21.05.2007

Comprehensive Analysis of Exported Proteins from *Mycobacterium tuberculosis* H37Rv

Hiwa Målen[1], Frode S. Berven[2,3], Kari E. Fladmark[3,4] and Harald G. Wiker[1]

Published in Proteomics April 19. 2007 vol 7 pp 1702-1718

http://www.interscience.wiley.com/

http://www.wiley-vch.de/publish/en/journals/alphabeticIndex/2120/

[1] Section for Microbiology and Immunology, the Gade Institute, University of Bergen, and Department of Microbiology and Immunology, Haukeland University Hospital, Bergen, Norway
[2] Institute of Medicine, University of Bergen and Laboratory of Clinical Biochemistry, Haukeland University Hospital
[3] Proteomics Unit at University of Bergen (PROBE), University of Bergen, Norway
[4] Department of Molecular Biology, University of Bergen

Corresponding author: Harald G. Wiker, Section for Microbiology and Immunology, The Gade Institute, University of Bergen, Bergen, Norway
E-mail: harald.wiker@gades.uib.no
Phone: +4755974650 Fax: +4755974689

Abstract

Proteins secreted by *Mycobacterium tuberculosis* play an essential role in the pathogenesis of tuberculosis. The culture filtrates of *M. tuberculosis* H37Rv made by Sadamu Nagai, Japan, are considerably enriched for secreted proteins compared to other culture filtrates. Complementary approaches were used to identify the secreted proteins in these culture filtrates: 1) 2D-PAGE combined with MALDI-TOF MS and 2) liquid chromatography coupled MS/MS. Peptides derived from a total of 257 proteins were identified of which 144 were identified by more than one peptide. Several members of the immunologically important Early Secretory Antigenic Target-6 (ESAT-6) family of proteins were found to be major components. The majority of the identified proteins, 159 (62%), were predicted to be exported through the general secretory pathway. We experimentally verified that the signal peptides, which mediate translocation through the cell membrane, had been removed in 41 of the identified proteins, and in 35 of those, there was an AXA motif N-terminally to the cleavage site, showing that this motif is important for recognition and cleavage of signal peptides in mycobacteria. A large fraction of the secreted proteins were unknowns, suggesting that we have mapped an unexplored part of the exported proteome of *M. tuberculosis.*

Keywords: *Mycobacterium tuberculosis* / Two-dimensional gel electrophoresis / Liquid chromatography-tandem mass spectrometry / Mass spectra / Secreted protein

1 Introduction

Tuberculosis is a major cause of morbidity and mortality due to a single bacterial pathogen, *Mycobacterium tuberculosis* [1]. According to the WHO, one-third of the world's population is latently infected and there are more than 10 million new cases and 2-3 million deaths annually. In 1998, the entire genome of *Mycobacterium tuberculosis* H37Rv was sequenced [2], followed by the genome of the clinical isolate, *M. tuberculosis* CDC1551 [3] and *Mycobacterium bovis* AF2122/97 [4]. The availability of whole genomic sequences of *M. tuberculosis* complex organisms has greatly facilitated the analysis of its proteome [5-10].

Bacille Calmette-Guerin (BCG) is a live attenuated strain of *M. bovis* used for vaccination [11]. In newborn children, vaccination with BCG prevents miliary and meningeal tuberculosis. The efficiency of BCG in preventing pulmonary tuberculosis in adults, which is mainly responsible for infectiousness at the community level, is however highly variable and in all low [12]. Therefore, there is an urgent need to develop new strategies for combating the pathogen. The genomes of the members of the *M. tuberculosis* complex, including virulent *M. tuberculosis*, *M. africanum*, *M. microti*, *M. bovis*, and attenuated *M. bovis* BCG are highly conserved [13]. Comparative genome analyses have revealed distinct regions in the genome of *M. tuberculosis* H37Rv that have been deleted in wild-type *M. bovis*, *M. bovis* BCG substrains and/or clinical isolates of *M. tuberculosis* [11, 14]. In addition, complementary proteomic analyses have revealed several differences in the protein composition between *M. tuberculosis* and *M. bovis* BCG [15, 16]. The consequences of these differences with respect to virulence and pathogenesis are still not fully understood.

A significant number of mycobacterial proteins inferred from the genome are predicted to be exported. In bacteria, the well-characterized general secretory (Sec) pathway transports unfolded proteins across the cytoplasmic membrane to the bacterial envelope and the extracellular environment. Proteins targeted to this system contain specific N-terminal signal sequences which consist of three distinct regions: the N-, H-, and C- regions. The hydrophobic core, also designated the H-region, consists of 10 to 15 amino acid residues. It is formed by a stretch of hydrophobic residues that seems to adopt an α-helical conformation in the membrane. At the N-terminal side, the H-region is flanked by a positively charged stretch of polar residues, the N-region. At the C-terminal side, the H-region is flanked by the C-region, a stretch of short uncharged polar residues, usually carrying the consensus sequence

AXA at position -1 to -3 relative to the cleavage site for Signal peptidase I [17, 18]. Upon export, the signal sequence is cleaved by a type I signal peptidase, releasing the mature protein at the external side of the cell membrane [19, 20].

An important feature of *M. tuberculosis* is its ability to survive and proliferate inside host macrophages and to inhibit the acidification and maturation of the phagosome [21-24]. Phagosome modification does not take place upon phagocytosis of dead bacilli, suggesting that secreted proteins play a key role in pathogenesis [25]. Secreted proteins are also considered to be key T-cell antigens of protective immune responses against *M. tuberculosis* [26, 27].

Several attempts have been made to define the secreted proteome of *M. tuberculosis*. In one of the original proteomic studies of *M. tuberculosis* H37Rv culture filtrate, 8 secreted proteins were purified and identified [28]. Relatively few additional proteins with predicted N-terminal signal peptide, which can be recognized as secreted through general secretory pathway, have been uncovered in subsequent studies by various authors [9, 10, 16, 29-31]. This is probably due to considerable contamination of regular culture filtrates with intracellular proteins [32]. In the present study we have analysed a unique culture filtrate of *M. tuberculosis* H37Rv with minimal presence of cytoplasmic proteins. By using state of the art proteomic approaches we were able to identify 257 proteins, 159 of them had predicted N-terminal signal peptides, out of which 25 also had a predicted transmembrane domain in the mature part, and 36 proteins were predicted to be lipoproteins. These identifications represents a major contribution to the exploration of the exported proteome of *M. tuberculosis* which is of great importance in order to understand the pathogenesis of tuberculosis.

2 Experimental procedures

2.1 Bacterial cultivation and sample preparation

M. tuberculosis H37Rv ATCC27294 from the National Institute of Health, Tokyo, Japan was cultured as surface pellicle on the wholly synthetic Sauton medium for 3 weeks without shaking. Bacteria were removed by filtration and the culture filtrate was concentrated by 80% ammonium sulphate precipitation. Precipitated proteins were dissolved in buffer and dialyzed against distilled water and lyophilised [28]. Three different batches, referred to as batch A, B and C, of the *M. tuberculosis* H37Rv culture filtrate were analysed to emphasise the

reproducibility of the protein profile of the culture filtrate proteins. These preparations were a generous gift from Sadamu Nagai, Osaka Japan.

2.2 One-dimensional gel electrophoresis

Fifty µg of *M. tuberculosis* H37Rv culture filtrate proteins were mixed with 25 µl sodium-dedocyl-sulphate (SDS) loading buffer and boiled for 5 minutes prior to separation on a 10 cm long, 1 mm thick 12% SDS-polyacrylamide (SDS-PAGE) gel. The protein migration was allowed to proceed until the blue dye had migrated to the bottom of the gel. The protein bands were visualized with Coomassie Brilliant Blue R-250 (CBB) (Bio-Rad, Hercules, CA, USA). The molecular mass standard, full-range-rainbow-RPN800 (Amersham Biosciences AB Uppsala, Sweden), was used to divide each lane into 10 segments. Each segment was cut into smaller pieces, destained, and in-gel digested with trypsin (Promega, Woods Hollow Road, Madison, U.S.A.).

2.3 Two-dimensional gel electrophoresis

Two mg of *M. tuberculosis* H37Rv culture filtrate proteins were mixed with 350 µl of rehydration buffer (7 M urea, 2 M thiourea, 4% 3-[(3-Cholamidopropyl)Dimethyl-Ammonio]-1-Propanesulfonate (CHAPS), 2% dithiothreitol (DTT), 2% Immobiline pH Gradient buffer (IPG buffer) pH 4-7). Isoelectric focusing was performed at 20°C on 18 cm immobiline dry strips (Amersham Biosciences AB, Uppsala, Sweden) with pH intervals (3-10; 4-5; 4.5-5.5; 5.3-6.5) using Multiphor II Electrophoresis System (Amersham Biosciences AB Uppsala, Sweden). Running conditions: Current 2 mA; power 5 W using the EPS 3501 XL Power Supply in gradient mode and with check option for current turned off. Prior to the second dimension, the strips were incubated for 15 minutes in equilibration buffer (6 M urea, 2% SDS, 0.375 M Tris, pH 8.8, 20% glycerol) with 130 mM DTT first and then with 135 mM iodoacetoamide. The equilibrated strip was then placed on an ExcelGel XL SDS 12-14% (Amersham Biosciences). The second dimension was run at: current 40 mA and power 40 W for 4 hours. Proteins were visualized with CBB. Protein spots were excised from the gel and digested with trypsin.

2.4 In-gel digestion

Sliced gel spots or bands were washed twice with 50% acetonitrile (ACN) in 25 mM ammonium bicarbonate (NH_4HCO_3) for 15 minutes at room temperature (RT). The gel pieces

were dehydrated by incubating them with 50 µl 100% ACN for 20 minutes at RT. Proteins were reduced using 10 mM DTT and alkylated with 55 mM iodoacetamide (IAA); both in 100 mM NH_4HCO_3. The gel pieces were dehydrated with ACN as described above, and rehydrated in 25 mmol/l NH_4HCO_3 containing 0.01 µg/µl modified trypsin (Promega). Proteins were digested by trypsin for 16–20 h at 37°C. Then, the tryptic peptides were eluted by incubating the gel pieces with 50 µl 1% trifluoroacetic acid (TFA) for 20 minutes at RT. The supernatant containing tryptic peptides were collected by centrifugation at 15700 *g* for 10 minutes. Additional peptides were extracted from gel pieces by incubation with 50 µl 0.1% TFA in 50% ACN for 20 minutes at RT, followed by centrifugation at 15700 *g*. The supernatant was collected and added to the previous one. Finally, the gel pieces were dehydrated by incubating the gel pieces with 50 µl 100% ACN for 20 minutes at RT, and the supernatant was collected by centrifugation as described above and added to the pool.

2.5 In-solution digestion

Proteins were precipitated from the *M. tuberculosis* H37Rv culture filtrate by 80% v/v acetone at -20ºC over night. The protein pellet obtained after centrifugation at 15700 *g* for 10 minutes was dissolved in aqueous 6 M urea (Ultragrade Fluka), 100 mM Tris (Merck) pH 8 and 10 mM DTT (Amersham Biosciences), and incubated for 1 hour at 37ºC. Iodoacetoamide (IAA) was added to a final concentration of 25 mM followed by an one hour incubation in the dark at 37ºC. To avoid unwanted protease alkylation, DTT corresponding to 2.5 mM was added followed by a 20 minute incubation at 37ºC. The sample volume was diluted 1:3 with MilliQ water (Sigma), and $CaCl_2$ was added to a final concentration of 1 mM. Trypsin (Promega) was added to obtain a protein:trypsin ratio of 1:50, and a 16 hour incubation at 37ºC was conducted. The reaction was quenched by adding TFA to obtain a pH<3 in the digested solution.

2.6 Matrix-assisted laser desorption ionization time-off-flight mass spectrometry (MALDI-TOF MS)

The tryptic peptides extracted from the gel slices were concentrated and desalted using $ZipTip_{C18}$ (Millipore, Billerica, MA, USA), and eluted with saturated alpha-cyano-4-hydoxy-cinnamic acid solution (CHCA) (Sigma) in 50% ACN and 0.1% TFA. One µl sample eluted from the matrix was applied to the steel target plate and analysed by MALDI-TOF MS (Autoflex & Ultraflex, Bruker Daltonics) and MALDI with tandem mass spectrometry (MS/MS) (Ultraflex, Bruker Daltonics). The obtained mass spectra were searched against the

M. tuberculosis complex database using MASCOT (http://www.matrixscience.com). The search parameters were: 100 ppm tolerance as maximum mass error, monoisotopic mass value, and fixed modification of cysteine by carboxymethyl. A protein was regarded identified if the matched peptide mass fingerprint covered 20% of the complete protein sequence. An assignment with sequence coverage below 20% was only accepted if one or more of the main peaks were identified by MALDI MS/MS with a significant MASCOT score (above 95% certainty).

2.7 Liquid chromatography (LC) Electrospray Ionisation (ESI) with MS/MS

On-line LC-MS/MS was performed using the Ultimate high pressure liquid chromatography (HPLC) equipment (LC Packings) and an ESI-Q-TOF Ultima Global instrument (Waters, Micromass, Manchester, UK). One of two different columns were used, either a capillary 0.3 x 150 mm C18 reverse phase column (LC Packings) or a nano 0.075 x 150 mm C18 reverse phase in-house packed column. The flow rate through the columns was 3 µl/min for the capillary column and 0.2 µl/min for the nano column. The solvent gradient went from 5% B to 60% B in 42 minutes, then from 60% B to 95% B in 10 minutes, ending with constant 95% B for 5 minutes. Solvent A was aqueous 2% ACN in 0.1% TFA, whereas solvent B was aqueous 90% ACN in 0.1% TFA. Proteolytic peptide profiles were acquired in data-dependent MS/MS mode from 3 to 60 minutes with a maximum of 40 seconds spectral accumulation time, 10 seconds for each of a maximum of four selected peptides from each scan window. The electrospray voltage was set to 3kV and 2.4kV, respectively for the capillary and nano setup. The collision energy was set to variable using the charge recognition option. The obtained data was searched against the publicly available *M. tuberculosis* complex database using MASCOT Deamon or against the Removed-Signal-Sequence database described below. The search results from the different LC-MS/MS analyses were stored and combined in a publicly available system (http://genesis.ugent.be/ms_lims/).

2.8 Determination of N-terminal start sites in predicted secreted proteins

The Neural Network method (SignalPNN) and the Hidden Markov Model (SignalPHMM) for predicting signal peptides are publicly available at the SignalP server (http://www.cbs.dtu.dk/services/SignalP/). All proteins in the proteome of *M. tuberculosis* H37Rv that were predicted to have a signal peptidase I cleavage site by either of the two SignalP methods (v 2.0) were compiled in a MASCOT searchable database after removal of

the predicted signal peptide. This database is referred to as the Removed-Signal-Sequence database and is available at (http://www.bioinfo.no/publications/wiker2006/)

One-sided binomial test (Clopper-Pearson) was used to investigate if there were any correlation between the AXA motif present in position -1 to -3 relative to the experimentally identified cleavage sites.

3 Results

3.1 Identification of *M. tuberculosis* culture filtrate proteins from Two-dimensional (2D) PAGE using MALDI

The major goal of this study was to do a comprehensive proteomic analysis of culture filtrates of *M. tuberculosis* H37Rv with minimal content of intracellular proteins in order to identify as many secreted proteins as possible. Three different batches, A, B and C, of 3-4 week old *M. tuberculosis* H37Rv culture filtrates were compared by both SDS-PAGE and 2D-PAGE. The batches had highly similar protein profiles, and the major secreted proteins like the antigen 85 components, MPT32, MPT63, MPT64 and MTC28 were found to constitute a substantial part of the total protein in all three batches. In contrast, many proteins thought to be intracellularly derived were found to be minor constituents. In particular we did not observe GroEL1 or GroEL2 proteins in these culture filtrates which is in accordance with previous observations on this type of culture filtrate [33]. A representative 2D-PAGE pattern of the culture filtrate from batch A is shown in figure 1 and the corresponding SDS-PAGE profile in figure 2. The majority of spots were found in the area between pH 4.0 and 6.5 (Fig. 1). The area between pH 6.5 and pH 10.0 did not contain CBB stained spots (results not shown). All the detected spots in the CBB-stained 2D-PAGE gels (Fig. 1) were excised and subjected to in-gel digestion followed by MALDI-TOF-MS and/or MALDI MS/MS analysis. Three-hundred and eighteen spots were identified as 118 unique gene products (Table 1). The identified proteins were classified according to the *M. tuberculosis* H37Rv gene annotation (http://sanger.ac.uk/projects/M_tuberculosis/Gene_list/).

3.2 Protein identification by LC-ESI-MS/MS

The *M. tuberculosis* H37Rv culture filtrate was also analysed using two different LC-MS/MS approaches. Firstly, the unfractionated culture filtrate of batch A was trypsinated in solution, and analysed directly with LC-MS/MS. The peptides to be fragmented by MS/MS were limited by only selecting peptides within specified molecular mass windows. The windows chosen were 300-400 Da, 400-500 Da, 500-600 Da, 600-700 Da, 700-800 Da, 800-900 Da,

900-1000 Da, 1000-1300 Da, 1300-2000 Da and 300-2000 Da. Secondly, in order to identify more proteins, the culture filtrates of batch A, B and C were prefractionated using SDS-PAGE prior to capillary or nano LC-MS/MS analysis. A total of 10 gel slices were cut and in-gel digested with trypsin (Fig. 2). In total, 199 distinct gene products were identified from the LC-MS/MS analysis based on at least one identified peptide per protein with a MASCOT confidence level above 95% (Table 1). Information about the reliability of each identification, like number of peptides matching each protein, is given in supplementary Table 1. The protein profile of the three batches showed that proteins with predicted signal peptide using Signal P constituted a major part of all three: batch A 60%, batch B 69% and batch C 83%.

3.3 Determination of signal peptidase I cleavage sites

Possible cleavage sites for signal peptidase I can be determined by various computational methods. The SignalP method for prediction of secreted proteins based on Neural Network method (NN) or the Hidden Markov Model (HMM) [34, 35] is one of the most reliable programs for this purpose [36]. Of the 257 proteins identified by 2D-PAGE/MALDI-TOF-MS and/or 1D-SDS-PAGE/LC-MS/MS, 159 had a predicted signal peptide by SignalP (Table 1 and supplementary table 2).

To identify the cleavage site for signal peptidase I of secreted proteins of *M. tuberculosis* H37Rv and to verify a cleavable signal peptide, each mass finger print was searched against the Removed-Signal-Sequence database using MASCOT. Out of the 159 identified proteins with a predicted signal peptide, 41 N-terminal peptides lying immediately C-terminally to a predicted cleavage site were identified (Table 2), which confirms the existence of a cleavable signal sequence in many of these proteins. The signal peptides of these proteins are shown in table 3.

Interestingly, the SignalP method predicted in total 188 cleavage sites in 127 of the proteins predicted to be processed by Signal peptidase I and 85 of these cleavage sites had an AXA motif at the N-terminal side (Supplementary table 2). However, among the 127 proteins, 35 out of 41 experimentally verified cleavage sites had an AXA motif (Table 2). The one-sided binomial test showed that this observation was highly signficant ($p<0.000003$).

Twenty-five of the identified N-terminal cleavage sites were in agreement with both the Neural Network method and the Hidden Markov Model predictions, while 10 corresponded only with the cleavage site predicted by the Hidden Markov Model, and 6 only with the Neural Network method. Furthermore, 13 of the identified N-terminal sequences started with aspartic acid in the (+1) position and proline in the (+2) position (the DP motif), which may serve as a sorting or recognition signal following translocation and cleavage by signal peptidase I [37]. Interestingly, the N-terminal cleavage site of signal peptidase I predicted by SignalP in 5 potential lipoproteins (Rv0526, Rv0999, Rv2911, Rv3668c and Rv3759c) were also detected; with the predicted lipobox further N-terminally to the detected cleavage sites. This finding opens for the possibility that lipoproteins might be alternatively cleaved and processed by signal peptidase I and signal peptidase II.

All the experimental data were searched against the NCBI database, but none of the identified peptides corresponded to potential signal peptides.

3.4 Functional distribution of the identified *M. tuberculosis* H37Rv culture filtrate proteins

The annotated *M. tuberculosis* H37Rv proteins have been classified into 12 distinct functional groups (http://genolist.pasteur.fr/TubercuList/). The 257 proteins identified by 2D-PAGE MALDI-TOF-MS and SDS-PAGE combined with LC-MS/MS in this study were distributed across ten of those functional groups (Fig. 3). Most of the identified proteins are involved in prokaryotic cell wall and cell processes (functional group 3, 40.2%) and intermediary metabolism and respiration (functional group 7, 25.0%).

3.5 Isoelectric point and molecular mass distribution of the identified *M. tuberculosis* H37Rv culture filtrate proteins

The methods used in this study for protein identification were able to cover wide p*I* and molecular mass ranges, from a p*I* value of 3.64 (possible resuscitation-promoting factor RpfA, Rv0867c) for protein identified by both 2D-PAGE MALDI and LC-MS/MS to p*I* value 12.60 which belonged to Rv3760 identified by LC-MS/MS. The majority of the proteins clustered between p*I* 4-6.5, which is in agreement with previous 2D-PAGE based studies performed on culture filtrate proteins (Fig. 4A) [9].

The protein with lowest molecular mass in this study was 9.41 kDa (Putative ESAT-6 like protein EsxN, Rv1793), as observed by 2D-PAGE/MALDI-TOF-MS. Probable respiratory nitrate reductase, *narG* (Rv1161) with a molecular mass of 136.92 kDa observed by LC-MS/MS represented the largest identified secreted protein. The majority of the proteins were found in the range between 10 to 50 kDa (Fig. 4B). In many cases, the same protein was identified from different 2D-PAGE spots, and in different SDS-PAGE fractions with different molecular mass and p*I*, possibly as a consequence of post-translational modifications or proteolytic processing. For example, Rv3587c was detected in three different spots (Fig. 1). For 45 secreted proteins the observed average molecular mass was 23.6 kDa by 2D-PAGE (Supplementary Table 1). The average theoretical mass of these proteins was 31.0 kDa and after removal of the signal peptides as predicted by the NN method, it was 26.5 kDa. With 16 potential lipoproteins the observed average molecular mass was 29.5 kDa. The theoretical masses were 34.4 kDa before and 32.2 kDa after predicted signal peptidase II cleavages. In 57 proteins without predicted signal peptide the observed average mass was 30.7 kDa while the theoretical mass was 32.2 kDa. The differences between average theoretical and average observed masses in groups of proteins with and without predicted signal peptides therefore supported that signal peptides had been removed from the secreted proteins.

3.6 Novel protein identification

Among the 257 proteins identified in this study, 92 proteins represent novel identifications. Seventy-two of the novel identifications had a N-terminal signal peptide predicted by one or both of the Neural Network (NN) and the Hidden Markov Model (HMM) methods, of those, 13 proteins were predicted to be lipoproteins by ScanProsite (http://au.expasy.org/prosite/). Fifty-nine of the novel proteins were without resemblance to other proteins that have been functionally characterised previously, showing that there are many proteins with unknown functions in the secreted subproteome of *M. tuberculosis*.

To our knowledge, these results represent the largest number of novel predicted secreted proteins in *M. tuberculosis* culture filtrate reported in one study (Fig. 5), revealing a new part of the *M. tuberculosis* culture filtrate proteome that may prove important for the pathogenesis of the bacteria.

