factor

Abstract

Mycobacterium tuberculosis remains the predominant bacterial scourge of mankind. Understanding of its biology and pathogenicity has been greatly advanced by the determination of whole genome sequences for this organism. Bacterial lipoproteins are a functionally diverse class of membrane-anchored proteins. The signal peptides of these proteins direct their post-translational export and lipid modification. These signal peptides are amenable to bioinformatic analysis, allowing the lipoproteins encoded in whole genomes to be catalogued. This review applies bioinformatic methods to the identification and functional characterisation of the lipoproteins encoded in the *M. tuberculosis* genomes. Ninety nine putative lipoproteins were identified and so this family of proteins represents ca. 2.5% of the *M. tuberculosis* predicted proteome. Thus, lipoproteins represent an important class of cell envelope proteins that may contribute to the virulence of this major pathogen.

Table of Contents

1. Introduction

At the start of the twenty-first century tuberculosis (TB) remains one of the major infectious diseases of man with ca. 8 million active cases and ca. 2 million deaths annually [1,2]. Furthermore, an estimated one third of the worlds population is infected with latent TB and face the possibility of reactivation disease. Disease in this population, especially when combined with the devasting impact of HIV co-infection, remains a significant issue in global disease control [1,2]. Thus obtaining a greater understanding of the molecular basis of the virulence of *Mycobacterium tuberculosis* and its disease pathogenesis presents an urgent challenge to the scientific community. In this regard, the publication of the genome sequence of *M. tuberculosis* strain H37Rv [3] represents a contemporary landmark in TB research. Another major achievement has been in the definition of the composition and organisation of the mycobacterial cell envelope, which has provided important insights into the physiology and pathogenicity of *M. tuberculosis* [4-7]. The mycobacterial cell envelope is a complex structure dominated by the peptidoglycan-arabinogalactan-mycolic acid wall skeleton in which the mycolic acids are orientated so as to provide the foundation for an external permeability barrier. As such it seems likely that this lipid permeability barrier represents the outer boundary of a 'pseudoperiplasmic' compartment, which is defined at its internal aspect by the plasma membrane. However, the nature of the protein components and metabolic activities that may be present within this subcellular compartment remain largely unknown.

All bacteria apparently localise specific proteins to their cell envelopes by post-translational lipid modification to produce membrane-anchored lipoproteins (Lpp; [8,9]). In mycobacteria lipid modification is likely to represent an important mechanism

by which proteins are localised within the cell envelopes, such that these proteins are located within the above described pseudoperiplasmic compartment. Indeed, Lpp of Gram-positive bacteria have been previously suggested to be functional equivalents of periplasmic proteins in Gram-negative bacteria [9,10]. Given the likely significance of Lpp to bacterial physiology and the expectation that some Lpp may be virulence factors we have used a bioinformatic strategy to compile an inventory of putative Lpp of *M. tuberculosis.* The likely functions of these Lpp are discussed with particular reference to the contribution some of them may make during pathogenesis.

2. The lipoprotein biosynthetic pathway in *M. tuberculosis*

Lpp biogenesis is dependant on the presence of specific type II signal peptide sequences [8,11-14). The signal peptide directs preprolipoprotein export through the plasma membrane whereupon a diacylglyceride unit is added by thioether linkage to a crucial cysteine residue. The enzyme that carries out this lipidation reaction, prolipoprotein diacylglyceryl transferase (Lgt), is apparently an essential enzyme in Gram-negative bacteria but is dispensable in the Gram-positive bacteria studied to date [15,16]. However, an *lgt* mutant of *Streptococcus pneumoniae* was attenuated in an animal model of disease [16]. Subsequent to the lipidation by Lgt, the signal peptide is cleaved by a specific prolipoprotein signal peptidase II enzyme (Lsp) at a cleavage site immediately preceding the lipidated cysteine, which consequently becomes the N-terminus of the mature Lpp. Lsp is also apparently dispensable for growth of Gram-positive bacteria in vitro [17-19] although *lsp* mutants of *Listeria monocytogenes* and *Staphylococcus aureus* have been shown to be attenuated in animal models [19-21].

Several mycobacterial Lpp with type II signal peptides have been experimentally characterised as Lpp [14,22] although in no case has there been direct chemical characterisation of the diglyceride-modified cysteine residue. However, consistent with the above biosynthetic pathway, putative *lgt* (*Rv1614*) and *lsp* (*Rv1539*) genes have been identified in the *M. tuberculosis* genome [3]. It is likely that the actions of these two enzymes are both necessary and sufficient for ensuring Lpp anchoring. In Gram-negative bacteria a third enzyme, lipoprotein N-acyl transferase (Lnt), adds an additional amide-linked fatty acid to the amino terminus of the mature Lpp [23]. However, the presence of this enzyme in *M. tuberculosis* needs to be clarified: although Rv2051c was originally annotated as a two-domain enzyme containing a putative Lnt domain, this protein has been characterised as a polyprenol monophosphomannose synthase (Ppm1, [24]). Although the putative Lnt domain is not needed for Ppm1 activity, on overexpression in *M. smegmatis* it appeared to enhance the mannosyltransferase activity. Interestingly, the two domains of Ppm1 are encoded by separate, adjacent open reading frames in the genomes of other mycobacteria and corynebacteria. Thus the role of Rv2051c and its homologues in mycobacterial Lpp biogenesis remains unclear.

3. Identification of putative lipoproteins of *M. tuberculosis*

In contrast to the cleavage sites of type 1 signal peptides, there is considerable sequence conservation in the amino acids that immediately precede the lipidated cysteine in Type II signal peptides [8,12-14]. This sequence is commonly referred to as the ‘lipobox’ and its sequence characteristics are described in the Prosite pattern PS00013 ([25], http://www.expasy.ch/prosite/). We have since derived a revised

sequence pattern (G+LPP) for the confident identification of putative Lpp sequences in the genomes of Gram-positive bacteria from the signal sequence features of thirty three experimentaly verified Gram-positive bacterial Lpp [14]. The recognition of a lipobox cysteine appropriately placed in relation to other typical signal peptide features has allowed the bioinformatic identification of genes encoding putative Lpp, revealing that Lpp are an abundant class of proteins, typically representing ca. 1.5% or more of the total predicted proteins in the sequenced bacterial genomes [13,14,19,26-28].

In the present analysis the G+LPP pattern was used in a taxon-restricted pattern search to retrieve *M. tuberculosis* sequences in the Swiss-Prot/TrEMBL database ([29], http://ca.expasy.org/sprot/), using the ScanProsite tool ([30], http://ca.expasy.org/tools/scanprosite/). Except where discrepancies are referred to in the text below, all sequences were common to both the *M. tuberculosis* strain H37Rv [3] and strain CDC1551 [31] genomes and are refered to herein by their original gene designations (Rv number, [3]). Sequences unique to strain CDC1551 are referred to as annotated (MT number; [31]). The N-terminal features of the 59 sequences identified as matching the G+LPP pattern were re-examined using SignalP ([32,33], http://www.cbs.dtu.dk/services/SignalP-2.0/) and prediction methods for identifying membrane-spanning domains (MSD), notably the TMHMM server ([34], http://www.cbs.dtu.dk/services/TMHMM-2.0/) and similar tools as described previously [14]. As well as the 3 previously described experimentally verified Lpp (Rv0432; Rv0934, Rv3763), the G+LPP pattern identified 51 other probable Lpp sequences (Table S1 in the online supplementary material). Two sequences that had ambiguous signal peptide features (Rv0847 and Rv2945c) were retained as possible

Lpp (Table S2 in the online supplementary material) whilst 3 sequences (MT2138.1, MT3476 and MT3814.1) wherein the predicted lipobox cysteine was inappropriately placed in relation to typical signal peptide features were considered false-positives, as described previously [14]. These sequences are notable in that all are subject to annotation discrepancies between the *M. tuberculosis* strain H37Rv and strain CDC1551 genomes [3,31].

Following the above analysis, a pattern search with the Prosite pattern PS00013 identified an additional 32 *M. tuberculosis* sequences in the Swiss-Prot/TrEMBL database that had a putative lipobox that did not match the more restrictive cleavage site in the G+LPP pattern. Further examination of the N-terminal features of these sequences as above suggested that a substantial proportion (25%) should be excluded as false-positives whilst 24 should be retained and considered possible Lpp (Table S2 in the online supplementary material). These analyses confirmed the previous observation [14] that the stringency of the G+LPP pattern compared to PS00013 gives greater confidence (i.e. fewer false positives) when predicting Lpp signal peptides. Moreover, the pattern search with the G+LPP pattern identified six putative Lpp that were not recognised by the PS00013 pattern.

Finally, a variety of strategies, notably reference to the H37Rv genome annotation and to DOLOP [27], were used to identify putative Lpp sequences that were not retrieved by either of the above pattern searches. This approach identified 19 further sequences (Tables S3 and S4 in the online supplementary material). These included a large subset of 10 sequences with hydrophobic h-regions and lipobox cleavage sites that match the G+LPP pattern but with anomalously long signal peptide n-regions. These sequences were considered ‘anomalous probable’ Lpp

(Table S3 in the online supplementary material). Rv2080, which has a typical lipobox cleavage site but a minor variation in its h-region sequence, was also included in this category (Table S3 in the online supplementary material). The remaining 8 sequences that had features consistent with anomalous type II signal peptides were considered 'anomalous possible' Lpp (Table S4 in the online supplementary material).

Analysis of the signal peptide features indicated that the proven and probable Lpp (Supplementary Table S1) have typical type II signal peptide characteristics with respect to their length (mean cysteine position 24.0±4.1; range 16-33 amino acids; n=54) and that most of the variation derives from variation in the length of the n-region (mean length 6.7±3.6; range 2-15 amino acids; n=54). Thus these signal peptides are typically shorter than those of exported proteins of *M. tuberculosis*, for which a median length of 32 amino acids (range 21-49; n=28) has been reported [35]. The comparative shortness of Lpp signal peptides has been noted previously [13,14]. Analysis of the frequency of amino acid usage in the lipobox cleavage site suggested some minor selectivity in comparison to the G+LPP pattern: a slightly increased preference for glycine at the –4 position (relative to the cysteine) was noted, whilst at the –3 position a slightly decreased preference for leucine was linked to an increase in the frequency of occurrence of valine and alanine (data not shown). As for Gram-positive bacterial Lpp generally, there was a notable preference for small amino acids at the +2 position following the lipobox cysteine, with alanine, glycine, serine or threonine present in 45/54 (83%) of the sequences. Thus, the putative Lpp differ from secreted proteins where a marked preference for proline (15 out of 28 [54%] sequences examined) at the +2 position has been reported [35].

The recognition of the subset of ten anomalous probable Lpp signal sequences (Supplementary Table S3) containing long n-regions was of interest. The Rv0179c and Rv2518c sequences identified by searching with PS00013 (Supplementary Table S2) also belong to this group. The length of the n-region (mean 20.8±5.6; range 16-33 amino acids; n=12) of these sequences notably affected the position of the lipobox cysteine (mean cysteine position 38.6±6.0; range 33-52; n=12). As the h-region features of these putative Lpp are comparable to those of the proven/probable Lpp (data not shown) the length of this region is probably decisive in orientating the lipobox cysteine such that this critical amino acid can interact with the membrane-bound Lgt enzyme.

Cumulatively, the above searches identified 99 sequences including 3 proven, 62 probable (Supplementary Tables S1 and S3) and 34 possible Lpp (Supplementary Tables S2 and S4). As such putative Lpp represent ca. 2.5% of the *M. tuberculosis* predicted proteome. Forty five of the 99 putative Lpp identified in the *M. tuberculosis* genome had a homologue (Supplementary Tables S1-S4) in the highly degenerated genome of *M. leprae* [36], including Rv0344c which has extensive amino acid identity (67% over 179 amino acids) to an unannotated *M. leprae* sequence. Twenty three out of the 99 putative Lpp identified were homologous to pseudogenes in the *M. leprae* genome (Supplementary Tables S1-S4). The putative Lpp sequences were subjected to functional categorisation following BLAST sequence analysis [37] and analysis of sequence motifs and patterns including reference to the annotation at Swiss-Prot/TrEMBL, the curated Tuberculist server (http://genolist.pasteur.fr/TubercuList/index.html) and the Pfam database ([38], http://www.sanger.ac.uk/Software/Pfam/).

