Published in final edited form as:

Proteomics. 2010 May ; 10(10): 1966–1984. doi:10.1002/pmic.200900836.

Descriptive proteomic analysis shows protein variability between closely related clinical isolates of *Mycobacterium tuberculosis*

Carolina Mehaffy¹, Ann Hess², Jessica E. Prenni³, Barun Mathema⁴, Barry Kreiswirth⁴, and Karen M. Dobos^{1,a}

¹Department of Microbiology, Immunology and Pathology, Colorado State University; Fort Collins, CO, 80523

²Center for bioinformatics and Department of Statistics, Colorado State University, Fort Collins, CO, 80523

³Proteomics and Metabolomics Facility, Colorado State University Fort Collins, CO, 80523

⁴Public Health Research Institute, Newark, NJ, 07103

Abstract

The use of isobaric tags such as iTRAQ allows the relative and absolute quantification of hundreds of proteins in a single experiment for up to eight different samples. More classical techniques such as 2-DE can offer a complimentary approach for the analysis of complex protein samples. In this study, the proteomes of secreted and cytosolic proteins of genetically closely related strains of *Mycobacterium tuberculosis* were analyzed. Analysis of 2-D gels afforded 28 spots with variations in protein abundance between strains. These were identified by tandem MS/MS. Meanwhile, a rigorous statistical analysis of iTRAQ data allowed the identification and quantification of 101 and 137 proteins in the secreted and cytosolic fractions respectively. Interestingly, several differences in protein levels were observed between the closely related strains BE, C28 and H6. Seven proteins related to cell wall and cell processes were more abundant in BE, while enzymes related to metabolic pathways (GltA2, SucC, Gnd1, Eno) presented lower levels in the BE strain. Proteins involved in iron and sulfur acquisition (BfrB, ViuB, TB15.3 and SseC2) were more abundant in C28 and H6. In general, iTRAQ afforded rapid identification of fine differences between protein levels such as those presented between closely related strains. This provides a platform from which the relevance of these differences can be assessed further using complimentary proteomic and biological modeling methods.

Keywords

Clinical isolates; iTRAQ; *Mycobacterium tuberculosis*

1. Introduction

Evaluation of protein abundance in both qualitative and quantitative terms offers important information on different aspects of cell physiology and biology with implications in

^aCorresponding author: Karen M. Dobos, Ph.D. Assistant Professor. Mycobacteria Research Laboratories, Department of Microbiology, Immunology and Pathology, 1682 Campus Delivery, Colorado State University, Fort Collins, CO, 80523-1682. Phone: 491-6549, Fax: 491-1815, Karen.Dobos@Colostate.Edu.

All of the authors declare no conflicts of interest.

medicine, infectious diseases and cell development. Through the years, the field of proteomics has made significant advances using a variety of techniques that allow not only the identification but also the quantification of specific proteins in a complex mix.

2-D gel electrophoresis (2-DE) has been the most commonly employed method to resolve and quantify (using image analysis software) individual proteins followed by identification by mass spectrometry. This technique allows for the separation and visualization of proteins in one single gel, which can be used to create maps representing the proteomes of whole cells or subcellular fractions.

In contrast to 2-DE, gel-free (a.k.a. “shotgun”) techniques allow for separation and identification of proteins in a complex mixture through the coupling of liquid chromatography to mass spectrometry. This shotgun approach, combined with the use of stable isotope labeling such as iTRAQ (isobaric tags for relative and absolute quantification), can be used to quantify hundreds of proteins for up to 4 [1] and more recently, 8 [2] different samples in a single mix. iTRAQ labeling has been successfully used for the analysis and identification of cancer biomarkers [3, 4], virulence factors and other molecular determinants, as well as host responses to pathogenic bacteria [5–7].

Mycobacterium tuberculosis (*Mtb*) is the causal agent of tuberculosis (TB) infection, which, in 2008, accounted for 1.8 million deaths [8]. In addition, the global incidence of TB is increasing at a rate of approximately 1% each year, with an estimation of more than 9 million new cases each year [8]. This makes TB the second leading cause of infectious disease-related death after HIV/AIDS [8].

Unfortunately, more than a century after the first description of tuberculosis infection, there is still a major need for more effective vaccines, diagnostic tests and drug treatments, crucial for the control of this disease [9].

Mycobacterium tuberculosis presents very limited genetic variation. However, there is a high degree of phenotypic variability among *Mtb* isolates, including differences in clinical outcome and epidemiological behavior. Both host and bacterial factors play an important role on this variability [10–12]. Proteomic analysis of *Mtb* strains may provide clues in relation to pathogenicity and prevalence of strains. Moreover, identification of differences in protein abundance between strains complements the development of vaccines, serodiagnostic tests, and choice of drug targets [9]. In this regard, the *Mtb* secreted proteome has been studied thoroughly by both 2-DE [13–22] and shotgun strategies [23]. A 2-DE database for *Mtb* is also publicly available (<http://www.mpiib-berlin.mpg.de/2D-PAGE/>) [24]. The shotgun approach when complemented with 2-DE provided the most complete proteome analysis of *Mtb* secreted proteins. In that study, 257 proteins were identified, increasing the annotation of exported proteins in *Mtb*. Cytosolic and cell wall proteomes [16, 25] as well as whole cell lysates [17, 26] and membranes [25, 27] have also been studied, although to a lesser extent. Similar to the studies performed on the secreted fraction, the most comprehensive analysis of the cytosol was obtained by a shotgun strategy in which 356 proteins were identified [25].

In the present study we used both 2-DE and iTRAQ approaches followed by tandem mass spectrometry (MS/MS) to identify and quantify secreted and cytosolic proteins from a group of closely related *Mtb* clinical isolates that appear to be very successful in causing disease [28]. These strains (i.e. BE, C28 and H6) comprise the S75 group which was identified based on the low copy number of the insertion fragment IS6110 in an epidemiologic study of clinical isolates from a TB cluster in New Jersey [28]. In addition, strain CDC1551, which is related to the S75 group but belongs to a different genetic subgroup was also included in the analysis.

2. Materials and Methods

2.1. *M. tuberculosis* strains and culture conditions

Glycerol stocks of clinical isolates BE, C28 and H6, as well as CDC1551 were plated in Middlebrook 7H11 (Difco), supplemented with OADC. After incubation for 2 weeks at 37°C colonies were inoculated in 100ml of Middlebrook 7H9 supplemented with OADC and 0.05% Tween. Cells were further cultured at 37°C in agitation for two weeks. Cells were washed twice with sterile PBS (Invitrogen) and inoculated in 1L of GAS media [29]. Cultures were then incubated at 37°C in agitation for 4weeks. All cultures were prepared in triplicate.

2.2. Culture Filtrate Proteins (CFP)

Each culture was carefully filtrated using a 0.2um zap-cap filter. Secreted proteins were recovered from the culture filtrate as described previously [30]. Briefly, 1L culture filtrates were concentrated to approximately 25ml of volume, using a 10KDa MWCO membrane (Millipore). The filtrate was further concentrated to approximately 300µl by centrifugation at 3000 rpm, 4°C using an Amicon Ultracell- 15 with a 10KDa molecular weight cut-off (MWCO). After this, proteins were subjected to buffer exchange three times using 10mM ammonium bicarbonate. Proteins were kept at 4°C until quantification by the bicinchoninic acid (BCA) assay (Pierce) and then at –20°C for longer storage.

2.3. Cytosol

Cells were pelleted by centrifugation at 3000rpm and washed twice with 10ml of sterile PBS (Invitrogen). Harvested cells were inactivated with 2.4 Mrad of cesium γ -irradiation for 24hr and death was confirmed by Alamar Blue assay (Invitrogen).