3.7 Major *M. tuberculosis* H37Rv culture filtrate proteins

The most abundant proteins in this *M. tuberculosis* H37Rv culture filtrate as observed by 2D-PAGE, were the secreted antigen 85 complex (85A (Rv3804c), 85B (Rv1886c), 85C (Rv0129c) and 85D (MPT51, Rv3803c)), immunogenic protein MPT63 (Rv1926c), immunogenic protein MPT64 (Rv1980c), alanine-, proline-rich secreted protein MPT32, Rv1860) and secreted proline-rich protein MTC28 (Rv0040c). In addition, protein members of the 6 kDa early secreted antigenic target-6 (ESAT-6) family were also among the abundant proteins (Fig. 1). Among the novel proteins identified by 2D-PAGE, Rv0063 (Possible oxidoreductase) and Rv3587c (Probable conserved membrane protein) were also relatively abundant. Rv3587c has only one predicted transmembrane region which is coincident with its predicted signal peptide, indicating that it is a secreted protein rather than a membrane protein.

3.8 Identification of ESAT-6 family proteins

ESAT-6 is the primary component of a family of small proteins without signal peptides secreted by an alternative mechanism [37, 38]. The protein members of this family are potent T-cell antigens which are essential for the pathogenicity of the bacterium [38]. The culture filtrate of *M. tuberculosis* H37Rv profiled in this study, contained many ESAT-6 proteins: EsxG (Rv0287), EsxL (Rv1198), EsxN (Rv1793), EsxO (Rv2346c), EsxB (Rv3874) and EsxA (Rv3875). One protein spot analysed in this study (Fig. 1, ¤) matched several ESAT-6 family member proteins that are highly homologous and difficult to resolve by 2D-PAGE or MS due to highly similar masses and isoelectric points. Therefore the mass fingerprint from this protein spot matched all these proteins: EsxJ (Rv1038c), EsxK (Rv1197), EsxP (Rv2347c) and EsxW (Rv3620c). We also identifed 3 peptides by LC-MS/MS that could be derived from any of these proteins or EsxM (Rv1792). Two peptides identified by LC-MS/MS matched all of these proteins: EsxI (Rv1037c), EsxL (1198), EsxN (Rv1793), EsxO (Rv2346c) and EsxV (Rv3619c). One peptide identified by LC-MS/MS matched both EsxK (Rv1197) and EsxP (Rv2347c).

3.9 Identifications of Lipoproteins

Lipoproteins (Lpp) represent a distinct class of proteins, associated with the membrane compartment of the bacteria by means of post-translational lipid modifications, and several are found to be surface accessible [39, 40]. These proteins are functionally diverse, and can be involved in interactions between the organism and the host [41]. We identified 36 potential

lipoproteins, predicted to have lipid modification sites by the ScanProsite program. Thirteen of these proteins had not been identified previously in *M. tuberculosis*.

4 Discussion

Proteins secreted by *M. tuberculosis* are believed to mediate important biological functions by interacting with host cells, notably macrophages, and are thus potentially important for virulence and pathogenesis [22, 42-45]. Many proteomic studies using culture filtrates from *M. tuberculosis* have been published [9, 10, 16, 29, 46, 47]. However, the number of identified proteins predicted to be secreted, as based on the presence of a signal peptide in the preprotein, has been relatively low, compared to the total number of proteins identified in these studies (Fig.5). This study provides a considerably more comprehensive picture of the secreted protein repertoire of *M. tuberculosis* H37Rv than previously shown. It was achieved by analysis of a unique type of culture filtrate with almost neglectible content of intracellularly derived proteins, produced by Sadamu Nagai (Osaka, Japan), from cultures of *M. tuberculosis* H37Rv on the wholly synthetic Sauton medium. The first proteomic study of *M. tuberculosis* by Nagai *et al.* [28], focused on secreted proteins in which 12 culture filtrate proteins were chromatographically purified and partially characterized. The presently analyzed culture filtrates had similar quality as the previously analysed culture filtrate in which the antigen 85 complex constitute a major part of the total protein. Looking further back in the literature, similar observations of the composition of culture filtrates of *M. tuberculosis* H37Rv were done by Yoneda and Fukui *et. al.*, 1965 [53] also working in Osaka Japan. They found that the α antigen (antigen 85) comprised 40% of the total protein which is in line with our present results. These results are different, and much higher than what is usually found in *M. tuberculosis* culture filtrates. The important question is whether the phenomenon is due to excellent culture technique or whether this is a strain specific phenomenon. It is not possible to determine this at the present stage, but we have previously worked with several BCG culture filtrates produced by Nagai in the same way as for *M. tuberculosis*. These culture filtrates had significant levels of intracellular proteins. It is therefore possible that the local *M. tuberculosis* strain of H37Rv used in Osaka, Japan has a distinct genetic feature, being resistant to the lysis regularly occurring in *M. tuberculosis* complex organisms. An autolytic toxin, MazF-like toxin gene, has been postulated in *M. tuberculosis* [54], which is proposed to be involved in cell response to starvation and may be

the prokaryotic equivalent of apoptosis. Mutations affecting this enzyme could explain the nature of Nagai`s culture filtrate.

The N-terminal peptides of mature secreted proteins were identified in 41 of 159 exported proteins predicted by SignalP. These results verify the existence of a cleavable signal sequence in those proteins, and shows that the SignalP program provides a correct identification of the signal peptide and its cleavage site in a large fraction of the predicted secreted proteins. However, a majority of the identified N-terminal peptides of mature secreted proteins had an AXA motif N-terminally to their cleavage sites, showing that the mycobacterial signal peptidase I, preferentially recognizes the AXA motif. This motif is underestimated by SignalP, possibly due to the presence of only a few mycobacterial proteins in SignalPs training set. This might partly explain the failure to identify the N-terminal peptide of the other predicted secreted proteins identified in this study. Our results shows that there is a need for a separate algorithm for prediction of secreted proteins in acid-fast bacilli. This has not been possible before because few signal peptidase I cleavage sites in mycobacteria have been known, but with the data presented in this paper this goal is much closer.

Exported lipoproteins have been shown to be exposed at the surface of *M. tuberculosis* and *M. bovis*. They are however vulnerable to proteolytic cleavage relatively close to the N-terminal cysteine and are subsequently released as soluble protein in the culture filtrate. Similar observations have also been reported in *Bacillus subtilis* [55]. The lipoproteins we identified in the culture filtrate may represent such proteolytically processed proteins. Our method for identification of N-terminal peptides of mature secreted proteins is not applicable to intact processed lipoproteins because information about the masses of secondary modification on the N-terminal cysteine is not part of the preferences in MASCOT. However, we found five potential N-terminal peptides among the 36 predicted lipoproteins. Closer inspection revealed that the cleavage site for signal peptidase I, predicted by SignalP and observed in our study, lies C-terminally from the predicted cleavage site for signal peptidase II. This finding suggests that some lipoproteins may be alternatively processed by signal peptidase I or II, and represent a mechanism for dual localization in A) the extracellular environment or B) as lipoprotein in the cell wall.

In conclusion, we have obtained a comprehensive picture of the *M. tuberculosis* H37Rv culture filtrate protein repertoire. Two-hundred-fifty-seven proteins were identified by a combination of 2D-PAGE/MALDI-TOF-MS and LC-MS/MS. Sixty-two percent of them had predicted N-terminal signal peptide, suggesting that the culture filtrate was particularly enriched with respect to secreted proteins. The 92 novel proteins identified in this study provide further insight into the *M. tuberculosis* secreted proteome, and reveal a large portion of previously unidentified proteins with unknown function, which might be involved in the pathogenesis of tuberculosis. We have improved the methods for identification of secreted proteins by referring mass fingerprints of proteins to a database of secreted proteins with predicted signal peptides removed. This method is also useful for determining the cleavage site for mycobacterial signal peptidase I, which is shown to preferentially recognize the AXA motif.

5 Reference List

[1] Kaufmann, S. H. *Nat. Med.* 2000, *6*, 955-960.

[2] Cole, S. T., Brosch, R., Parkhill, J., Garnier, T. *et al., Nature* 1998, *393*, 537-544.

[3] Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L. *et al., J. Bacteriol.* 2002, *184*, 5479-5490.

[4] Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N. *et al., Proc. Natl. Acad. Sci. U.S.A.* 2003, *100*, 7877-7882.

[5] Xiong, Y., Chalmers, M. J., Gao, F. P., Cross, T. A., Marshall, A. G. *J. Proteome. Res.* 2005, *4*, 855-861.

[6] Mawuenyega, K. G., Forst, C. V., Dobos, K. M., Belisle, J. T. *et al., Mol. Biol. Cell* 2005, *16*, 396-404.

[7] Schmidt, F., Donahoe, S., Hagens, K., Mattow, J. *et al., Mol. Cell Proteomics* 2004, *3*, 24-42.

[8] Gu, S., Chen, J., Dobos, K. M., Bradbury, E. M. *et al., Mol. Cell Proteomics* 2003, *2*, 1284-1296.

[9] Rosenkrands, I., King, A., Weldingh, K., Moniatte, M. *et al., Electrophoresis* 2000, *21*, 3740-3756.

[10] Rosenkrands, I., Weldingh, K., Jacobsen, S., Hansen, C. V. *et al., Electrophoresis* 2000, *21*, 935-948.

[11] Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H. *et al., Science* 1999, *284*, 1520-1523.

[12] Fine, P. E. *Novartis. Found. Symp.* 1998, *217*, 57-69.

[13] Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N. *et al., Proc. Natl. Acad. Sci. U.S.A.* 1997, *94*, 9869-9874.

[14] Behr, M. A. *Lancet Infect. Dis.* 2002, *2*, 86-92.

[15] Jungblut, P. R., Schaible, U. E., Mollenkopf, H. J., Zimny-Arndt, U. *et al., Mol. Microbiol.* 1999, *33*, 1103-1117.

[16] Mattow, J., Schaible, U. E., Schmidt, F., Hagens, K. *et al., Electrophoresis* 2003, *24*, 3405-3420.

[17] Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S., van Dijl, J. M. *Microbiol. Mol. Biol. Rev.* 2000, *64*, 515-547.

[18] Tjalsma, H., Antelmann, H., Jongbloed, J. D., Braun, P. G. *et al., Microbiol. Mol. Biol. Rev.* 2004, *68*, 207-233.

[19] van Wely, K. H., Swaving, J., Freudl, R., Driessen, A. J. *FEMS Microbiol. Rev.* 2001, *25*, 437-454.

[20] Ton-That, H., Marraffini, L. A., Schneewind, O. *Biochim. Biophys. Acta* 2004, *1694*, 269-278.

[21] Stewart, G. R., Patel, J., Robertson, B. D., Rae, A., Young, D. B. *PLoS. Pathog.* 2005, *1*, 269-278.

[22] Malik, Z. A., Thompson, C. R., Hashimi, S., Porter, B. *et al., J. Immunol.* 2003, *170*, 2811-2815.

[23] Fratti, R. A., Chua, J., Vergne, I., Deretic, V. *Proc. Natl. Acad. Sci. U.S.A.* 2003, *100*, 5437-5442.

[24] Pieters, J., Gatfield, J. *Trends Microbiol.* 2002, *10*, 142-146.

[25] Beatty, W. L., Russell, D. G. *Infect. Immun.* 2000, *68*, 6997-7002.

[26] Roberts, A. D., Sonnenberg, M. G., Ordway, D. J., Furney, S. K. *et al., Immunology* 1995, *85*, 502-508.

[27] Andersen, P. *Infect. Immun.* 1994, *62*, 2536-2544.

[28] Nagai, S., Wiker, H. G., Harboe, M., Kinomoto, M. *Infect. Immun.* 1991, *59*, 372-382.

[29] Sonnenberg, M. G., Belisle, J. T. *Infect. Immun.* 1997, *65*, 4515-4524.

[30] Beatty, W. L., Russell, D. G. *Infect. Immun.* 2000, *68*, 6997-7002.

[31] Weldingh, K., Rosenkrands, I., Jacobsen, S., Rasmussen, P. B. *et al., Infect. Immun.* 1998, *66*, 3492-3500.

[32] Wiker, H. G. *Scand. J. Immunol.* 2001, *54*, 82-86.

[33] Wiker,H.G., Harboe,M., Nagai, S. *J. Gen. Microbiol.* 1991, *137*, 875-884.

[34] Nielsen, H., Brunak, S., von Heijne, G. *Protein Eng.* 1999, *12*, 3-9.

[35] Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G. *Protein Eng.* 1997, *10*, 1-6.

[36] Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L. *J. Mol. Biol.* 2001, *305*, 567-580.

[37] Wiker, H. G., Wilson, M. A., Schoolnik, G. K. *Microbiology* 2000, *146*, 1525-1533.

[38] Brodin, P., Rosenkrands, I., Andersen, P., Cole, S. T., Brosch, R. *Trends Microbiol.* 2004, *12*, 500-508.

[39] Harboe, M., Whelan, A. O., Ulvund, G., McNair, J. *et al., Scand.J.Immunol.* 2002, *55*, 82-87.

[40] Young, D. B., Garbe, T. R. *Res. Microbiol.* 1991, *142*, 55-65.

[41] Sutcliffe, I. C., Harrington, D. J. *FEMS Microbiol. Rev.* 2004, *28*, 645-659.

[42] Beatty, W. L., Russell, D. G. *Infect. Immun.* 2000, *68*, 6997-7002.

[43] Stewart, G. R., Patel, J., Robertson, B. D., Rae, A., Young, D. B. *PLoS. Pathog.* 2005, *1*, 269-278.

[44] Fratti, R. A., Chua, J., Vergne, I., Deretic, V. *Proc. Natl. Acad. Sci. U.S.A.* 2003, *100*, 5437-5442.

[45] Pieters, J., Gatfield, J. *Trends Microbiol.* 2002, *10*, 142-146.

[46] Beatty, W. L., Russell, D. G. *Infect. Immun.* 2000, *68*, 6997-7002.

[47] Bahk, Y. Y., Kim, S. A., Kim, J. S., Euh, H. J. *et al., Proteomics* 2004, *4*, 3299-3307.

[48] Manca, C., Lyashchenko, K., Wiker, H. G., Usai, D. *et al., Infect. Immun.* 1997, *65*, 16-23.

[49] Harboe, M., Nagai, S., Wiker, H. G., Sletten, K., Haga, S. *Scand. J. Immunol.* 1995, *42*, 46-51.

[50] Wiker, H. G., Harboe, M., Nagai, S. *J. Gen. Microbiol.* 1991, *137*, 875-884.

[51] Coler, R. N., Skeiky, Y. A., Vedvick, T., Bement, T. *et al., J. Immunol.* 1998, *161*, 2356-2364.

[52] Covert, B. A., Spencer, J. S., Orme, I. M., Belisle, J. T. *Proteomics* 2001, *1*, 574-586.

[53] Yoneda, M., Fukui, Y. *Am. Rev. Respir. Dis*. 1965, *92*, 9-18.

[54] Mendum, T., Wagley, S., Wall, S., McFadden, J. Abstract nr. 2080, Identification of a suicide gene in *Mycobacterium tuberculosis*, Keystone Symposia 2-7 April, 2005.

[55] Antelmann, H., Tjalsma, H., Voigt, B., Olhmeier, S. *et al., Genome Res.* 2001, *11*, 1484-1502.

Figure legends

Figure 1. 2D-PAGE of culture filtrate proteins from *M. tuberculosis* H37Rv**.** Two mg of culture filtrate proteins were applied to 18 cm IPG strips with different pH range in the first dimension isoelectric focusing, followed by SDS-PAGE in the second dimension. Proteins were visualized by CBB staining. Predicted secreted proteins are indicated by arrows and their Rv number. Molecular weight markers are indicated to the left. Note: ¤ indicates protein spots identified as different ESAT-6 family member proteins (Rv1038c, Rv1197, Rv1792, Rv2347c and Rv3620c).

Figure 2. One-dimensional SDS-PAGE of *M. tuberculosis* H37Rv culture filtrate proteins. The molecular weight standard is shown on the left, and the fraction numbers on the right. Explanation of the fraction numbers: (1) >160 kDa, (2) ranges from 105-160 kDa, (3) ranges from 75-105 kDa, (4)) ranges from 50-75 kDa, (5) ranges from 35-50 kDa, (6) ranges from 30-35 kDa, (7) ranges from 25-30 kDa, (8) ranges from 15-25 kDa, (9) ranges from 15-10 kDa, (10) <10 kDa.

Figure 3. Functional categorization of the identified *M. tuberculosis* H37Rv culture filtrate proteins. Explanation of functional category numbers: (0) virulence, detoxification, and adaptation, (1) lipid metabolism, (2) information pathway, (3) cell wall and cell processes, (6) PE/PPE family member proteins, (9) regulatory proteins, (10) conserved hypothetical proteins, and (16) conserved hypothetical proteins with an orthologue in *M. bovis*. Functional group codes are taken from the web server (http://genolist.pasteur.fr/TubercuList/).

Figure 4. p*I* (A) and molecular weight (B) distribution of the identified *M. tuberculosis* H37Rv culture filtrate proteins.

Figure 5. A comparison between the number of identified *M. tuberculosis* culture filtrate proteins and the total number of protein identifications in selected previous studies. a) The total number of proteins with predicted N-terminal signal peptide in *M. tuberculosis* genome predicted by Hidden Markov Model (HMM) method, b) Culture filtrate protein identified by Rosenkrands, I. (2000) [9], c) Culture filtrate proteins identified by Mattow, J. (2003) [16], d) culture filtrate proteins identified in this study. The black part of the figure represents number of proteins with predicted signal peptide, while the white part represents proteins without predicted signal peptide.

Acknowledgement

This work was supported by grants from Helse Vest (911077 and 911117) and Indo-Norwegian Institutional Cooperation Programme. The present study was supported by The National Programme for Research in Functional Genomics in Norway (FUGE) funded by The Norwegian Research Council. We are indebted to Morten Harboe for critical comments to the manuscript and to Sadamu Nagai for providing *M. tuberculosis* H37Rv culture filtrate protein. Kalaiarasy Kugarajh and Karen Egeland are greatly acknowledged for technical assistance.

Table 1: List of *M. tuberculosis* H37Rv culture filtrate proteins identified by 2D-PAGE combined with MALDI-TOF MS and liquid chromatography coupled MS/MS.

Gene number	Gene name	Identification method	Functional group[b]	Signal peptide prediction by NN[c]	Signal peptide prediction by HMM[d]	Protein identity	References
Rv0170	*mce1B*	LC-MS/MS	0	0.75	0.65	Mce family protein	e
Rv0172	*mce1D*	LC-MS/MS	0	0.66	0.98	Mce family protein	f
Rv0174	*mce1F*	LC-MS/MS	0	0.95	0.88	Mce family protein	g
Rv0350	*dnaK*	MS&LC-MS/MS	0	0.14	0.01	70 kDa heat shock protein	h,i,j,k,l
Rv0563	*htpY*	LC-MS/MS	0	0.879	0.973	Probable transmembrane heat shock protease	f
Rv1477	-	LC-MS/MS	0	0.9	1	Hypothetical invasion protein	
Rv1908c	*katG*	MS&LC-MS/MS	0	0.1	0	Catalase peroxidase peroxynitritase T	k,m
Rv1932	*tpx*	MS	0	0.18	0.01	Probable thiol peroxidase	h,i,j,l
Rv2006	*otsB1*	LC-MS/MS	0	0.19	0	Probable trehalose-6-phosphate phosphatase	g
Rv2190c	-	LC-MS/MS	0	0.74	1	Conserved hypothetical protein	
Rv2299	*htpG*	LC-MS/MS	0	0.09	0	Probable chaperone protein	g
Rv3418c	-	LC-MS/MS	0	0	0	10 kDa Chaperonin	h,i,j
Rv3759c	*proX*	LC-MS/MS	0	0.92	1	Possible osmoprotectant binding lipoprotein	e
Rv3846	*sodA*	LC-MS/MS	0	0.05	0	Superoxide dismutase	i,m
Rv0129c	*fbpC*	MS&LC-MS/MS	1	0.67	1	Secreted antigen 85C	j,k,m
Rv0242c	*fabG4*	LC-MS/MS	1	0.18	0	Probable 3-oxoacyl reductase (acyl-carrier protein)	i
Rv0244c	*fadE5*	MS	1	0.16	0	Probable acyl-CoA dehydrogenase	e
Rv0436c	*pssA*	LC-MS/MS	1	0.66	0.99	Probable CDP-Diacylglycerol-serine O-phosphatidyltransferase	
Rv1074c	*fadA3*	MS	1	0.42	0	Probable β-ketoacetyl CoA thiolase	h,i,n
Rv1323	*fadA4*	MS	1	0.25	0	Probable acetyl-CoA acetyltransferase	j
Rv1886c	*fbpB*	MS&LC-MS/MS	1	0.81	1	Secreted antigen 85B	h,j,k,l,m
Rv2831	*echA16*	MS&LC-MS/MS	1	0.24	0	Probable enoyl-CoA hydratase	h

Rv3803c	*fbpD*	MS&LC-MS/MS	1	0.93	1	Secreted antigen MPT51	h,j,k,l,m
Rv3804c	*fbpA*	MS&LC-MS/MS	1	0.68	1	Secreted antigen 85A (Myocolyl transferase)	h,j,k,l,m
Rv0002	*dnaN*	MS	2	0.25	0	DNA polymerase III (β-chain, DNA nucleotidyltransferase)	
Rv0009	*ppiA*	MS	2	0.19	0	Possible iron-regulated peptidyl-prolyl cis-trans isomerase	h,j,k,l,m
Rv0054	*ssb*	MS	2	0.19	0	Possible single-strand binding protein	j
Rv0631c	*recC*	LC-MS/MS	2	0.15	0	Probable exonuclease V	g
Rv3201c	-	LC-MS/MS	2	0.46	0.84	Possible ATP-dependant DNA helicase	g
Rv0012	-	LC-MS/MS	3	0.5	0.67	Probable conserved membrane protein	g
Rv0040c	*mtc28*	MS&LC-MS/MS	3	0.66	1	Secreted proline-rich protein	o
Rv0064	-	LC-MS/MS	3	0.68	0.64	Probable conserved transmembrane protein	g
Rv0116c	-	LC-MS/MS	3	0.79	1	Possible conserved membrane protein	
Rv0173	*mce1E*	LC-MS/MS	3	0.81	0.996	Possible Mce-family lipoprotein	f
Rv0203	-	LC-MS/MS	3	0.86	1	Possible exported protein	
Rv0219	-	LC-MS/MS	3	0.505	0.625	Probable conserved transmembrane protein	
Rv0237	*lpqI*	MS	3	0.84	1	Probable conserved lipoprotein	f
Rv0265c	*fecB2*	LC-MS/MS	3	0.526	0.996	Probable periplasmic iron-transport lipoprotein	e
Rv0283	-	LC-MS/MS	3	0.252	0	Possible conserved membrane protein	f,g
Rv0287	*esxG*	MS/MS&LC-MS/MS	3	0.32	0.01	ESAT-6 like protein	h
Rv0309	-	LC-MS/MS	3	0.89	1	Possible conserved exported protein	
Rv0398c	-	MS&LC-MS/MS	3	0.83	1	Possible secreted protein	
Rv0402c	*mmpL1*	LC-MS/MS	3	0.8	0.9	Probable conserved transmembrane transport protein	g
Rv0411c	*glnH*	MS&LC-MS/MS	3	0.71	1	Probable glutamine-binding lipoprotein	
Rv0446c	-	LC-MS/MS	3	0.7	0.81	Possible conserved transmembrane protein	
Rv0477	-	MS/MS&LC-MS/MS	3	0.9	1	Possible conserved secreted protein	
Rv0506	*mmpS2*	LC-MS/MS	3	0.873	0.879	Probable conserved membrane protein	
Rv0559c	-	LC-MS/MS	3	0.83	1	Possible conserved secreted protein	
Rv0583c	*lpqN*	LC-MS/MS	3	0.64	1	Probable conserved lipoprotein	e
Rv0677c	*mmpS5*	MS&LC-MS/MS	3	0.84	0.99	Possible conserved membrane protein	
Rv0680c	-	MS/MS	3	0.82	1	Probable conserved transmembrane protein	
Rv0732	*secY*	LC-MS/MS	3	0.69	0.82	Probable preprotein translocase	