GenTHREADER ([39], http://bioinf.cs.ucl.ac.uk/psipred/) was also used to support some predictions. Sequence alignments were made using ClustalW [40] accessed at Pôle Bio-Informatique Lyonnais (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_clustalw.html).

4. Functional categorisation of putative Lpp in *M. tuberculosis*

4.1. Solute binding proteins (SBP) of ABC transport systems

Lpp SBP are abundant in the genomes of Gram-positive bacteria as components of ABC transport systems [14,28,41] and those of *M. tuberculosis* have been catalogued previously [42]. SBP are unique to prokaryotic ABC importer systems and deliver substrates to membrane-located permeases prior to transport into the cell [41-43]. The present analysis identified 17 likely Lpp SBPs (Table 1) including 3 (Rv1166; Rv1244 and Rv2585c) that had not been identified by Braibant et al. [42]. 11 of these SBP genes were adjacent to genes encoding permease and ATP-binding components of typical putative ABC transport systems [42]. Several of these SBP have been investigated in relation to their potential role in virulence. PstS-1 and PstS-3 have received attention as vaccine candidates, following delivery as either protein or as DNA vaccines [44-47] whilst the ModA SBP is of interest as a signature-tagged transposon mutagenesis study found that *modA* disruption attenuated survival in the mouse lung [48]. Inactivation of the SubI SBP in *Mycobacterium bovis* BCG did not significantly attenuate the growth of mutants during mouse infection studies [49,50]. However, analysis of the growth of transposon mutants has indicated that *subI* is essential for optimal growth on defined media [51].

This study also suggested that putative SBPs for peptides (Rv1166; Rv3666c) and sugars (Rv2041; Rv2833) are needed for optimal growth in vitro.

Of the six apparently 'orphan' putative SBP that are not clearly associated with typical permease systems, GlnH (Rv0411c) has been shown to be necessary for growth in vitro [51] and may interact with ABC transporter components (Rv2563-GlnQ or Rv0072-Rv0073) located elsewhere on the chromosome [42]. The permease systems with which the three newly identified SBPs may interact cannot be identified, although it is intriguing that all are putative SBPs for amino acid or peptide substrates: as three complete ABC transport systems and the GlnH system(s) are present for such substrates [42] these SBPs could possibly interact with components encoded by other chromosomal loci.

The remaining two orphan SBPs (FecB2/Rv0265c and FecB/Rv3044) are both putative SBPs for iron(III)-siderophore substrates and, in the absence of other ABC transporter components for this substrate family, their functions remain unclear. However, it is possible that these Lpp could interact with either the membrane-associated mycobactin or the secreted carboxymycobactin forms of the *M. tuberculosis* iron-chelating siderophores and thereby participate in either iron transport or other aspects of iron homeostasis [52].

4.2. Lipoprotein Enzymes

4.2.1. Enzymes involved in cell wall metabolism

Given the predicted localisation of Lpp at the interface of the cell membrane and the peptidoglycan layer, it was unsurprising that seven putative Lpp were identified

with functions that may relate to cell wall metabolism (Table 1). These include the putative transpeptidase (penicillin-binding protein) Rv2864c and the putative peptidases Rv0399c and Rv1922. Rv0838 (LpqR) is a putative D-Ala D-Ala dipeptidase that is homologous to proteins in the VanX family that participate in glycopeptide resistance, including VanX of the vancomycin producer *Amycolatopsis orientalis* [53]. However, Rv0838 is not encoded as part of a typical glycopeptide resistance gene cassette and so its role remains unclear. Cumulatively, it can be speculated that these four putative Lpp may play roles in peptidoglycan crosslinking and remodelling and so it is notable that *Rv0399c* has been shown to be necessary for optimal growth in defined media [51].

The three other putative Lpp included in this category may contribute to cell wall metabolism or resistance to β-lactam antibiotics. *Rv2068c* encodes a previously characterised β-lactamase [54]. β-lactamase Lpp are present in other Gram-positive bacteria [9,10]. Rv3593 is homologous to ORF12 from the clavulanic acid biosynthetic cluster of *Streptomyces clavuligerus* [55] and contains an SxxK motif typical of the acyltransferase superfamily [56]. As for Rv0399c, a likely housekeeping role for this protein is suggested by the recent observation that *Rv3593* is necessary for optimal growth in vitro [51]. Likewise, Rv2905 is noted to exhibit similiarities to β-lactamases and contains an SxxK motif [56].

4.2.2. Degradative enzymes

Several putative Lpp are predicted to be degradative enzymes including esterases (Rv0671 and Rv3298c, which are 36% identical to each other);

proteases/peptidases (Rv0418, Rv0419, Rv2224c and Rv2672); a putative phosphorylase (Rv2293c) and a family 3 glycosyl hydrolase (Rv0237). Although the specific substrates of these putative enzymes remain to be determined, their predicted localisation to the *M. tuberculosis* pseudoperiplasm suggests they could be involved in nutrient metabolism. The Rv2224c putative protease is 38% identical to SlpD, an apparently essential putative Lpp protease of *Streptomyces lividans* [57] and it is notable that transcription of *Rv2224c* was induced ca. four-fold within infected macrophages [58].

4.2.3. Other enzymes and metabolic activities

One of the few well characterised Lpp of *M. tuberculosis* is the Cu,Zn superoxide dismutase (SodC, Rv0432; [59,60]). This protein was demonstrated to be lipidated by radiolabelling of the protein with palmitic acid when expressed in *E. coli* and by comparison of the recombinant protein and the native protein on non-denaturing activity gel electrophoresis [60]. Consistent with this SodC has been shown by immunogold electron microscopy to be localised to the cell envelope in *M. tuberculosis* [59] although it is difficult to determine if this represents localisation to the plasma membrane or the surface layers of the organism. SodC was also shown to be induced following phagocytotic uptake by macrophages. Investigation of a *sodC* deletion mutant has shown that this enzyme contributes to the ability of *M. tuberculosis* to resist oxidative stress either in culture or following phagocytosis into induced murine peritoneal macrophages in vitro [61], although a *sodC* mutant was unaffected for growth in activated murine bone marrow macrophages and in a guinea pig model of infection [62]. Cumulatively, these data suggest that cell envelope-

localised SodC could represent an important front-line defence against oxidative stress during intramacrophage growth. However, it is also notable that *sodC* transcription may be switched off as *M. tuberculosis* enters the persistant phase associated with time points longer than 20 days in the mouse model of lung infection [63].

Rv3390 is a putative phosphoglycerate/bisphosphoglycerate mutase family member (Pfam PF00300). The *M. tuberculosis* H37Rv genome encodes nine putative members of this family and, aside from the participation of phosphoglycerate mutase in glycolysis, the functions of these proteins are unclear. The majority are cytoplasmic but the Rv3390 putative Lpp and Rv0754, a member of the PE-PGRS family [3], are predicted to be exported.

Rv2394 is a putative Lpp homologue of the periplasmic γ-glutamytransferase of *E.coli* which catalyses the cleavage of glutathione [64]. Glutathione is a highly abundant thiol tripeptide that acts as an antioxidant in both bacterial and mammalian cells. Cleavage of excreted glutathione by periplasmic γ-glutamytransferases may act as a mechanism for recycling the amino acid constituents as nutrients [64]. However, in actinomycetes such as *M. tuberculosis*, the most abundant low molecular weight thiol antioxidant is mycothiol and glutathione is absent [65]. Thus Rv2394, with its predicted pseudoperiplasmic location, could be involved in metabolising glutathione derived from the host. This could represent a nutrient acquisition or signalling pathway as the adjacent ORF *Rv2395* encodes a putative integral membrane protein of the OPT superfamily, members of which are possibly transporters for peptide signalling molecules [66].

Several putative Lpp have likely roles in redox reactions (Table 1). Of these, the Rv0132c possible Lpp has been annotated as a putative F_{420}-dependent glucose-6-phosphate dehydrogenase due to its homology with an *M. smegmatis* enzyme [67]. However, the most significant homology resides in the N-terminal domain which is likely to interact with the F_{420} coenzyme [67] whereas their C-terminal regions show much lower homology. Thus it would seem prudent to consider this enzyme a putative F_{420}-dependent oxidoreducase until its substrate specificity has been directly demonstrated.

The *M. tuberculosis* genome encodes two putative thioredoxin Lpp of unknown function. Firstly, the possible Lpp Rv0526 is a thioredoxin-like protein and the adjacent ORF, Rv0525 encodes a putative cytochrome c biogenesis protein. Thus, these two domains may interact to allow transfer of electrons across the cytoplasmic membrane and onto, as yet unknown, acceptors in the pseudoperiplasmic compartment, in a manner analogous to the action of the β and γ domains within the integral membrane protein DsbD in *E. coli* [68]. The *Rv0524-Rv0529* locus has been shown to be necessary for growth in vitro[51]. The other putative Lpp thioredoxin (Rv1677, DsbF; thiol:disulphide interchange protein) exhibits high homology (54% identity over amino 137 acids) with the secreted antigen MPT53 (Rv2878c), which also contains a CXXC active site and which has recently been demonstrated to act as an oxidant [69].

M. tuberculosis possesses a putative Lpp multi-copper binding oxidase, Rv0846c. Recently, evidence has been presented that periplasmic multi-copper oxidases of gram-negative bacteria may act as ferroxidases, oxidising Fe(II) to Fe(III)

prior to Fe(III) uptake [70]. Combined ferroxidase/Fe(III) transport systems could thus represent an important route for iron acquisition by pathogens. Thus Rv0846c may be localised to the *M. tuberculosis* pseudoperiplasm in order to participate in iron metabolism (see also Section 4.1. above).

4.3. Lpp with putative roles in adhesion and cell invasion

4.3.1. MPT83 (Rv2873), a putative adhesin

One of the most extensively studied mycobacterial Lpp is the MPB83 protein of *Mycobacterium bovis* BCG which has been characterised by Harboe, Wiker and co-workers. MPB83 exhibits extensive sequence homology to the secreted antigen MPB70 [71]. However, the *mpb83* gene sequence was noted to encode a putative Lpp [72] and convincing evidence for lipidation of MPB83 has been presented [73,74]. The *mpb83* gene was also recognised to be located close to the *mpb70* gene in the *M. bovis* BCG genome and their levels of expression are linked [72,73,75]. MPB83 has been immunolocalised to the surface of *M. bovis* BCG by both electron microscopy [73] and flow cytometry [74]. Thus, the MPB83 and MPB70 proteins represent highly homologous but differently localised proteins [74]. However, it has also been noted that MPB83 is released from the cells as both a mature, lipidated 25-26 kDa form and as a hydrophilic 22-23 kDa form [71,74-76]. N-terminal sequencing of secreted forms of MPB83 from *M. bovis* culture supernatants indicates proteolytic cleavage of the MPB83 following residues Ser3 and Val23 of the mature Lpp sequence [77].

M. tuberculosis produces identical homologues of both MPB83 and MBP70, designated MPT83 (Rv2873, Table 1) and MPT70 (Rv2875) respectively [75]. Thus the evidence that MPB83 is a Lpp provides strong evidence that MPT83 is also likely to be a Lpp and it is notable that recombinant MPT83 remained cell-associated after cloning into *M. smegmatis* [75]. The previously noted correlation in the level of expression of MPB83 and MPB70 has been confirmed by genetic analysis of transcription on the *Rv2871-Rv2875* locus [78].

Several structural features of MPT83/MPB83 are notable in comparison to their MPT70/MPB70 homologues. Firstly, the mature MPT83/MPB83 proteins are not only anchored by lipid modification but also possess an N-terminal sequence extension of 32 amino acids. This sequence, which is absent from their secreted homologues, contains a threonine motif ($T_{48}T_{49}$) that permits O-linked trimannosylation in MPB83 [77]. Secondly, both the MPT83 and MPT70 sequences contain cysteine residues that could form an internal disulphide bond [79] and their highly homologous regions belong to the β-Ig-H3/fascilin family (Pfam PF02469). This domain occurs in a variety of proteins from diverse taxa (e.g. man, plants, fungi, bacteria) and in many cases is present in multiple repeats [80]. MPT83 and MPT70 each contain a single copy of the domain. Several eukaryotic members of the fascilin family are adhesins, notably osteoblast-specific factor 2, OSF-2. MPT83 exhibits ca. 30% amino acid identity with domains in OSF-2 and consequently it has been suggested that MPT83/MPT70 may be an adhesin involved in bone tropism [75,81]. However, no direct role for MPT83 in bone adhesion has been demonstrated and the proposed link between MPB70/MPB83 and post-BCG vaccination osteitis [75,81] has not yet been substantiated. Thus MPT83 should be considered a putative adhesin for an as yet

unknown ligand. Whether MPT83/MPT70 function involves other cotranscribed genes such as the gene product of the intervening *Rv2874*, which is a DipZ family protein [78], also remains to be determined.