Cells were lyophilized and lipids removed by three sequential extractions with chloroform:methanol (2:1 vol:vol) followed by extraction with methanol:chloroform: water (10:10:3 vol:vol:vol) twice. Finally, cells were dried and resuspended in 10ml of breaking buffer (1mM EDTA-PBS supplemented with one tablet of protease inhibitor (Roche Diagnostics)) per 50ml of buffer and broken by sonication on 50% duty cycle (12 times, 60sec with intervals of 90sec on ice). After sonication, breaking buffer was added to a final volume of 40ml and unbroken cells were removed by centrifugation at 3000rpm for 5min, 4°C. Supernatant was further centrifuged for one hour at 27,000xg, 4°C to separate the cell wall (pellet) from cytosol and membrane fractions (supernatant). Finally, membrane proteins were harvested by ultracentrifugation at 100,000xg for 8 hours (2x 4h). The supernatant or cytosol fraction was concentrated by centrifugation using Amicon 15 tubes with a 10KDa MWCO. Buffer exchange using 10mM ammonium bicarbonate was performed three times. Finally, cytosol proteins were quantified and kept at 4°C until further analysis.

2.4. 2-DE

All reagents and accessories were obtained from Invitrogen, unless otherwise specified. 250 µg of each CFP were dried under vacuum and solubilized in 200µl of rehydration buffer (1% CHAPS, 1X NuPAGE® Sample reducing agent, 0.75% ZOOM® Ampholytes pH 4–7, 0.25% ZOOM® Ampholytes pH 3–10, 8M urea). Urea for buffer had been previously deionized with 5% AG-501-X8 resin (Bio-rad) for 2h at RT. Samples were solubilized for at least 8h at 4°C with intervals of 5min sonication every 2h. A ZOOM® IPG strip (pH 4 – 7) was rehydrated with each sample and held at 4°C overnight. Isoelectric focusing (IEF) was performed in a ZOOM® IPG Runner Mini-cell chamber using the following parameters: 1. 250V, 4mA, 10min; 2. 450V, 3mA, 10min; 3. 750V, 2mA, 10min; 4. 1000V, 1mA, 10min; 5. 2000V, 1mA, 2hrs. Following IEF separation, strips were equilibrated twice in NuPage® LDS sample buffer containing 1X Sample Reducing Agent for 15min and placed in a

NuPage® Novex 4–12% Bis-Tris ZOOM® minigel (8×8 cm). Electrophoresis in the second dimension was performed at 200V for 40min. Proteins were visualized with 0.25% coomassie brilliant blue R-250 solution in 50% Methanol, 10% acetic acid. 2-D gels from each strain were compared using PDquest 2-D analysis software (Bio-Rad). Density of spots was calculated and normalized against total density in each gel.

2.5. In gel digestion of proteins

Spots from 2-D gels were excised and destained in 60% acetonitrile, 0.2M Ammonium bicarbonate at 37°C for 30min, three times. Gel pieces were covered with 2–5µl of trypsin solution (12µg/µl, 0.2M ammonium bicarbonate) and digestion was performed overnight at 37°C. Digestion reaction was quenched with 10% Trifluoroacetic acid (TFA). Peptides were extracted twice with 100µl of 60% acetonitrile, 0.1% TFA at 37°C for 40min. Finally, the extract containing digested peptides is dried and peptides are resuspended in 10µl of buffer A (5% acetonitrile, 0.1% acetic acid). Samples are stored at –20C until MS/MS analysis.

2.6. iTRAQ (Isobaric Tag for Relative and Absolute Quantification) labeling

100µg of CFP, and 25µg of cytosolic proteins of each strain were labeled using the iTRAQ system (Applied Biosystems) following manufacturer's instructions. Tag 114 was always used to label CDC15151, while BE, C28 and H6 were always labeled with 115, 116 and 117 respectively. After labeling, samples for each subcellular fraction were combined in a single tube and cleaned-up using the Applied Biosystems Cation-Exchange Cartridge System following the manufacturer's instructions. After cation exchange purification, samples were further purified and desalted in a C-18 Sep-Pak column. After clean up was performed, each sample was dried under vacuum and reconstituted in Buffer A (2% acetonitrile, 0.5% acetic acid) to a final concentration of 2µg/µl. Samples were stored at –20C until MS/MS analysis.

2.7. Mass Spectrometry

Peptides from the in-gel digestion were analyzed by mass spectrometry using LTQ linear ion trap (Thermo Scientific). Briefly, 1 µl of purified sample was injected onto a Zorbax 300SB-C18 reverse phase nanospray column (3.5 µm, 150×0.3 mm; Agilent) using an Agilent 2D nanoHPLC. Peptides were eluted directly into the mass spectrometer using a 42 minute linear gradient from 25%–55% buffer B (90% ACN, 0.1% formic acid) at a flow rate of 300 nanoliters/min. Spectra were collected over a m/z range of 200–2000 Da using a dynamic exclusion limit of 2 MS/MS spectra of a given peptide mass for 30 s (exclusion duration of 90 s). Compound lists of the resulting spectra were generated using Bioworks 3.2 software (Thermo Scientific) with an intensity threshold of 5,000 and 1 scan/group. Spectra were subjected to interrogation against the *M. tuberculosis* genome (Genbank accession #AL123456, R9, 3,991 entries) using SEQUEST (Release 27, rev12). All searches were performed assuming trypsin digestion, 4 missed cleavages, fragment ion mass tolerance of 1.00 Da, and a parent ion tolerance of 3.00 Da. Oxidation of methionine (+16) and the acrylamide adduct of cysteine (+70) were specified as variable modifications. All in-gel analyses were compiled in Scaffold (version 2.02, Proteome Software Inc, Portland, OR) in order to validate MS/MS based peptide and protein identification. Peptide identifications were accepted if they could be established at greater than 90% probability as specified by the Peptide Prophet algorithm [31]. Protein identifications were only accepted if they could be established at greater than 90% probability assigned by the Protein Prophet algorithm [31], and contained at least 1 unique identified peptide. Sequence coverage for each protein in each spot was calculated using the Mudpit option in Scaffold (version 2.02, Proteome Software Inc, Portland, OR) compiling MS/MS results from each spot/strain.

iTRAQ labeled peptides from secreted and cytosolic fractions from the three biological replicates for each strain were analyzed on a quadrupole time-of-flight mass spectrometer

(Q-TOF MS, Agilent) directly coupled to a HPLC-chip. 0.5 µl of each sample was injected onto a reverse phase nanospray column (C18 SB-Zorbax, 75µm×43 mm analytical column, 5µm, 300A with a 40 nL enrichment cartridge). The ramped collision energy was set to a slope of 5V/100Da and offset of 2.2V. Peptides were eluted directly into the mass spectrometer using a 50 minute gradient (15%–90% buffer B (90% ACN, 0.1% formic acid) at a flow rate of 600 nanoliters/min. Tandem mass spectra are collected over a m/z range of 50–2000 Da using a dynamic exclusion limit of 2 MS/MS spectra of a given peptide mass for 30 s. Compound lists of the resulting spectra were generated using Spectrum Mill software (Agilent) with an intensity threshold of 3,000 and 5 scan/group. All biological samples were injected three times to obtain technical replicates.

Tandem mass spectra were searched against the *Mycobacterium tuberculosis* database (as above) using the Mascot search engine (Matrix Science, Version 2.1) and assuming trypsin digestion. In addition, the following parameters were applied in the search: Fixed modifications: Methylthio (C), iTRAQ4plex (N-term), iTRAQ4plex (K), Variable modifications: Carbamidomethyl (C), Carboxymethyl (C), Oxidation (M), Monoisotopic masses, Peptide mass tolerance: 2.0 Da, Fragment ion tolerance: 0.8 Da, and Missed Cleavages: 4.