Rv0774c	-	LC-MS/MS	3	0.488	0.999	Probable conserved exported protein	
Rv0835	*lpqQ*	MS&LC-MS/MS	3	0.86	1	Possible lipoprotein	
Rv0838	*lpqR*	MS&LC-MS/MS	3	0.86	1	Probable conserved lipoprotein	
Rv0867c	*rpfA*	LC-MS/MS	3	0.67	1	Possible resuscitation-promoting factor	
Rv0876c	-	LC-MS/MS	3	0.24	0	Possible conserved transmembrane protein	g
Rv0928	*pstS3*	MS&LC-MS/MS	3	0.79	1	Periplasmic phosphate-binding lipoprotein	n
Rv0932c	*pstS2*	MS&LC-MS/MS	3	0.91	1	Periplasmic phosphate-binding lipoprotein	e
Rv0934	*pstS1*	MS&LC-MS/MS	3	0.93	1	Periplasmic phosphate-binding lipoprotein	h,k,l
Rv0996	-	LC-MS/MS	3	0.67	0.99	Probable conserved transmembrane protein	
Rv1037	*esxI group*	MS/MSLC-MS/MS	3	0.13	0	Putative ESAT-6 like proteins: EsxI (Rv1037c), EsxL (Rv1198), EsxO (Rv2346c) and EsxV (Rv3619c)	h
Rv1038c	*esxJ group*	MS&LC-MS/MS	3	0.13	0	Putative ESAT-6 like proteins: EsxJ (Rv1038c), EsxK (Rv1197), EsxP (Rv2347c) and EsxW (Rv3620c)	i
Rv1075c	-	LC-MS/MS	3	0.77	1	Conserved exported protein	g
Rv1166	*LpqW*	LC-MS/MS	3	0.91	1	Probable conserved lipoprotein	
Rv1174c	*TB8.4*	LC-MS/MS	3	0.8	1	Low molecular weight T-cell antigen	p
Rv1183	*mmpL10*	LC-MS/MS	3	0.69	0.97	Probable conserved transmembarne transport protein	g
Rv1197	*esxK*	LC-MS/MS	3	0.11	0	Putative ESAT-6 like protein	h
Rv1198	*esxL*	MS/MS&LC-MS/MS	3	0.11	0	Putative ESAT-6 like protein	h
Rv1252c	*lprE*	LC-MS/MS	3	0.617	0.932	Probable lipoprotein	
Rv1269c	-	MS	3	0.7	1	Probable conserved secreted protein	e,n
Rv1270c	*lprA*	LC-MS/MS	3	0.75	1	Possible lipoprotein	f,g
Rv1371	-	MS	3	0.1	0	Probable conserved membrane protein	g
Rv1382	-	LC-MS/MS	3	0.7	0.77	Probable export or membrane protein	
Rv1435c	-	LC-MS/MS	3	0.73	1	Probable conserved proline-, glycine-, valine-rich secreted protein	
Rv1488	-	LC-MS/MS	3	0.758	0.999	Possible conserved exported protein	e
Rv1541c	*LprI*	LC-MS/MS	3	0.79	1	Possible lipoprotein	
Rv1793	*esxN*	MS/MS&LC-MS/MS	3	0.1	0	Putative ESAT-6 like protein	n
Rv1845c	-	LC-MS/MS	3	0.905	0.994	Conserved hypothetical transmembrane protein	g

Rv1860	*apa*	MS/MS&LC-MS/MS	3	0.82	1	Alanine-, proline-rich secreted protein	h,j,k,l,m
Rv1884c	*rpfC*	LC-MS/MS	3	0.56	0.12	Possible resucitation-promoting factor	
Rv1899c	*lppD*	LC-MS/MS	3	0.679	0.989	Possible lipoprotein	e
Rv1910c	-	LC-MS/MS	3	0.89	1	Possible exported protein	
Rv1911c	*lppC*	LC-MS/MS	3	0.61	1	Probable lipoprotein	
Rv1926c	*mpt63*	MS&LC-MS/MS	3	0.92	1	Immunogenic protein MPT63	g,h,i,m
Rv1980c	*mpt64*	MS&LC-MS/MS	3	0.93	1	Immunogenic protein MPT64	h,i,j,k,l,m
Rv1984c	*cfp21*	LC-MS/MS	3	0.86	1	Probable cutinase	h,i,l
Rv2060	-	LC-MS/MS	3	0.7	0.98	Possible conserved integral membrane protein	g
Rv2080	*lppJ*	MS/MS	3	0.78	1	Possible lipoprotein	
Rv2224c	-	LC-MS/MS	3	0.74	1	Possible exported protease	f,g
Rv2253	-	MS&LC-MS/MS	3	0.94	1	Possible unknown secreted protein	
Rv2301	*cut2*	MS/MS&LC-MS/MS	3	0.83	1	Possible cutinase	i,l
Rv2346c	*esxO*	MS/MS&LC-MS/MS	3	0.09	0	Putative ESAT-6 like protein	g
Rv2376c	*cfp2*	MS/MS&LC-MS/MS	3	0.95	1	Low-molecular-weight antigen	i
Rv2450c	*rpfE*	LC-MS/MS	3	0.77	1	Possible resucitation-promoting factor	
Rv2544	*lppB*	LC-MS/MS	3	0.577	0.603	Probable conserved lipoprotein	
Rv2563	-	LC-MS/MS	3	0.662	0.954	Probable glutamine-transport transmembrane protein	f
Rv2575	-	LC-MS/MS	3	0.28	0.01	Possible conserved glycine-rich membrane protein	m
Rv2576c	-	LC-MS/MS	3	0.58	0.94	Possible conserved membrane protein	
Rv2585c	-	LC-MS/MS	3	0.883	1	Possible conserved lipoprotein	g
Rv2668	-	MS&LC-MS/MS	3	0.87	1	Possible exported alanine-, valine-rich protein	
Rv2693c	-	LC-MS/MS	3	0.65	0.95	Probable conserved alanine-, leucine-rich integral membrane protein	
Rv2721c	-	MS&LC-MS/MS	3	0.76	1	Possible alanine-, glycine-rich integral transmembrane protein	e
Rv2799	-	MS&LC-MS/MS	3	0.62	0.89	Probable membrane protein	
Rv2873	*mpt83*	MS/MS&LC-MS/MS	3	0.89	1	Immunogenic cell surface lipoprotein, MPT83	q
Rv2875	*mpt70*	MS/MS&LC-MS/MS	3	0.88	1	Major secreted immunogenic protein MPT70	r
Rv2878c	*mpt53*	MS&LC-MS/MS	3	0.69	1	Soluble secreted antigen MPT53	h,i,j,k,l,m
Rv2905	*lppW*	LC-MS/MS	3	0.89	1	Probable conserved alanine-rich lipoprotein	

Rv2911	*dacB2*	MS&LC-MS/MS	3	0.94	1	Probable D-alanyl, D-alanine carboxypeptidase	
Rv2945c	*lppX*	MS	3	0.75	1	Probable conserved lipoprotein	e
Rv2994	-	LC-MS/MS	3	0.6	0.53	Probable conserved integral membrane protein	
Rv3004	*cfp6*	LC-MS/MS	3	0.75	1	Low-molecular-weight protein antigen	
Rv3006	*lppZ*	MS&LC-MS/MS	3	0.87	1	Probable conserved lipoprotein	e
Rv3016	*lpqA*	MS/MS	3	0.78	1	Probable lipoprotein	
Rv3036c	*TB22.2*	LC-MS/MS	3	0.84	1	Probable conserved secreted protein	h,i,j,k
Rv3044	*fecB*	MS&LC-MS/MS	3	0.65	1	Probable FeIII-dicitrate-binding periplasmic lipoprotein	s
Rv3193c	-	LC-MS/MS	3	0.83	1	Probable conserved transmembrane protein	f
Rv3194c	-	LC-MS/MS	3	0.932	1	Possible conserved secreted protein	f
Rv3240c	*secA1*	LC-MS/MS	3	0.229	0	Probable preprotein translocase subunit	f
Rv3244c	*lpqB*	MS&LC-MS/MS	3	0.85	1	Probable conserved lipoprotein	g
Rv3402c	-	LC-MS/MS	3	0.36	0.97	Conserved hypothetical protein	
Rv3495c	*lprN*	LC-MS/MS	3	0.77	1	Possible Mce-family lipoprotein	g
Rv3584	*lpqE*	LC-MS/MS	3	0.88	1	Possible conserved lipoprotein	e
Rv3587c	-	MS&LC-MS/MS	3	0.52	0.98	Probable conserved membrane protein	
Rv3629c	-	LC-MS/MS	3	0.72	0.99	Probable conserved integral membrane protein	
Rv3682	*ponA2*	MS&LC-MS/MS	3	0.76	1	Probable bifunctional membrane-associated penicillin-binding protein	e
Rv3693	-	LC-MS/MS	3	0.842	0.997	Possible conserved membrane protein	f
Rv3760	-	LC-MS/MS	3	0.608	0.924	Possible conserved membrane protein	
Rv3835	-	MS	3	0.45	0	Probable conserved membrane protein	g
Rv3874	*esxB*	MS&LC-MS/MS	3	0.15	0	10 kDa culture filtrate antigen (CFP10)	h,i,j
Rv3875	*esxA*	MS/MS&LC-MS/MS	3	0.13	0	6 kDa early secretory antigenic target (ESAT-6)	h,i
Rv3917c	*parB*	LC-MS/MS	3	0.573	0.991	Probable chromosome partitioning protein	e
Rv3428c	-	LC-MS/MS	5	0.121	0	Possible transposase	f
Rv0285	*PE5*	LC-MS/MS	6	0.58	0.99	PE family protein	
Rv0453	*PPE11*	LC-MS/MS	6	0.326	0.833	PPE family protein	g
Rv1386	*PE15*	LC-MS/MS	6	0.61	0.98	PE family protein	
Rv1759c	*wag22*	LC-MS/MS	6	0.44	0.87	PE-PGRS family protein	

Rv2430c	*PPE41*	MS/MS&LC-MS/MS	6	0.2	0	PPE family protein	
Rv2431c	*PE25*	MS&LC-MS/MS	6	0.17	0	PE family protein	
Rv3872	*PE35*	LC-MS/MS	6	0.26	0.03	PE family-related protein	
Rv0062	*celA*	MS	7	0.6	0.18	Possible cellulase (Endoglucanase)	
Rv0063	-	MS&LC-MS/MS	7	0.65	1	Possible oxidoreductase	
Rv0066	*icd2*	LC-MS/MS	7	0.263	0.007	Probable isocitrate dehydrogenase	e,g
Rv0075	-	LC-MS/MS	7	0.091	0	Probable aminotransferase	e
Rv0125	*pepA*	MS/MS&LC-MS/MS	7	0.93	1	Probable serine protease	t
Rv0211	*pckA*	LC-MS/MS	7	0.094	0	Probable iron-regulated phosphoenol pyruvate carboxykinase	f,g
Rv0291	*mycP3*	MS&LC-MS/MS	7	0.73	1	Probable membrane-anchored mycosin	e
Rv0315	-	MS&LC-MS/MS	7	0.71	1	Possible β-1,3-glucanase precursor	n
Rv0363c	*fba*	MS	7	0.16	0	Probable fructose-bisphosphate aldolase	h,j
Rv0408	*pta*	LC-MS/MS	7	0.43	0	Probable phosphate acetyltransferase	
Rv0462	*lpd*	MS&LC-MS/MS	7	0.21	0.03	Dihydrolipoamide dehydrogenase	g
Rv0501	*galE2*	MS	7	0.28	0	Possible UDP-glucose 4-epimerase	g
Rv0526	-	MS&LC-MS/MS	7	0.87	1	Possible thioredoxin protein	e
Rv0843	-	LC-MS/MS	7	0.1	0	Probable dehydrogenase	
Rv0851c	-	LC-MS/MS	7	0.168	0.573	Probable short-chain-type dehydrogenase/reductase	
Rv0884c	*serC*	MS	7	0.16	0	Possible phosphoserine aminotransferase	h,i
Rv1050	-	LC-MS/MS	7	0.28	0.97	Probable oxidoreductase	g
Rv1077	*cbs*	MS	7	0.18	0.04	Probable cystathionine β-synthase	h,j,n
Rv1098c	*fum*	LC-MS/MS	7	0.15	0	Probable fumarase	f,g
Rv1161	*narG*	LC-MS/MS	7	0.08	0	Probable respiratory nitrate reductase	f,g
Rv1310	*atpD*	LC-MS/MS	7	0	0	Possible ATP-synthase, β-chain	f,g
Rv1415	*ribA2*	MS	7	0.1	0	Probable riboflavin biosynthesis protein	
Rv1436	*gap*	MS	7	0.16	0	Probable glyceraldehyde 3-phosphate dehydrogenase	j
Rv1437	*pgk*	MS	7	0.16	0	Probable phosphoglycerate kinase	e
Rv1438	*tpi*	MS	7	0.21	0	Probable triosephosphate isomerase	j
Rv1448c	*tal*	MS&LC-MS/MS	7	0.22	0.08	Probable transaldolase	g
Rv1449c	*tkt*	LC-MS/MS	7	0.15	0	Probable transketolase	g

Rv1454c	*qor*	MS	7	0.07	0	Probable quinone reductase	h,j
Rv1475c	*acn*	MS&LC-MS/MS	7	0.37	0	Probable iron-regulated aconitate hydratase	h,j
Rv1812c	-	LC-MS/MS	7	0.6	0.56	Possible dehydrogenase	m
Rv1833c	-	LC-MS/MS	7	0.09	0	Possible haloalkane dehalogenase	e
Rv1837c	*glcB*	MS	7	0.23	0	Probable malate synthase	j,n
Rv1869c	-	LC-MS/MS	7	0.506	0.538	Probable reductase	g
Rv1876	*bfrA*	MS	7	0.11	0	Probable bacterioferritin	n
Rv2068c	*blaC*	MS&LC-MS/MS	7	0.66	1	Class A β-lactamase	n
Rv2110c	*prcB*	MS	7	0.2	0	Proteasome (beta subunit)	h,i
Rv2192c	*TrpD*	LC-MS/MS	7	0.22	0.08	Probable anthranilate phosphoribosyltransferase	e
Rv2200c	*ctaC*	LC-MS/MS	7	0.526	0.843	Probable transmembrane cytochrome C oxidase	f
Rv2201	*asnB*	LC-MS/MS	7	0.606	0.995	Probable asparagine synthetase	f
Rv2220	*glnA1*	MS&LC-MS/MS	7	0.1	0	Glutamine synthetase	h,i,k
Rv2236c	*cobD*	MS	7	0.39	0.12	Probable cobalamin biosynthesis transmembrane protein	g
Rv2241	*aceE*	LC-MS/MS	7	0.161	0	Probable pyruvate dehydrogenase, E1 component	f
Rv2251	-	LC-MS/MS	7	0.492	0.767	Possible flavoprotein	g
Rv2334	*cysK1*	MS	7	0.21	0	Probable cysteine synthase A	e
Rv2445c	*ndkA*	MS	7	0.21	0	Probable nucleoside diphosphate kinase	i,j
Rv2465c	-	MS	7	0.14	0	Probable isomerase	h
Rv2672	-	MS&LC-MS/MS	7	0.84	1	Possible secreted protease	f
Rv2766c	*fabG5*	LC-MS/MS	7	0.463	0.784	Probable short-chain type dehydrogenase/reductase	e
Rv2848c	*cobB*	LC-MS/MS	7	0.573	0.618	Probable cobyrinic acid A,C-diamine synthase	g
Rv2874	*dipZ*	LC-MS/MS	7	0.57	0.96	Possible integral membrane C-type cytochrome biogenesis protein	g
Rv3106	*fprA*	LC-MS/MS	7	0.52	0.997	NADPH adrenodoxin oxidoreductase	
Rv3111	*moaC1*	MS	7	0.26	0	Probable molybdenum cofactor biosynthesis protein	
Rv3158	*nuoN*	LC-MS/MS	7	0.522	0.563	Probable NADH dehydrogenase I	f
Rv3248c	*sahH*	LC-MS/MS	7	0.16	0	Probable adenosyl-homocysteinase	g
Rv3310	-	MS/MS&LC-MS/MS	7	0.79	1	Possible acid phosphatase V	n
Rv3356c	*folD*	MS	7	0.15	0	Probable bifunctional protein: methylene-tetrahydrofolate	h,i

						dehydrogenase/cyclohydrolase	
Rv3397c	*phyA*	LC-MS/MS	7	0.178	0	Probable phytoene synthase	
Rv3485c	-	LC-MS/MS	7	0.51	0.042	Probable short-chain type dehydrogenase/reductase	e,h
Rv3509c	*ilvX*	MS	7	0.22	0	Probable acetohydroxyacid synthase	j
Rv3668c	-	LC-MS/MS	7	0.87	1	Possible protease	
Rv3671c	-	MS	7	0.65	0.49	Possible membrane associated serine protease	f
Rv3710	*leuA*	LC-MS/MS	7	0.11	0	2-isopropylmalate synthase	h,i
Rv3725	-	LC-MS/MS	7	0.34	0.97	Possible oxidoreductase	
Rv3841	*bfrB*	MS/MS	7	0.08	0	Possible bacterioferritin	e
Rv3914	*trxC*	MS&LC-MS/MS	7	0.14	0	Thioredoxin (MPT46)	h,i,j,k,m
Rv0015c	*pknA*	MS	9	0.47	0.03	Transmembrane serine-, threonine-protein kinase A	g
Rv0472c	-	LC-MS/MS	9	0.129	0	Probable transcriptional regulatory protein	
Rv0490	*senX3*	LC-MS/MS	9	0.881	0.999	Putative two component sensor histidine kinase	
Rv0982	*mprB*	LC-MS/MS	9	0.913	0.964	Probable two component sensor kinase	g
Rv0019c	-	LC-MS/MS	10	0	0	Conserved hypothetical protein	g
Rv0140	-	LC-MS/MS	10	0.122	0	Conserved hypothetical protein	
Rv0164	*TB18.5*	MS/MS	10	0.47	0	Conserved hypothetical protein	h,i,j
Rv0192	-	LC-MS/MS	10	0.17	0	Conserved hypothetical protein	n
Rv0340	-	LC-MS/MS	10	0.18	0.01	Conserved hypothetical protein	
Rv0455c	-	MS&LC-MS/MS	10	0.9	1	Conserved hypothetical protein	
Rv0674	-	LC-MS/MS	10	0.88	0.94	Conserved hypothetical protein	
Rv1158c	-	LC-MS/MS	10	0.82	0.97	Conserved hypothetical alanine-, proline-rich protein	
Rv1186c	-	LC-MS/MS	10	0.429	0.12	Conserved hypothetical protein	g
Rv1352	-	MS/MS&LC-MS/MS	10	0.83	1	Conserved hypothetical protein	
Rv1498A	-	LC-MS/MS	10	0.112	0.002	Conserved hypothetical protein	t
Rv1729		LC-MS/MS	10	0.25	0.02	Conserved hypothetical protein	g
Rv1784		LC-MS/MS	10	0.147	0	Conserved hypothetical protein	
Rv1804c	-	MS	10	0.5	1	Conserved hypothetical protein	
Rv1810	-	LC-MS/MS	10	0.78	1	Conserved hypothetical protein	
Rv1815	-	MS/MS&LC-MS/MS	10	0.87	1	Conserved hypothetical protein	k

Rv1827	*cfp17*	MS	10	0.09	0	Conserved hypothetical protein	h,i,l
Rv1891	-	LC-MS/MS	10	0.4	1	Conserved hypothetical protein	
Rv1906c	-	LC-MS/MS	10	0.78	1	Conserved hypothetical protein	
Rv2074	-	MS	10	0.19	0	Conserved hypothetical protein	
Rv2140c	*TB18.6*	MS&LC-MS/MS	10	0.34	0.11	Conserved hypothetical protein	h,i,j
Rv2314c	-	MS	10	0.15	0	Conserved hypothetical protein	e
Rv2469c	-	LC-MS/MS	10	0.573	0.95	Conserved hypothetical protein	
Rv2631	-	MS	10	0.4	0.05	Conserved hypothetical protein	
Rv2823c	-	LC-MS/MS	10	0.147	0.001	Conserved hypothetical protein	g
Rv3031	-	LC-MS/MS	10	0.31	0	Conserved hypothetical protein	g
Rv3267	-	MS&LC-MS/MS	10	0.87	1	Conserved hypothetical protein	g
Rv3354	-	LC-MS/MS	10	0.88	1	Conserved hypothetical protein	
Rv3369	-	LC-MS/MS	10	0	0	Conserved hypothetical protein	u
Rv3484	*cpsA*	MS	10	0.67	1	Possible conserved protein	g
Rv3627c	-	LC-MS/MS	10	0.87	1	Conserved hypothetical protein	f
Rv3705c	-	LC-MS/MS	10	0.83	1	Conserved hypothetical protein	
Rv3722c	-	MS&LC-MS/MS	10	0.22	0	Conserved hypothetical protein	n
Rv3881c	-	MS&LC-MS/MS	10	0.11	0	Conserved hypothetical alanine-, glycine-rich protein	m
Rv3899c	-	LC-MS/MS	10	0.277	0.748	Conserved hypothetical protein	
Rv0787	-	MS&LC-MS/MS	16	0.68	0	Hypothetical protein	e
Rv0999	-	LC-MS/MS	16	0.62	0.8	Hypothetical protein	e
Rv1419	-	LC-MS/MS	16	0.5	1	Hypothetical protein	
Rv1887	-	LC-MS/MS	16	0.18	0.07	Hypothetical protein	
Rv2401	-	MS	16	0.34	0	Hypothetical protein	
Rv3033	-	LC-MS/MS	16	0.679	1	Hypothetical protein	n
Rv3413c	-	LC-MS/MS	16	0.07	0	Hypothetical alanine-, proline-rich protein	
Rv3491	-	MS/MS&LC-MS/MS	16	0.63	1	Hypothetical protein	
Rv3572	-	MS&LC-MS/MS	16	0.84	1	Hypothetical protein	
Rv3849	-	MS	16	*(0.78)*	0	Conserved hypothetical protein	j
MT0066.1[a]	-	MS/MS		0.577	1	Hypothetical protein	

MT2420	-	MS/MS	0.109	0	Hypothetical protein
MT3437.1	-	MS	0.097	0	Hypothetical protein

[a]Proteins not annotated in *M. tuberculosis* H37Rv strain.