4.3.2. Putative lipoprotein members of the Mce family

In 1995 Arruda et al. [82] identified a *M. tuberculosis* locus, *mce*, that conferred on *E. coli* the ability to invade mammalian cells. Subsequently the *M. tuberculosis* genome sequence revealed that this locus was in fact part of an operon (*Rv0167-Rv0174*) comprising eight genes of which *mce* was the third. Moreover, three comparable operons were also evident elsewhere in the genome (*Rv0587-RvRv0594*; *Rv1964-Rv1972* and *Rv3501c-Rv3494c* respectively; [3]). In each case, the *mce* homologue is preceded by genes encoding putative membrane proteins and is followed by five genes that may have signal peptides or N-terminal hydrophobic sequences including a putative Lpp as the penultimate gene of each operon (Table 1). Moreover, the corresponding gene in the single *M. leprae* Mce operon (Mce1E; ML2593) encodes a putative Lpp [83].

All the MceA-F proteins from each operon contain an N-terminal region of sequence homology that is documented in the PFAM database entry PF02470. However, the functional inter-relationships of these proteins and their contribution to virulence remains unclear. The archetypal Mce1A is clearly linked with a mammalian cell invasion phenotype [8284,85] and deletion of part of the N-terminal region common to the Mce family abolished the ability of recombinant Mce1A to direct uptake of latex beads into HeLa cells [85]. However, latex beads coated in Mce2A were not taken up into HeLa cells [85]. Moreover, Mce operons are present in both

non-pathogenic mycobacteria [86] and other bacteria, notably *S. coelicor* (Sco2414-Sco2421; [87]). Although the link between the function(s) of the Mce operons and virulence is not clear, it is apparent that the putative Mce Lpp are expressed and antigenic in vivo as sera from 4/10 tuberculosis patients (but none of 10 sera from BCG-vaccinated healthy controls) cross-reacted with recombinant Mce1E [88]. Although the Mce3 operon is deleted in *M. tuberculosis* complex organisms other than *M. tuberculosis* and *Mycobacterium canetti* [89], the *Mycobacterium avium* homologue of Mce3E (Rv1970) has also been shown to be expressed early during growth in macrophages [90].

4.4. Lipoproteins with putative roles in signalling and sensory functions

4.4.1. Rv1368 (LprF) and Rv1690 (LprJ), accessory proteins to the KdpD potassium sensor kinase

Using post-genomic technologies, the putative lipoproteins LprF (Rv1368) and LprJ (Rv1690) were recently identified as proteins that interact with the KdpD histidine kinase in the potassium-dependent sensing of osmotic stress [91]. Mutagenesis of the LprJ lipobox was reported to affect LprJ localisation, providing evidence that this protein is indeed a Lpp. However, Steyn et al. [91] predicted topologies for LprF and LprJ that are not supported upon re-analysis with a wider range of prediction tools: each can be instead predicted to have a typical Lpp topology i.e. a membrane anchor and an extracytoplasmic domain (our data not shown). Such topology is as consistent with the fusion protein localisation data of Steyn et al. as those proposed previously [91]. Moreover, it allows for the predicted formation of an intramolecular disulphide in

LprJ and its homologues (see below). Cumulatively, it may be hypothesised that LprF and LprJ are membrane localised and that these proteins possibly act as sensors/receptors for signals in the extracytoplasmic region that interact with the integral membrane protein KdpD. This proposed interaction between extracytoplasmic components and the sensor kinase is analagous to that proposed for the interaction of the KapB lipoprotein with the KinB sensor kinase in *B. subtilis* [92]. However, it is remains difficult to resolve why LprF and LprJ were found to interact specifically with the putative cytoplasmic N-terminal domain of KdpD, as shown by yeast hybrid screens and SELDI-TOF mass spectrometry [91]. Clearly, the topologies of KdpD, LprF and LprJ require clarification.

The LprJ and LprF lipoproteins are themselves each representative of a family of *M. tuberculosis* proteins. LprF belongs to a family [93,94] including three other putative Lpp (Rv1270c, Rv1411c and Rv2945c) which exhibit relatively low overall amino acid sequence homology (data not shown). The *M. bovis* homologue of Rv2945c has been cell surface localised by flow cytometry [95] and Rv1411c (LprG) is processed as a lipoprotein when cloned in *E. coli* [96]. Given the proposed sensor role of LprF described above, it is interesting that Rv1411c (LprG) has previously proposed to act as a sensor for the Rv1410c P55 antibiotic efflux pump as the genes for these two proteins are located in an operon [94,97]. A defined mutant in this operon has been demonstrated to be attenuated in a BALB/C mouse model of infection [98]. Rv1411c induced strong delayed-type hypersensitivity and Th1-type immune responses in immunised mice [96]. However, mice vaccinated with Rv1411c gave an unfavourable response in subsequent infectious challenge experiments [96].

Likewise, DNA vaccination of mice with the *M. bovis* homologue of Rv2945c was unsuccesful in conferring protective immunity [95].

LprJ is a member of a family of eight proteins in *M. tuberculosis* that exhibit very low amino acid sequence homology but which are characterised by the presence of two conserved cysteines. In contrast to LprJ, the other seven members of this family are predicted to be exported proteins. In each case, one of the conserved cysteines (Cys73 in LprJ) is central to the mature protein sequence whilst the other (Cys113 in LprJ) is part of a conserved YCP motif near the C-terminus. It is clearly a possibility that these two conserved cysteines may form a disulphide bond that would have a major influence on protein folding and, in the case of LprJ, topology. In this regard, it is significant that fusion proteins containing only amino acids 1-99 of LprJ (i.e. fusions that would disrupt the proposed Cys73-Cys113 disulphide bond) gave ambiguous localisation results [91].

4.4.2. The Rv1009 growth-promoting factor

Study of the Rpf resuscitation-promoting factor of *Micrococcus luteus* has led to the understanding that actinomycete bacteria secrete proteinaceous growth stimulating factors [99,100]. The *M. tuberculosis* genome contains five genes which encode proteins containing highly conserved Rpf-like domains [101]. Unsurprisingly, four of these are predicted to be secreted proteins, whereas RpfB (Rv1009) is a putative Lpp. The mature domains of all five proteins were produced as recombinant proteins and shown to stimulate in vitro growth of *M. bovis* BCG when added exogenously in picomolar quantities and when late stationary phase cells were used as an inoculum [101,102]. The mature protein sequence of Rv1009 contains three

tandem repeats of a domain of unknown function (DUF348; Pfam PF03990) in the N-terminal half of the protein and the ca. 70 amino acid Rpf-domain represents the extreme C-terminus. It is plausible that the growth stimulatory properties of this domain are dependant in vivo on its release by proteolytic cleavage. The inter-relationship between these proteins and their growth promoting activities is clearly an area that demands further study, especially in the context of the ability of *M. tuberculosis* to enter and subsequently emerge from a persistant (latent) state in vivo. However, whilst *rv1009* belongs to the sub-set of *M. tuberculosis* genes whose mutation by transposon-insertion mutagenesis led to slow growth on defined laboratory media [51], a subsequent study of a defined *rv1009* deletion mutant revealed a small colony phenotype but no observable defect in growth and persistence in a mouse model of infection [103].

4.4.3. Other putative Lpp that may be involved in cell signalling or sensory systems

Rv1911c belongs to a family of microbial proteins that include the *E. coli* periplasmic protein YcbL and which are similar to eukaryotic phosphatidyl ethanolamine-binding proteins that are important in cellular signalling [104]. Rv1911c is highly homologous to the product of the adjacent *Rv1910c* gene, which encodes a predicted exported protein. The substrates bound by the different members of this bacterial protein family are likely to vary but are probably phosphorylated substrates, including phospholipid head-groups [104,105]. Thus the Rv1911c and Rv1910c proteins may be associated with the outer face of the plasma membrane.

M. tuberculosis was one of the first bacteria noted to utilise a family of eukaryotic-like serine/threonine protein kinases that are proposed to respond to environmental signals via extracytoplasmic C-terminal sensory domains [3,106]. The putative sensory domain of one of these kinases, PknH, exhibits significant sequence homology to the Rv2403c and Rv3576 (PknM) putative Lpp. Thus these two proteins may act as receptor proteins that interact with signalling systems.

4.5. Lipoproteins of unknown function

4.5.1. Rv3763, the 19 kDa antigen

Probably the most extensively studied Lpp of *M. tuberculosis* is the 19 kDa antigen (Rv3763, Table ; [107]), for which convincing experimental evidence of lipidation has been presented [22]. This antigen has homologues in other mycobacteria including *M. avium*, M*. intracellulare* and *M. leprae* [108,109]. However, the 19-kDa antigen was found to be absent from 2 out of 9 strains of *M. tuberculosis* [110]. In addition to its lipidation, evidence has been presented that the 19-kDa antigen is glycosylated and that this is dependant on five threonine residues located within an 8 amino acid sequence (T13-T20) at the N-terminus of the mature protein [111,112]. Glycosylation may be a protective mechanism for retaining the 19 kDa antigen in its membrane-associated location ([112]; see section 6 below). Despite extensive study of the immunobiology of the 19-kDa antigen, including its interaction with Toll-like receptor 2 and intramacrophage trafficking of released antigen (for examples see [113-118]), the function of this protein remains unknown.

4.5.2. Lipoprotein homologues of known mycobacterial antigens

The Rv0583c (LpqN) putative Lpp of *M. tuberculosis* is highly homologous to the MK35 antigen of *Mycobacterium kansasii* [119]. This latter protein was suggested to be a Lpp by Triton X-114 detergent partitioning studies and was shown to be strongly immunogenic in guinea pig delayed-type hypersensitivity tests. Despite the sequence homology between Rv0583c and MK35, it was noted that no reaction to recombinant MK35 was observed when the guinea pigs had been sensitised with *M. tuberculosis* H37Rv. A second *M. tuberculosis* putative Lpp, Rv1016c , exhibits distant sequence homology to Rv0583c and significant homology with the exported proline-rich antigen MTC28 [120]. Similarly, Rv2116 (LppK) exhibits significant sequence homology to the MTB12 (Rv2376c) secreted antigen [121].

4.5.3. The LppA paralogue family and other inter-related lipoproteins of unknown function

Although *M. tuberculosis* has been considered to exhibit relatively little genetic diversity it is clear from whole genome comparisons that this diversity may have been underestimated [3,31]. One interesting polymorphism identified involves a family of closely related Lpp paralogues. The *M. tuberculosis* H37Rv genome encodes two adjacent putative Lpp, Rv2543 and Rv2544 (LppA and LppB respectively), that are 87% identical to each other. The corresponding region of the *M. tuberculosis* CDC1551 genome contains both of these paralogues (MT2618 and MT2620) and a third, MT2619. It appears that these three paralogues have arisen by gene duplication followed by loss of the MT2619 sequence from the H37Rv genome [31]. These three

sequences are also distantly related to Rv2796c, another putative Lpp of unknown function (Table 1).

Two putative Lpp, Rv0483 and Rv2518c , belong to the recently described ErfK/YbiS/YcfS/YnhG family (Pfam PF03734). The function of these proteins remains unknown but may relate to the presence of a conserved region containing histidine and cysteine residues.

The Rv0604 (LpqO) and Rv2999 (LppY) probable Lpp are closely related proteins that also have significant homologies with other conserved hypothethical proteins. These proteins appear to consist of a fusion of duplicated domains but their function remains unknown. Aligning these *M. tuberculosis* proteins and their homologues identified a motif F-X(10,14)-G-[DE]-X(6)-E-X(18)-H-X-H-X(5)-P-X(5)-H which is conserved in both the N and C-terminal domains and in three clostridial proteins that each contain only a single copy of the domain.

4.5.4. Other lipoproteins of unknown function

The Rv0679c putative Lpp belongs to a family containing other non-Lpp proteins of *M. tuberculosis*. Thus the protein exhibits significant homology to both the Rv0680c putative exported protein encoded by the adjacent gene and also to the Rv0314c protein. This latter protein is predicted to be a membrane protein with a cytoplasmic N-terminal domain and a C-terminal domain which is homologous to the Rv0679c/Rv0680c proteins. Thus, the predicted localisation of the conserved domains of these three proteins is consistent with an inter-related pesudoperiplasmic function.