All peptide ratios were calculated against strain BE (i.e. the reporter ion intensity for BE, 115, was the denominator). The results from the Mascot search were imported into Scaffold to allow for manual validation, which was performed in all cases where only one or two peptides were mapped to a protein. Criteria for manual validation included: (1) Mascot score over minimum threshold of 29, (2) minimum of 80% coverage of theoretical y or b ions (at least 5 in consecutive order), (3) absence of prominent unassigned peaks greater than 10% of the minimum intensity (i.e. noise level), (4) Indicative residue specific fragmentation, such as intense ions N-terminal to proline and immediately C-terminal to aspartate and glutamate. Single isobarically tagged peptides used for identification and quantification of a protein had to be present in at least two of the three biological replicates and represented by multiple spectra. This increased the confidence level for proteins which identification was based on unique peptide and eliminated single peptide, single sample based identification. Only ratios for manually validated peptides and those with a Mascot score greater than 27 were exported to excel for further statistical processing. Following statistical processing, all identified proteins were validated via Scaffold using the parameters described above. In addition, the filtered data set was subjected to spectral count analysis using Scaffold. Counts were normalized, taking into consideration the protein length, as longer proteins may result in higher numbers of MS/MS data that may not be directly related to the absolute abundance of that protein in a complex sample [32].

2.8. Statistical analysis

2-DE: Normalized density of each spot of biological replicates for each strain was analyzed by the Student's t-test assuming equal variances. P-values <0.05 were considering statistically significant.

iTRAQ data: Reporter ion ratios for strains CDC1551 (114), C28 (116) and H6 (117) for all technical replicates were calculated using the value of BE (115) from the same technical replicate as the denominator. After removing negative and missing values, ratios were converted to the log₂ scale and the average was calculated for replicate peptides from each technical replicate and strain. The ratios were normalized such that the median log ratio value (for each strain within each technical replicate) is set to zero. The data was filtered at both the protein and peptide levels. Proteins appearing in only a single biological replicate for a given experiment (Cytosol or CFP) were removed from further analysis. Peptides that were matched to more than one protein (by Mascot) were also removed from further

analysis. Spectra for proteins represented by only a single peptide were manually inspected and in some cases dropped. For each experiment and each protein, a mixed model was fit to the data using SAS proc mixed. The normalized $\log_2$ ratio was the response variable. Fixed effects included strain, peptide and strain-peptide interaction. Random effects included biological replicates and technical replicates. After model fitting, proteins with peptides that showed significant differences in opposite directions (up- versus down-regulation) were tagged and their spectrums were manually inspected. Some peptides were dropped from the analysis at this point and model fitting was repeated for the affected proteins. This analysis allowed the direct comparison between each of the strain for each protein. Differences with p-values <0.05 were considered statistically significant. Further manual validation (as described above) was performed on all of the proteins exhibiting statistically significant changes.

2.9. Native purification of Cfp2

Briefly, CFP from H37Rv was subjected to 40% ammonium sulfate and proteins precipitated by centrifugation at $27,000 \times g$. The pellet was saved and the supernatant was subjected to additional extraction with ammonium sulfate to reach a saturation level of 70%. The proteins were again precipitated by centrifugation and the pellet was dialyzed for 48h against 10mM Ammonium bicarbonate. After dialysis, proteins were quantified and aliquots of 30mg were obtained and subjected to buffer exchange into cation exchange loading buffer (50mM Sodium acetate, pH 4.5) using the Amicon system. The sample was applied to an SP-sepharose column (1.5ml) that had been previously packed as described by the manufacturer (Amersham) and equilibrated into loading buffer. The column was washed with 2 column volumes (CV) of loading buffer, followed by washing with 2CV of 10% elution buffer (50mM sodium acetate, pH 5.5, 1M NaCl). Bound proteins were eluted using a linear gradient from 10% – 50% elution buffer in 12 CV, and fractions collected. Fractions were resolved by SDS-PAGE and probed with anti-Cfp2 antisera. Fractions containing Cfp2 were pooled together and protein polishing accomplished by size exclusion chromatography (Superdex 200, GE Healthcare), using the Waters 600 HPLC system. Fractions were recovered, quantified and analyzed by SDS-PAGE. Finally, Cfp2 containing fractions were pooled and dialyzed against 10mM Ammonium bicarbonate.

2.10. Sequence analysis of cfp2

DNA from each strain was extracted as described elsewhere [33, 34]. The entire coding region of cfp2 for each strain was amplified using the following primers (5'-ATGAAGATGGTGAAATCGATCGCCG-3' and 5'-TCAGTTCCTGCGGCCTGCA-3'). Platinum Taq DNA polymerase High Fidelity (Invitrogen) was used for PCR amplification as suggested by the manufacturer and using the following thermal profile: 5min at 94°C, 35cycles of: a. 94°C for 30sec, b. 60°C for 30sec and c. 68°C for 1min, and a final elongation at 68°C for 3min. PCR products were agarose-gel purified using the PureLink Quick extraction kit (Invitrogen) followed by ligation into pGEM-T-easy vector (Promega) and cloning into *E.coli* DH5- α competent cells (Invitrogen). DNA Sequencing was performed at the Macromolecular Resources Laboratory (Colorado State University) and obtained sequences were analyzed using Vector NTI software (Invitrogen)

3. Results

3.1. 2-DE (CFP)

2-D gels for each biological replicate were analyzed and composite proteome maps representing each strain were created. Approximately 120 spots were visualized in each sample. Spots were quantified by densitometry and twenty eight of them that presented some extent of variability between strains were selected for identification by MS/MS

(Figure 1 Table 1). Identification of the protein was achieved for the majority of the spots and the identity was, in most cases, the same regardless of the strain from which spot was originated. MS/MS data for spots # 16 and #18 was only obtained for CDC1551. These two spots were present in the other three strains but in low quantity. These spots also presented high variability and therefore the differences between strains were not statistically significant. Spot # 9 was identified as Rv1926c in strain CDC1551, while in the other three strains it was identified as Rv1906c. Since both proteins have similar molecular mass and isoelectric point after cleavage of the predicted signal peptide, it could be possible that spot #9 of CDC1551 contains in fact both proteins, but Rv1926c is present in higher quantities. Similarly, Rv1926c could be present in spot #9 of strains BE, C28 and H6 but in quantities that are not detectable by LTQ-MS/MS.

Spot # 25 was identified as FbpB, however, western blot analysis using monoclonal antibodies against FbpB, FbpC and FbpA indicated that this spot corresponds to FbpA. The data obtained from MS/MS analysis is probably due to contamination with spot # 2 which is adjacent to spot # 25 and corresponds to FbpB, one of the most abundant proteins in the secreted fraction of *Mtb*.

Fifteen of the spots were identified as more than one protein. In addition, several proteins such as SapM, Rv1906c, FbpC1, Cfp2, GroES, HspX and Mpt64 were present in more than one spot, probably corresponding to specific species of these proteins. In this regard, other authors have previously reported the presence of some proteins such as HspX and GroES, in multiple spots [9, 17, 21, 22, 35].

In the present study, the protein with the most different chemical species was the low molecular antigen Cfp2, also called Mbt12, which was present in 8 different spots, followed by GroES in 5 and HspX in 4 spots. Interestingly, the only spot identified as Cfp2 (Spot # 17) that did not contain any other proteins was only found in CDC1551. Puzzled by this finding, native Cfp2 was purified and resolved in a 2-DE (Data not shown). Different spots were subjected to N-terminal sequencing and it was possible to determine that the Cfp2 species that was present in the majority of the spots had the predicted N-terminal sequence after removal of the signal sequence (DPASAPD). However, the spot that is only present in CDC1551 lacks the first two amino acids (DP) from the mature protein and therefore the N-terminal sequence for that Cfp2 species corresponds to ASAPDVP. Differences between Cfp2 spots bearing the predicted N-terminal sequence are probably due to other post-translational modifications that result in different isoelectric points.

After statistical analysis, it was determined that 15 of the identified spots presented significant differences between at least two of the four strains (Table 1).

Cfp2, Rv1810, and a spot containing Mpt64 and LprA presented higher abundance in CDC1551 when compared to the closely related strains BE, C28 and H6. Meanwhile, Ag85A was absent from CDC1551 and Ag85B abundance was lower in this strain than in the closely related strains BE, C28 and H6. The lower abundance of Ag85 complex in CFP of CDC1551 was confirmed by western blot analysis (Figure S1).

Two spots both corresponding to Rv1906, were found to be significantly more abundant in strain BE when compared to the other three strains.