[b]Explanation of functional group: (0) virulence, detoxification, and adaptation, (1) lipid metabolism, (2) information pathway, (3) cell wall and cell processes, (6) PE/PPE family member proteins, (9) regulatory proteins, (10) conserved hypothetical proteins, and (16) conserved hypothetical proteins with an orthologue in *M. bovis*.

[c]Probability to have N-terminal signal peptide predicted by Neural Network (NN).

[d]Probability to have N-terminal signal peptide predicted by Hidden Markov Model (HMM).

These proteins have been observed by others previously: [e][8], [f][5], [g][6], [h][9], [i][10], [j][15], [k][29], [l][31], [m][28], [n][7], [o][48], [p][51], [q][49], [r][50], [s][52], [t][16], [u][47]

Rv3849: The signal peptide prediction by the NN method predicts a very short signal peptide, only 6 amino acids, which is unlikely.

MT0066.1: The nucleotide seqence of MT0066.1 is identical in M. tuberculosis H37Rv and found in positions 65012..65392 on the minus (-) strand overlapping with Rv0061 which is annotated as a conserved hypothetical protein (questionable ORF). 7.

MT2420: The sequence of the peptide identified to belong to MT2420 is not encoded in the M. tuberculosis H37Rv genome. The peptide is highly homologous to a peptide found in Rv1793, the only difference being an A in MT2420 instead of an S. The peptide may therefore be derived from Rv1793 being modified posttranslationally by removal of a hydroxyl group to change S to A.

MT3437.1: The nucleotide seqence of MT3437.1 is identical in M. tuberculosis H37Rv and found in positions 3720757..3721236 on the pluss (+) strand overlapping slightly with Rv3333c (position 3720782), extending in the gap between this gene and Rv3334 which starts at position 372125

Table 2: Proteins with experimentally identified N-terminal peptide C-terminally to the SignalP predicted N-terminal signal sequence.

Gene number	Gene name	-1 to -3 position	N-terminal sequence of mature protein	Identification method	Signal peptide prediction method
Rv0040c	*mtc28*	ASA	↓DPLLPPPPIPAPVSAPATVPPVQNLTALPGGSSNR	MS	NN[a]
Rv0063	-	ATA	↓DPAASGWEALSSALGGK	LC-MS/MS	NNHMM[b]
Rv0125	*pepA*	AQA	↓APPALSQDR	LC-MS/MS	NNHMM
Rv0129c	*fbpC*	AGA	↓FSRPGLPVEYLQVPSASMGR	MS & LC-MS/MS	NNHMM
Rv0285	*PE5*	ASA	↓APVITAVVPPAADPVSLQTAAGFSAQGVEHAVVTAEGVEELGR	LC-MS/MS	NNHMM
Rv0291	*mycP3*	AWA	↓IGPPVVDAAAQPPSGDPGPVAPMEQR	MS	NN
Rv0398c	-	AGA	↓EPTGALPPMTSSGSGPVIGDGDAALR	MS	NNHMM
Rv0455c	-	AAA	↓DSTEDFPIPR	MS & LC-MS/MS	NNHMM
Rv0477	-	AQA	↓DPEADPGAGEANYGGPPSSPR	LC-MS/MS	NNHMM
Rv0526	-	AVA	↓QGGTFEFVSPGGK	MS	HMM[c]
Rv0559c	-	AQA	↓DDYDAPFNNTIHR	LC-MS/MS	NNHMM
Rv0867c	*rpfA*	ATA	↓ATDGEWDQVAR	LC-MS/MS	HMM
Rv0999	-	ASS	↓TTASTGDIAK	LC-MS/MS	NN
Rv1158c	-	AHA	↓DPAPAPAPAPNIPQQLISSAANAPQILQNLATALGATPPLSAPK	LC-MS/MS	NN
Rv1174c	*TB8.4*	ASA	↓DPVDAVINTTCNYGQVVAALNATDPGAAAQFNASPVAQSYLR	LC-MS/MS	NNHMM
Rv1269c	-	ANA	↓ADVYGAIAYSGNGSWGR	MS	NNHMM
Rv1352	-	ARA	↓ETGEQFPGDGVFLVGTDIAPGTYR	LC-MS/MS	NNHMM
Rv1477	-	ATA	↓DPQTDTIAALIADVAK	LC-MS/MS	NNHMM
Rv1804c	-	AHA	↓GPSGDDAVFLASLER	MS	NNHMM
Rv1810	-	GKA	↓DPTGDDAAFLAALDQAGITYADPGHAITAAK	LC-MS/MS	NNHMM
Rv1815	-	ASA	↓DPVLVFPGMEIR	LC-MS/MS	NNHMM
Rv1845c	-	LLA	↓RATWPLR	LC-MS/MS	HMM
Rv1886c	*fbpB*	AGA	↓FSRPGLPVEYLQVPSPSMGR	MS	NNHMM
Rv1906c	-	AGA	↓DPEPAPTPK	LC-MS/MS	NNHMM
Rv1910c	-	RKA	↓APLAPKAAALGRSMPETPTGDVLTISSPAFADGAP	LC-MS/MS	HMM

Rv1926c	*mpt63*	ALA	↓AYPITGK	MS	NNHMM
Rv1984c	*cfp21*	AHA	↓DPCSDIAVVFAR	LC-MS/MS	HMM
Rv2190	-	VLA	↓DPADDALAK	LC-MS/MS	NNHMM
Rv2253	-	AVA	↓AAEPSWNGQYLVTLSANAK	MS	NN
Rv2301	*cut2*	ATA	↓ACPDAEVVFAR	LC-MS/MS	NN
Rv2576c	-	ARA	↓DPVGHQVTYTVTTTSDLMANIR	LC-MS/MS	NNHMM
Rv2668	-	AWA	↓GDAPIGHIGDTLR	MS & LC-MS/MS	HMM
Rv2693	-	SSA	↓AGLLPAIGFALSMAGLILLWRLLR	LC-MS/MS	HMM
Rv2911	*dacB2*	AWA	↓DADVQPAGSVPIPDGPAQTWIVADLDSGQVLAGR	MS	HMM
Rv3354	-	AQA	↓NPVDDAFIAALNNAGVNYGDPVDAK	LC-MS/MS	NNHMM
Rv3491	-	ARA	↓LGPPPDGSYSFNQAGVSGVTWTITALCDQPSGTR	MS/MS	NNHMM
Rv3668c	-	AAA	↓DDKLPLGGGAGIVVNGDTMCTLTTIGHDK	LC-MS/MS	NNHMM
Rv3705c	-	ADA	↓HPSEPGVVSYAVLGK	LC-MS/MS	NNHMM
Rv3759c	*proX*	ANA	↓DPLGSATGSVK	LC-MS/MS	NNHMM
Rv3803c	*fbpD*	AKA	↓APYENLMVPSPSMGR	MS & LC-MS/MS	HMM
Rv3804c	*fbpA*	AGA	↓FSRPGLPVEYLQVPSPSMGR	MS	HMM

↓ The experimentally identified signal peptidase I cleavage site.

[a]N-terminal signal peptide predicted only by Neural Network method (NN).

[b]N-terminal signal peptide predicted by both Neural Network method and Hidden Markov Model (NNHMM).

[c]N-terminal signal peptide predicted only by Hidden Markov Model (HMM).

Table 3: Signal peptides of *M. tuberculosis* H37Rv culture filtrate proteins with experimentally determined N-terminal mature peptides

Gene number	Signal sequence	Mature sequence
Rv0040c	[a]MIQIARTWRVFAGGMATGFIGVVLVTAGKASA	[b]↓DPL
Rv0063	MAREISRQTFLRGAAGALAAGAVFGSVRATA	↓DPA
Rv0125	MSNSRRRSLRWSWLLSVLAAVGLGLATAPAQA	↓APP
Rv0129c	MTFFEQVRRLRSAATTLPRRLAIAAMGAVLVYGLVGTFGGPATAGA	↓FSR
Rv0285	MTLRVVPEGLAAASAAVEALTARLAAAHASA	↓APV
Rv0291	MIRAAFACLAATVVVAGWWTPPAWA	↓IGP
Rv0398c	MGVIARVVGVAACGLSLAVLAAAPTAGA	↓EPT
Rv0455c	MSRLSSILRAGAAFLVLGIAAATFPQSAAA	↓DST
Rv0477	MKALVAVSAVAVVALLGVSSAQA	↓DPE
Rv0526	MQSRATRRSGALTMRRLVIAAAVSALLLTGCSGRDAVA	↓QGG
Rv0559c	MKGTKLAVVVGMTVAAVSLAAPAQA	↓DDY
Rv0867c	MSGRHRKPTTSNVSVAKIAFTGAVLGGGGIAMAAQATA	↓ATD
Rv0999	MRPPLAPQFAADLLVKTVSTLRSSGAALGRLTTMRKAVLAVGSVCWLVGCSSGASS	↓TTA
Rv1158c	MPTIWTFVRAAAVLVGSSAALLTGGIAHA	↓DPA
Rv1174c	MRLSLTALSAGVGAVAMSLTVGAGVASA	↓DPV
Rv1269c	MTTMITLRRRFAVAVAGVATAAATTVTLAPAPANA	↓ADV
Rv1352	MARTLALRASAGLVAGMAMAAITLAPGARA	↓ETG
Rv1477	MRRNRRGSPARPAARFVRPAIPSALSVALLVCTPGLATA	↓DPQ
Rv1804c	MRVVSTLLSIPLMIGLAVPAHA	↓GPS
Rv1810	MQLQRTMGQCRPMRMLVALLLSAATMIGLAAPGKA	↓DPT
Rv1815	MVRLVPRAFAATVALLAAGFSPATASA	↓DPV
Rv1845c	MSALAFTILAVLLAGPTPALLA	↓RAT
Rv1886c	MTDVSRKIRAWGRRLMIGTAAAVVLPGLVGLAGGAATAGA	↓FSR
Rv1906c	MRLKPAPSPAAAFAVAGLILAGWAGSVGLAGA	↓DPE
Rv1910c	MAHAFHRFALAILGLALPVALVAYGGNGDSRKA	↓APL

Rv1926c	MKLTTMIKTAVAVVAMAAIATFAAPVALA ↓AYP
Rv1984c	MTPRSLVRIVGVVVATTLALVSAPAGGRAAHA ↓DPC
Rv2190	MRLDQRWLIARVIMRSAIGFFASFTVSSGVLAANVLA ↓DPA
Rv2253	MSGHRKKAMLALAAASLAATLAPNAVA ↓AAE
Rv2301	MNDLLTRRLLTMGAAAAMLAAVLLLTPITVPA ↓GYP
Rv2576c	MPAGVGNASGSVLDMTSVRTVPSAVALVTFAGAALSGVIPAIARA ↓DPV
Rv2668	MRHWLIVLATLLVAAAGVAAANDVPRAWA ↓GDA
Rv2693	MNANRTSAQRLLAQAGGVSGLVYSSLPVVTFVVASSA ↓AGL
Rv2911	MRKLMTATAALCACAVTVSAGAAWA ↓DAD
Rv3354	MNLRRHQTLTLRLLAASAGILSAAAFAAPAQA ↓NPV
Rv3491	MNIRCGLAAGAVICSAVALGIALHSGDPARA ↓LGP
Rv3668c	MQTAHRRFAAAFAAVLLAVVCLPANTAAA ↓DDK
Rv3705c	MRIAAAVVSIGLAVIAGFAVPVADA ↓HPS
Rv3759c	MRMLRRLRRATVAAAVWLATVCLVASCANA ↓DPL
Rv3803c	MKGRSALLRALWIAALSFGLGGVAVAAEPTAKA ↓APY
Rv3804c	MQLVDRVRGAVTGMSRRLVVGAVGAALVSGLVGAVGGTATAGA ↓FSR

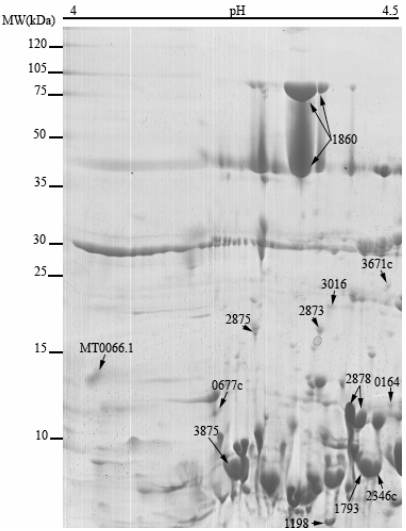
[a] ↓ The experimentally identified signal peptidase I cleavage site.

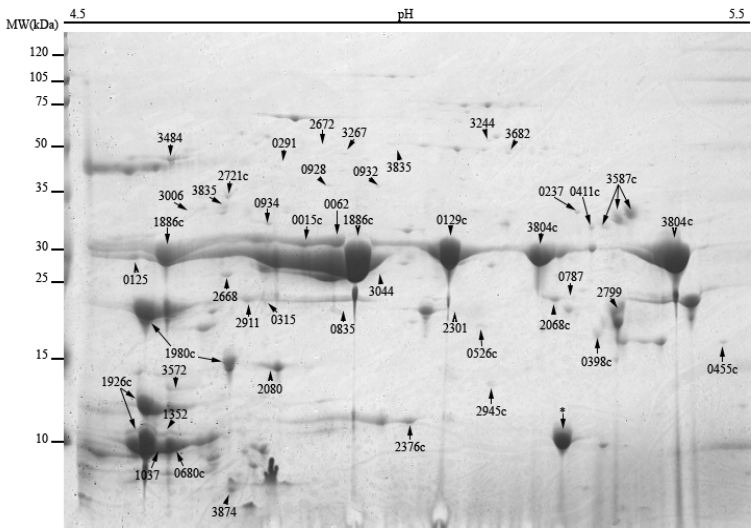
Supplentary material

Supplementary table 1. A detailed list of identified *M. tuberculosis* H37Rv culture filtrate proteins by 2D-PAGE combined with MALDI-TOF MS and liquid chromatography coupled with MS/MS.

Supplementary table 2. List of predicted secreted or exported *M. tuberculosis* H37Rv culture filtrate proteins identified in this study, with their possible retention peptides.

Figure 1


MW(kDa)
4
pH
4.5
120
105
75
50
35
30
25
15
10
1860
3671c
3016
2875
2873
MT0066.1
0677c
2878
0164
3875
2546c
1793
1198


MW(kDa)
4.5
pH
5.5
120
105
75
50
35
30
25
15
10
3484
2672
3267
0291
3244
3682
0928
0932
3835
2721c
3006
3835
0934
0062
0237
0411c
3587c
1886c
0015c
1886c
0129c
3804c
3804c
0125
0787
2668
3044
2799
2911
0315
0835
2301
2068c
1980c
0526c
0398c
0455c
3572
1926c
2080
2945c
1352
2376c
1037
0680c
3874

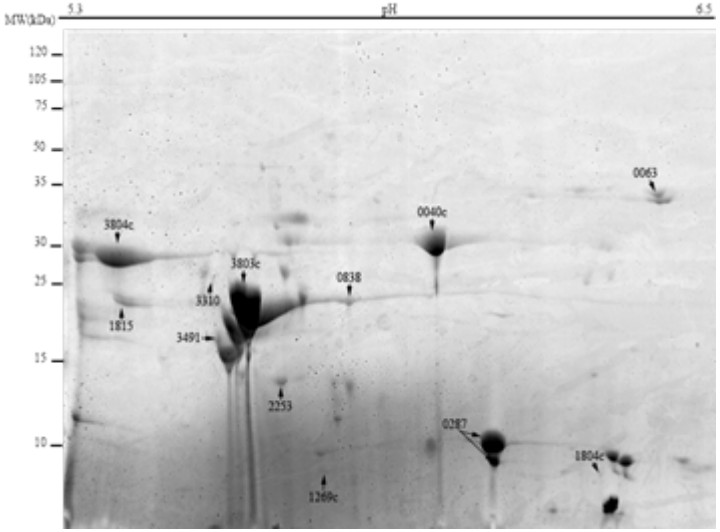

MW(kDa)
5.3
pH
6.5
120
105
75
50
35
30
25
15
10
0063
0040c
3804c
3803c
0838
3310
1815
3491
2253
0287
1804c
1269c

Figure 2.

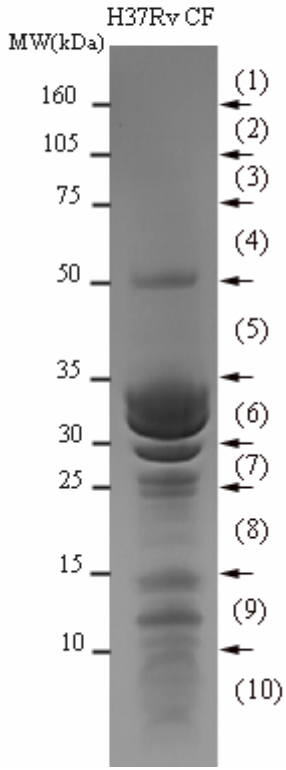


Figure 3.

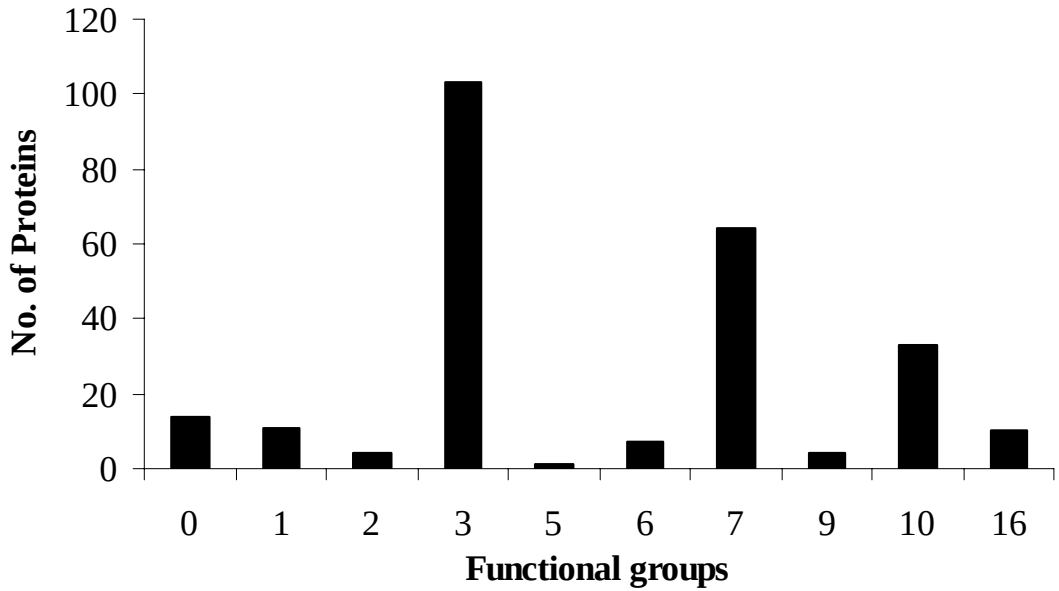

120
100
80
60
40
20
0
No. of Proteins
0
1
2
3
5
6
7
9
10
16
Functional groups

Figure 4.

a)

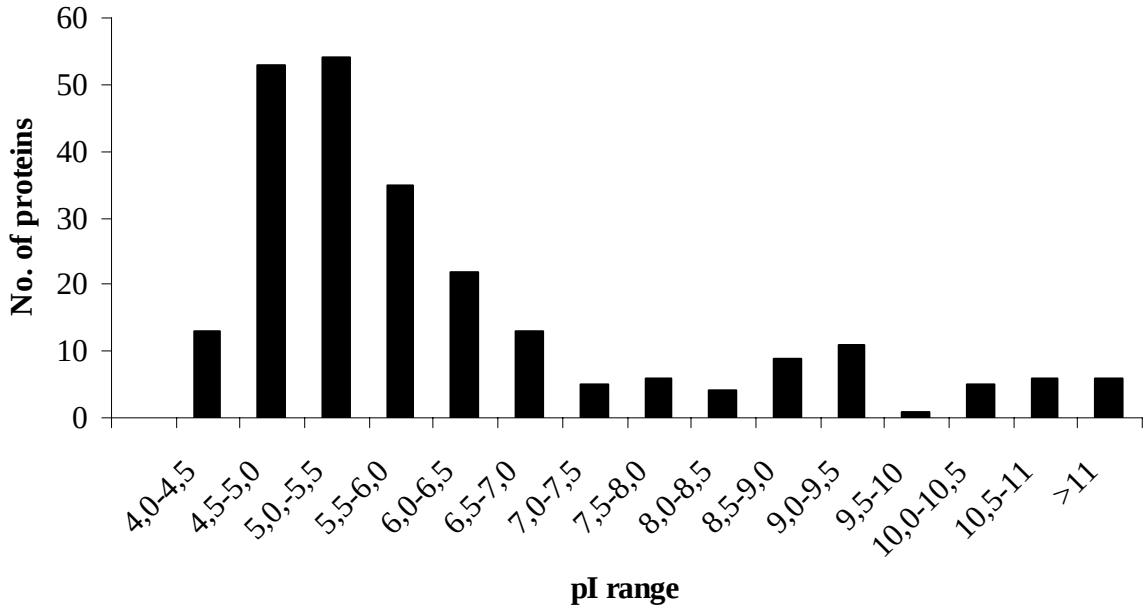

60
50
40
30
20
10
0
No. of proteins
4,0-4,5
4,5-5,0
5,0,-5,5
5,5-6,0
6,0-6,5
6,5-7,0
7,0-7,5
7,5-8,0
8,0-8,5
8,5-9,0
9,0-9,5
9,5-10
10,0-10,5
10,5-11
>11
pI range

Figure 4.

b)

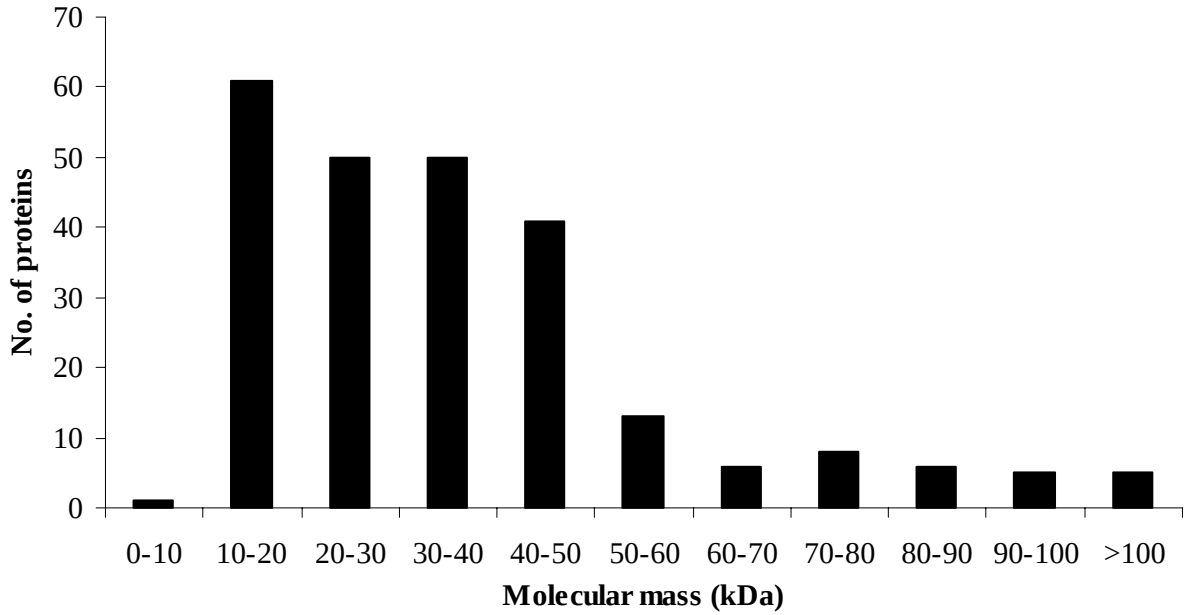
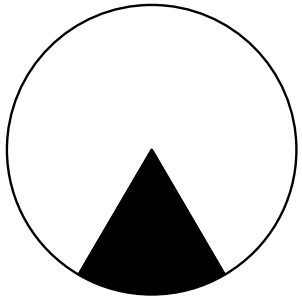