In addition to the above, the *M. tuberculosis* genome contains 28 additional putative Lpp of unknown function (Supplementary Tables S1-S4). Of these the majority are conserved hypotheticals although in many cases it is noted that they share significant homology only with sequences within other mycobacterial genomes. Two sequences (Rv0962c and Rv1799) are apparently unique to *M. tuberculosis*. Moreover, four of these proteins of unknown function (Rv1274; Rv2138; Rv2999 and Rv3244c) were shown to be necessary for optimal growth in vitro [51]. Similarly, *Rv1252c* has been identified as an iron-induced gene of unknown function that is independent of IdeR regulation [122].

5. Lipoprotein localisation

It is clear from the above analyses that putative Lpp represent an abundant and functionally diverse sub-set of the *M. tuberculosis* proteome. It is predicted that the lipid modification of these proteins serves to anchor them to the outer face of the plasma membrane, as in other Gram-positive bacteria. However, in the context of contemporary models of the more complex architecture of the mycobacterial cell envelope [4-7], these proteins can be viewed as having a pseudoperiplasmic location, which draws parallels with the cell envelope organisation of Gram-negative bacteria. It remains to be determined if significant numbers of Lpp are associated with the outer mycolate-based lipid layers of the mycobacterial cell envelope. It is, however, noted that there is evidence for secretion or a peripheral localisation of all the well characterised Lpp in *M. tuberculosis*. Thus, the Rv0432 SOD has been immunolocalised to the peripheral layers of *M. tuberculosis* H37Rv [59] and the MPB83 homologue of MPT83 has been localised to the surface of *M. bovis* BCG (see

section 4.3.1 above; [73,74]). Moreover, the PstS SBPs have been detected at the cell surface of *M. bovis* BCG by flow cytometry [123] and of *M. tuberculosis* by immunogold electron microscopy [124]. Since these SBPs are believed to deliver phosphate to plasma membrane-associated ABC permease components [42], the apparent surface association of these proteins is unexpected. Likewise the *M. bovis* BCG homologue of Rv2945c (LppX; section 4.4.1.) has been surface-localised by flow cytometry and this antigen is also released into culture supernatants [95]. Cumulatively these results may reflect the release of acylated Lpp ('shedding') or proteolytic cleavage downstream of the lipidated cysteine ('shaving'), as proposed previously for *B. subtilis* Lpp [125]. Evidence for proteolytic release of the Pst-S1 Lpp has been obtained by N-terminal sequencing of antigen recovered from culture supernatants, which suggested a cleavage site preceding Ser3 of the mature Lpp sequence [126]. In addition, proteolytic release of the MPB83 antigen has been observed in *M. bovis* [77] and proteolytic release of the 19 kDa antigen was observed after cloning into *M. smegmatis* [112]. However, it is also clear that intact 19 kDa antigen can be released from cells since sub-cellular trafficking of this antigen within macrophages was directed by an acylation-dependant pathway distinct from that followed by live mycobacteria [117]. Thus, lipid modification may serve the primary purpose of retaining proteins at the mycobacterial plasma membrane but there may also be alternative pathways of Lpp processing that lead to their localisation within other subcellular compartments of either the bacterium or host cells.

It is also apparent from the above analyses that a significant number of putative Lpp have significant paralogues in the *M. tuberculosis* secretome (Table 2). Whether

this represents a form of functional compartmentalisation is an interesting question for future study.

6. Glycosylation of lipoproteins

It is now clear that, like various other bacteria, *M. tuberculosis* is able to glycosylate proteins [77,112,127] and these glycoproteins include the 19 kDa antigen (section 4.5.1) and the MPB83 Lpp (Section 4.3.1). In each of the chemically-characterised mycobacterial glycoproteins O-mannosylated threonine residues are found in the proximity of proline residues but a precise sequence motif that directs glycosylation cannot yet be defined [77,127]. Post-translational glycosylation may be linked to protein export as all three well-characterised glycoproteins of *M. tuberculosis* are exported. Moreover, an *M. smegmatis* expression system and Concanavilin A lectin binding assays provided experimental support for bioinformatic predictions that many other putative Lpp may also be glycosylated [128]. These data suggested that the Rv0432 superoxide dismutase, the Pst-S1 SBP and four other putative Lpp (Rv0411c; Rv1541c; Rv2270; Rv2341) contain sequence motifs that can direct mannosylation in a heterologous mycobacterial host.

The function of glycosylation remains unclear at present. It has been proposed that glycosylation of the 19 kDa antigen protects a proteolytically sensitive cleavage site so that the intact protein is retained by its lipid anchor [112]. However, glycosylation at the $T_{48}T_{49}$ motif does not prevent the proteolytic release of the 23-kDa form of the MPB83 antigen [77].

7. Concluding comments

The availability of genome sequences for strains of *M. tuberculosis* [3,31] provides a major resource to underpin research on the mechanisms of virulence in this devastating bacterial pathogen. The present study suggests that putative Lpp could represent as much as 2.5% of the *M. tuberculosis* proteome, consolidating and extending the original analysis of Cole et al. [3]. Thus, Lpp are likely to represent a significant class of cell envelope proteins involved in interactions between the organism and the host. This study has provided a functional categorisation of these putative Lpp which identifies many inter-related protein sequences, including both putative Lpp families and relationships between putative Lpp and exported proteins. It is hoped that this analysis will provide the basis for novel lines of investigation into the biology of *M. tuberculosis*.

Acknowledgements

The authors are grateful to the Wellcome Trust and the Horserace Betting Levy Board for their support of work in our laboratories on bacterial lipoproteins. We thank Noel Carter (University of Sunderland) for extensive discussions on the topologies of LprF and LprJ.

Addendum

The significance of Lpp to the virulence of *M. tuberculosis* has been confirmed recently by the important findings of Sander et al. [129] that inactivation of the *Rv1539 Lsp* gene (encoding the lipoprotein signal peptidase) impaired the ability of the mutant to replicate in cultured mouse macrophages and led to attenuation in a BALB/c

mouse model of infection. Processing of the signal peptides of Rv1411c (LprG) and MPT83 were also disrupted in the *lsp* mutant, confirming the predictions that these proteins are indeed Lpp.

References

[1] Dye, C., Scheele, S., Dolin, P., Pathania, V. and Raviglione, M.C. for the WHO Global Surveillance and Monitoring Project (1999). Global burden of tuberculosis. Estimated incidence, prevalence, and mortality by country. J.A.M.A. 282, 677-686.

[2] Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C. and Dye, C. (2003) The growing burden of tuberculosis - Global trends and interactions with the HIV epidemic. Arch. Intern. Med. 63, 1009-1021.

[3] Cole S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry ,C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature, 393 537-544.

[4] Minnikin, D. E. (1982). Lipids: Complex lipids, their chemistry, biosynthesis and roles. In: The Biology of the Mycobacteria (Ratledge, C. and Stanford, J.L., Eds.), pp 95-184. Academic Press, London

[5] Brennan, P.J. and Nikaido, H. (1995). The envelope of mycobacteria. Ann. Rev. Biochem. 64, 29-63.

[6] Daffe, M. and Draper, P. (1998). The envelope layers of mycobacteria with reference to their pathogenicity. Adv. Microb. Physiol. 39, 131-203.

[7] Dmitriev, B.A, Ehlers, S., Rietschel, E.T. and P. J. Brennan (2000). Molecular mechanics of the mycobacterial cell wall: from horizontal layers to vertical scaffolds. Int. J. Med. Microbiol. 290, 251-258.

[8] Braun, V. and Wu, H. C. (1994). Lipoproteins, structure, function, biosynthesis and model for protein export. New Comprh. Biochem. 27, 319-341.

[9] Sutcliffe, I.C. and Russell, R. R. B. (1995). Lipoproteins of Gram-positive bacteria. J. Bacteriol. 177, 1123-1128.

[10] Nielsen, J.B.K. and J.O.Lampen (1982). Glyceride-cysteine lipoproteins and secretion by Gram-positive bacteria. J. Bacteriol. 152, 315-322.

[11] Klein, D., Somorja, R. L. and Lau, P. C. K. (1988). Distinctive properties of signal sequences from bacterial lipoproteins. Prot. Eng. 2, 15-20.

[12] von Heijne, G. (1989). The structure of signal peptides from bacterial lipoproteins. Prot. Eng. 2, 531-534.

[13] Tjalsma, H., Bolhuis, A., Jongbloed, J. D.H., Bron, S. and van Dijl, J. M. (2000). Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol. Mol. Biol. Rev. 64, 515-547.

[14] Sutcliffe, I.C. and Harrington, D.J. (2002). Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes. Microbiol. 148, 2065-2077.

[15] Leskelä, S., Wahlstrom, E., Kontinen, V. P. and Sarvas, M. (1999). Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in *Bacillus subtilis*: characterization of the *lgt* gene. Mol. Microbiol. 31, 1075-1085.

[16] Petit, C. M., Brown, J. R., Ingraham, K., Bryant, A. P. and Holmes, D. J. (2001). Lipid modification of prelipoproteins is dispensable for growth in vitro but essential for virulence in *Streptococcus pneumoniae*. FEMS Microbiol. Lett. 200, 229-233.

[17] Venema, R., Tjalsma, H., van Dijl, J.M., de Jong, A., Leenhouts, K., Buist, G. and Venema, G. (2003). Active lipoprotein precursors in the Gram-positive eubacterium *Lactococcus lactis*. J. Biol. Chem. 278, 14739-14746.

[18] de Greeff, A., Hamilton, A., Sutcliffe, I.C., Buys, H., van Alphen, L. and Smith, H.E. (2003). Lipoprotein signal peptidase of *Streptococcus suis* serotype 2. Microbiol. 149, 1399-1407.

[19] Réglier-Poupet, H., Frehel, C., Dubail, I., Beretti, J-L., Berch, P., Charbit, A. and Raynaud, C. (2003). Maturation of lipoproteins by Type II signal peptidase is required for phagosomal escape of *Listeria monocytogenes*. J. Biol. Chem. 278, 49469-49477.

[20] Coulter, S.N., Schwan, W.R., Ng, E.Y.W., Langhorne, M.H., Ritchie, H.D., Westbrock-Wadman, S., Hufnagle, W.O., Folger, K.R., Bayer, A.S. and Stover, C.K. (1998). *Staphylococcus aureus* genetic loci impacting growth and survival in multiple infection models. Mol. Microbiol. 30, 393-404.

[21] Mei, J-M., Nourbakhsh, F., Ford, C.W. and Holden, D.W. (1998). Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Mol. Microbiol. 26, 399-407.

[22] Young, D. B. and Garbe, T. R. (1991). Lipoprotein antigens of *Mycobacterium tuberculosis*. Res. Microbiol. 142, 55-65.

[23] Gupta, S.D., Gan, K., Schmid, M.B. and Wu, H.C. (1993). Characterization of a temperature-sensitive mutant of salmonella-typhimurium defective in apolipoprotein n-acyltransferase. J.B. Chem. 268, 16551-16556.

[24] Gurcha, S.S., Baulard, A.R., Kremer, L., Locht, C., Moody, D.B., Muhlecker, W., Costello, C.E., Crick, C.R., Brennan, P.J. and G.S. Besra (2002). Ppm1, a novel polyprenol monophosphomannose synthase from of *Mycobacterium tuberculosis*. Biochem. J. 365, 441-450.

[25] Falquet, L., Pagni M., Bucher P., Hulo N., Sigrist C.J, Hofmann K. and Bairoch A. (2002) The PROSITE database, its status in 2002. Nucleic Acids Res. 30, 235-238.

[26] Chambaud, I., Wróblewski, H. and Blanchard, A. (1999). Interactions between mycoplasma lipoproteins and the host immune system. Trends Microbiol. 7, 493-499.

[27] Babu, M.M. and Sankaran, K. (2002). DOLOP - database of bacterial lipoproteins. Bioinformatics 8, 641-643.

[28] Sutcliffe, I.C. and Harrington, D. J. (2004). Putative lipoproteins of *Streptococcus agalactiae* identified by bioinformatic genome analysis. Anton. van Leeuwen. 85, 305-315.

[29] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucl. Acids Res. 31, 365-370.

[30] Gattiker, A., Gasteiger, E. and Bairoch A. (2002). ScanProsite: a reference implementation of a PROSITE scanning tool. Appl. Bioinform. 1:107-108.