3.2. iTRAQ

BE was arbitrarily chosen as the reference strain for the iTRAQ analysis in order to highlight the differences between the closely related isolates (i.e. BE, C28 and H6). Data analysis in which CDC1551 (tag 114) was designated as the reference strain revealed similar

results (data not shown). In addition, the mixed model used for the statistical analysis of the iTRAQ data allowed the direct comparison of each of the strains against each other regardless of the choice of reference strain.

193 proteins were identified in the cytosol (Cyt) fraction. 137 proteins of this fraction were retained after data filtering and were used for quantification and statistical analysis (Table 2 and S1). Finally, 63 of these proteins (45.99%) presented statistically significant variations between at least two of the four strains (Tables 2 and S2).

The global analysis of secreted proteins by iTRAQ allowed the identification of 146 proteins. Protein quantification and statistical analysis was performed for the 101 proteins that were retained after data filtering (Table 3 and S3). CFP proteins presenting significant variation (p -value < 0.05) between at least two of the four strains represented 43.56% ($n=44$) of the quantified CFP data set (Table 3 and S2).

28 of the quantified proteins were found in both fractions, corresponding to 27.77% and 20.43% of the CFP and cytosol respectively (Figure S2).

To confirm that the iTRAQ results were representing real differences between protein levels, several proteins (i.e. Mpt64, HspX, ModD, HbhA, Cfp10, GroEL2 and Cfp2) were analyzed by western blot and compared to iTRAQ quantitative results. Band intensities observed in fluorescent western blots followed the same pattern of relative protein levels obtained by iTRAQ for most of these individual proteins and corroborated the proteomics findings (Figure S3).

3.3. Variation in protein expression levels between Mtb strains revealed by iTRAQ

In total, 107 proteins with differences in expression in the CFP or cytosol accounted for 228 statistically significant comparisons (Table 2, 3 and S2). The majority of the variability was found between CDC1551 and the S75 group (i.e. BE, C28 and H6).

When variability between the closely related S75 group isolates was analyzed, few differences were found when comparing the H6 to C28 isolate. This analysis revealed only 13 differentially expressed proteins, 3 of which were secreted and 10 cytosolic proteins (Figure S4).

In contrast, BE presented the greatest variation in proteins when compared to C28 and H6. Specifically, 32 CFP (31.68%) and 35 cytosolic (25.55%) proteins were significantly different between BE and at least one of the other two members of the S75 group. Of these, 17 CFP and 7 cytosolic proteins were differentially expressed for both comparisons (C28 vs BE and H6 vs BE) (Figure 2). Overall, differences between BE and the two S75 strains were responsible for 75.7% and 55.5% of the changes in protein expression for secreted and cytosolic fractions respectively.

Since CDC1551 was used as an outside member for comparison purposes, differences between CDC1551 and all members of the S75 group were evaluated. Several proteins, (i.e. 27 and 50 in secreted and cytosolic fraction respectively) presented significant differences (p -value < 0.05) between CDC1551 and at least one of the other three strains. For review purposes, table S4 shows only proteins with p -value < 0.01 for differential proteins between CDC1551 and at least one S75 strain or with p -values < 0.05 for differential proteins between CDC1551 and all members of the S75 group.

Sixteen proteins, 5 in the CFP and 11 in the cytosol, presented similar levels within the S75 strains but differ significantly from CDC1551. Mpt83, Rv0398c, Wag31, Rv0854, Isr2, Rv0566c, AcpP, Rv2111c, and Cfp2 presented higher levels in CDC1551. From these,

Mpt83, and Cfp2 were at least two fold more abundant in CDC1551, while Rv2111c was four times more abundant in CDC1551 versus the S75 group isolates. On the other hand, Rv2204c, ProC, RpsQ, RpsG and RpsA presented lower levels in CDC1551 (Table S4).

3.4. Relative abundance of identified proteins in cytosol and secreted fractions

Identification of the most abundant proteins in *Mtb* secreted and cellular fractions has important implications in the context of the initial host-pathogen interactions, diagnosis, and vaccines. For this reason, spectral counts of identified cytosolic and secreted proteins were obtained by merging the MS/MS data acquired from all iTRAQ experiments including biological and technical replicates [32]. From this composite analysis, the most abundant protein in the secreted fraction was found to be GroES, followed by Rv0559c, and AcpP. Classical CFP-proteins such as GlnA1, SodA, Cfp10, HspX, ModD and PstS1 were also abundant proteins present in this fraction. In the cytosol the most abundant protein also corresponded to GroES, followed by Rv1211 and AcpP (Figure 3).

3.5. Molecular weight (Mr) and isoelectric point (IP) distribution in iTRAQ labeled proteins

The IP range detected by the iTRAQ approach in the CFP fraction was 3.68 – 11.02 with an average of 5.08. However, only 4 proteins had an IP higher than 7.0, indicating that most proteins in the CFP have an IP between 3.7 and 6.9. The proteins with the highest IP in the CFP fraction corresponded to the putative integration host factor (MihF, Rv1388) (IP= 11.02) and the invasion protein (Rv1477) (IP=9.33). In the cytosol, the IP range was much broader (3.63–12.48) and the average was higher (6.82). The protein with the highest IP in this fraction corresponded to the DNA-binding protein HU (HupB, Rv2986c). In general, proteins with the higher IP present in the cytosol fraction were mainly ribosomal proteins, which constituted 22.3% of the identified/quantified proteins in the cytosol. This set of ribosomal proteins had an IP range of 4.3 – 12.33 with an average of 10.29. Similarly, the Mr range in the cytosol (i.e. 5.91–100.87 KDa) was broader than in the CFP (i.e. 6.87–94.25 KDa). Nonetheless, the Mr average was higher in the CFP (31.99 KDa) than in the cytosol (21.07KDa).

3.6. Functional category of identified proteins

Most of the identified proteins in the cytosol belonged to the functional groups 2, 3, and 10, corresponding to information pathways (26.67%) cell wall and cell processes (17.95%), and conserved hypotheticals (31.28%) respectively. In contrast, secreted proteins were mainly grouped in functional groups 3 (21.92%), 7(39.04%) and 10 (21.23%) which correspond to cell wall and cell processes, intermediary metabolism and adaptation, and conserved hypotheticals respectively (Figure S5). Interestingly, regulatory proteins (functional group 9) from the cytosol fraction and proteins involved in cell wall and cell processes (functional group 3) from the secreted fraction presented the least variation between strains (Figure S6).

4. Discussion

In recent years, the global study of the full set of proteins encoded in a genome also referred as proteomics has become a useful tool in the study of microbial physiology, in the context of virulence, pathogenesis, and environmental adaptation. Proteomic analysis via 2-DE has been widely used to study many biological systems. In mycobacteriology, this technique has been used to assess differences in protein levels between different mycobacterial species [21, 35, 36], clinical isolates [9, 22, 37–39] and in response to certain stimuli [40–44]. Similarly, over the last decade, developments in labeling procedures, mass spectrometry instrumentation, and data analysis/bioinformatic tools have opened the proteomics field to include analyses of differences in proteomes using gel-free or shotgun techniques [23, 25, 45].

Early studies employing the analysis of proteins by 2-DE afforded global identification and annotation of mycobacterial proteins [16, 17, 21, 22, 24]. In these studies, a combination of several factors including use of large gels, visualization by silver staining, and the use of multiple gels with different isoelectric point ranges allowed the recognition of thousands of spots from different mycobacterial fractions. Such approaches are essential to obtain a comprehensive annotation of the *Mtb* proteome. However, the use of larger gels and multiple isoelectric point ranges are very time consuming for both protein separation and image analysis. This makes their implementation more difficult when multiple *Mtb* strains are compared. In addition, in most cases, the number of identified proteins by mass spectrometry corresponds to only 10 – 30% of the number of total spots, probably due to the small amount of protein present in faint spots as well as interference of silver staining with mass spectrometry analysis.