70
60
50
40
30
20
10
0
No. of proteins
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
>100
Molecular mass (kDa)

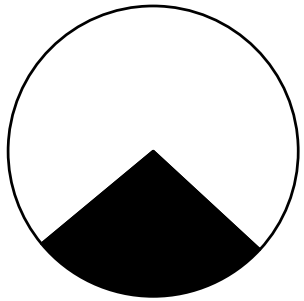
Figure 5

a) 3267



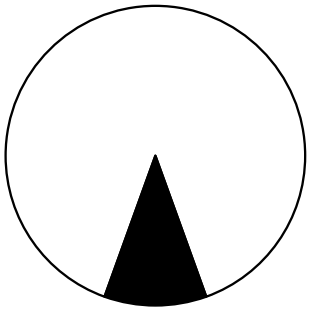
657

b) 38



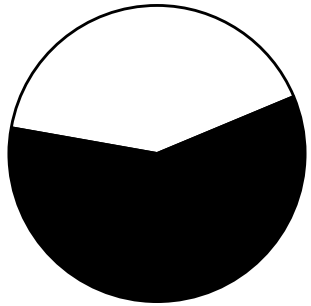
14

c) 135



17

d) 98



159

Supplementary table 1. A detailed list of *M. tuberculosis* H37Rv culture filtrate proteins identified by 2D-PAGE combined with MALDI-TOF MS and liquid chromatography coupled MS/MS

Gene number [a]	Theoretical molecular mass (kDa) [c]	Observed molecular mass (kDa) [d]	Theoretical isoelectric point (pI) [e]	Observed isoelectric point (pI) [f]	Identification method [g]	Sequence coverage [h]	Peptide sequence [i]	Fraction number [j]	Batch number [k]	Score	Minimum significance score (p<0.05)	Charge
Rv0002	42.1	50	4.76	4.79	MS	7/20%			A	52	51	
Rv0009	19.2	17	5.81	6.28	MS	7/57%			A	89	51	
Rv0012	23.22/23.22		6.57/6.57		LC-MS/MS	1	IDSVHGRSVDTALAAMQR	6	A	17	11	3
Rv0015c	43.34/-	31	10.33/-	4.88	MS	6/20%			A	54	51	
Rv0019c	10.33/-		8.38/-		LC-MS/MS	1	QGLVLQLTR	10	B	20	13	2
Rv0040c	28.92/27.29	30	8.68/8.91	5.8	MS	8/37%			A	105	43	
					LC-MS/MS	7	ALDITLPMPPR	10	A,C	20	12	2
					LC-MS/MS		WTQVPDPNVPDAFVVIADR	6	A,B,C	33	18	2
					LC-MS/MS		ENDMTLNTSR	6	A	37	21	2
					LC-MS/MS		HVIATSGADK	6	A	15	14	2
					LC-MS/MS		LGNSVYTSNAQLVVYR	6	A,B,C	97	23	2
					LC-MS/MS		LIGDFDPAEAITHGYIDSQK	6	A,B,C	51	21	2
					LC-MS/MS		LLAWQTTNASMANFDGFPSSIIEGTYR	8	A	80	19	3
Rv0054	17.35	20	5.12	5.2	MS	7/73%			A	143	51	
Rv0062	32.62/-	31.5	4.96/-	4.87	MS	9/64%			A	97	51	
Rv0063	46.21/46.21	35	6.81/6.81	6.83	MS	8/25%			A	54	51	
					LC-MS/MS	8	AMAFAAANNLK	7	A	67	22	2
					LC-MS/MS		AYSVGGYVNYLEVNQPPAR	4	A,C	41	24	2
					LC-MS/MS		DPAASGWEALSSALGGK	5	C	105	11	2
					LC-MS/MS		ILATCPAGSGGSVAAAIVSAVGTQPTG-TENHTFNYLDLVR	7	A	21	20	3
					LC-MS/MS		VLQPDDGPQFATAK	7	A,C	64	22	2
					LC-MS/MS		VTVTPATGLYAMHQVLAAAGR	3	A,C	15	12	3

					LC-MS/MS		YFGPNLSR	4	A,C	42	22	2
					LC-MS/MS		YLAVGNLNPSPLGYVGGSDVFTTITPA-TAQGIASAVDAFPR	7	A	70	19	3
Rv0064	103.44/102.95		6.26/6.38		LC-MS/MS	1	YELLSSGRK	5	A	21	13	1
Rv0066	82.54		5.7		LC-MS/MS	1	VPDNLAELGR	3	A	40	19	2
Rv0075	37.7		5.1		LC-MS/MS	1	ANRDHLAR	3	A	19	16	2
Rv0116c	23.88/22.13		7.36/7.36		LC-MS/MS	1	TVVMDSR	10	A	19	13	2
Rv0125	31.50/31.50	29	4.53/4.53	4.5	MS/MS	5	TQDVAVLQLR		A	64	51	1
					LC-MS/MS		TQDVAVLQLR	10	A,C	53	10	2
					LC-MS/MS		TGNVTLAEGPPA	8	A,C	18	11	2
					LC-MS/MS		GAGGLPSAAIGGGVAVGEPVVAMGNSGGQGGTPR	8	A,C	73	12	3
					LC-MS/MS		SGGGSPTVHIGPTAFLGLGVVDNNGNGAR	10	C	23	11	3
					LC-MS/MS		APPALSQDR	6	A	26	11	2
Rv0129c	32.02/32.02	32.5	4.99/4.99	5.6	MS	9/21%			A	80	51	
					LC-MS/MS	13	ADIQHVLNGATPPAAPAAPAA	8	A,C	18	10	2
					LC-MS/MS		DTYAADGGR	10	A	51	21	2
					LC-MS/MS		EMPAWLQANK	10	A,B,C	39	14	2
					LC-MS/MS		FLEGLTLR	10	A,B,C	36	13	2
					LC-MS/MS		FSRPGLPVEYLQVPSASMGR	8	A,B,C	43	16	3
					LC-MS/MS		IWVYCGNGTPSDLGGDNIPAK	6	A,B,C	10	9	2
					LC-MS/MS		LVANNTR	8	A	41	24	2
					LC-MS/MS		NDPMVQIPR	5	A,C	36	22	2
					LC-MS/MS		NGVFNFPPNGTHSWPYWNEQLVAMK	5	A,B,C	46	23	3
					LC-MS/MS		RNDPMVQIPR	5	A	38	22	2
					LC-MS/MS		TNQTFR	8	A	26	21	2
					LC-MS/MS		VQFQGGGPHAVYLLDGLR	1	A,B,C	97	10	3
					LC-MS/MS		WETFLTR	10	A,C	41	22	2
Rv0140	13.76		4.17		LC-MS/MS	1	MSNRIVLEPSADHPITIEPTNR	3	A	18	17	3
Rv0164	17.06/-	13	5.34/-	4.47	MS/MS	1	YPEWNEGVK		A	56	51	1
Rv0170	34.19/34.61		7.93/8.50		LC-MS/MS	1	ALHLVDGGRR	3	A	11	11	3

Rv0172	53.08/52.95		4.85/4.85		LC-MS/MS	1	TQVPTEWDELR	6	A	32	24	2
Rv0173	38.81[b]		6.01[b]		LC-MS/MS	1	LLAYVGGRSEVLNR	10	A,C	12	13	2
Rv0174	51.71/51.23		5.16/5.16		LC-MS/MS	2	GTVPSEIGPALDNSNR	8	A	28	13	2
					LC-MS/MS		VTAVEPTDQGAR	6	A	54	23	2
Rv0192	38.9		7.24		LC-MS/MS	1	LTVSDAVR	4	A	26	24	2
Rv0203	10.27/10.34		6.23/6.29		LC-MS/MS	1	AHFEANPK	10	A	31	23	2
Rv0211	67.25		4.68		LC-MS/MS	1	ALHSVGAPLEPGQK	3	A	33	17	3
Rv0219	16.27/16.27		10.96/10.96		LC-MS/MS	1	AGCSRVDAIDEE	10	C	15	12	3
Rv0237	37.67[b]	32	5.27[b]	5.2	MS	11/45%			A	76	51	
Rv0242c	46.83		6.4		LC-MS/MS	2	VVVVGGTPEAAASTNER	7	A	36	22	2
					LC-MS/MS		GQTNYATTK	5	A	39	22	2
Rv0244c	66.01	34	5.19	4.48	MS	6/21%			A	53	51	
Rv0265c	32.73[b]		5.43[b]		LC-MS/MS	1	AVLDAADVLIWMTESPEDEK	10	A	7	6	3
Rv0283	55.94		7.32		LC-MS/MS	1	SPIDLADHAVTSGLGLGADVPAPR	10	A	20	19	3
Rv0285	6.59/6.59		4.00/4.00		LC-MS/MS	2	AGVGVGESGASYLAGDAAAAATYGVVGG	8	A	67	23	3
					LC-MS/MS		APVITAVVPPAADPVSLQTAAGFSA-QGVEHAVVTAEGVEELGR	10	A	65	23	3
Rv0287	9.7	10	5.99	6	MS/MS	3	SLLDAHIPQLVASQSAFAAK		A	86	51	1
					LC-MS/MS		SLLDAHIPQLVASQSAFAAK	10	A	77	22	3
					LC-MS/MS		HTIGQAEQAAMSAQAFHQGESSAAFQAAHAR	10	A,C	78	19	4
					LC-MS/MS		VNTLLDVAQANLGEAAGTYVAADAAAASTYTGF	10	A	60	20	3
Rv0291	43.48/42.59	40	5.17/5.30	4.7	MS	7/21%			A	65	51	
					LC-MS/MS	2	GEGQLVAIIDTGVQPGPR	5	B	36	19	2
					LC-MS/MS		LVALSGTSYAAGYVSGVAALVR	5	B	35	19	3
Rv0309	19.09/18.62		8.48/8.49		LC-MS/MS	2	MDVYQR	10	A	31	22	2
					LC-MS/MS		HSVVMGVNK	10	A	27	23	2
Rv0315	28.81/27.11	22	4.79/4.77	4.75	MS	6/36%			A	66	51	
					LC-MS/MS	9	EWPFNDPGYK	6	C	43	24	2
					LC-MS/MS		FNCLAPGMWPAWWLSNDDPGR	9	C	13	9	3
					LC-MS/MS		GGIGTTWEAR	9	A,C	41	23	2

					LC-MS/MS		VFPVLNLAVGGSGGGDPATGSYPQEMLVDWVR	9	A,C	78	19	3
					LC-MS/MS		QNVFLDGNSNLVLR	9	C	65	22	2
					LC-MS/MS		SGEIDLIEWYGNGTWPSGTTVHANP-DGTAFETCPIGVDGGWHNWR	9	A	44	17	4
					LC-MS/MS		VTWNPSGMYFWLDYADGIEPYFSVPATGIEDLNEPIR	10	A	21	20	3
					LC-MS/MS		WQVSNHR	9	A	44	22	2
							TPIKNPVGFDRPQFFGQYR	9	C	20	14	2
					LC-MS/MS		YFGGLVHGLWR	8	C	42	23	2
Rv0340	18.35		4.53		LC-MS/MS	2	ANSLLDFVISLVR	7	A	38	22	2
					LC-MS/MS		SIAEAHLTDVTR	8	A	33	28	2
Rv0350	66.83	65	4.85	4.79	MS	6/36%			A	232	51	
					LC-MS/MS	7	NTTIPTKR	5	A	52	24	2
					LC-MS/MS		MPAVTDLVK	5	A	24	23	2
					LC-MS/MS		NQAVTNVDR	3	A	43	23	2
					LC-MS/MS		YTAPEISAR	3	A	41	22	2
					LC-MS/MS		NQAETLVYQTEK	3	A	54	21	2
					LC-MS/MS		AALGGSDISAIK	3	A	42	23	2
					LC-MS/MS		TTPSIVAFAR	3	A,C	34	23	2
Rv0363c	36.54	37	5.49	6	MS	10/44%			A	74	51	
Rv0398c	19.16/19.16	17	5.36/5.36	5.4	MS	10/48%			A	97	51	
					LC-MS/MS	5	DVASVFLPLQR	7	A,C	33	22	2
					LC-MS/MS		ISQQLFSFGDPTVQEVDGSDAAQFI-TAAAAVADRDVASVFLPLQR	10	A	32	18	4
					LC-MS/MS		VLGCQQNTAGSGAGFGAR	10	A,C	92	21	2
					LC-MS/MS		ISQQLFSFGDPTVQEVDGSDAAQFITAAAAVADR	10	A	35	19	3
					LC-MS/MS		RGEEGYFVLLAGTASDFCSAPNANYR	6	A	49	9	3
Rv0402c	100.65/99.77		6.31/6.21		LC-MS/MS	1	VADLSTLTDQLQRMIDITQR	7	A	12	12	4
Rv0408	72.94		5.2		LC-MS/MS	2	KIDTALELMDR	8	A	27	23	3
					LC-MS/MS		LRDSPVAGR	3	A	17	17	2
Rv0411c	32.87[b]	31	5.04[b]	5.25	MS	11/33%			A	91	51	

					LC-MS/MS	3	ILSAAER	3	A	16	15	2
					LC-MS/MS		TMSITCER	8	A	18	12	2
					LC-MS/MS		EIAPPPVIVSVVNWADCLVALQQR	1	A	28	11	3
Rv0436c	27.55/26.03		10.01/10.06		LC-MS/MS	1	ILDAQSR	8	A	18	14	2
Rv0446c	26.41/26.41		10.57/10.57		LC-MS/MS	1	GATPVQALRK	10	A	18	14	2
Rv0453	-/49.88		-/4.59		LC-MS/MS	1	MPMLPGAWDLGTWDR	4	C	13	11	3
Rv0455c	13.70/13.70	16	5.72/5.72	5.3	MS	5/41%			A	68	51	
					LC-MS/MS	2	DSTEDFPIPR	10	C	55	14	2
					LC-MS/MS		DTSPVYYQR	10	A	33	22	2
Rv0462	49.2	50	5.7	5.6	MS	11/30%			A	98	51	
					LC-MS/MS	1	ALPNEDADVSK	6	A	42	23	2
Rv0472c	26.38		7.72		LC-MS/MS	1	MLAVMLR	2	A,C	21	15	2
Rv0477	13.52/13.52	12.5	4.56/4.56	4.5	MSMS	4	LVDHTEWAQWGSLPSLR		A	75	51	1
					LC-MS/MS		DPEADPGAGEANYGGPPSSPR	10	A	44	13	2
					LC-MS/MS		VYPSQVGR	10	A	36	23	2
					LC-MS/MS		LVDHTEWAQWGSLPSLR	9	A	27	12	3
Rv0490	42.77/42.77		6.07/6.07		LC-MS/MS	1	SRATGGSGLGLAIVK	7	A	17	11	3
Rv0501	41.06	32	10.38	5.2	MS	8/22%			A	54	51	
Rv0506	13.44/12.52		6.99/6.31		LC-MS/MS	1	IAYLDPDAR	9	A	29	19	2
Rv0526	20.06[b] 19.66/19.30	18	5.17[b] 5.00/5.16	5.5	MS	11/60%			A	112	51	
					LC-MS/MS	3	GAGVSFLGIDVR	10	A	26	23	2
					LC-MS/MS		VAEEEPSGR	10	A	59	24	2
					LC-MS/MS		AEVSQLQR	4	A	45	24	2
Rv0559c	9.75/9.75		7.01/7.01		LC-MS/MS	5	GTTQGQAFQFLGAAIDHYCPEHVGVLQR	9	A	73	12	4
					LC-MS/MS		GVDGDAYK	10	A	30	21	2
					LC-MS/MS		SATFLQR	4	A	24	23	2
					LC-MS/MS		FGIYGPQDYNAWLAK	10	A	59	22	2
					LC-MS/MS		DDYDAPFNNTIHR	10	A,C	27	10	2
Rv0563	26.58/26.18		9.45/9.45		LC-MS/MS	1	IVRELATSAHQPMPR	6	C	14	10	3

Rv0583c	21.83[b]		4.39[b]		LC-MS/MS	1	LLPESSR	9	A	14	13	2
Rv0631c	119.5		6.39		LC-MS/MS	1	TRNHIAR	4	A	17	16	2
Rv0674	24.28/24.28		7.02/7.02		LC-MS/MS	1	SADGYRLSDR	5	A	13	9	2
Rv0677c	12.98/10.32	12	4.57/4.26	4.24	MS	5/61%			A	57	51	
					LC-MS/MS	2	ITVDGEVKDER	10	A	37	23	2
					LC-MS/MS		VFADDPEPFDPK	9	C	15	10	2
Rv0680c	10.33/10.33	9	4.65/4.65	4.6	MS/MS	1	NGDPFIWDR		A	56	51	1
Rv0732	43.93/41.85		9.23/9.32		LC-MS/MS	1	FEELR	6	A	12	11	1
Rv0774c	24.84/27.36		5.41/5.90		LC-MS/MS	1	ASGEDAGAMVLNELIPLLDTQR	6	A	24	10	2
Rv0787	31.70/-	23	4.78/-	5.18	MS	6/31%			A	58	51	
					LC-MS/MS	1	ACQLGAPLQSPSVTDDEPTR	10	A	98	21	2
Rv0835	20.83[b]	21	4.82[b]	4.83	MS	6/27%			A	52	51	
					LC-MS/MS	1	EPPEADTNVPGPCR	7	A	21	13	2
Rv0838	24.98[b]	24	5.97[b]	5.78	MS	6/33%			A	105	43	
					LC-MS/MS	3	AAGLVDVR	9	A	40	25	2
					LC-MS/MS		GVVPDAAIDLR	6	C	31	4	2
					LC-MS/MS		SVDVTFASAQR	7	C	18	9	3
Rv0843	35.72		5.87		LC-MS/MS	1	LIAAGTTR	8	A	33	24	2
Rv0851c	-/26.76		-/5.53		LC-MS/MS	1	GARVVLGDVDKPGLR	10	C	6	10	3
Rv0867c	36.75/36.31		3.64/3.64		LC-MS/MS	2	ATDGEWDQVAR	4	A	12	12	2
					LC-MS/MS		GLSNATPR	5	A	44	24	2
Rv0876c	57.93		11.35		LC-MS/MS	1	VMPPTIDLVR	3	A	21	20	2
Rv0884c	40.23	47.5	4.77	4.8	MS	8/37%			A	72	51	
Rv0928	35.84[b]	35	5.26[b]	4.8	MS	8/27%			A	56	51	
					LC-MS/MS	3	YPDSQVGTAVK	6	A	21	12	2
					LC-MS/MS		SDESGTTDNFQR	7	A	36	22	2
					LC-MS/MS		RPGSYPIVLATYEIVCSK	6	B	32	19	3
Rv0932c	35.73[b]	35	4.76[b]	5.9	MS	5/30%			A	66	58	
					LC-MS/MS	2	SGTSDNFQK	6	A	20	9	2
					LC-MS/MS		AFMQAAIGPGQEGLDQYGSIPLPK	6	B	20	17	3

Rv0934	35.90[b]	32	4.82[b]	4.75	MS	9/37%			A	65	51		
					LC-MS/MS	6	ASFLDQVHFQPLPPAVVK	5	B	45	18	3	
					LC-MS/MS		DAATAQTLQAFLHWAITDGNK	5	B	29	21	2	
					LC-MS/MS		GLMNIALAISAQQVNYNLPGVSEHLK	5	B	37	18	3	
					LC-MS/MS		SDGSGDTFLFTQYLSK	6	B	24	17	2	
					LC-MS/MS		SDGSGDTFLFTQYLSKQDPEGWGK	5	B	63	17	3	
					LC-MS/MS		TWDDPQIAALNPGVNLPGTAVVPLHR	7	B	50	18	3	
Rv0982	48.69/48.69		7.23/7.23		LC-MS/MS	1	MALNLMDNAAKWSPPGGHVGVR	9	C	10	9	3	
Rv0996	35.97/36.16		4.62/4.60		LC-MS/MS	1	LGVENTR	9	A	15	14	2	
Rv0999	20.88[b] 20.30/20.02		6.13[b] 6.07/6.08		LC-MS/MS	1	TTASTGDIAK	10	A	23	13	2	
Rv1037c*	9.8	10	4.48	4.6	MS/MS	2	NFQVIYEQANAHGQK		A	110	51	1	
					LC-MS/MS		TINYQFGDVDAHGAMIR	9	A	24	24	2	
Rv1038c*	11	10	5.17	5.2	MS	4/33%			A	59	51		
					LC-MS/MS	4	FEVHAQTVEDEAR	10	A,C	96	23	2	
					LC-MS/MS		FEVHAQTVEDEARR	9	A,C	53	18	4	
					LC-MS/MS		NIVNMLHGVR	9	A	51	22	2	
					LC-MS/MS		FMTDPHAMR	10	A	69	45	2	
Rv1050	-/29.46		-/10.41		LC-MS/MS	1	SGRIMNMSSVVGR	10	A	14	11	3	
Rv1074c	42.65	47	4.92	4.8	MS	11/45%			A	113	51		
Rv1075c	30.88/29.31		10.52/10.49		LC-MS/MS	1	ALAHTRGVR	7	A	11	9	2	
Rv1077	48.63	50	5.17	5.2	MS	8/28%			A	74	51		
Rv1098c	50.14		5.18		LC-MS/MS	1	TAANSFEAQAAR	6	A	35	23	2	
Rv1158c	18.54/15.12		4.23/4.56		LC-MS/MS	2	VDLPQLPYLPLQVPQQLSLPADLPALASGV-IPAAPIAPTPPAPGAPALPPGPPSLLAALP	4	A	9	6	4	
					LC-MS/MS		DPAPAPAPAPNIPQQLISSAANAPQILQN-LATALGATPPLSAPK	10	A	23	10	3	
Rv1161	136.92		6.67		LC-MS/MS	1	GVPVR	5	A	24	17	1	
Rv1166	63.48[b]		5.14[b]		LC-MS/MS	1	VENIDPQR	5	A	13	11	2	
Rv1174c	8.34/8.34		4.23/4.23		LC-MS/MS	3	NFLAAPPPQR	10	A,C	50	21	2	