[31] Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs, W.R., Venter, J.C. and Fraser, C.M. (2002). Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. J. Bacteriol.184, 5479-5490.

[32] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Prot. Eng. 10, 1-6.

[33] Nielsen, H. and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. In: Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pp.122-130. AAAI Press, California.

[34] Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305: 567-580.

[35] Wiker, H.G., Wilson, M.A. and Schoolnik, G.K. (2000). Extracytoplasmic proteins of Mycobacterium tuberculosis - mature secreted proteins often start with aspartic acid and proline. Microbiol. 146, 1525-1533.

[36] Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor. K., Whitehead, S., Woodward, J.R. and Barrell BG. (2001). Massive gene decay in the leprosy bacillus. Nature, 409, 1007-1011.

[37] Altschul, S. F., Madden, T.L., Schaffer, A. A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids. Res. 25, 3389-3402.

[38] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L.L. (2002). The Pfam Protein Families Database. Nucl. Acids Res. 30, 276-280.

[39] McGuffin, L.J. and Jones, D.T. (2003). Improvement of the GenTHREADER method for genomic fold recognition. Bioinformatics 19, 874-881.

[40] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22, 4673-4680.

[41] Quentin, Y., Fichant, G. and Denizot, F. (1999). Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. J. Mol. Biol. 287, 467-484.

[42] Braibant, M., Gilot, P. and Content, J. (2000). The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis. FEMS Microbiol. Rev. 24, 449-467.

[43] Higgins, C.F. (2001). ABC transporters: physiology, structure and mechanism – an overview. Res. Microbiol. 152, 205-210.

[44] Tanghe, A., Lefevre, P., Denis, O., D'Souza, S., Braibant, M., Lozes, E., Singh, M., Montgomery, D., Content, J. and Huygen, K. (1999). Immunogenicity and protective efficacy of tuberculosis DNA vaccines encoding putative phosphate transport receptors. J. Immunol. 162, 1113-1119.

[45] Fonseca, D.P.A.J., Benaissa-Trouw, B., Van Engelen, M., Kraaijeveld, C.A., Snippe, H., Verheul, A.F.M. (2001). Induction of cell-mediated immunity against *Mycobacterium tuberculosis* using DNA vaccines encoding cytotoxic and helper T-cell epitopes of the 38-kilodalton protein. Infect. Immun. 69, 4839-4845.

[46] da Fonseca, D.P.A.J., Frerichs, J., Singh, M., Snippe, H. and Verheul, A.F.M. (2001). Induction of antibody and T-cell responses by immunization with ISCOMS containing the 38-kilodalton protein of Mycobacterium tuberculosis. Vaccine 19, 122-131.

[47] D'Souza, S., Rosseels, V., Denis, O., Tanghe, A., De Smet, N., Jurion, F., Palfliet, K., Castiglioni, N., Vanonckelen, A., Wheeler, C. and Huygen, K. (2002). Improved tuberculosis DNA vaccines by formulation in cationic lipids. Infect. Immun. 70, 3681-3688.

[48] Camacho, L.R., Ensergueix, D., Perez, E., Gicquel, B. and Guilhot, C. (1999). Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol. Microbiol. 34, 257-267.

[49] McAdam, R.A., Weisbrod, T.R., Martin, J., Scuderi, J.D., Brown, A.M, Cirillo, J.D., Bloom, B.R. and Jacobs, W.R. (1995). In-vivo growth-characteristics of leucine and methionine auxotrophic mutants of *Mycobacterium bovis* BCG generated by transposon mutagenesis. Infect. Immun. 63, 1004-1012.

[50] Wooff, E., Michell, S.L., Gordon, S.V., Chambers, M.A., Bardarov, S., Jacobs, W.R., Hewinson, R.G. and Wheeler, P.R. (2002). Functional genomics reveals the sole sulphate transporter of the *Mycobacterium tuberculosis* complex and its relevance to the acquisition of sulphur in vivo. Mol. Micro. 43, 653-663.

[51] Sassetti, C.M., Boyd, D.H., Rubin, E.J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. 48, 77-84.

[52] Rodriguez, G.M. and Smith, I. (2003). Mechanisms of iron regulation in mycobacteria: role in physiology and virulence. Mol. Micro. 47, 1485-1494.

[53] Marshall, C.G., Lessard, I.A.D., Park, I.S. and Wright, G.D. (1998). Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. Anti. Ag. Chemother. 42, 2215-2220.

[54] Hackbarth, C.J., Unsal, I. and Chambers, H.F. (1997). Cloning and sequence analysis of a class A ß-lactamase from *Mycobacterium tuberculosis* H37Ra. Anti. Ag. Chemother. 41, 1182-1185.

[55] Mellado, E., Lorenzana, L.M., Rodríguez-Sáiz, M., Díez, B., Liras, P. and Barredo, J.L. (2002). The clavulanic acid biosynthetic cluster of *Streptomyces clavuligerus*: genetic organization of the region upstream of the *car* gene. Microbiol. 148, 1427-1438.

[56] Goffin, C. and Ghuysen, J-M. (2002). Biochemisty and comparative genomics of SxxK superfamily acyltransferases offer a clue to the mycobacterial paradox: presence of penicillin-susceptible target proteins versus lack of efficiency of penicillin as therapeutic agent. Microbiol. Mol. Biol. Rev. 66, 702-738.

[57] Binnie, C., Butler, M.J., Aphale, J.S., Bourgault, R., DiZonno, M.A., Krygsman, P., Liao, L., Walczyk, E. and Malek, L.T. Isolation and characterization of two genes encoding proteases associated with the mycelium of *Streptomyces lividans* 66. J. Bacteriol. 177, 6033-6040.

[58] Dubnau, E., Fontan, P., Manganelli, R., Soares-Appel, S. and Smith, I. (2002). *Mycobacterium tuberculosis* genes induced during infection of human macrophages. Infect. Immun. 70, 2787-2795.

[59] Wu, C.H.H., Tsai-Wu, J-J., Huang, Y-T., Lin, C-Y., Lioua, G-G. and Lee, F-J.S. (1998). Identification and subcellular localization of a novel Cu,Zn superoxide dismutase of *Mycobacterium tuberculosis*. FEBS Letts. 439, 192-196.

[60] D'Orazio, M., Folcarelli, S., Mariani, F., Colizzi, V., Rotilio, G. and Battistoni, A. (2001). Lipid modification of the Cu,Zn superoxide dismutase from *Mycobacterium tuberculosis*. Biochem. J. 359, 17-22.

[61] Piddington, D.L., Fang, F.C., Laessig, T., Cooper, A.M., Orme, I.M. and Buchmeier, N.A. (2001). Cu,Zn superoxide dismutase *of Mycobacterium tuberculosis* contributes to survival in activated macrophages that are generating an oxidative burst. Infect. Immun. 69, 4980-4987.

[62] Dussurget, O., Stewart, G., Neyrolles, O., Pescher, P., Young, D. and Marchal, G. (2001). Role of *Mycobacterium tuberculosis* copper-zinc superoxide dismutase. Infect. Immun. 69, 529-533.

[63] Shi, L.B., Jung Y-J., Tyagi, S., Gennaro, M.L. and North, R.J. (2003). Expression of Th1-mediated immunity in mouse lungs induces a *Mycobacterium tuberculosis* transcription pattern characteristic of nonreplicating persistence. Proc.Nat.Acad.Sci. USA 100, 241-246.

[64] Suzuki, H., Hashimoto, W. and H. Kumagai (1999). Glutathione metabolism in *Escherichia coli*. J. Mol. Catal. B: Enzym. 6, 175-184.

[65] Newton, G.L., Arnold, K., Price, M.S., Sherrill, C., Delcardayre, S.B., Aharonowitz, Y., Cohen, G., Davies, J., Fahey, R.C. and Davis, C. (1996). Distribution of thiols in microorganisms: Mycothiol is a major thiol in most actinomycetes. J. Bacteriol. 178, 1990-1995.

[66] Lubkowitz, M.A., Barnes, D., Breslav, M., Burchfield, A., Naider, F. and Becker, J.M. (1998). *Schizosaccharomyces pombe isp4* encodes a transporter representing a novel family of oligopeptide transporters. Mol. Microbiol. 28, 729-41.

[67] Purwantini, E. and Daniels, L. (1998). Molecular analysis of the gene encoding F-420-dependent glucose-6-phosphate dehydrogenase from *Mycobacterium smegmatis*. J. Bacteriol. 180, 2212-2219.

[68] Collet, J.F. and Bardwell, J.C.A. (2002). Oxidative protein folding in bacteria, Mol. Microbiol. 44, 1-8.

[69] Goulding, C.W., Apostol, M.I., Gleiter, S., Parseghiani, A., Bardwell, J., Gennaro, M. and Eisenberg, D. (2004). Gram-positive DsbE proteins function differently from Gram-negative DsbE homologs. A structure to function

analysis of DsbE from *Mycobacterium tuberculosis*. J. Biol. Chem. 279, 3516-3524.

[70] Huston, W.M., Jennings, M.P. and McEwan, A.G. (2002). The multicopper oxidase of *Pseudomonas aeruginosa* is a ferroxidase with a central role in iron acquisition. Mol Microbiol. 45, 1741-1750.

[71] Harboe, M., Nagai, S., Wiker, H.G., Sletten, K. and Haga, S. (1995). Homology between the MPB70 and MPB83 proteins of *Mycobacterium bovis* BCG. Scand. J. Immunol. 42, 46-51.

[72] Matsuo, T., Matsuo, H., Ohara, N., Matsumoto, S., Kitaura, H., Mizuno, A. and Yamada, T. (1996). Cloning and sequencing of an MPB70 homologue corresponding to MPB83 from *Mycobacterium bovis* BCG. Scand. J. Immunol. 43, 483-489.

[73] Vosloo, W., Tippoo, P., Hughes, J. E., Harriman, N., Emms, M., Beatty, D. W., Zappe, H. and Steyn, L. M. (1997). Characterisation of a lipoprotein in *Mycobacterium bovis* (BCG) with sequence similarity to the secreted protein MPB70. Gene 188, 123-128.

[74] Harboe, M., Wiker, H.G., Ulvund, G., Lund-Pedersen, B., Andersen, A.B., Hewinson, R.G. and Nagai, S. (1998). MPB70 and MPB83 as indicators of protein localization in mycobacterial cells. Infect Immun. 66, 289-296.

[75] Hewinson, R.G., Michell, S.L., Russell, W.P., McAdam, R.A. and Jacobs, W.R. (1996). Molecular characterization of MPT83: A seroreactive antigen of *Mycobacterium tuberculosis* with homology to MPT70. Scand. J. Immunol. 43, 490-499.

[76] Harboe, M., Whelan, A.O., Ulvund, G., McNair, J., Pollock, J.M., Hewinson, R.G. and Wiker, H.G. (2002). Generation of antibodies to the signal peptide of the MPT83 lipoprotein of *Mycobacterium tuberculosis* Scand. J. Immunol. 55, 82-87.

[77] Michell, S.L., Whelan, A.O., Wheeler, P.R., Partico, M., Easton, R.L., Etienne, A.T., Haslam, S.M., Dell, A., Morris, H.R., Reason, A.J., Herrmann, J.L., Young, D.B. and Hewinson, R.G. (2003). The MPB83 antigen from *Mycobacterium bovis* contains O-linked mannose and (1→3)-mannobiose moieties. J. Biol. Chem. 278, 16423-16432.

[78] Juárez, M.D., Torres, A., Espitia, C. (2001). Characterization of the *Mycobacterium tuberculosis* region containing the *mpt83* and *mpt70* genes. FEMS Microbiol. Lett. 203, 95-102.

[79] Wiker, H.G., Lyashchenko, K.P., Aksoy, A.M., Lightbody, K.A., Pollock, J.M., Komissarenko, S.V., Bobrovnik, S.O., Kolesnikova, I.N., Mykhalsky, L.O.,

Gennaro, M.L. and Harboe M (1998). Immunochemical characterization of the MPB70/80 and MPB83 proteins of *Mycobacterium bovis*. Infect Immun. 66, 1445-1452.

[80] Kawamoto, T., Noshiro, M., Shen, M., Nakamasu, K., Hashimoto, K., Kawashima-Ohya, Y., Gotoh, O. and Kato, Y. (1998). Structural and phylogenetic analyses of RGD-CAP/beta ig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. Biochim. Biophys. Acta 1395, 288-292.