In this present study, a less sensitive, however, more straightforward 2-DE technique was used; allowing the resolution of the second dimension in a short period of time. Visualization of protein spots was performed by staining with coomassie brilliant blue, which does not interfere with upstream mass spectrometry analysis. 2-DE data was also complemented with the iTRAQ method which offers a higher throughput analysis in a streamlined procedure.

Combination of these techniques allowed the identification of several differences in the protein levels of secreted and cytosolic proteins of four different *Mtb* clinical isolates. Three of these isolates (i.e. BE, C28 and H6) are genetically very closely related and have been previously grouped in an *Mtb* cluster, the S75 group [28]. Within this group, BE is the most common strain, identified in 73.2% of the isolates, followed by H6 and C28 (23.2% and 3.6% respectively).

In addition to their genetic relatedness, their genetic variation is assumed to be small, as seen by the low copy number of the insertion fragment IS6110, previously associated with major genomic variations in clinical isolates [46].

These strains belong to the Euro-American lineage (principal genetic group 2, cluster IV). For this reason, strain CDC1551, which has been extensively characterized in terms of its biology [47–51] and belongs to the same group but different cluster (i.e. cluster V) was included in this study as a reference for S75 strains.

When compared to other *Mtb* strains such as reference strain H37Rv and hypervirulent clinical isolate HN878, *Mtb* CDC1551 has been shown to induce higher levels of Th-1 cytokines (i.e. tumor necrosis factor-alpha, (TNF-a), interleukin (IL)-6, IL-12) both in vivo and in vitro infection in mouse and human monocytes respectively [50]. This has led others to classify this strain as hyperimmunogenic [52]. In agreement with this idea, CDC1551 presented higher levels of several immunogenic proteins (Figure 1 Table S4) that could be involved with this previously reported hyperimmunogenic phenotype. For instance, Mpt83 and Cfp2 levels were considerably higher in strain CDC1551 versus the other strains. The expression of the highly immunogenic lipoprotein Mpt83 in *Mtb* strains has been reported to be considerably low during in vitro culture, but abundantly expressed in vivo, based on the immune responses of mice infected with *Mtb* [53]. At the same time, Mpt83 has been shown to be highly expressed in certain strains of *M. bovis* BCG and *M. bovis* clinical isolates with a mutation in Rv0444, the anti-sigma factor for SigK [54, 55]. CDC1551 does not bear any mutation in either sigK (Rv0445) or anti-sigK (Rv0444), nonetheless, it would be interesting to study the expression of the sigK regulon, including Mpt83, and assess any alterations in CDC1551. The expression of Mpt83 in clinical isolates of *Mtb* is important, as Mpt83 is currently being evaluated as a subunit vaccine in conjunction with Ag85 and Mpt64 [56].

Differential expression of Mpt83 in CDC1551 and S75 strains suggests that protective response elicited by Mpt83 could vary depending on the *Mtb* strain involved in a natural infection.

Interestingly, Rv2111c showed 4 times higher levels in CDC1551 (Table 2). This is a small protein recently named Pup for prokaryotic ubiquitin-like protein. This protein was recently described by Pearce and colleagues (2008) and has homology to the eukaryote ubiquitin carboxylterminal di-glycine-glutamine motif [57]. Pup is specifically conjugated to proteasome substrates, targeting them for degradation. Thus it is possible for CDC1551 to present an increased level of protein degradation compared to BE, C28 and H6. Interestingly, protein degradation by the proteasome unit has been shown to be essential for *Mtb* virulence and nitric oxide resistance [58, 59]. In addition, Pearce and colleagues (2008) were unable to identify Pup in its unconjugated form, suggesting that this protein is either rapidly degraded or is very efficient in forming interactions to proteasome substrates [57]. This contrasts our findings, and may indicate a loss of accessory molecules or products required for proper Pup interaction in CDC1551, resulting in the observed increase in Pup expression in CDC1551.

Several differences between protein levels were observed between members of the *Mtb* cluster S75 in a manner that is consistent with the genetic relatedness between these isolates. For instance protein levels between clinical isolates C28 and H6 were very similar, with only a few proteins demonstrating significant variation between these two strains (Figure S4). In contrast, several secreted and cytosolic proteins vary in strain BE when compared to strains C28 and H6 (Figure 2 Table 2 and 3). Strains H6 and C28 are thought to evolve from BE [28]; therefore, this may have an impact on the similarities or differences observed in the proteomes of these strains. Interestingly, C28 and H6 present, with some exceptions, higher levels of proteins involved in virulence, detoxification and adaptation (DnaK, HspX, Rv2493, Cfp29 and CspA) and lower levels of proteins related to the cell wall and cell processes (Cfp2, SecE2, Rv2345, PstS1, Rv3587, Rv0559c, Rv0431, PonA2, LppZ, Rv3312A) (Tables 2 and 3). Further, the differential expression of proteins belonging to group 7 (Intermediary metabolism and respiration) is found when the data is divided into subgroups. Specifically, proteins involved in metabolic pathways (i.e. Eno, SucC, GltA2, Gnd1), and nutrient acquisition (BfrB, ViuB, Sse2) are, in general, more abundant in strains C28 and H6, while proteins involved in protein degradation (PrcB, PepD, PepN) and amino acid synthesis (GlnA1, SerC) are present in lower levels in these two strains. A lower abundance of several immunogenic proteins in strains C28 and H6 versus strain BE was also observed (Tables 2 and 3). This trend was not universal; however, as BfrB and Cfp29, B- and T-cell antigens respectively [60, 61] were both more abundant in these two strains compared to BE.

The value of using different proteomic approaches in a single study, as well as the advantages and disadvantages of 2-DE and shotgun methods have been clearly shown elsewhere [26, 62, 63]. Therefore, the use of both 2-DE and iTRAQ techniques in this present study was performed in an effort to obtain a comprehensive and complementary analysis of *Mtb* clinical isolates proteomes rather than to compare the utility of each technique.

Accordingly, even though the quantification and assessment of protein differences between isolates is straightforward when iTRAQ method is used, 2-DE analysis provided important information than otherwise would have been missing. For instance, both, 2-DE and iTRAQ analyses clearly showed the increase abundance of Cfp2 in CDC1551 (Figures 1 and 4, Tables 1 and 3). However, only 2-DE showed the presence of different species of this protein, from which just one (Figure 1 and 4, spot #17) seems to account for the majority of

variation in Cfp2 levels between isolates. Interestingly, the differences between density levels of this Cfp2 species were statistically significant between CDC1551 and all three other isolates. However, comparison of Cfp2 quantity by iTRAQ labeling, which by default will show the total sum of the abundance of the different species of this protein, was only statistically significant between CDC1551 and BE (p-value = 0.018) while the differences were not significant when CDC1551 was compared to either C28 (p-value = 0.17) or H6 (p-value = 0.11). Sequence analysis of cfp2 identified a SNP in CDC1551. This SNP corresponded to an amino acid change and thus loss of this peptide as part of the iTRAQ comparative analysis. This subsequently reflected less significant differences between the MS/MS spectra of the remaining peptides by iTRAQ. Nonetheless, the pattern of Cfp2 quantity between strains was similar between 2-DE and iTRAQ. Similarly, different protein species of Rv1906c were identified by 2-DE; nonetheless, levels for this protein among strains were similar between 2-DE and iTRAQ (Figure 1, Tables 1 and 3). Differences in the quantity of individual protein species but not overall protein quantity may explain patterns of enhanced or reduced protein function. For example, the expression of modified variants of CadF in *Campylobacter jejuni* as observed by 2-DE correlates to differences in immunogenicity and virulence between strains [64]. Similarly, differences in immunogenicity to Cfp2 and Rv1906 from CDC1551 versus the clinical isolates tested here maybe explored to assess the role of protein species in relation to protein function.

Likewise, a general agreement of protein level patterns between techniques was obtained for proteins identified in single spots such as CspA (Rv3648c) and GlnA1 (Rv2220).