					LC-MS/MS		DPVDAVINTTCNYGQVVAALNATDPGA-AAQFNASPVAQSYLR	2	A	70	10	4
					LC-MS/MS		AAMAAQLQAVPGAAQYIGLVESVAGSCNNY	10	A	37	12	3
Rv1183	104.06/103.78		8.98/8.88		LC-MS/MS	2	GAQPNTSLADASISMSGYPVMLRDIR	9	A	14	9	3
					LC-MS/MS		DNGTLDKVVGLAR	7	C	17	12	3
Rv1186c	57.52		5.9		LC-MS/MS	1	HPSDSVVAGAVR	7	A	17	17	2
Rv1197*	10.96		5.02		LC-MS/MS	1	MWASAQNISGAGWSGMAEATSLDTMAQMNQAF	10	A	32	20	2
Rv1198	9.9	5	4.97	4.5	MS/MS	2	AQAGLLEAEHQAIIR		A	78	51	1
					LC-MS/MS		AQAGLLEAEHQAIIR	10	A	73	23	2
Rv1252c	17.71[b]		5.21[b]		LC-MS/MS	1	SEQPWNPEPLAGNYNECAQLSAVVIK	10	C	11	10	3
Rv1269c	9.08/9.08	8	6.26/6.26	5.7	MS	3/32%			A	45	43	
Rv1270c	22.60[b]		4.93[b]		LC-MS/MS	7	DASVAGSQQADGVATTK	9	A	21	13	2
					LC-MS/MS		GLANLLANLK	7	B	38	19	2
					LC-MS/MS		ITGNSSADDIATLAGSR	7	B	92	18	2
					LC-MS/MS		LAVTGDVPNLR	7	A,B	33	19	2
					LC-MS/MS		LEGDISNTPQTVATGSATLLVGNK	7	B	20	18	3
					LC-MS/MS		LEGDISNTPQTVATGSATLLVGNKSEDAK	7	B	22	17	3
					LC-MS/MS		TVPTTVWIASDGSSHLVQIQIAPTK	7	B	26	18	3
Rv1310	53		4.6		LC-MS/MS	1	NFGGTSVFAGVGER	3	B	19	18	3
Rv1323	40.08	40	4.91	5	MS	11/35%			A	71	51	
Rv1352	9.98/9.98	11	4.63/4.63	4.58	MS/MS	2	TEGPSNPLILVFGR		A	127	51	1
					LC-MS/MS		TEGPSNPLILVFGR	10	A,B,C	92	23	2
					LC-MS/MS		ETGEQFPGDGVFLVGTDIAPGTYR	10	A	44	9	3
Rv1371	55.12	10	9.44	4.33	MS	10/5%			A	55	51	
Rv1382	14.43/15.83		5.21/9.29		LC-MS/MS	1	AVLTRYPSGIMVER	8	A	18	15	2
Rv1386	6.39/7.03		4.63/4.63		LC-MS/MS	2	SGVGVAESGASYAAR	3	A	36	12	2
					LC-MS/MS		DALAAASYLSGGL	9	A	24	14	2
Rv1415	46.01	42.5	5.46	4.75	MS	9/33%			A	53	51	
Rv1419	13.65/13.65		4.18/4.18		LC-MS/MS	2	WNLTDDR	10	A	44	23	2
					LC-MS/MS		LQPCVNWISQHWTVQPDGLVK	9	C	12	11	3

Rv1435c	15.22/15.22		3.76/3.76		LC-MS/MS	1	DALTDPAPAGGPVPGQPVLPGPSASAPAGAR	10	A	33	13	3
Rv1436	35.95	35	5.19	5.25	MS	7/37%			A	52	51	
Rv1437	42.51	45	4.83	4.81	MS	9/36%			A	89	51	
Rv1438	27.4	30	5.54	5.66	MS	6/37%			A	53	51	
Rv1448c	40.72	50	4.87	4.88	MS	9/34%			A	101	51	
					LC-MS/MS	3	GQAGVANAR	5	A	29	23	2
					LC-MS/MS		ALSEGHTYDAQIAELAAR	5	B	32	19	3
					LC-MS/MS		TVTTDDVR	5	A	31	22	2
Rv1449c	75.55		4.78		LC-MS/MS	1	EDVLTHTR	4	A	24	22	2
Rv1454c	34.04	24	5.37	4.52	MS	12/72%			A	108	51	
Rv1475c	102.49	105	4.95	4.95	MS	32/44%			A	228	51	
					LC-MS/MS	6	IDTPGEADYYR	3	A	25	22	2
					LC-MS/MS		SNLIGMGVIPLQFPEGK	3	A	30	23	2
					LC-MS/MS		AVNDNDLSVTAVLSGNR	3	A	68	22	2
					LC-MS/MS		NGGILQYVLR	3	A	35	23	2
					LC-MS/MS		NEDGSNITK	2	A	27	23	2
					LC-MS/MS		AVIAESFER	2	A	26	22	2
Rv1477	45.72/45.72		6.07/6.07		LC-MS/MS	2	DANAAIAAAQHR	8	A	42	14	2
					LC-MS/MS		DPQTDTIAALIADVAK	7	C	22	11	2
Rv1488	37.88/37.88		6.15/6.15		LC-MS/MS	1	VARVELR	7	A	13	13	2
Rv1498A	7.62		6.23		LC-MS/MS	2	ALDWFEVQSIR	5	B,C	71	18	2
					LC-MS/MS		GHLVDGAVAHFQVTMK	5	B	26	18	3
Rv1541c	20.01[b]		6.20[b]		LC-MS/MS	1	DACAQDTDPR	3	A	11	11	3
Rv1729c	33.62		4.55		LC-MS/MS	1	TISNPFRCHGVDVDLASLVYTGPR	10	B	21	18	3
Rv1759c	-/71.30		-/5.07		LC-MS/MS	1	AGLYGNGGDGGAGGDGATSGKGGAG-GNAVVIGNGGNGGNAGK	10	A	13	10	3
Rv1784	101.47		5.46		LC-MS/MS	1	DVPVKPGR	4	A	17	16	2
Rv1793	10	9	4.76	4.45	MS/MS	1	AQAASLEAEHQAIVR		A	80	51	1
					LC-MS/MS		AQAASLEAEHQAIVR	7	A	68	19	3
Rv1804c	8.93/8.93	8.5	6.29/6.29	6.2	MS	3/54%			A	49	43	

Rv1810	8.43/8.43		4.50/4.50		LC-MS/MS	3	FAAIASGAYCPEHLEHHPS	10	A	39	12	2
					LC-MS/MS		DPTGDDAAFLAALDQAGITYADPGHAITAAK	10	A	97	9	3
					LC-MS/MS		DYNPGLTMDSAAK	10	A	31	13	2
Rv1812c	40.00/39.93		6.49/6.49		LC-MS/MS	1	TGVSVAAVSPGGVTLSSGERLAAATVVWCAGMR	8	A	12	6	4
Rv1815	20.18/20.18	24	5.33/5.33	5.4	MS/MS	5	QDNHVCTLGYVDPALK		A	121	51	1
					LC-MS/MS		QDNHVCTLGYVDPALK	9	A	43	16	3
					LC-MS/MS		STSEQVHADLGVTPLA	10	A	50	21	3
					LC-MS/MS		DNTPSGSTVATHELIADYEAIVLADDVTASNILPSGR	10	A	80	21	3
					LC-MS/MS		DPVLVFPGMEIR	10	A,C	54	13	2
					LC-MS/MS		GDSGGPVYLAPDGGPAQIVGIFNSVWGGFPAAVSWR	10	A	43	20	2
Rv1827	17.25	20	4.29	4.2	MS	4/39%			A	52	51	
Rv1833c	32.15		9.34		LC-MS/MS	1	TIIPR	6	A	19	19	1
Rv1837c	80.4	75	5.3	5.5	MS	23/37%			A	143	76	
Rv1845c	27.39/30.60		11.66/11.78		LC-MS/MS	1	RATWPLR	1	C	13	12	2
Rv1860	28.78/26.85	37	4.52/4.64	4.4	MS/MS	9	TTGDPPFPGQPPPVANDTR		A	65	51	1
					LC-MS/MS		TTGDPPFPGQPPPVANDTR	4	A,B,C	69	21	2
					LC-MS/MS		ALAESIRPLVAPPPAPAPAPAEPAPAP-APAGEVAPTPTTPTPQR	6	A,B	41	19	4
					LC-MS/MS		FSDPSKPNGQIWTGVIGSPAANAPDAGPPQR	7	A,B,C	45	12	3
					LC-MS/MS		IDNPVGGFSFALPAGWVESDAAHFDYGSALLSK	7	A	27	11	3
					LC-MS/MS		INQETVSLDANGVSGSASYYEVK	5	A,B	55	9	2
					LC-MS/MS		LYASAEATDSKAAAR	5	B	20	18	3
					LC-MS/MS		LGSDMGEFYMPYPGTR	4	A,B,C	51	10	2
					LC-MS/MS		LYASAEATDSK	5	A,B	65	23	2
					LC-MS/MS		WFVVWLGTANNPVDK	4	A,B,C	67	13	2
Rv1869c	41.61/42.11		4.92/4.93		LC-MS/MS	1	YDKLLLATGSAPR	4	C	11	12	2
Rv1876	18.34	18	4.5	4.48	MS	12/74%			A	145	51	
Rv1884c	11.21/-		8.01/-		LC-MS/MS	1	EQQIAVANR	5	A	32	23	2
Rv1886c	30.66/30.66	30	4.87/4.87	4.62	MS	8/41%			A	83	51	
					LC-MS/MS	10	AGCQTYKWETFLTSELPQWLSANR	8	A,B	57	20	3

					LC-MS/MS		FQDAYNAAGGHNAVFNFPPNGTHSWEYWGAQLNAMK	10	A	36	20	4
					LC-MS/MS		FQDAYNAAGGHNAVFNFPPNGTHSWEYWGAQLNAMKPDLQR	8	A	52	16	4
					LC-MS/MS		FSRPGLPVEYLQVPSPSMGR	8	A,B,C	54	22	3
					LC-MS/MS		LWVYCGNGTPNELGGANIPAEFLENFVR	8	A,B	41	13	3
					LC-MS/MS		NDPTQQIPK	4	A	30	13	2
					LC-MS/MS		VQFQSGGNNSPAVYLLDGLR	1	A,B,C	42	18	3
					LC-MS/MS		WETFLTSELPQWLSANR	8	A,B,C	32	19	2
					LC-MS/MS		AADMWGPSSDPAWER	6	A,B,C	50	12	2
					LC-MS/MS		AADMWGPSSDPAWERNDPTQQIPK	6	B	23	17	3
Rv1887	39.53		5.85		LC-MS/MS	1	VTSGEALTEPNPPEEQPNASAPQQDR	9	A	41	23	3
Rv1891	-/11.92		-/4.56		LC-MS/MS	1	GWQPGWFTGAGFFPPEP	10	A	16	12	2
Rv1899c	32.36[b]		9.79[b]		LC-MS/MS	1	HAGGVAAAIAR	1	C	12	13	2
Rv1906c	12.56/12.56		4.68/4.68		LC-MS/MS	2	KPQVVTIEPTDK	6	A	70	10	2
					LC-MS/MS		DPEPAPTPK	10	A	38	12	2
Rv1908c	80.7	45	5.13	5.5	MS	6/25%			A	61	51	
					LC-MS/MS	2	ALVEVYGADDAQPK	4	A	37	23	2
					LC-MS/MS		SSYGTGTGK	3	A	41	25	2
Rv1910c	17.57/16.49		5.71/5.38		LC-MS/MS	2	FTLYHLPAVPPLAGLAGTQAAR	10	A	28	22	3
					LC-MS/MS		APLAPKAAALGRSMPETPTGDVLTISSPAFADGAP	10	A	28	22	3
Rv1911c	17.57[b]		5.30[b]		LC-MS/MS	2	FTLYHLPVALQLPPGATGVQAAQAIAQAASGQAR	10	A	52	21	4
					LC-MS/MS		QGYFGPCPPAGTGTHHYR	6	A	43	10	3
Rv1926c	13.66/13.66	10.5	4.50/4.50	4.6	MS	5/52%			A	63	51	
					LC-MS/MS	7	GSVTPAVSQFNAR	10	A,B,C	50	10	2
					LC-MS/MS		IYFDVTGPSPTIVAMNNGMEDLLIWEP	10	A	50	23	2
					LC-MS/MS		LGSELTMTDTVGQVVLGWK	10	A,B,C	51	19	2
					LC-MS/MS		SSTAVIPGYPVAGQVWEATATVNAIR	10	A,B,C	63	17	3
					LC-MS/MS		SSTAVIPGYPVAGQVWEATATVNAIRGSVTPAVSQFNAR	10	A	35	21	3
					LC-MS/MS		TADGINYR	10	A	33	22	2
					LC-MS/MS		VLWQAAGPDTISGATIPQGEQSTGK	10	A,B,C	63	20	3
Rv1932	16.89	17	4.37	4.34	MS	7/61%			A	67	51	

Rv1980c	22.43/22.43	15	4.60/4.60	4.6	MS	5/26%			A	61	51		
					LC-MS/MS	10	AFDWDQAYR	6	A,B,C	26	16	2	
					LC-MS/MS		EAPYELNITSATYQSAIPPR	7	A,B,C	68	21	2	
					LC-MS/MS		FLSAATSSTPR	9	A	35	23	2	
					LC-MS/MS		GTDTGQACQIQMSDPAYNINISLPSYYPDQK	6	C	41	17	3	
					LC-MS/MS		GTQAVVLK	10	A	23	16	2	
					LC-MS/MS		KPITYDTLWQADTDPLPVVFPIVQGELSK	10	A	13	11	3	
					LC-MS/MS		QTGQQVSIAPNAGLDPVNYQNFAVTNDGVIF-FFNPGELLPEAAGPTQVLVPR	10	A	43	15	4	
					LC-MS/MS		SLENYIAQTR	6	A,B,C	42	22	2	
					LC-MS/MS		VYQNAGGTHPTTTYK	10	A	76	22	3	
					LC-MS/MS		DKFLSAATSSTPR	6	A,C	41	12	3	
Rv1984c	19.01/18.66		4.85/4.69		LC-MS/MS	4	ASASNGSDDASAHIQR	7	A	76	24	2	
					LC-MS/MS		GTHQASGLGDVGEAFVDSLTSQVGGR	8	C	36	21	3	
					LC-MS/MS		DPCSDIAVVFAR	10	A	35	22	2	
					LC-MS/MS		SIGVYAVNYPASDDYR	8	C	73	10	2	
Rv2006	14.57		6.25		LC-MS/MS	1	RLQVAGVR	4	A	30	23	2	
Rv2060	8.03/10.28		11.42/11.88		LC-MS/MS	1	GVPVR	5	A	24	17	1	
Rv2068c	30.02[b]	23	5.26[b]	5.1	MS	6/42%				74	51		
					LC-MS/MS	2	AGGGYDAEPR	9	A	32	23	2	
					LC-MS/MS		LITYTSDDIR	3	A	44	11	2	
Rv2074	15	12.5	9.5	4.47	MS	5/33%			A	55	51		
Rv2080	17.34[b]	15	5.19[b]	4.72	MS/MS	1	QIVAAADLQAVR		A	76	51	1	
Rv2110c	30.3	24	4.65	4.7	MS	8/40%			A	77	51		
Rv2140c	18.63	18	5.41	5.32	MS	5/76%			A	156	51		
					LC-MS/MS	1	TTSPDPYAALPK	10	A	34	22	2	
Rv2190c	35.71/35.71		5.74/5.74		LC-MS/MS	1	DPADDALAK	6	A	20	15	2	
Rv2192c	37.7		6.4		LC-MS/MS	1	HAAAVR	3	A	12	9	2	
Rv2200c	35.12/35.32		5.76/5.62		LC-MS/MS	1	GELAPQPVG	7	C	11	10	2	
Rv2201	70.21/70.21		6.45/6.45		LC-MS/MS	1	SFSGAQLR	2	C	10	9	2	

Rv2220	53.57	57	5.4	5	MS	5/45%			A	53	51	
					LC-MS/MS	1	IPITGSNPK	6	A	28	22	2
Rv2224c	52.65[b]		5.38[b]		LC-MS/MS	1	GVASSRPAIWCNSDADNDRLR	10	A	17	12	2
Rv2236c	33	25	10.85	5.3	MS	4/20%			A	55	51	
Rv2241	100.18		5.84		LC-MS/MS	1	IVPIIPDEAR	2	A	18	15	2
Rv2251	44.02/47.37		5.79/5.98		LC-MS/MS	1	ATLDPAGILNPGKLIP	10	A	20	13	2
Rv2253	15.21/15.14	14	5.59/5.59	6	MS	6/62%			A	47	43	
					LC-MS/MS	5	TGTSMAANRPEYPHK	10	A	13	11	3
					LC-MS/MS		ANYTFSSR	10	A	50	24	2
					LC-MS/MS		NEFIPRPIEYTWNGTQWVR	7	A,C	35	10	3
					LC-MS/MS		SITAYTPGQYGILTGVFHTDIASGTCK	7	A	29	13	3
					LC-MS/MS		GNVDMPVSAKPIVG	7	A	39	13	2
Rv2299	72.96		4.51		LC-MS/MS	1	LRIEALR	4	A	18	18	2
Rv2301	20.62/19.67	21	4.96/4.96	5.5	MS/MS	3	FEPPGIGTVGNAFVSALR		A	92	51	1
					LC-MS/MS		FEPPGIGTVGNAFVSALR	8	A,C	78	21	2
					LC-MS/MS		ACPDAEVVFAR	6	A	30	16	2
					LC-MS/MS		GRFEPPGIGTVGNAFVSALR	10	A,B	78	21	3
Rv2314c	48.71	50	4.93	4.96	MS	10/23%			A	61	51	
Rv2334	32.75	32.5	5.2	52	MS	11/53%			A	100	51	
Rv2346c	9.9	8	4.76	44.2	MS/MS	2	AQAGLLEAEHQAIVR		A	81	51	1
					LC-MS/MS		AQAGLLEAEHQAIVR	7	A	99	18	2
					LC-MS/MS		DVLAAGDFWGGAGSVACQEFITQLGR	9	A	103	20	3
Rv2376c	14.74/11.61	12	5.10/5.51	4.82	MS/MS	3	GSLVEGGIGGTEAR		A	71	51	1
					LC-MS/MS		GSLVEGGIGGTEAR	9	A,B,C	69	18	2
					LC-MS/MS		ASAMELLQAAGN	9	A,B	17	12	2
					LC-MS/MS		AAEHGDLPLSFSVTNIQPAAAGSATADVSVSGPK	10	A	32	22	3
Rv2401	11.1	15	5.17	4.28	MS	4/33%			A	54	51	
Rv2430c	21.98	27	4.77	4.98	MS/MS	2	APPPIAHSTVLVAPVSPSTASSR		A	91	51	1
					LC-MS/MS		APPPIAHSTVLVAPVSPSTASSR	9	A	40	18	3
					LC-MS/MS		SLDVEMTAVQR	9	A	44	22	2

Rv2431c	10.6	10.8	5.76	5.23	MS	4/69%			A	58	51	
					LC-MS/MS	1	YATAEADNIK	10	A	54	21	2
Rv2445c	14.5	17	5.34	5.26	MS	6/53%			A	65	51	
Rv2450c	14.78/14.78		4.12/4.12		LC-MS/MS	1	VAENVLR	10	A	17	14	2
Rv2465c	17.27	15.5	6.14	6.2	MS	6/40%			A	70	51	
Rv2469c	21.94/20.99		9.86/9.73		LC-MS/MS	1	ADTVDHVVPR	7	C	16	12	2
Rv2544	20.74[b]		5.25[b]		LC-MS/MS	1	FNDDSYGQDFYRNGSLCK	3	A	13	11	3
Rv2563	31.11/31.41		8.39/6.80		LC-MS/MS	1	TVDSMGVDAFVVK	1	C	16	14	2
Rv2575	30.8		4.93		LC-MS/MS	1	IQQQTTGR	9	A	33	24	2
Rv2576c	11.56/11.56		4.93/4.93		LC-MS/MS	3	YMITLHTPIAGGQPLVYTATLANPSQWAIVTASGGLR	10	A	44	19	3
					LC-MS/MS		DPVGHQVTYTVTTTSDLMANIR	10	A,C	46	10	3
					LC-MS/MS		YMSADPPSMAAFNADSSK	10	C	35	21	2
Rv2585c	55.43[b]		4.96[b]		LC-MS/MS	1	ALALCVPR	1	C	13	6	4
Rv2631	45.52	50	6.99	4.95	MS	11/21%			A	53	51	
Rv2668	16.22/15.31	25.5	4.90/4.87	4.67	MS	6/39%			A	66	51	
					LC-MS/MS	5	GDAPIGHIGDTLR	10	A	40	13	3
					LC-MS/MS		DPVSVVVLLDEK	10	A	63	25	2
					LC-MS/MS		GGVYWDAYRDPVSVVVLLDEK	10	A	40	22	3
					LC-MS/MS		VDTGTYVADVTVSSVVPVDPPPGFGYTR	10	A	67	20	3
					LC-MS/MS		SFPDSSVTR	10	A	41	23	2
Rv2672	51.15[b]	55	4.82[b]	4.81	MS	19/54%			A	169	51	
					LC-MS/MS		SGDMNLLSALINR	6	A	30	23	2
Rv2693c	17.10/20.10		9.62/10.01		LC-MS/MS	1	AGLLPAIGFALSMAGLILLWRLLR	10	A	13	12	3
Rv2721c	68.94/68.94	35	4.42/4.42	4.68	MS	15/28%			A	108	51	
					LC-MS/MS	2	AAGGAAGPLGAK	6	A	40	22	2
					LC-MS/MS		EFTTVPAVLAEQLK	6	A	21	13	2
Rv2766c	22.73/24.19		5.14/5.42		LC-MS/MS	1	VNAICPGVVRTR	2	A	17	14	2
Rv2799	19.89/17.11	23	5.24/5.22	5.4	MS	7/33%			A	67	51	
					LC-MS/MS	4	ANDLVPYYR	10	C	30	21	2
					LC-MS/MS		DIPFDVIQR	10	A,C	20	12	2

					LC-MS/MS		LGLAYTPPEAEEGLR	10	A,C	81	22	2
					LC-MS/MS		TYAQTLPPDAIETTIAGHR	10	A,C	14	12	3
					LC-MS/MS		AAQYWVR	10	A	22	13	2
Rv2823c	90.74		6.7		LC-MS/MS	1	WVYFLTR	4	A,C	33	18	2
Rv2831	26.63	30	5.82	5.75	MS	12/53%			A	157	51	
					LC-MS/MS		TSGDTIAANR	9	A	33	23	2
Rv2848c	44.05/45.55		5.96/6.11		LC-MS/MS	1	FVAHAACNTPRA	4	C	14	13	2
Rv2873	19.80[b]	17	4.67[b]	4.38	MS/MS	2	IDGTHQTLQGADLTVIGAR		A	98	51	1
					LC-MS/MS		LPAATIDQLK	10	A	25	23	2
Rv2874	71.16/71.16		9.64/9.64		LC-MS/MS	1	CGYHSHLTGGEFDVNR	6	A	5	3	4
Rv2875	16.31/16.31	16.7	4.31/4.31	4.36	MS/MS	5	QTLQGASVTVTGQGNSLK		A	57	51	
					LC-MS/MS		QTLQGASVTVTGQGNSLK	8	A	26	23	2
					LC-MS/MS		LPASTIDELKTNSSLLTSILTYHVVAGQTSPANVVGTR	10	A	22	20	4
					LC-MS/MS		LPASTIDELK	10	A	22	20	2
					LC-MS/MS		VGNADVVCGGVSTANATVYMIDSVLMPPA	10	A	24	20	3
					LC-MS/MS		TNSSLLTSILTYHVVAGQTSPANVVGTR	7	A	44	10	3
Rv2878c	14.62/14.62	12.5	4.57/4.57	4.2	MS	5/38%			A	55	51	
					LC-MS/MS	4	ADGTSTFVNNPTAAMSQDELSGR	10	A,B,C	48	22	3
					LC-MS/MS		ADVGAMQSFVSK	10	A,B,C	70	22	2
					LC-MS/MS		YNLNFTNLNDADGVIWAR	10	A,B,C	43	11	2
					LC-MS/MS		YNVPWQPAFVFYR	9	A,B,C	54	13	2
Rv2905	31.18[b]		4.96[b]		LC-MS/MS	1	YGLRSTAPPSDGR	7	C	14	11	3
Rv2911	28.38[b] 27.62/27.30	22	4.75[b] 4.75/4.75	4.7	MS	9/40%			A	54	51	
					LC-MS/MS	6	AATLGATSTHATTPSGLDGPGGSGASTAHDLVVIFR	7	B	20	17	4
					LC-MS/MS		DQNVAHPPASTIK	6	A	54	22	2
					LC-MS/MS		KTFVGAAAR	8	A	36	24	2
					LC-MS/MS		VLLALVALDELDLNSTVVADVADTQAECNCVGVKPGR	1	A	65	19	4
					LC-MS/MS		YPGAIGGK	8	A	51	23	2
					LC-MS/MS		QLLDGLLLVSGNDAANTLAHMLGGQDVTVAK	1	A	60	11	4