[81] Ulstrup, J.C., Jeansson, S., Wiker, H.G. and Harboe, M. (1995). Relationship of secretion pattern and MPB70 homology with osteoblast-specific factor-2 to osteitis following *Mycobacterium bovis* BCG vaccination. Infect. Immun. 63, 672-675.

[82] Arruda, S., Bomfim, G., Knights, R., Huimabyron, T. and Riley, L.W. (1995). Cloning of an *Mycobacterium tuberculosis* DNA fragment associated with entry and survival inside cells. Science 261, 1454-1457.

[83] Wiker, H.G., Spierings, E., Kolkman, M.A.B., Ottenhoff, T.H.M. and Harboe, M. (1999). The mammalian cell entry operon 1 (mce1) of *Mycobacterium leprae* and *Mycobacterium tuberculosis*. Microb. Pathog. 27, 173-177.

[84] Flesselles, B., Anand, N.N., Remani, J., Loosmore, S.M. and Klein, M.H. (1999). Disruption of the mycobacterial cell entry gene of *Mycobacterium bovis* BCG results in a mutant that exhibits a reduced invasiveness for epithelial cells. FEMS Microbiol. Lett. 177, 237-242.

[85] Chitale, S., Ehrt, S., Kawamura, I., Fujimura, T., Shimono, N., Anand, N., Lu, S., Cohen-Gould, L. and Riley, L.W. (2001). Recombinant *Mycobacterium tuberculosis* protein associated with mammalian cell entry. Cell. Microbiol. 3, 247-254.

[86] Haile, Y., Caugant, D.A., Bjune, G. and Wiker, H.G. (2002). *Mycobacterium tuberculosis* mammalian cell entry operon (*mce*) homologs in *Mycobacterium* other than tuberculosis (MOTT). FEMS Immunol. Med. Microbiol. 33, 125-132.

[87] Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood DA (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417, 141-147.

[88] Ahmad, S., Akbar, P.K., Wiker, H.G., Harboe, M. and Mustafa, A.S. (1999). Cloning, expression and immunological reactivity of two mammalian cell entry proteins encoded by the mce1 operon of *Mycobacterium tuberculosis*. Scand. J. Immunol. 50, 510-518.

[89] Cole, S.T. (2002). Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. Microbiol. 148, 2919-2928.

[90] Hou, J.Y., Graham, J.E., Clark-Curtiss, J.E. (2002). *Mycobacterium avium* genes expressed during growth in human macrophages detected by selective capture of transcribed sequences (SCOTS). Infect. Immun. 70, 3714-3726.

[91] Steyn, A.J.C., Joseph, J., Bloom, B.R. (2003). Interaction of the sensor module of *Mycobacterium tuberculosis* H37Rv KdpD with members of the Lpr family. Mol. Microbiol. 47, 1074-1089.

[92] Dartois, V., Djavakhishvili, T. and Hoch, J. A. (1997). KapB is a lipoprotein required for KinB signal transduction and activation of the phosphorelay to sporulation in *Bacillus subtilis.* Mol. Microbiol. 26, 1097-1108.

[93] Oftung, F., Wiker, H.G., Deggerdal, A. and Mustafa, A.S. (1997). A novel mycobacterial antigen relevant to cellular immunity belongs to a family of secreted lipoproteins. Scand. J. Immunol. 46, 445-451.

[94] Bigi, F., Alito, A., Romano, M.I., Zumarraga, M., Caimi, K. and Cataldi, A. (2000). The gene encoding P27 lipoprotein and a putative antibiotic-resistance gene form an operon in *Mycobacterium tuberculosis* and *Mycobacterium bovis*. Microbiol. 146, 1011-1018.

[95] Lefèvre, P., Denis, O., De Wit, L., Tanghe, A., Vandenbussche, P., Content, J. and Huygen, K. (2000). Cloning of the gene encoding a 22-kilodalton cell surface antigen of *Mycobacterium bovis* BCG and analysis of its potential for DNA vaccination against tuberculosis. Infect. Immun. 68, 1040-1047.

[96] Hovav, A-H., Mullerad, J., Davidovitch, L., Fishman, L., Bigi, F., Cataldi, A. and Bercovier, H. (2003). The *Mycobacterium tuberculosis* recombinant 27-kilodalton lipoprotein induces a strong Th1-type immune response deleterious to protection. Infect. Immun. 71, 3146-3154.

[97] Silva, P.E.A., Bigi, F., Santangelo, M.D., Romano, M.I., Martin, C., Cataldi, A. and Ainsa, J.A. (2001). Characterization of P55, a multidrug efflux pump in *Mycobacterium bovis* and *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. 45, 800-804.

[98] Bigi, F., Gioffré, A., Kleep, L., Santangelo, M.D.L.P., Alito, A., Caimi, K., Meikle, V., Zumárraga, M., Taboga, O., Romano, M.I. and Cataldi, A. (2004). The knockout of the *lprG-Rv1410* operon produces strong attenuation of *Mycobacterium tuberculosis.* Microb. Infect. 6, 182-187.

[99] Mukamolova, G.V., Kaprelyants, A.S., Young, D.I., Young, M. and Kell, D.B. (1998). A bacterial cytokine. Proc. Nat. Acad. Sci. USA 95, 8916-8921.

[100] Kell, D.B. and Young, M. (2000). Bacterial dormancy and culturability: the role of autocrine growth factors. Curr. Opin. Microbiol. 3, 238-243.

[101] Mukamolova, G.V., Turapov, O.A., Young, D.I., Kaprelyants, A.S., Kell, D.B. and Young, M. (2002). A family of autocrine growth factors *in Mycobacterium tuberculosis*. Mol. Microbiol. 46, 623-635.

[102] Zhu, W., Plikaytis, B.B. and Shinnick, T.M. (2003). Resuscitation factors from mycobacteria: homologs of *Micrococcus luteus* proteins. Tuberculosis 83, 261-269.

[103] Tufariello, J.M., Jacobs Jr, W.R. and J. Chan (2004). Individual *Mycobacterium tuberculosis* resuscitation-promoting factor homologues are dispensable for growth in vitro and in vivo. Infect. Immun. 72, 515-526.

[104] Serre, L., de Jesus, K.P., Zelwer, C., Bureaud, N., Schoentgen, F. and Bénédetti, H. (2001). Crystal structures of YBHB and YBCL from *Escherichia coli*, two bacterial homologues to a Raf kinase inhibitor protein. J. Mol. Biol. 310, 617-634.

[105] Vallée, B.S., Tauc, P., Brochon, J.C., Maget-Dana, R., Lelièvre, D., Metz-Boutigue, M.H., Bureaud, N. and Schoentgen, F. Behaviour of bovine phosphatidylethanolamine-binding protein with model membranes. Evidence of affinity for negatively charged membranes. Eur. J. Biochem. 268, 5831-5841.

[106] Av-Gay, Y. and Everett, M. (2000). The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. Trends Microbiol. 8, 238-244.

[107] Ashbridge, K. R., Booth, R. J., Watson, J. D. and Lathigra, R. (1989). Nucleotide sequence of the 19 kDa antigen gene from *Mycobacterium tuberculosis*. Nucl. Acids Res. 17, 1249.

[108] Nair, J., Rouse, D.A.and Morris, S.L (1992) Nucleotide sequence analysis and serologic characterization of the *Mycobacterium intracellulare* homolog of the *Mycobacterium tuberculosis* 19-kDa antigen. Mol. Microbiol. 6, 1431-1439.

[109] Booth, R.J., Williams, D.L., Moudgil, K.D., Noonan, L.C., Grandison, P.M., Mckee, J.J., Prestidge, R.L. and Watson, J.D. (1993). Homologs of *Mycobacterium leprae* 18-kilodalton and *Mycobacterium tuberculosis* 19-kilodalton antigens in other mycobacteria. Infect. Immun. 61, 1509-1515.

[110] Lathigra, R., Zhang, Y., Hill, M., Garcia, M.J., Jackett, P.S. and Ivanyi, J. (1996). Lack of production of the 19-kDa glycolipoprotein in certain strains of *Mycobacterium tuberculosis*. Res. Microbiol. 147, 237-249.

[111] Garbe, T., Harris, D., Vordermeier, M., Lathigra, R., Ivanyi, J. and Young, D. (1993). Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in *Mycobacterium smegmatis*: immunological analysis and evidence of glycosylation. Infect. Immun. 61, 260-267.

[112] Herrmann, J.L., O'Gaora, P., Gallagher, A., Thole, J.E.R. and Young, D.B. (1996). Bacterial glycoproteins: A link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. EMBO J. 15, 3547-3554.

[113] Brightbill, H.D., Libraty, D.H., Krutzik, S.R., Yang, R-B., Belisle, J.T., Bleharski, J.R., Maitland, M., Norgard, M.V., Plevy, S.E., Smale, S.T., Brennan, P.J., Bloom, B.R., Godowski, P.J. and Modlin, R.L. (1999). Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. Science 285, 732-736.

[114] Fonseca, D.P.A.J., Joosten, D., Snippe, H. and Verheul, A.F.M. (2000). Evaluation of T-cell responses to peptides and lipopeptides with MHC class I binding motifs derived from the amino acid sequence of the19-kDa lipoprotein of *Mycobacterium tuberculosis*. Mol. Immunol. 37, 413-422.

[115] Yeremeev, V.V., Stewart, G.R., Neyrolles, O., Skrabal, K., Avdienko, V.G., Apt, A.S. and Young, D.B. (2000). Deletion of the 19kDa antigen does not alter the protective efficacy of BCG. Tubercle Lung Dis. 80, 243-247.

[116] Neufert, C., Pai, R.K., Noss, E.H., Berger, M., Boom, W.H. and Harding, C.V. (2001). *Mycobacterium tuberculosis* 19-kDa lipoprotein promotes neutrophil activation. J. Immunol. 167, 1542-1549.

[117] Neyrolles, O., Gould, K., Gares, M.P., Brett, S., Janssen, R., O'Gaora, P., Herrmann, J.L., Prévost, M.C., Perret, E., Thole, J.E.R. and Young, D. (2001). Lipoprotein access to MHC class I presentation during infection of murine macrophages with live mycobacteria. J. Immunol. 166, 447-457.

[118] Lopez, M., Sly, L.M., Luu, Y., Young, D., Cooper, H. and Reiner, N.E. (2003). The 19-kDa *Mycobacterium tuberculosis* protein induces macrophage apoptosis through toll-like receptor-2. J. Immunol. 170, 2409-2416.

[119] Armoa, G.R.G., Rouse, D.A., Nair J., Mackall, J.C and Morris, S.L. (1995). A highly immunogenic putative *Mycobacterium kansasii* lipoprotein. Microbiol. 141, 2705-2712.

[120] Manca, C., Lyashchenko, K., Colangeli, R. and Gennaro, M.L. (1997). MTC28, a novel 28-kilodalton proline-rich secreted antigen specific for the *Mycobacterium tuberculosis* complex. Infect. Immun. 65, 4951-4957.

[121] Webb, J.R., Vedvick, T.S., Alderson, M.R., Guderian, J.A., Jen, S.S., Ovendale, P.J., Johnson, S.M., Reed, S.G. and Skeiky, Y.A.W. (1998). Molecular cloning, expression, and immunogenicity of MTB12, a novel low-molecular-weight antigen secreted by *Mycobacterium tuberculosis*. Infect. Immun. 66, 4208-4214.

[122] Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K. and Smith, I. (2002). *ideR*, an essential gene in *Mycobacterium tuberculosis*: Role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. 70, 3371-3381.

[123] Lefèvre, P., Braibant, M., DeWit, L., Kalai, M., Röeper, D., Grötzinger, J., Delville, J.P., Peirs, P., Ooms, J., Huygen, K. and Content, J. (1997). Three different putative phosphate transport receptors are encoded by the *Mycobacterium tuberculosis* genome and are present at the surface of *Mycobacterium bovis* BCG. J. Bacteriol. 179, 2900-2906.

[124] Espitia, C., Elinos, M., Hernández-Pando, R. and Mancilla, R. (1992). Phosphate starvation enhances expression of the immunodominant 38-kilodalton protein antigen of *Mycobacterium tuberculosis*: demonstration by immunogold electron-microscopy. Infect. Immun. 60, 2998-3001.

[125] Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J.M. and Hecker, M. (2001). A proteomic view of genome-based signal peptide predictions. Genome Res. 11, 1484-1502.