In addition to the identification of protein species, 2D-GE analysis increased the coverage of identified proteins. For instance, Rv0425c, Rv2878, Rv3310, Rv3803 and Rv3804 were all identified by 2D-GE but not by iTRAQ. Meanwhile, other proteins for which relative quantification was obtained by 2D-GE were only present in one biological replicate in the iTRAQ approach and therefore not included in the quantitative analysis by this last method (i.e. Rv1886c, Rv1270c, Rv3705c).

Rv1886c corresponds to Antigen 85B (Ag85B), which in previous studies [23, 65, 66], as well as in our 2-DE and those reported by others (<http://www.mpiib-berlin.mpg.de/2D-PAGE/>), appears to be among the most abundant proteins in the CFP. Puzzlingly, Ag85B is not only absent from two of the three biological replicates in the iTRAQ dataset, but also the iTRAQ signal in the remaining biological replicate is below the cutoff. This is very interesting because it suggests that iTRAQ labeling might not be completely effective regardless of the abundance of some proteins. The presence of modifications such as acetylation of the primary amino group or the formation of pyroglutamic acid in peptides with N-terminal glutamine or glutamic acid has been previously reported to result in lack of labeling by the iTRAQ reagents [67] and could explain the absence of Ag85B in the iTRAQ dataset. Further studies are required to confirm the presence of these modifications in Ag85B and other proteins identified by 2D-GE but not by iTRAQ analysis.

Comparison of iTRAQ and 2D-GE data was not possible for many of the proteins identified by both techniques (i.e. Rv1860, Rv0462, Rv1980c, Rv1080c, Rv3628, Rv3418, Rv3874 and Rv2031c) due to the presence of more than one protein in a single spot, making it difficult to determine the relative amount of these proteins by 2D-GE. In these instances, iTRAQ analysis is a more valuable tool.

In addition, iTRAQ labeling was shown to be more sensitive than traditional 2D-GE, allowing the identification of several differences in the protein levels of closely related *Mtb* strains.

In the present study, 101 and 137 proteins were identified and quantified in the *Mtb* secreted (CFP) and cytosolic fraction respectively. Of these, only 28 proteins were identified in both fractions (Figure S2). The majority of them have previously been shown to be present in the culture filtrates of *Mtb* [17, 21, 23, 24], indicating a low level of contamination from cytosolic and other subcellular fractions in the CFP prepared for this study. This is also supported by the distinct protein distribution in functional groups between fractions. For instance, cytosolic, but not the secreted fraction, contained a considerable amount of ribosomal proteins as well as other proteins involved in regulation pathways (Functional category 2) (Figure S5). Meanwhile, most of the proteins in the secreted fraction belong to functional categories 3 (cell wall and cell processes) and 7 (intermediary metabolism), both of which have been reported to constitute the majority of proteins in this fraction [23]. Differences in the biochemical properties between secreted and cytosolic fractions were also observed. Particularly, the isoelectric point of cytosolic proteins tends to be higher than of secreted proteins which could explain the relatively low coverage of cytosolic proteins by 2D-GE in previous studies in which an IP range of 4–7 was used [17].

The iTRAQ approach described in this study contributed to the annotation of 53 secreted and 68 cytosolic proteins for which no previous proteomic identification was available.

Variation in expression levels between proteins that were found in both fractions was observed (Figure S2). For instance, when a value of ± 0.2 is used as an arbitrary cutoff to determine variation between fractions, only 7 (25%), 9 (32.14%) and 10 (35.71%) proteins presented similar levels between cytosol and secreted fractions in CDC1551, C28 and H6 respectively (Data not shown). The variation between fractions within one single strain illustrates the utility of proteomics over mRNA arrays on the ability to localize and quantify protein levels in different cell compartments, and to identify different protein species that result from post translational modifications.

Interestingly, conserved and hypothetical proteins with unknown function corresponded to a significant portion of both secreted (21.23% (n=21)) and cytosolic (31.28% (n=31)) fractions from which 7 and 19 proteins presented significant variation between at least two of the four analyzed strains. This remarks the importance of continuing current and future research focused on characterization of *Mtb* proteins at the functional level which will increase our understanding of mycobacterial physiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NIH grant NIH, NIAID, N01-AI-40091

Abbreviations

Mtb	Mycobacterium tuberculosis
CFP	Culture filtrate proteins

References

1. Ross PL, Huang YN, Marchese JN, Williamson B, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004; 3:1154–1169. [PubMed: 15385600]

2. D'Ascenzo M, Choe L, Lee KH. iTRAQPak: an R based analysis and visualization package for 8-plex isobaric protein expression data. *Brief Funct Genomic Proteomic*. 2008; 7:127–135. [PubMed: 18272547]
3. Chen Y, Choong LY, Lin Q, Philp R, et al. Differential expression of novel tyrosine kinase substrates during breast cancer development. *Mol Cell Proteomics*. 2007; 6:2072–2087. [PubMed: 17855441]
4. Gagne JP, Ethier C, Gagne P, Mercier G, et al. Comparative proteome analysis of human epithelial ovarian cancer. *Proteome Sci*. 2007; 5:16. [PubMed: 17892554]
5. Drummelsmith J, Winstall E, Bergeron MG, Poirier GG, Ouellette M. Comparative proteomics analyses reveal a potential biomarker for the detection of vancomycin-intermediate *Staphylococcus aureus* strains. *J Proteome Res*. 2007; 6:4690–4702. [PubMed: 17997515]
6. Lenco J, Link M, Tambor V, Zakova J, et al. iTRAQ quantitative analysis of *Francisella tularensis* ssp. *holarctica* live vaccine strain and *Francisella tularensis* ssp. *tularensis* SCHU S4 response to different temperatures and stationary phases of growth. *Proteomics*. 2009; 9:2875–2882. [PubMed: 19415661]
7. Zeaiter Z, Huynh HQ, Kanyo R, Stein M. CagA of *Helicobacter pylori* alters the expression and cellular distribution of host proteins involved in cell signaling. *FEMS Microbiol Lett*. 2008; 288:227–234. [PubMed: 18811656]
8. Dye C. Global epidemiology of tuberculosis. *Lancet*. 2006; 367:938–940. [PubMed: 16546542]
9. Pheiffer C, Betts JC, Flynn HR, Lukey PT, van Helden P. Protein expression by a Beijing strain differs from that of another clinical isolate and *Mycobacterium tuberculosis* H37Rv. *Microbiology*. 2005; 151:1139–1150. [PubMed: 15817781]
10. Reed MB, Domenech P, Manca C, Su H, et al. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature*. 2004; 431:84–87. [PubMed: 15343336]
11. Shimono N, Morici L, Casali N, Cantrell S, et al. Hypervirulent mutant of *Mycobacterium tuberculosis* resulting from disruption of the *mce1* operon. *Proc Natl Acad Sci U S A*. 2003; 100:15918–15923. [PubMed: 14663145]
12. Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere YO, et al. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J Clin Microbiol*. 2005; 43:3185–3191. [PubMed: 16000433]
13. Weldingh K, Rosenkrands I, Jacobsen S, Rasmussen PB, et al. Two-dimensional electrophoresis for analysis of *Mycobacterium tuberculosis* culture filtrate and purification and characterization of six novel proteins. *Infect Immun*. 1998; 66:3492–3500. [PubMed: 9673225]
14. Urquhart BL, Cordwell SJ, Humphery-Smith I. Comparison of predicted and observed properties of proteins encoded in the genome of *Mycobacterium tuberculosis* H37Rv. *Biochem Biophys Res Commun*. 1998; 253:70–79. [PubMed: 9875222]
15. Sonnenberg MG, Belisle JT. Definition of *Mycobacterium tuberculosis* culture filtrate proteins by two-dimensional polyacrylamide gel electrophoresis, N-terminal amino acid sequencing, and electrospray mass spectrometry. *Infect Immun*. 1997; 65:4515–4524. [PubMed: 9353028]
16. Rosenkrands I, Weldingh K, Jacobsen S, Hansen CV, et al. Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. *Electrophoresis*. 2000; 21:935–948. [PubMed: 10768780]
17. Rosenkrands I, King A, Weldingh K, Moniatte M, et al. Towards the proteome of *Mycobacterium tuberculosis*. *Electrophoresis*. 2000; 21:3740–3756. [PubMed: 11271494]
18. Mollenkopf HJ, Mattow J, Schaible UE, Grode L, et al. *Mycobacterial* proteomes. *Methods Enzymol*. 2002; 358:242–256. [PubMed: 12474391]
19. Mattow J, Schmidt F, Hohenwarter W, Siejak F, et al. Protein identification and tracking in two-dimensional electrophoretic gels by minimal protein identifiers. *Proteomics*. 2004; 4:2927–2941. [PubMed: 15378758]
20. Mattow J, Jungblut PR, Muller EC, Kaufmann SH. Identification of acidic, low molecular mass proteins of *Mycobacterium tuberculosis* strain H37Rv by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry. *Proteomics*. 2001; 1:494–507. [PubMed: 11681203]