Rv2945c	22.36[b]	13	4.89[b]	5.3	MS	7/48%			A	62	51		
Rv2994	43.49/42.58		11.22/11.23		LC-MS/MS	1	CHAWPNGPR	7	A	18	12	2	
Rv3004	7.60/8.70		10.61/11.32		LC-MS/MS	2	LRTLADER	9	A	14	14	2	
					LC-MS/MS		WDDIDGLRFHR	8	A	15	12	2	
Rv3006	36.41[b]	33	4.71[b]	4.61	MS	9/36%			A	75	51		
					LC-MS/MS	1	LAPSTGAVTGEPDVVR	7	A	60	23	2	
Rv3016	20.23[b]	22	4.56[b]	4.38	MS/MS	1	TTFQDRPDGSLISEAAAAYR			61	51	1	
Rv3031	57.79		7.2		LC-MS/MS	1	NRLLSESER	3	A	29	22	2	
Rv3033	15.32/15.32		4.73/4.73		LC-MS/MS	1	VGARPDSVTCPDNLKGVEGAK	7	A	35	10	3	
Rv3036c	22.08/20.94		4.91/4.91		LC-MS/MS	4	DGFVNVAQGSPLR	7	A,B,C	78	19	2	
					LC-MS/MS		DQPYQMDATSEQHSSGQPPQATR	10	A,C	60	21	3	
					LC-MS/MS		FFQDLGGAHPSTWYK	1	A,B	42	16	2	
					LC-MS/MS		AFNYNLATSQPITFDTLFVPGTTPLDSIYPIVQR	10	A	36	19	3	
Rv3044	35.20[b]	27	5.09[b]	4.97	MS	15/32%			A	130	51		
					LC-MS/MS	3	AAAAADPGPPTRPAHNAAGVSPEMVQVPAEAQR	3	A	33	6	4	
					LC-MS/MS		IAAVDALITGFAEHATQVGTK	4	A,C	68	13	3	
					LC-MS/MS		AYIEIGTTAADLAK	3	A	34	12	2	
Rv3106	46.76/46.76		5.37/5.37		LC-MS/MS	1	AAGEPHGRPR	4	A	13	11	2	
Rv3111	17.84	32	8.76	5.26	MS	6/29%			A	52	51		
Rv3158	48.68/52.08		7.97/9.25		LC-MS/MS	1	LLSQEAAMK	4	C	14	11	2	
Rv3193c	103.24/103.24		8.53/8.53		LC-MS/MS	2	LDEAITK	2	A	17	12	2	
					LC-MS/MS		AVVLARLR	4	A	17	12	2	
Rv3194c	32.14/32.03		5.05/5.05		LC-MS/MS	1	VGQIGGITHK	3	A	13	6	2	
Rv3201c	112.33/112.33		5.91/5.91		LC-MS/MS	2	LAWAALR	4	A	16	14	2	
					LC-MS/MS		AEADGVKPPTAAVLVR	3	A	19	10	3	
Rv3240c	105.98		5.09		LC-MS/MS	1	FLGLQVGVILATMTPDERR	1	A	21	16	3	
Rv3244c	59.58[b]	57	5.12[b]	5.9	MS	11/24%			A	82	51		
					LC-MS/MS	3	IPVDSTAVASR	5	A	44	24	2	
					LC-MS/MS		MPEQTAAAVSR	5	A	51	23	2	
					LC-MS/MS		NTLYFADPTGK	9	A,C	17	13	2	

Rv3248c	54.32		4.85		LC-MS/MS	1	IADLSLADFGR	6	A	32	23	2
Rv3267	48.06/47.79	50	4.87/4.87	4.82	MS	10/30%			A	59	51	
					LC-MS/MS	1	ADDLGAQQVAK	10	A	20	14	2
Rv3310	28.03/27.14	25	5.40/5.40	5.55	MS/MS	2	TNNSLLVVTWDEDDGSSR		A	90	51	1
					LC-MS/MS		SQAAIIGNK	9	A	27	24	2
Rv3354	9.70/9.70		3.72/3.72		LC-MS/MS	2	ALGQSVCPILAEPGGSFNTAVASVVAR	10	A	28	13	3
					LC-MS/MS		NPVDDAFIAALNNAGVNYGDPVDAK	10	A	76	13	3
Rv3356c	29.48	32	5.97	6.2	MS	8/32%			A	75	51	
Rv3369	15.7		5.69		LC-MS/MS	2	FGLTEAIAAYSTR	10	B	69	19	2
					LC-MS/MS		LTSDLYGWLTTVAR	9	B	47	18	2
Rv3397c	33.14		7.37		LC-MS/MS	1	MEIDWTGCRDFDELIVYCR	7	A	8	6	4
Rv3402c	-/42.90		-/6.45		LC-MS/MS	1	TGMADAGVR	10	A	19	10	2
Rv3413c	31.24		4.52		LC-MS/MS	2	QDLINEVNLLNTK	7	A	40	22	2
					LC-MS/MS		VEQMIAQGQWAEAQDELAEVSSTVQAVTDGSR	10	A	58	41	3
Rv3418c	10.7		4.62		LC-MS/MS	5	IPLDVAEGDTVIYSK	1	B	19	19	2
					LC-MS/MS		RIPLDVAEGDTVIYSK	9	B	32	19	3
					LC-MS/MS		VNIKPLEDKILVQANEAETTTASGLVIPDTAK	9	A	69	16	4
					LC-MS/MS		YGGTEIKYNGEEYLILSAR	9	A	36	20	3
					LC-MS/MS		YNGEEYLILSAR	10	A	48	19	2
Rv3428	45.46		9.19		LC-MS/MS	1	ERVTVPR	3	A	19	19	2
Rv3484	49.50/48.03	48	4.72/4.68	4.61	MS	9/23%			A	78	51	
Rv3485c	31/31		5.98/5.98		LC-MS/MS		VCTPLPR	8	C	20	19	2
Rv3491	17.40/17.40	17	5.47/5.47	5.5	MS/MS		LGPPPDGSYSFNQAGVSGVTWTITALCDQPSGTR		A	46	43	3
					LC-MS/MS	4	NMNDYSDPIVWAFNCALNVVSTTPQQITR	10	A	88	19	3
					LC-MS/MS		LQNFSGR	10	A	23	22	2
					LC-MS/MS		QPFSLQLIGPPPSPVQR	10	A,C	60	12	2
Rv3495c	39.14[b]		4.80[b]		LC-MS/MS	1	ALDTLPDAVR	10	A	12	11	2
Rv3509c	52.07	64	4.76	4.7	MS	16/40%			A	116	51	
Rv3572	16.29/16.22	13	4.75/4.75	4.55	MS	4/30%			A	55	51	
					LC-MS/MS	3	AGGSDVITTVYFGEGPPDK	10	A,C	9	7	2

					LC-MS/MS		LAYLDAHATSQFER	8	A,C	24	22	3
					LC-MS/MS		TPDGPTGFPPGLWAR	10	A,C	58	22	2
Rv3584	15.76[b]		8.11[b]		LC-MS/MS	4	IQAVQTSDFIQPGK	7	A	23	13	2
					LC-MS/MS		LTINNVLLR	7	A	67	7	2
					LC-MS/MS		AVDLVLVAVNQSPDVSDR	7	A	58	12	2
					LC-MS/MS		LPASGMLFVGTPDGQIVAPGPLPSNQAAK	7	A	24	10	3
Rv3587c	22.18/22.18	32	5.79/5.79	4.91	MS	7/32%			A	82	51	
					LC-MS/MS	5	DVGAAVLAAYVYSLDNK	6	A,B	21	13	2
					LC-MS/MS		DVGAAVLAAYVYSLDNKR	5	A,B,C	98	22	3
					LC-MS/MS		EGDDCPDSTLAVK	7	A	35	22	2
					LC-MS/MS		GLTNAPQYYVGDQPK	10	A,B,C	25	11	2
					LC-MS/MS		TFSPGEQVTTAVTWTGMGS	10	A	52	12	2
Rv3627c	43.89/41.33		5.71/4.96		LC-MS/MS	1	SRTPALDAGR	10	A	12	11	2
Rv3629c	34.21/36.57		9.41/8.80		LC-MS/MS	1	WIEVPFAR	7	A	12	11	2
Rv3668c	20.99[b] 20.18/20.18		4.63[b] 4.63/4.63		LC-MS/MS	4	VTPVAVFNGFAINGIGPDPSFGQIACK	10	A	18	10	3
					LC-MS/MS		DDKLPLGGGAGIVVNGDTMCTLTTIGHDK	10	A	22	11	4
					LC-MS/MS		YIPLHTPAVVMSINADLADINAK	6	A	72	10	3
					LC-MS/MS		NRPGAGFVPVPA	6	A	28	12	2
Rv3671c	36.45/-	25	6.38/-	4.45	MS	5/21%			A	52	51	
Rv3682	81.94/80.38	50	5.26/5.20	5.1	MS	12/20%			A	60	51	
					LC-MS/MS	1	LKDAGFQVADQTNSVNSSAK	10	A	25	21	3
Rv3693	42.92/42.92		11.30/11.30		LC-MS/MS	1	RVVIVLDTGR	9	C	14	10	2
Rv3705c	19.99/19.99		4.91/4.91		LC-MS/MS	3	QAVGVFASNDAADR	7	A,C	31	22	2
					LC-MS/MS		HPSEPGVVSYAVLGK	10	A,C	41	12	2
					LC-MS/MS		CFVQTR	10	A	27	21	2
Rv3710	70.11		5.2		LC-MS/MS	1	GGVAYIK		A	23	22	2
Rv3722c	47.34	40	5.63	5.65	MS	9/26%			A	69	51	
					LC-MS/MS		HQQILAPK	5	A	29	23	2
Rv3725	-/30.11		-/9.53		LC-MS/MS	1	CPIPR	7	A	31	23	2

Rv3759c	30.46[b] 30.10/30.10		5.20[b] 5.20/5.20		LC-MS/MS	1	DPLGSATGSVK	8	A	12	11	2
Rv3760	5.56/5.56		12.60/12.60		LC-MS/MS	1	KTHAAALR	4	A	15	11	2
Rv3803c	28.48/27.81	25	5.52/5.51	5.5	MS	9/41%			A	58	51	
					LC-MS/MS	9	DIPVAFLAGGPHAVYLLDAFNAGPDVSNW- VTAGNAMNTLAGK	9	A	48	6	4
					LC-MS/MS		GISVVAPAGGAYSMYTNWEQDGSK	10	A	43	23	2
					LC-MS/MS		GLAPGGHAAVGAAQGGYGAMALAAFHPDR	9	A,B,C	21	11	4
					LC-MS/MS		MFYNQYR	9	A,C	36	22	2
					LC-MS/MS		QWDTFLSAELPDWLAANR	6	A,B,C	58	21	2
					LC-MS/MS		SVGGHNGHFDFPASGDNGWGSWAPQLGAMSGDIVGAIR	9	A,C	39	19	4
					LC-MS/MS		VWVWSPTNPGASDPAAMIGQAAEAMGNSR	9	A,B,C	40	21	3
					LC-MS/MS		APYENLMVPSPSMGR	10	A,C	34	3	2
					LC-MS/MS		WHDPWVHASLLAQNNTR	6	A,C	38	10	3
Rv3804c	32.34/31.65	30	5.32/5.32	5.6	MS	7/21%			A	58	43	
					LC-MS/MS	11	ALGATPNTGPAPQGA	8	A,C	20	12	2
					LC-MS/MS		ASDMWGPK	8	A	48	24	2
					LC-MS/MS		EDPAWQR	4	A	26	23	2
					LC-MS/MS		FLEGFVR	6	A,B,C	25	12	2
					LC-MS/MS		LIANNTR	10	A	30	24	2
					LC-MS/MS		NDPLLNVGK	10	A,C	33	22	2
					LC-MS/MS		VQFQSGGANSPALYLLDGLR	6	A,B,C	36	23	3
					LC-MS/MS		VWVYCGNGKPSDLGGNNLPAK	5	A,B,C	30	21	3
					LC-MS/MS		WETFLTSELPGWLQANR	6	A,B,C	27	21	2
					LC-MS/MS		AGCQTYKWETFLTSELPGWLQANR	8	A,B,C	46	16	3
					LC-MS/MS		ASDMWGPKEDPAWQR	6	A,B	26	18	3
Rv3835	40.76/-	34	5.06/-	4.91	MS	10/23%			A	70	51	
Rv3841	20.44	20	4.73	4.81	MS/MS	1	AGANLFELENFVAR		A	88	51	1
Rv3846	23.3		5.96		LC-MS/MS	3	YAAATSQTK	10	A	59	22	2
					LC-MS/MS		AKEDHSAILLNEK	10	A	42	21	3

					LC-MS/MS		NLSPNGGDKPTGELAAAIADAFGSFDKFR	6	A	49	20	4
Rv3849	14.71	9	8.82	5.7	MS	5/45%			A	52	51	
Rv3872	9.27		3.95		LC-MS/MS	1	AGEAVQDVAR	7	A	50	45	2
Rv3874	10.79	10	4.59	4.5	MS	6/67%			A	55	51	
					LC-MS/MS		ADEEQQQALSSQMGF	10	A	56	22	2
					LC-MS/MS		AEMKTDAATLAQEAGNFER	9	A,C	47	25	2
					LC-MS/MS		FQEAANK	10	A	27	22	2
					LC-MS/MS		GAAGTAAQAAVVR	10	A	42	23	2
					LC-MS/MS		ISGDLK	10	A	24	24	2
					LC-MS/MS		QAGVQYSR	10	A	42	23	2
					LC-MS/MS		QELDEISTNIR	10	A	49	42	2
					LC-MS/MS		QKQELDEISTNIR	9	A,C	49	42	2
					LC-MS/MS		TDAATLAQEAGNFER	10	A,B,C	47	25	2
					LC-MS/MS		TQIDQVESTAGSLQGQWR	9	A,B	61	41	3
					MS/MS		TQIDQVESTAGSLQGQWR		A	114	51	1
Rv3875	9.9	9	4.48	4.25	MS/MS	2	WDATATELNNALQNLAR		A	75	51	1
					LC-MS/MS		WDATATELNNALQNLAR	10	A,B	38	26	2
					LC-MS/MS		LAAAWGGSGSEAYQGVQQK	10	A	67	41	2
Rv3881c	47.59	60	4.75	4.55	MS	10/39%			A	124	76	
					LC-MS/MS	10	ANEVEAPMADPPTDVPITPCELTAAK	5	A	22	20	3
					LC-MS/MS		DQILPVYAEYQQR	6	A,C	63	21	2
					LC-MS/MS		EHPTYEDIVGLER	5	A,C	26	22	2
					LC-MS/MS		EYLAAGAK	7	A	43	23	2
					LC-MS/MS		EAAALSGDVAVK	3	A	46	22	2
					LC-MS/MS		LYAENPSAR	6	A	39	23	2
					LC-MS/MS		TQSQTVTVDQQEILNR	5	A	97	22	2
					LC-MS/MS		VATAGEPNFMDLK	6	A,C	30	21	2
					LC-MS/MS		VLTEYNNK	6	A	32	22	2
					LC-MS/MS		AALEPVNPPKPPPAIK	4	A,C	25	23	3
Rv3899c	-/36.08		-/5.72		LC-MS/MS	1	RDSLPR	2	C	16	13	1

Rv3914	12.54	12	5.6	4.9	MS	4/39%			A	59	51		
					LC-MS/MS	5	ATDLTVAKLDVDTNPETAR	9	B	77	17	3	
					LC-MS/MS		MVAPVLEEIATER	9	A,B	61	18	2	
					LC-MS/MS		NFQVVSIPTLILFK	10	A,B,C	38	24	2	
					LC-MS/MS		ATDLTVAK	9	B	29	24	2	
					LC-MS/MS		NFQVVSIPTLILFKDGQPVK	10	B	48	19	3	
Rv3917c	34.63/34.46		5.96/6.12		LC-MS/MS	1	SLAGSQTGVRYQIVMGER	3	A	15	13	3	
MT0066,1 [a]	9.02/9.02	13.7	4.20/4.20	4	MS/MS	1	WGFGDLAVCDGEK		A	58	51	1	
MT2420 [a]	10.00	10	4.97	4.6	MS/MS	1	AQAAALEAEHQAIVR		A	131	51	1	
MT3437,1 [a]	17.35	9	11.07	5.18	MS	7/41%			A	52	51		

[a]Gene number as annotated for *M. tuberculosis* H37Rv (Rv) or *M. tuberculosis* CDC1551 (MT). Three proteins were only identified with the annoted genes for *M. tuberculosis* CDC1551. The nucleotide sequence of MT0066.1 is identical in *M. tuberculosis* H37Rv and found in positions 65012..65392 on the minus (-) strand overlapping with Rv0061 which is annotated as a conserved hypothetical protein (questionable ORF). The sequence of the peptide identified to belong to MT2420 is not encoded in the *M. tuberculosis* H37Rv genome. The peptide is highly homologous to a peptide found in Rv1793, the only difference being an A in MT2420 instead of an S. The peptide may therefore be derived from Rv1793 being modified posttranslationally by removal of a hydroxyl group to change S to A. The nucleotide sequence of MT3437.1 is identical in *M. tuberculosis* H37Rv and found in positions 3720757..3721236 on the plus (+) strand overlapping slightly with Rv3333c (position 3720782), extending in the gap between this gene and Rv3334 which starts at position 3721257.

[b]Theoretical molecular mass (kDa) or isoelectric point (p*I*) calculated after removing the predicted signal peptide as determined by prediction of lipoprotein consensus motif using Compute pI/MW tool, publicly available at (http://au.expasy.org/tools/pi_tool.html). See also supplementary table 2.

[c]Theoretical molecular mass (kDa) calculated after removing the predicted signal peptide by Neural Network/Hidden Markov model, using Compute pI/MW tool, publicly available at (http://au.expasy.org/tools/pi_tool.html). Only one figure is given when the predictions were concordant. If one of the methods did not predict a signal peptide it is designated with a "-". See also supplementary table 2.

[d]Observed molecular masses as determined by 2D-PAGE.

[e]Theoretical isoelectric point (p*I*) calculated after removing the predicted signal peptide by Neural Network/Hidden Markov model, Compute pI/MW tool, publicly available at (http://au.expasy.org/tools/pi_tool.html). If one of the methods did not predict a signal peptide it is designated with a "-". See also supplementary table 2.

[f]Observed p*I* as determined by 2D-PAGE.

[g]Peptides were identified by Matrix-assisted laser desorption ionization time-of-flight mass spectrometry of spots collected from 2D-PAGE (MS), Matrix-assisted laser desorption ionization time-off-flight mass spectrometry combined with tandem mass spectrometry (MS/MS) or Liquid chromatography Electrospray Ionisation with tandem mass spectrometry (LC- MS/MS)

[h]None of the identified peptides by MS or MS/MS were predicted signal peptides,

[i]Sequence coverage for individual proteins is given as number of peptides/percentage of sequence coverage for MS data, or as number of identified peptides for MS/MS and LC-MS/MS data.

[j] Fraction number of identified peptide after SDS-PAGE performed as shown in Figure 2. Explanation of the fraction numbers: (1) >160 kDa, (2) ranges from 105-160 kDa, (3) ranges from 75-105 kDa, (4)) ranges from 50-75 kDa, (5) ranges from 35-50 kDa, (6) ranges from 30-35 kDa, (7) ranges from 25-30 kDa, (8) ranges from 15-25 kDa, (9) ranges from 15-10 kDa, (10) <10 kDa. Many peptides were observed several times and also in more than one fraction. In such cases the fraction with most observations or most reliable observation is given.

[k] Batch number refers to the three different batches (A,B or C) of 3 - 4 week old *M. tuberculosis* H37Rv culture filtrates analysed.

* Two peptides identified by LC-MS/MS matched all of these proteins: EsxI (Rv1037c), EsxL (1198), EsxN (Rv1793), EsxO (Rv2346c) and EsxV (Rv3619c),

** One protein spot (Fig, 1, ¤) fingerprint matched all of these proteins: EsxJ (Rv1038c), EsxK (Rv1197), EsxP (Rv2347c) and EsxW (Rv3620c)

*** One peptide identified by LC-MS/MS matched both EsxK (Rv1197) and EsxP (Rv2347c).

Supplementary table 2. List of predicted secreted and exported proteins of *M. tuberculosis* H37Rv culture filtrate proteins identified in this study, with their possible retention peptides.