[126] Andersen,A.B., Ljungqvist, l. and Olsen, M. (1990). Evidence that protein antigen b of *Mycobacterium tuberculosis* is involved in phosphate metabolism. J. Gen. Microbiol. 136, 477-480.

[127] Dobos, K.M., Khoo, K.H., Swiderek, K.M., Brennan, P.J. and Belisle, J.T. (1996). Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. J. Bacteriol. 178, 2498-2506.

[128] Herrmann, J.L., Delahay, R., Gallagher, A. Robertson, B. and Young, D. (2000). Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. FEBS Lett. 473, 358-362.

[129] Sander, P., Rezwan, M., Walker, B., Rampini, S.K., Kroppenstedt, R.M., Ehlers, S., Keller, C., Keeble, J.R., Hagemeier, M., Colston, M.J., Springer, B. and Böttger, E.C. (2004). Lipoprotein processing is required for virulence of *Mycobacterium tuberculosis*. Mol. Microbiol. 52, 1543-1552.

Table 1. Functional categorisation of *M. tuberculosis* putative Lpp.

Functional category	Sub-category andORF
SBPs in ABC transport systems (categorised by their predicted substrates)	**Iron:** Rv3044 (FecB); Rv0265c (FecB2) **Molybdenum and Phosphate:** Rv0928 (PstS3); Rv0932c (PstS2); Rv0934 (PstS1); Rv1857 (ModA) **Peptides:** Rv0411c; Rv1166; Rv1244; Rv1280c; Rv2585c; Rv3666c; Rv3759c **Sugars:** Rv1235; Rv2041c; Rv2833c **Sulphates:** Rv2400c
Enzymes predicted to be involved in cell wall metabolism	Rv0399c; Rv0838; Rv1922; Rv2068c; Rv2864c; Rv2905; Rv3593
Enzymes predicted to be involved in degradative processes	**Esterases:** Rv0671; Rv3298c **Glycosyl hydrolase:** Rv0237 **Phosphorylase:** Rv2293c **Proteinase/peptidases:** Rv2224c, Rv2672, Rv0418; Rv0419
Other enzymes and metabolic activities	**Copper oxidase:** Rv0846c **FAD-linked oxidase:** Rv2251 **γ-glutamyl transferase:** Rv2394 **Oxidoreductases and Thioredoxins:** Rv0132c, Rv0526, Rv1677; Rv3006 **Phosphoglycerate mutase:** Rv3390 **Superoxide dismutase:** Rv0432
Putative Lpp with roles in adhesion and cell invasion	**Adhesin:** Rv2873 (MPT83) ***mce* operon proteins:** Rv0173 (Mce1E); Rv0593 (Mce2E); Rv1970 (Mce3E); Rv3495c (Mce4E)
Putative roles in signalling and related functions	Rv1009 (RpfB); Rv1270c; Rv1368 (LprF); Rv1411c; Rv1690 (LprJ); Rv1911c; Rv2403c; Rv2945; Rv3576
Unknown function	**Inter-related Lpp of unknown function:** Rv0483 and Rv2518c; Rv0583c and Rv1016c; Rv0604 and Rv2999; Rv1228 and 2341; The LppA paralogue family (Rv2543; Rv2544; Rv2796c; MT2619) **Other:** Rv0179c; Rv0344c; Rv0381c; Rv0460; Rv0679c; Rv0847; Rv0962c; Rv1064c; Rv1252c; Rv1274; Rv1275; Rv1418; Rv1541c; Rv1799; Rv1881c; Rv1921c; Rv2046; Rv2080; Rv2116; Rv2138; Rv2171; Rv2270; Rv2290; Rv2330c; Rv2784c; Rv2843; Rv3016; Rv3244c; Rv3584; Rv3623; Rv3763 (19 kDa antigen); MT2627.1

Table 2. Putative Lpp with significant homologies to exported proteins of *M. tuberculosis*

Putative Lpp	**Exported homologue**	**% identity (sequence alignment length)**
Rv2873 (MPT83)	Rv2875 (MPT70)	73% (164)
Rv1911c	Rv1910c	63% (206)
Rv1009 RpfB	Rv0867c RpfA[a]	62% (74)
Rv1677	Rv2878c (MPT53)	54% (137)
Rv1690	Rv3354 (YCP family)[a]	46% (115)
Rv1016c	Rv0040c (MTC28)	35% (167)
Rv2116	Rv2376c (MTB12)	29% (162)
Rv3495c (Mce4E)	Rv3496c (Mce4F)[a]	25% (247)
Rv0679c	Rv0680c	38% (96)

[a] representative of several exported proteins

SUPPLEMENTARY MATERIAL FOR ONLINE PUBLICATION

Table S1. Proven/probable Lpp sequences identified by pattern searching with the G+LPP pattern

Protein	Name	Signal peptide sequence	Functional categorisation	ML[a]
Rv0432	SOD	MP**K**PA**D**H**R**NHAAVSTSVLSALFLGAGAALLSACS	Superoxide dismutase	ML1925 71% [b]
Rv0934	PstS1	M**K**I**R**LHTLLAVLTAAPLLLAAAGCG	SBP phosphate	pg[c]
Rv3763	LpqH	M**KR**GLTVAVAGAAILVAGLSGCS	19 kDa antigen	ML1966 43%
Rv0237	LpqI	MAFP**R**TLAILAAAAALVVACS	Putative glycosyl hydrolase; 126/404 (31%) identical to *Streptomyces thermoviolaceus* NagA	ML2569 73%
Rv0344c	LpqJ	M**R**LSLIA**R**GMAALLAATALVAGCN	Conserved hypothetical; 121/179 (67%) identical to an unannotated ML sequence	see text
Rv0381c		V**R**ILVAWATCGAVVLSGLTGCS	Function unknown; homologues present in other mycobacterial genomes	pg
Rv0399c	LpqK	MPVL**RR**LGCSVLALGLLAGCA	Putative peptidase (PBP); 78/303 (25%)identical to *Bacillus cereus*	- [c]

			alkaline D-peptidase; similar to Rv1922	
Rv0583c	LpqN	M**K**HFTAAVATVALSLALAGCS	143/228 (62%) identical to MK35 Lpp of *M. kansasii*	pg
Rv0604	LpqO	MI**RRR**GA**R**MAALLAAAALALTACA	Function unknown; 166/312 (53%) identical to Rv2999	-
Rv0928	PstS3	M**K**LN**R**FGAAVGVLAAGALVLSACG	Phosphate binding protein (PstS-1 paralogue)	ML2095 77%
Rv0932c	PstS2	M**K**FA**R**SGAAVSLLAAGTLVLTACG	Phosphate binding protein (PstS-1 paralogue)	ML2095 63%
Rv0962c	LprP	M**KR**TS**R**SLTAALLGIAALLAGCI	Function unknown; sequence unique to *M. tuberculosis* and *M. bovis*	pg
Rv1009	RpfB	ML**R**LVVGALLLVLAFAGGYAVAACK	Resuscitation promoting factor	ML0240 82%
Rv1016c	LpqT	MAG**RR**CPQDSV**R**PLAVAVAVATLAMSAVACG	Function unknown; 59/167 (35%) identical to 28 kDa Pro-rich antigen Rv0040c	ML0246 67%
Rv1064c	LpqV	M**R**PS**R**YAPLLCAMVLALAWLSAVAGCS	Function unknown; homologues present in other mycobacterial genomes	-
Rv1166	LpqW	MGVPSPV**RR**VCVTVGALVALACMVLAGCT	Putative SBP (family 5 peptides) but not part of an identifiable ABC system	ML1497 80%

Rv1235	LpqY/ MalE	MVMS**R**G**R**IP**R**LGAAVLVALTTAAAACG	Putative SBP (family 1 sugars)	ML1086 77%
Rv1244	LpqZ	M**R**IT**R**ILALLLAVLLAVSGVAGCS	Putative SBP but not part of an identifiable ABC system	ML1093 73%
Rv1252c	LprE	MPGVWSPPCPTTP**R**VGVVAALVAATLTGCG	Function unknown; 55/201 (27%) identical to Rv3483c non-Lpp	ML1099 69%
Rv1270c	LprA	M**K**HPPCSVVAAATAILAVVLAIGGCS	See section 4.4.1. Belongs to a family including LppX, LprF and LprG	ML0557 38%
Rv1274	LprB	M**RRK**V**RR**LTLAVSALVALFPAVAGCS	Function unknown; homologues present in other mycobacterial genomes	ML1115 78%
Rv1411c	LprG	M**R**TP**RR**HC**RR**IAVLAAVSIAATVVAGCS	100% identical to *M. bovis* P27 antigen; See section 4.4.1. Belongs to a family including LppX, LprA and LprF	ML0557 68%
Rv1418	LprH	MACLG**R**PGC**R**GWAGASLVLVVVLALAACT	Function unknown; homologues present in other mycobacterial genomes	-
Rv1541c	LprI	M**R**WIGVLVTALVLSACA	Conserved hypothetical, significant homologies with XCC3075 and XAC3204s of pathogenic *Xanthomonas* sp.	pg
Rv1677	DsbF	MTHS**R**LIGALTVVAIIVTACG	Thioredoxin (putative Thiol:disulfide	-

			interchange protein)	
Rv1857	ModA	M**R**WIGLSTGLVSAMLVAGLVACG	Putative SBP, probably for molybdate	pg
Rv1881c	LppE	MCN**R**LVTVTGVAMVVAAGLSACG	Low homology (26% identity) to *M. intracellulare* MI22 Lpp	pg
Rv1911c	LppC	MESPMTSTLH**R**TPLATAGLALVVALGGCG	Similar to eukaryotic phosphatidyl ethanolamine-binding proteins; 62% identical to adjacent non-Lpp Rv1910c.	-
Rv1921c	LppF	MV**R**LIPSLLAMATVLGGVIGCS	Function unknown; 41% identity with a *Rhodococcus erythropolis* putative Lpp	-
Rv1922		MDSTVTASI**RR**MLGLLAATLLLGGCT	Putative peptidase; 102/315 (32%) identical to *Sphingomonas* sp. ACM-3962 MlrB peptidase. 107/385 (27%) identical to Rv0399c Lpp	-
Rv2046	LppI	M**R**IAALVAVSLLIAGCS	Function unknown; homologues present in other mycobacterial genomes	pg
Rv2116	LppK	M**RR**NI**R**VTLGAATIVAALGLSGCS	Conserved hypothetical; 48/162 (29%) identical to Rv2376c MTB12 secreted antigen	ML1315 50%

Rv2138	LppL	MLTGN**K**PAVQ**RR**FIGLLMLSVLVAGCS	Conserved hypothetical; homologous to corynebacterial putative Lpp e.g. *C. glutamicum* Cgl1517	pg
Rv2224c		MGM**R**LS**RR**D**K**IA**R**MLLIWAALAAVALVLVGCI	Putative hydrolase; 186/488 (38%) identical to *Streptomyces lividans* SlpD proteinase; 49% identity to Rv2223c non-Lpp and 23% identity to Rv2672 Lpp	ML1633 84%
Rv2251		M**R**W**R**ASSAPSISAPPIATGCC	Putative FAD-linked oxidase	pg
Rv2270	LppN	M**R**LPG**R**HVLYALSAVTMLAACS	Conserved hypothetical; 37/137 (27%) identical to *C. glutamicum* Cgl0837 putative Lpp	-
Rv2341	LppQ	MPVGG**R**QHVFE**K**LASILGLVAAPLMLLGLSACG	Function unknown; 36/116 (31%) identical to LpqX (Rv1228)	-
Rv2394		MSVWL**R**AGALVAAVMLSLSGCG	γglutamyl transferase; 192/582 (32%) to *E. coli* Ggt	-
Rv2403c	LppR	MTN**R**W**R**WVVPLFAVFLAAGCT	Function unknown. Low homology to Cterminal domain of Rv1266c pknH	-
Rv2585c		MAP**RRRR**HT**R**IAGL**R**VVGTATLVAATTLTACS	Putative SBP (family 5 peptides) but not part of an identifiable ABC system	ML0489 78%