21. Jungblut PR, Schaible UE, Mollenkopf HJ, Zimny-Arndt U, et al. Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol Microbiol.* 1999; 33:1103–1117. [PubMed: 10510226]
22. Betts JC, Dodson P, Quan S, Lewis AP, et al. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. *Microbiology.* 2000; 146(Pt 12):3205–3216. [PubMed: 11101678]
23. Malen H, Berven FS, Fladmark KE, Wiker HG. Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics.* 2007; 7:1702–1718. [PubMed: 17443846]
24. Mollenkopf HJ, Jungblut PR, Raupach B, Mattow J, et al. A dynamic two-dimensional polyacrylamide gel electrophoresis database: the mycobacterial proteome via Internet. *Electrophoresis.* 1999; 20:2172–2180. [PubMed: 10493122]
25. Mawuenyega KG, Forst CV, Dobos KM, Belisle JT, et al. *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. *Mol Biol Cell.* 2005; 16:396–404. [PubMed: 15525680]
26. Schmidt F, Donahoe S, Hagens K, Mattow J, et al. Complementary analysis of the *Mycobacterium tuberculosis* proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology. *Mol Cell Proteomics.* 2004; 3:24–42. [PubMed: 14557599]
27. Gu S, Chen J, Dobos KM, Bradbury EM, et al. Comprehensive Proteomic Profiling of the Membrane Constituents of a *Mycobacterium tuberculosis* Strain. *Mol Cell Proteomics.* 2003; 2:1284–1296. [PubMed: 14532352]
28. Mathema B, Bifani PJ, Driscoll J, Steinlein L, et al. Identification and evolution of an IS6110 low-copy-number *Mycobacterium tuberculosis* cluster. *J Infect Dis.* 2002; 185:641–649. [PubMed: 11865421]
29. Takayama K, Schnoes HK, Armstrong EL, Boyle RW. Site of inhibitory action of isoniazid in the synthesis of mycolic acids in *Mycobacterium tuberculosis*. *J Lipid Res.* 1975; 16:308–317. [PubMed: 806645]
30. Dobos KM, Swiderek K, Khoo KH, Brennan PJ, Belisle JT. Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *Infect Immun.* 1995; 63:2846–2853. [PubMed: 7622204]
31. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem.* 2002; 74:5383–5392. [PubMed: 12403597]
32. Zybilov B, Mosley AL, Sardi ME, Coleman MK, et al. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res.* 2006; 5:2339–2347. [PubMed: 16944946]
33. Ausubel, FM.; Brent, R.; Kingston, RE.; Moore, DD.; Smith, JA.; Seidman, JG.; Struhl, K. *Current protocols in molecular biology.* Brooklyn, NY: Greene Publishing Associates; 1988.
34. van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol.* 1991; 29:2578–2586. [PubMed: 1685494]
35. Mattow J, Schaible UE, Schmidt F, Hagens K, et al. Comparative proteome analysis of culture supernatant proteins from virulent *Mycobacterium tuberculosis* H37Rv and attenuated *M. bovis* BCG Copenhagen. *Electrophoresis.* 2003; 24:3405–3420. [PubMed: 14595687]
36. Monahan IM, Betts J, Banerjee DK, Butcher PD. Differential expression of mycobacterial proteins following phagocytosis by macrophages. *Microbiology.* 2001; 147:459–471. [PubMed: 11158363]
37. Jiang X, Zhang W, Gao F, Huang Y, et al. Comparison of the proteome of isoniazid-resistant and -susceptible strains of *Mycobacterium tuberculosis*. *Microb Drug Resist.* 2006; 12:231–238. [PubMed: 17227207]
38. Jiang X, Zhang Y, Zhang W, Gao F, et al. Identification of unique genetic markers in Rv0927c among *Mycobacterium tuberculosis* W-Beijing strains. *Microbes Infect.* 2007; 9:241–246. [PubMed: 17224291]

39. Ryoo SW, Park YK, Park SN, Shim YS, et al. Comparative proteomic analysis of virulent Korean *Mycobacterium tuberculosis* K-strain with other mycobacteria strain following infection of U-937 macrophage. *J Microbiol.* 2007; 45:268–271. [PubMed: 17618234]
40. Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol.* 2002; 43:717–731. [PubMed: 11929527]
41. Garbe TR, Hibler NS, Deretic V. Response of *Mycobacterium tuberculosis* to reactive oxygen and nitrogen intermediates. *Mol Med.* 1996; 2:134–142. [PubMed: 8900541]
42. Rosenkrands I, Slayden RA, Crawford J, Aagaard C, et al. Hypoxic response of *Mycobacterium tuberculosis* studied by metabolic labeling and proteome analysis of cellular and extracellular proteins. *J Bacteriol.* 2002; 184:3485–3491. [PubMed: 12057942]
43. Starck J, Kallenius G, Marklund BI, Andersson DI, Akerlund T. Comparative proteome analysis of *Mycobacterium tuberculosis* grown under aerobic and anaerobic conditions. *Microbiology.* 2004; 150:3821–3829. [PubMed: 15528667]
44. Wong DK, Lee BY, Horwitz MA, Gibson BW. Identification of fur aconitase, and other proteins expressed by *Mycobacterium tuberculosis* under conditions of low and high concentrations of iron by combined two-dimensional gel electrophoresis and mass spectrometry. *Infect Immun.* 1999; 67:327–336. [PubMed: 9864233]
45. Radosevich TJ, Reinhardt TA, Lippolis JD, Bannantine JP, Stabel JR. Proteome and differential expression analysis of membrane and cytosolic proteins from *Mycobacterium avium* subsp. paratuberculosis strains K-10 and 187. *J Bacteriol.* 2007; 189:1109–1117. [PubMed: 17142399]
46. Ho TB, Robertson BD, Taylor GM, Shaw RJ, Young DB. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. *Yeast.* 2000; 17:272–282. [PubMed: 11119304]
47. Valway SE, Sanchez MP, Shinnick TF, Orme I, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N Engl J Med.* 1998; 338:633–639. [PubMed: 9486991]
48. Bishai WR, Dannenberg AM Jr, Parrish N, Ruiz R, et al. Virulence of *Mycobacterium tuberculosis* CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. *Infect Immun.* 1999; 67:4931–4934. [PubMed: 10456953]
49. Manca C, Reed MB, Freeman S, Mathema B, et al. Differential monocyte activation underlies strain-specific *Mycobacterium tuberculosis* pathogenesis. *Infect Immun.* 2004; 72:5511–5514. [PubMed: 15322056]
50. Manca C, Tsenova L, Barry CE 3rd, Bergtold A, et al. *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. *J Immunol.* 1999; 162:6740–6746. [PubMed: 10352293]
51. Tsenova L, Ellison E, Harbacheuski R, Moreira AL, et al. Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. *J Infect Dis.* 2005; 192:98–106. [PubMed: 15942899]
52. Barczak AK, Domenech P, Boshoff HI, Reed MB, et al. In vivo phenotypic dominance in mouse mixed infections with *Mycobacterium tuberculosis* clinical isolates. *J Infect Dis.* 2005; 192:600–606. [PubMed: 16028128]
53. Hewinson RG, Michell SL, Russell WP, McAdam RA, Jacobs WR Jr. Molecular characterization of MPT83: a seroreactive antigen of *Mycobacterium tuberculosis* with homology to MPT70. *Scand J Immunol.* 1996; 43:490–499. [PubMed: 8633206]
54. Charlet D, Mostowy S, Alexander D, Sit L, et al. Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in sigK. *Mol Microbiol.* 2005; 56:1302–1313. [PubMed: 15882422]
55. Said-Salim B, Mostowy S, Kristof AS, Behr MA. Mutations in *Mycobacterium tuberculosis* Rv0444c, the gene encoding anti-SigK, explain high level expression of MPB70 and MPB83 in *Mycobacterium bovis*. *Mol Microbiol.* 2006; 62:1251–1263. [PubMed: 17064366]
56. Cai H, Yu DH, Hu XD, Li SX, Zhu YX. A combined DNA vaccine-prime, BCG-boost strategy results in better protection against *Mycobacterium bovis* challenge. *DNA Cell Biol.* 2006; 25:438–447. [PubMed: 16907641]