Gene number	Signal peptide[a]	Signal peptidase	Retention signal	Predicted localization of protein
Rv0012	MRLTHPTPCPENGETMIDRRRSAWRFSVPLVCLLAGLLLAATHGVSG$^{\downarrow}$GTE	SPase I[c]	-	Secreted protein
Rv0019c	MQGLVLQLTRAGFLMLLWVFIWSVLRILKTDIYAPTGAVMMRRGLALRGTLLGARQRRHA**ARY	SPase I	-	Secreted protein
Rv0040c	MIQIARTWRVFAGGMATGFIGVVLVTAGKASA**DPLLPPPPIPAPVSAPA*TVP	SPase I	-	Secreted protein
Rv0063	MAREISRQTFLRGAAGALAAGAVFGSVRATA$^{\downarrow}$DPA[f]	SPase I	-	Secreted protein
Rv0116c	MRRVVRYLSVVVAITLMLTAESVSIATA**AVPPLQPIPGVASVSPANG*AVV	SPase I	-	Secreted protein
Rv0125	MSNSRRRSLRWSWLLSVLAAVGLGLATAPAQA$^{\downarrow}$APP	SPase I	-	Secreted protein
Rv0129c	MTFFEQVRRLRSAATTLPRRLAIAAMGAVLVYGLVGTFGGPATAGA$^{\downarrow}$FSR	SPase I	-	Secreted protein
Rv0164	MTAISCSPRPRYASRMPVLSKTVEVTADA**ASI	SPase I	-	Secreted protein
Rv0170	MKITGTVVKLGIVSVVLLFFTVMIIVIFG*QMR**FDR	SPase I	-	Secreted protein
Rv0172	MSTIFDIRNLRLPQLSRASVVIGSLVVVLALA**AG*IVG	SPase I	-	Secreted protein
Rv0174	MLTRFIRRQLILFAIVSVVA**IVVLG*WYY	SPase I	-	Secreted protein
Rv0203	MKTGTATTRRRLLAVLIALALPGAAVALLAEPSATG*A**SDP	SPase I	-	Secreted protein
Rv0285	MTLRVVPEGLAAASAAVEALTARLAAAHASA$^{\downarrow}$APV	SPase I	-	Secreted protein
Rv0309	MSRLLALLCAAVCTGCVAVVLAPVSLAVVNPWFA**NSVGN*ATQ	SPase I	-	Secreted protein
Rv0315	MLMPEMDRRRMMMMAGFGALAAALPAPTAWA**DPSRPAAPAGPTPAPAAPA*AAT	SPase I	-	Secreted protein
Rv0398c	MGVIARVVGVAACGLSLAVLAAAPTAGA$^{\downarrow}$EPT	SPase I	-	Secreted protein
Rv0453	MTSALIWMASPPEVHSALLSSGPGPGPVLA*AAT	SPase I	-	Secreted protein
Rv0455c	MSRLSSILRAGAAFLVLGIAAATFPQSAAA$^{\downarrow}$DST	SPase I	-	Secreted protein
Rv0477	MKALVAVSAVAVVALLGVSSAQA$^{\downarrow}$DPE	SPase I	-	Secreted protein
Rv0490	MTVFSALLLAGVLSALALAVGG$^{\downarrow}$AVG	SPase I	-	Secreted protein
Rv0506	MRMISVSGAVKRMWLLLAIVVVA**VVGGLGIYR*LHS	SPase I	-	Secreted protein
Rv0559c	MKGTKLAVVVGMTVAAVSLAAPAQA$^{\downarrow}$DDY	SPase I	-	Secreted protein
Rv0674	MPAMTARSVVLSVLLGAHPAWA$^{\downarrow}$TAS	SPase I	-	Secreted protein
Rv0677c	MIGTLKRAWIPLLILVVVAIA**GFTVQRIRTFFGSEGILVTPKVFA*DDP	SPase I	-	Secreted protein

Rv0774c	MPIRPNVHGMMARMPELSRRAVLGLGAGTVLGATSAYAIDM*LLQPRTSHAAPAAAIGTNVPLAPTPA**LDP	SPase I	-	Secreted protein
Rv0787	MHRPPWLAQLRRRLRIGVQL**GSR	SPase I	-	Secreted protein
Rv0851	MDGFPGRGAVITGGASGIGLATGTEFA*RRG	SPase I	-	Secreted protein
Rv0867c	MSGRHRKPTTSNVSVAKIAFTGAVLGGGGIAMA**AQATA*ATD	SPase I	-	Secreted protein
Rv0982	MWWFRRRDRAPLRATSSLSLRWRVMLLAMSMVAMVVVLMSFAVYAVISA↓ALY	SPase I	-	Secreted protein
Rv1050	MARQRFRDQVVLITGASSGIGEATAKAFA*REG	SPase I	-	Secreted protein
Rv1075c	MPRRSTIALATAGALASTGTA**YLGARNLLVGQATHA*RTV	SPase I	-	Secreted protein
Rv1158c	MPTIWTFVRAASSAALLTGGIAHA**DPAPAPAPAPNIPQQLISSAANAPQILQNLATALG*ATP	SPase I	-	Secreted protein
Rv1174c	MRLSLTALSAGVGAVAMSLTVGAGVASA↓DPV	SPase I	-	Secreted protein
Rv1269c	MTTMITLRRRFAVAVAGVATAAATTVTLAPAPANA↓ADV	SPase I	-	Secreted protein
Rv1352	MARTLALRASAGLVAGMAMAAITLAPGARA↓ETG	SPase I	-	Secreted protein
Rv1382	MNSGTLAGSLIFAAVLVMLIAVLA*RLMMRGWRRR**SER	SPase I	-	Secreted protein
Rv1386	MTLRVVPESLAGASAAIEAVTARLAAAHA*AAPFIA**AVI	SPase I	-	Secreted protein
Rv1419	MGELRLVGGVLRVLVVVGAVFDVAVLNAGAASA↓DGP	SPase I	-	Secreted protein
Rv1435c	MTLMAIVNRFNIKVIAGAGLFAAAIALSPDAAA↓DPL	SPase I	-	Secreted protein
Rv1477	MRRNRRGSPARPAARFVRPAIPSALSVALLVCTPGLATA↓DPQ	SPase I	-	Secreted protein
Rv1488	MQGAVAGLVFLAVLVIFAIIVVAKSVALIPQAEA↓AVI	SPase I	-	Secreted protein
Rv1759c	MSFVIAVPETIAAAATDLADLGSTIAGANAAAA*ANT	SPase I	-	Secreted protein
Rv1804c	MRVVSTLLSIPLMIGLAVPAHA↓GPS	SPase I	-	Secreted protein
Rv1810	MQLQRTMGQCRPMRMLVALLLSAATMIGLAAPGKA↓DPT	SPase I	-	Secreted protein
Rv1812c	MTRVVVIGSGFAGLWAALGA**A*RRL	SPase I	-	Secreted protein
Rv1815	MVRLVPRAFAATVALLAAGFSPATASA↓DPV	SPase I	-	Secreted protein
Rv1860	MHQVDPNLTRRKGRLAALAIAAMASASLVTVAVPATANA**DPEPAPPVPTTAASPPSTAAA*PPA	SPase I	-	Secreted protein
Rv1869c	MASSTTFVIVGGGLAGA*KAVEA**LRR	SPase I	-	Secreted protein
Rv1884c	MHPLPADHGRSRCNRHPISPLSLIGNASATSGDMSSMTRIAKPLIKSAMAAGLVTASMSLSTAVAHA**GPS	SPase I	-	Secreted protein
Rv1886c	MTDVSRKIRAWGRRLMIGTAAAVVLPGLVGLAGGAATAGA↓FSR	SPase I	-	Secreted protein
Rv1891	MIRELVTTAAITGAAIGGAPVAGA*DPQ	SPase I	-	Secreted protein
Rv1906c	MRLKPAPSPAAAFAVAGLILAGWAGSVGLAGA↓DPE	SPase I	-	Secreted protein
Rv1910c	MAHAFHRFALAILGLALPVALV**AYGGNGDSRKA*APL	SPase I	-	Secreted protein
Rv1926c	MKLTTMIKTAVAVVAMAAIATFAAPVALA↓AYP	SPase I	-	Secreted protein

Rv1980c	MRIKIFMLVTAVVLLCCSGVATA↓APK	SPase I	-	Secreted protein
Rv1984c	MTPRSLVRIVGVVVATTLALVSAPAGGRA**AHA*DPC	SPase I	-	Secreted protein
Rv2190c	MRLDQRWLIARVIMRSAIGFFASFTVSSGVLAANVLA↓DPA	SPase I	-	Secreted protein
Rv2201	MCGLLAFVAAPAGAAGPEGADA↓ASA	SPase I	-	Secreted protein
Rv2251	MRWRASSAPSISAPPIATGCCTPA*ASPPQTCCGAKTPVSRMRPTRCCCPAAPTGEDA**VAD	SPase I	-	Secreted protein
Rv2253	MSGHRKKAMLALAAASLAATLAPNAVA**A-*AEP	SPase I	-	Secreted protein
Rv2301	MNDLLTRRLLTMGAAAAMLAAVLLLTPITVPAGYPGAVAPATA**ACPDAEVVFAR*GRF	SPase I	-	Secreted protein
Rv2376c	MKMVKSIAAGLTAAAAIGAAAA**GVTSIMAGGPVVYQMQPVVFGAPLPLDPASA*PDV	SPase I	-	Secreted protein
Rv2450c	MKNARTTLIAAAIAGTLVTTSPAGIANA↓DDA	SPase I	-	Secreted protein
Rv2469c	MAHGKKRRGHRSSGVAAGVTGPASC**LHSVHSHR*LAS	SPase I	-	Secreted protein
Rv2576c	MPAGVGNASGSVLDMTSVRTVPSAVALVTFAGAALSGVIPAIARA↓DPV	SPase I	-	Secreted protein
Rv2668	MRHWLIVLATLLVAAAGVAAA**NDVPRAWA*GDA	SPase I	-	Secreted protein
Rv2766c	MTSLDLTGRTAIITGASRGIGLAIAQQLAA*AGAHVVLTARRQEA**ADE	SPase I	-	Secreted protein
Rv2799	MYTPGKGPPRAGGVVFTRVRLIGGLGALTA**AVVVVGTVGWQGIPPAPTGGDAVQLRSTA*APM	SPase I	-	Secreted protein
Rv2848c	MRVSAVAVAAPASGSG*KTTIATGLIGALRQA**GHT	SPase I	-	Secreted protein
Rv2875	MKVKNTIAATSFAAAGLAALAVAVSPPAAA↓GDL	SPase I	-	Secreted protein
Rv2878c	MSLRLVSPIKAFADGIVAVAIAVVLMFGLANTPRAVA↓ADE	SPase I	-	Secreted protein
Rv3004	MAHFAVGFLTLGLLVPVLTWPVSAPLLVIPVALS*ASIIRLRTLA**DER	SPase I	-	Secreted protein
Rv3033	MAHSIVRTLLASGAATALIAIPTACSFSIGTSHSHSVSKA↓EVA	SPase I	-	Secreted protein
Rv3036c	MRYLIATAVLVAVVLVGWPAAGA**PPSCAGLGGTVQA*GQI	SPase I	-	Secreted protein
Rv3106	MRPYYIAIVGSGPSAFFAAASLLKA↓ADT	SPase I	-	Secreted protein
Rv3158c	MILPAPHVEYFLLAPMLIVFSVAVAGVLAEA*FLPRRWRYGAQVTLALGGSAVALIAVIVVARS**IHG	SPase I	-	Secreted protein
Rv3194c	MNRRILTLMVALVPIVVFGVLLAVVTVPFVA**L*GPG	SPase I	-	Secreted protein
Rv3201	MTQTAAPARYSPAELACALGLFPPTAEQAAVIAAPPGPLVVIAGA↓GAG	SPase I	-	Secreted protein
Rv3267	MMSAQRVVRTVRTARAISTALAVAIVLGTGVA**WS*SVR	SPase I	-	Secreted protein
Rv3310	MLRGIQALSRPLTRVYRALAVIGVLAASLLASWVGA**VPQVGLAASA*LPT	SPase I	-	Secreted protein
Rv3354	MNLRRHQTLTLRLLAASAGILSAAAFAAPAQA↓NPV	SPase I	-	Secreted protein
Rv3402	MKIRTLSGSVLEPPSAVRA*TPG	SPase I	-	Secreted protein
Rv3484	MARSEGNRPRHRAVPQPSRIRKRLSRGVMTLVSVVALLMTGAGY**WVAHGALGGITISQA*LTP	SPase I	-	Secreted protein
Rv3485c	MNSRAPRNLAVSSPSAQVTGR**MVQ	SPase I	-	Secreted protein

Rv3491	MNIRCGLAAGAVICSAVALGIALHSGDPARA↓LGP	SPase I	-	Secreted protein
Rv3572	MTRLIPGCTLVGLMLTLLPAPTSA**A*GSN	SPase I	-	Secreted protein
Rv3587c	MLDLEPRGPLPTEIYWRRRGLALGIAVVVVGIAVAIVIAFVDSSAGA↓KPV	SPase I	-	Secreted protein
Rv3627c	MGPTRWRKSTHVVVGAAVLAFVAVVVAAA**ALVTTGGHRAGVRAPAPPPRPPTVKA*GVV	SPase I	-	Secreted protein
Rv3682	MPERLPAAITVLKLAGCCLLASVVATA**LTFPFAGGLGLMSNR*ASE	SPase I	-	Secreted protein
Rv3693	MILTGRTGLLALICVLPIALSPWPARAFVMLLVALAVAVTVDTLLA↓AST	SPase I	-	Secreted protein
Rv3705c	MRIAAAVVSIGLAVIAGFAVPVADA↓HPS	SPase I	-	Secreted protein
Rv3725	MQNATMRVLVTGGTGFVGGWTAKAIA*DAG	SPase I	-	Secreted protein
Rv3803	MKGRSALLRALWIAALSFGLGGVAVA**AEPTAKA*APY	SPase I	-	Secreted protein
Rv3804	MQLVDRVRGAVTGMSRRLVVGAVGAALVSGLVGA**VGGTATAGA*FSR	SPase I	-	Secreted protein
Rv3835	MLDAPEQDPVDPGDPASPPHGEAEQPLPGPRWPRALRASATRRALLLTALGGLLIAGLVTA**IPA	SPase I	-	Secreted protein
Rv3899c	MVTGQPAAAGAHSLSEGAMTAMQSGSVPPPQATPPITTPPVVSAPTMAAG*IEA	SPase I	-	Secreted protein
Rv3917c	MTQPSRRKGGLGRGLAALIPTGPA**DG*ESG	SPase I	-	Secreted protein
MT0066.1	MESAESIQRLTEFEMKLKFARLSTAILGCAAALVFPASVASA↓DPP	SPase I	-	Secreted protein
Rv0015c	MSPRVGVTLSGRYRLQRLIA**TGG	SPase I	1 TM[e]	Transmembrane protein
Rv0062	MTRRTGQRWRGTLPGRRPWTRPAPATCRRHLAFVELRHYFARVMSSAIGSVARWIVPLLGVAAVA**SIG	SPase I	1 TM	Transmembrane protein
Rv0064	METGSPGKRPVLPKRARLLVTAGMGMLALLLFGPRLV**DIYV*DWL	SPase I	6 TM	Transmembrane protein
Rv0219	MFDIATRFKNSYGSGPLHLLAMVSGFALLGYIVATA↓RPS	SPase I	4 TM	Transmembrane protein
Rv0291	MIRAAFACLAATVVVAGWWTPPAWA**IGPIGPPVVDAAA*QPP	SPase I	1 TM	Transmembrane protein
Rv0402c	MRSQRLAGHLSAAARTIHALSLPIILFWVALTIVVNVVA**PQLQSVAR*THS	SPase I	11 TM	Transmembrane protein
Rv0436c	MIGKPRGRRGVNLQILPSAMTVLSICAGLTAIKFA**LEHQPKAAMALIAAA*AIL	SPase I	4 TM	Transmembrane protein
Rv0446c	MVTSVSALAVAVVHSVAFA↓IGR	SPase I	4 TM	Transmembrane protein
Rv0563	MTWHPHANRLKTFLLLVGMSALIVAVGALFGRTALMLA**ALFA*VGM	SPase I	4TM	Transmembrane protein
Rv0680c	MKWNTVAASLAAGVITIAVALAAPPPAAHA↓KNG	SPase I	1 TM	Transmembrane protein
Rv0732	MLSAFISSLRTVDLRRKILFTLGIVILYRVGAA**LPSPGVNFPNVQQCIKEASA*GEA	SPase I	8 TM	Transmembrane protein
Rv0996	MPSIPQSLLWISLVVLWLFVLVPMLISKR*DA**VVR	SPase I	2 TM	Transmembrane protein
Rv1183	MVGCWVALALVLPMAVPSLAEMA**QR*HPV	SPase I	10 TM	Transmembrane protein
Rv1845c	MSALAFTILAVLLAGPTPALLA*RATWPLRAPRAAMVLWQAIALAAVLSSFSA**GIA	SPase I	4 TM	Transmembrane protein
Rv2060	MLTVVCLLVVTVLAICYRPLLFATVDPEVAAA*RGVPVRALGIVFAALMGVVAAQA**VQI	SPase I	3 TM	Transmembrane protein
Rv2200c	MTPRGPGRLQRLSQCRPQRGSGGPARGLRQLALAAMLGALAVTVSGCSWS*EA**LGI	SPase I	3TM	Transmembrane protein

Rv2563	MLFAALRDVQWRKRRLVIAIVSTGLVFAMTLVLTGLVNGFR*VEA**ERT	SPase I	4 TM	Transmembrane protein
Rv2693c	MNANRTSAQRLLAQAGGVSGLVYSSLPVVTFVVASSA*AGLLPAIGFALSMAGLILLWRLLRRESA**RPV	SPase I	5 TM	Transmembrane protein
Rv2721c	MNGQRGQLSTLIGRTLLGLAATAVTAVLLAPTVAA↓SPM	SPase I	1 TM	Transmembrane protein
Rv2874	MVESRRAAAAASAYASRCGIAPATSQRSLA↓TPP	SPase I	4 TM	Transmembrane protein
Rv2994	MSRDPTGVGARWAIMIVSLGVTASSFLFINGVAF**LIPRLENA*RGT	SPase I	7 TM	Transmembrane protein
Rv3193c	MGMRSAARMPKLTRRSRILIMIALGVIVLLLAGPRLIDA↓YVD	SPase I	6 TM	Transmembrane protein
Rv3629c	MSTFRIFGFSLLMTVVALVTGYLHG*GPTALFLLAVLALLEVSLSFDNA**IIA	SPase I	9 TM	Transmembrane protein
Rv3671c	MTPSQWLDIAVLAVAFIAAISGWRAGALGSMLSFGGVLLGATA**GVL	SPase I	3 TM	Transmembrane protein
Rv3760	MPGSVPGKAPEEPPVKFTRAAAVWSALIVGFLILILLLIFIAQNTASA↓QFA	SPase I	3 TM	Transmembrane protein
Rv0526	MQSRATRRSGALTMRRLVIAAAVSALLLTG↓↓CSGRDAVA*QGG**DAV	SPase I/II	Lipid	Secreted protein / Surface lipoprotein
Rv0999	MRPPLAPQFAADLLVKTVSTLRSSGAAGRLTTMRKAVLAVGSVCWLVG↓↓CSSGASS**TTA*STG	SPase I/II	Lipid	Secreted protein / Surface lipoprotein
Rv2911	MRKLMTATAALCA↓↓CAVTVSAGA**AWA*DAD	SPase I/II	Lipid	Secreted protein / Surface lipoprotein
Rv3668c	MQTAHRRFAAAFAAVLLAVV↓↓CLPANTAAA↓DDK	SPase I/II	Lipid	Secreted protein / Surface lipoprotein
Rv3759c	MRMLRRLRRATVAAAVWLATVCLVAS↓↓CANA↓DPL	SPase I/II	Lipid	Secreted protein / Surface lipoprotein
Rv0173	MMSVLARMRVMRHRAWQGLVLLVLALLLSS↓↓CGW	SPase II[d]	Lipid	Surface lipoprotein
Rv0237	MAFPRTLAILAAAAALVVA↓↓CSH	SPase II	Lipid	Surface lipoprotein
Rv0265c	MRQGCSRRGFLQVAEAAAATGLFAG↓↓CSS	SPase II	Lipid	Surface lipoprotein
Rv0411c	MTRRALLARAAAPLAPLALAMVLAS↓↓CGH	SPase II	Lipid	Surface lipoprotein
Rv0583c	MKHFTAAVATVALSLALAG↓↓CSF	SPase II	Lipid	Surface lipoprotein
Rv0835	MCCSTAAKSAVIVCCAAIATTA↓↓CSF	SPase II	Lipid	Surface lipoprotein
Rv0838	MRLIGRLRLLMVGLVVICGACA↓↓CDR	SPase II	Lipid	Surface lipoprotein
Rv0928	MKLNRFGAAVGVLAAGALVLSA↓↓CGN	SPase II	Lipid	Surface lipoprotein
Rv0932c	MKFARSGAAVSLLAAGTLVLTA↓↓CGG	SPase II	Lipid	Surface lipoprotein
Rv0934	MKIRLHTLLAVLTAAPLLLAAAG↓↓CGS	SPase II	Lipid	Surface lipoprotein
Rv1166	MGVPSPVRRVCVTVGALVALACMVLAG↓↓CTV	SPase II	Lipid	Surface lipoprotein
Rv1252c	MPGVWSPPCPTTPRVGVVAALVAATLTG↓↓CGS	SPase II	Lipid	Surface lipoprotein
Rv1270c	MKHPPCSVVAAATAILAVVLAIGG↓↓CST	SPase II	Lipid	Surface lipoprotein

Rv1541c	MRWIGVLVTALVLSA↓↓CAA	SPase II	Lipid	Surface lipoprotein
Rv1899c	MSRAAGLPRLSWFAGLTWFAGGSTGAG↓↓CAA	SPase II	Lipid	Surface lipoprotein
Rv1911c	MTSTLHRTPLATAGLALVVALGG↓↓CGG	SPase II	Lipid	Surface lipoprotein
Rv2068c	MRNRGFGRRELLVAMAMLVSVTG↓↓CAR	SPase II	Lipid	Surface lipoprotein
Rv2080	MPHSTADRRLRLTRQALLAAAVVPLLAG↓↓CAL	SPase II	Lipid	Surface lipoprotein
Rv2224c	MGMRLSRRDKIARMLLIWAALAAVALVLVG↓↓CIR	SPase II	Lipid	Surface lipoprotein
Rv2544	MIAPQPIPRTLPRWQRIVALTMIGISTALIGG↓↓CTM	SPase II	Lipid	Surface lipoprotein
Rv2585c	MAPRRRRHTRIAGLRVVGTATLVAATTLTA↓↓CSG	SPase II	Lipid	Surface lipoprotein
Rv2672	MATVVGMSRPMTSTAMLVALTCSATVLAA↓↓CVP	SPase II	Lipid	Surface lipoprotein
Rv2873	MINVQAKPAAAASLAAIAIAFLAG↓↓CSS	SPase II	Lipid	Surface lipoprotein
Rv2905	MRARPLTLLTALAAVTLVVVAG↓↓CEA	SPase II	Lipid	Surface lipoprotein
Rv2945c	MNDGKRAVTSAVLVVLGA↓↓CLA	SPase II	Lipid	Surface lipoprotein
Rv3006	MWTTRLVRSGLAALCAAVLVSSG↓↓CAR	SPase II	Lipid	Surface lipoprotein
Rv3016	MVGLTRPLLLCGATLLIAA↓↓CTR	SPase II	Lipid	Surface lipoprotein
Rv3044	MRSTVAVAVAAAVIAASSG↓↓CGS	SPase II	Lipid	Surface lipoprotein
Rv3244c	MRLTILLFLGAVLAG↓↓CAS	SPase II	Lipid	Surface lipoprotein
Rv3495c	MNRIWLRAIILTASSALLAG↓↓CQF	SPase II	Lipid	Surface lipoprotein
Rv3584	MNRCNIRLRLAGMTTWVASIALLAAALSG↓↓CGA	SPase II	Lipid	Surface lipoprotein

[a] Sec-type signal peptides were identified by using SignalP, publicly available at (http://www.cbs.dtu.dk/services/SignalP-3.0/).
↓ Signal peptidase I cleavage site concordantly predicted by Neural Network method and Hidden Markov model.
* Signal peptidase I cleavage site predicted only by Hidden Markov Model.
** Signal peptidase I cleavage site predicted only by Neural Network.
↓↓ Predicted Signal peptidase II cleavage sites. Lipoprotein signal peptides were identified by Lipo-tool, which is publicly available at (http://www.bioinfo.no/tools/lipo) or by using the ProSite PS00013 lipoprotein concensus motif.
[b] Identified export signals are Sec-type signal peptides (Sec), and lipoprotein signal peptides (Lipo).
[c] Signal peptidase I
[d] Signal peptidase II
[e] Transmembrane (TM) domains present in the mature part of the protein after processing by the Signal peptidase.
[f] The first 3 amino acids of verified mature sequences are underlined. See also table 2.