Rv2672		MATVVGMS**R**PMTSTAMLVALTCSATVLAACV	Putative hydrolase; 140/548 (25%) identical to *Streptomyces lividans* SlpD proteinase; 23% identical to Rv2224c; deleted in *M. bovis*	ML1339 74%
Rv2784c	LppU	M**R**AWLAAATTALFVVATGCS	Function unknown; homologues present in other mycobacterial genomes	-
Rv2796c	LppV	MM**R**WPTAWLLALVCVMATGCG	Function unknown; very low homology with LppA/LppB family	-
Rv2864c		MVT**K**TTLASATSGLLLLAVVAMSGCT	Putative transpeptidase penicillin binding protein	ML1577 81%
Rv2873	MPT83	MINVQA**K**PAAAASLAAIAIAFLAGCS	MPT83 antigen	-
Rv2905	LppW	M**R**A**R**PLTLLTALAAVTLVVVAGCE	Function unknown; 38/138 (27%) identical to *S. coelicor* Sco6382	pg
Rv2999	LppY	*MAGA**K**HAG**R**IVAITTAAAVILAACS*	Function unknown; 166/312 (53%) identical to Rv0604	-
Rv3016	LpqA	MVGLT**R**PLLLCGATLLIAACT	Function unknown; homologues present in other mycobacterial genomes	-
Rv3244c	LpqB	ME**R**LM**R**LTILLFLGAVLAGCA	Conserved hypothetical; homologous to other putative Lpp in actinomycete	ML0775 86%

			genomes	
Rv3390	LpqD	MA**KR**TPV**RK**ACTVLAVLAATLLLGACG	Putative phosphoglycerate mutase; 58/203 (28%) identical to *Amycolatopsis methanolica*	pg
Rv3495c	LprN	MN**R**IWL**R**AIILTASSALLAGCQ	Mce4 operon member; homologous to Rv1970; Rv0173 and Rv0593	ML2593 31%
Rv3584	LpqE	MN**R**CNI**R**L**R**LAGMTTWVASIALLAAALSGCG	Function unknown; homologues present in other mycobacterial genomes	ML0319 63%
Rv3593	LpqF	MGPA**R**LHN**RR**AG**RR**MLALSAAAALIVALASGCS	122/404 (29%) identical to *Streptomyces clavuligerus* orf12 involved in clavanulic acid biosynthesis	ML1923 79%
Rv3623	LpqG	MI**R**LV**R**HSIALVAAGLAAALSGCD	Conserved hypothetical; deleted in *M. bovis* BCG	pg

[a] ML, protein number of homologues in the *M. leprae* genome

[b] % amino acid sequence identity

[c] pg, pseudogene; - homologue absent

Table S2. Additional possible Lpp sequences identified by pattern searching with the PS00013 pattern

Protein	Name	Signal peptide sequence	Functional categorisation	ML[a]
Rv0132c		MTGIS**RR**TFGLAAGFGAIGAGGLGGGCS	Putative F420-dependent dehydrogenase; 118/328 (35%) identical to *M. smegmatis* Fgd	ML0269 35%[b]
Rv0173	LprK	MMSVLA**R**M**R**VM**R**H**R**AWQGLVLLVLALLLSSCG	Mce1 operon member; homologous to Rv0593 (65%), Rv1970 (40%) and Rv3495c (34%)	ML2593 76%
Rv0179c	LprO	MWI**R**AE**R**VAVLTPTASL**RR**LTACYAALAVCAALACT	Function unknown; homologues present in other mycobacterial genomes	pg[c]
Rv0265c	FecB2	M**R**QGCS**RR**GFLQVA**E**AAAATGLFAGCS	Putative SBP for Fe/siderophores but not part of an identifiable ABC system	pg
Rv0411c	GlnH	MT**RR**ALLA**R**AAAPLAPLALAMVLASCG	Putative SBP for amino acids such as glutamate but not part of an identifiable ABC system	ML0303 79%
Rv0418	LpqL	MVN**K**S**R**MMPAVLAVAVVVAFLTTGCI	Putative peptide hydrolase; 179/466 (38%) identical to NapH *Streptomyces alboniger* N-acetylpuromycin N-acetylhydrolase	-[c]
Rv0460		MIPLP**R**SWQLTSAMLVGNAIGLLAGVACS	Function unknown; induced within	pg

			macrophages	
Rv0526		MQS**R**AT**RR**SGALTM**RR**LVIAAAVSALLLTGCS M**RR**LVIAAAVSALLLTGCS	Putative thioredoxin thiol:disulfide oxidoreductase; 90/194 (46%) identical to Cgl0439	ML2412 76%
Rv0671	LpqP	ML**RR**VAILLAAVLAFAGCS	Putative hydrolase/esterase; 36% identical to Rv3298c LpqC; 79/285 (27%) identical to *Piromyces equi* EstA cinnamoyl ester hydrolase	ML0715 36%
Rv0838	LpqR	M**R**LIG**R**L**R**LLMVGLVVICGACACD	Putative D-Ala-D-Ala dipeptidase; 59/225 (26%) identical to *Salmonella typhimurium* PcgL	-
Rv0847	LpqS	MGHVESGHVVWM**R**SAIVAVALGVTVAAVAAACW	Function unknown; homologue present in the draft *M. marinarum* genome	-
Rv1228	LpqX	MS**R**QWHWLAATLLLITTAACS	Function unknown; 36/116 (31%) identical to LppQ (Rv2341)	-
Rv1799	LppT	MSV**K**S**K**NG**R**LAA**R**VLVALAALFAMIALTGSACL	Function unknown; sequence unique to *M. tuberculosis* and *M. bovis*	-
Rv1970	LprM	M**R**IGLTLVMIAAVVASCG	Mce3 operon member; homologous to Rv0173, Rv0593, and Rv3495c	ML2593 37%

Rv2041c		MVN**K**PFE**RR**SLL**R**GAGALTAASLAPWAAGCA	Putative Family 1 (sugars) SBP	ML1427 77%
Rv2068c	BlaC	M**R**N**R**GFG**RRE**LLVAMAMLVSVTGCA	β-lactamase	-
Rv2293c		MGAPL**R**HCLLVAAALSLGCG	Putative phosphorylase; 44/131 (33%) identical to Caulobacter crescentus CC2266	-
Rv2330c	LppP	M**RR**Q**R**SAVPILALLALLALLALIVGLGASGCA	Function unknown; homologues present in other mycobacterial genomes	pg
Rv2518c	LppS	MP**K**VGIAAQAG**R**T**R**V**RR**AWLTALMMTAVMIGAVACG	Conserved hypothetical; ErfK (PF03734) family member	ML0426 82%
Rv2833c	UgpB	MDPLN**RR**QFLALAAAAAGVTAGCA	Putative SBP family 1 (sugars)	-
Rv2945c	LppX	MNDG**KR**AVTSAVLVVLGACLALWLSGCS	See section 4.4.1. Belongs to a family including LprA, LprF and LprG	ML0136 76%
Rv3006	LppZ	MWTT**R**LV**R**SGLAALCAAVLVSSGCA	Conserved hypothetical, putative oxidoreductase. 69/234 (29%) identity to *S. coelicor* Sco5508	ML1699 87%
Rv3044	FecB	M**R**STVAVAVAAAVIAASSGCG	Putative SBP for iron; 93/348 (26%) identical to *C. pseudotuberculosis* FagD; not part of an identifiable ABC system	ML1729 75%

Rv3576	PknM	MG**K**QLAALAALVGACMLAAGCT	Function unknown. 68/196 (35%) identical to the C-terminal domain of Rv1266c PknH	pg
Rv3759c	ProX	M**R**ML**RR**L**RR**ATVAAAVWLATVCLVASCA	Putative SBP for amino acids/glycine betaine	pg
MT2627.1[d]		MD**RRRR**GGVAACLLVTGVSCR	Function unknown; homologues present in other mycobacterial genomes	-

[a] ML, protein number of homologues in the *M. leprae* genome

[b] % amino acid sequence identity

[c] pg, pseudogene; - homologue absent

[d] sequence is present but not annotated in the *M. tuberculosis* H37Rv genome

Table S3. Anomalous probable Lpp sequences that were not identified using either G+LPP or PS00013 in pattern searches

Protein	Name	Signal peptide sequence	Functional categorisation	ML[a]
Rv0419	LpqM	MHG**R**G**R**Y**R**PLV**R**CV**R**P**RR**VAASV**R**TPIACLAAVVVIAGCT	Putative zinc metallopeptidase	-
Rv0483	LprQ	MVI**R**VLF**R**PVSLIPVNNSSTPQSQGPIS**RR**LALTALGFGVLAPNV LVACA	Function unknown; ErfK (PF03734) family member	ML2446 78%[b]
Rv0846c		MPELATSGNAFD**KRR**FS**RR**GFLGAGIASGFALAACA	Putative multicopper oxidase; 158/614 (26%) identical to PcoA *Pseudomonas aeruginosa*	pg[c]
Rv1275	LprC	M**R**H**R**Q**R**TDPPIDCEFEM**RR**VLVGAAALITALLVLTGCT	Function unknown; homologues present in other mycobacterial genomes	ML1116 76%
Rv1280c	OppA	MAD**R**GQ**RR**GCAPGIASAL**R**ASFQG**K**S**R**PWTQT**R**YWAFALLTPLVV AMVLTGCS	Putative SBP family 5 (peptides)	ML1121 75%
Rv1368	LprF	MNGLISQACGSH**R**P**RR**PSSLGAVAILIAATLFATVVAGCG	See section 4.4.1. Belongs to a family including LppX, LprA and LprG.	ML0557 31%

Rv2080	LppJ	MPHSTAD**RR**L**R**LT**R**QALLAAAVVPLLAGCA	Function unknown; homologues present in other mycobacterial genomes	-[c]
Rv2171	LppM	M**R**ASYAPPSSQGS**R**VA**R**T**RRR**GMLAIAMLLMLVPLATGCL	Function unknown; homologues present in other mycobacterial genomes	ML0902 71%
Rv2543	LppA	MIAPQPIS**R**TLP**R**WQ**R**IVALTMIGISTALIGGCT	Function unknown; part of a family including LppB and MT2619	-
Rv2544	LppB	MIAPQPIP**R**TLP**R**WQ**R**IVALTMIGISTALIGGCT	Function unknown; part of a family including LppA and MT2619	-
Rv2843		ML**R**AAPVIN**R**LTN**R**PIS**RR**GVLAGGAALAALGVVSACG	Conserved hypothetical; 52/169 (30%) identical to *S. coelicor* Sco5702	ML1560 67%

[a] ML, protein number of homologues in the *M. leprae* genome

[b] % amino acid sequence identity

[c] pg, pseudogene; - homologue absent

Table S4. Anomalous possible Lpp sequences that were not identified using either G+LPP or PS00013 in pattern searches

Protein	Name	**Signal peptide sequence**	**Functional categorisation**	**ML** [a]
Rv0593	LprL	M**R**CGVSAGSANG**K**PN**R**WTL**R**CGVSAGH**R**GSVFLLAVLLAPVVLTSCT	Mce2 operon member; homologous to Rv1970, Rv0173 and Rv3495c	ML2593 61% [b]
Rv0679c		MVE**K**PL**R**AD**R**ATHS**R**LATFALALAAAALPLAGCS	Function unknown; 40/126 (34%) identical to adjacent Rv0680c	pg [c]
Rv1690	LprJ	MTAHTHDGT**R**TW**R**TG**R**QATTLLALLAGVFGGAASCA	Function unknown; belongs to the 'YCP' family of exported proteins (section 4.4.1.)	ML0676 33%
Rv1899c	LppD	MAAM**R**AHA**RRR**HPHALMS**R**AAGLP**R**LSWFAGLTWFAGGSTGAGCA [d]MS**R**AAGLP**R**LSWFAGLTWFAGGSTGAGCA	Conserved hypothetical; contains a conserved C-terminal domain of unknown function (PF01661)	pg
Rv2290	LppO	MTDP**R**HTV**R**IAVGATALGVSALGATLPACS	Function unknown; some similarity to the 19 kDa antigen	- [c]
Rv2400c	SubI	MLSLTLSEASCIASAS**R**W**R**HIIPAGVVCALIAGIGVGCH	SBP for sulphate	ML0615 76%

Rv3298c	LpqC	MPWA**R**MLSLIVLMVCLAGCG	Putative esterese/lipase; 36% identical to Rv0671 LpqP; 60/223 (26%) identical to *P. equi* EstA cinnamoyl ester hydrolase	ML0715 72%
Rv3666c	DppA	MV**R**QM**R**AALAALATGLLVLAPVAGCG	Putative family 5 (peptides) SBP	pg

[a] ML, protein number of homologues in the *M. leprae* genome

[b] % amino acid sequence identity

[c] pg, pseudogene; - homologue absent

[d] alternative signal peptide as annotated in Tuberculist