57. Pearce MJ, Mintseris J, Ferreyra J, Gygi SP, Darwin KH. Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science*. 2008; 322:1104–1107. [PubMed: 18832610]
58. Darwin KH, Ehrt S, Gutierrez-Ramos JC, Weich N, Nathan CF. The proteasome of *Mycobacterium tuberculosis* is required for resistance to nitric oxide. *Science*. 2003; 302:1963–1966. [PubMed: 14671303]
59. Pearce MJ, Arora P, Festa RA, Butler-Wu SM, et al. Identification of substrates of the *Mycobacterium tuberculosis* proteasome. *EMBO J*. 2006; 25:5423–5432. [PubMed: 17082771]
60. Rosenkrands I, Rasmussen PB, Carnio M, Jacobsen S, et al. Identification and characterization of a 29-kilodalton protein from *Mycobacterium tuberculosis* culture filtrate recognized by mouse memory effector cells. *Infect Immun*. 1998; 66:2728–2735. [PubMed: 9596740]
61. Sartain MJ, Slayden RA, Singh KK, Laal S, Belisle JT. Disease state differentiation and identification of tuberculosis biomarkers via native antigen array profiling. *Mol Cell Proteomics*. 2006; 5:2102–2113. [PubMed: 16899542]
62. Choe LH, Aggarwal K, Franck Z, Lee KH. A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in *Escherichia coli* cells. *Electrophoresis*. 2005; 26:2437–2449. [PubMed: 15924362]
63. Sophie, Alvarez; Berla, Bertram M.; Sheffield, Jeanne; Cahoon, Rebecca E., et al. Comprehensive analysis of the *Brassica juncea* root proteome in response to cadmium exposure by complementary proteomic approaches. *Proteomics*. 2009; 9:2419–2431. [PubMed: 19343712]
64. Scott NE, Marzook NB, Deutscher A, Falconer L, et al. Mass spectrometric characterization of the *Campylobacter jejuni* adherence factor CadF reveals post-translational processing that removes immunogenicity while retaining fibronectin binding. *Proteomics*. 10:277–288. [PubMed: 19941310]
65. Nagai S, Wiker HG, Harboe M, Kinomoto M. Isolation and partial characterization of major protein antigens in the culture fluid of *Mycobacterium tuberculosis*. *Infect Immun*. 1991; 59:372–382. [PubMed: 1898899]
66. Wiker HG, Harboe M, Nagai S. A localization index for distinction between extracellular and intracellular antigens of *Mycobacterium tuberculosis*. *J Gen Microbiol*. 1991; 137:875–884. [PubMed: 1906925]
67. Boehm AM, Putz S, Altenhofer D, Sickmann A, Falk M. Precise protein quantification based on peptide quantification using iTRAQ. *BMC Bioinformatics*. 2007; 8:214. [PubMed: 17584939]
68. Skjot RL, Oettinger T, Rosenkrands I, Ravn P, et al. Comparative evaluation of low-molecular-mass proteins from *Mycobacterium tuberculosis* identifies members of the ESAT-6 family as immunodominant T-cell antigens. *Infect Immun*. 2000; 68:214–220. [PubMed: 10603390]
69. Berthet FX, Rasmussen PB, Rosenkrands I, Andersen P, Gicquel B. A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). *Microbiology*. 1998; 144(Pt 11):3195–3203. [PubMed: 9846755]
70. Florczyk MA, McCue LA, Stack RF, Hauer CR, McDonough KA. Identification and characterization of mycobacterial proteins differentially expressed understanding and shaking culture conditions, including Rv2623 from a novel class of putative ATP-binding proteins. *Infect Immun*. 2001; 69:5777–5785. [PubMed: 11500455]
71. Jungblut PR, Muller EC, Mattow J, Kaufmann SH. Proteomics reveals open reading frames in *Mycobacterium tuberculosis* H37Rv not predicted by genomics. *Infect Immun*. 2001; 69:5905–5907. [PubMed: 11500470]

Figure 1.
Representative 2D gels from the secreted fraction of each analyzed strain (Clockwise:
CDC1551, BE, H6, C28). Numbers correspond to proteins in Table 1. pH range is 4 – 7.

Figure 2.

Differential CFP (Panel A) and Cytosolic (Panel B) proteins between BE and both, C28 and H6 (p-value <0.05) obtained by iTRAQ labeling followed by MS/MS analysis. Note: PrcB, Rv0569 and Rv2493 abundance values were also significantly different between BE and CDC1551 in the CFP. All cytosolic proteins, with exception of Rv2302 and Rv0020c also presented significantly different abundance values between BE and CDC1551.

Figure 3. Relative abundance of proteins identified in the CFP (secreted) or cytosolic fraction by iTRAQ labeling followed by MS/MS analysis. NSAF: Normalized Spectral-counts Abundance Factor. Note: Spectral counts were compiled from both biological and technical replicates.

Figure 4.
Comparison of Cfp2 levels between A: 2D-GE and B: iTRAQ. Notes: Please refer to Figure 1 and Table 1 for spot identification. Note that several peptides were used for identification and quantification of Cfp2 by iTRAQ. Only two are shown in this figure.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Identified proteins of selected 2D-GE spots and differences observed between *Mtb* strains.

Spot #	NCBI Accession number	Protein	Gene name	Number of unique peptides	Sequence coverage	pMr/mMr (kDa) ^a	pIP/mIP ^b	Comparison between strains ^c (p-value <0.05)	Reference
1	gi157116926 gi15607603	Rv1860 Rv0462	apa_modD lpdC	7 2	36% 4.3%	32.7/28.78 49.2/nss	4.7/4.52 5.7/nss	No statistically significant	[16, 17, 22, 23] [16, 17, 21, 23, 24]
2	gi15609023	Rv1886c	Ag85B, fbpB	6	24%	34.58/30.5	5.71/4.87	↓CDC1551 vs ↑BE, H6	[16, 17, 21, 23, 24]
3	gi15609117 gi15608410	Rv1980c Rv1270c	mp64 lprA	13 2	55% 11%	24.82/23.43 24.87/21.59	4.59/4.6 5.21/5.04	↑CDC1551 vs ↓BE, C28, H6	[16, 17, 21, 23, 24] [23]
4	gi15610841	Rv3705c ^d	--	3	11%	22.36/19.99	4.73/4.91	No statistically significant	[23]
5	gi15610446 gi157117159	Rv3310 Rv3803c ^d	sapM fbpC1, fbpD	3 1	14% 2.3%	31.8/27.1 31.08/27.8	6.6/5.4 6.63/5.51	↓CDC1551 vs ↑BE, C28	[23] [16, 17, 21, 23, 24]
6	gi157117159	Rv3803c ^d	fbpC1, fbpD	6	34%	31.08/27.8	6.63/5.51	No statistically significant	[16, 17, 21, 23, 24]
7	gi15608220 gi15610764	Rv1080c Rv3628	greA ppa	2 2	14% 14%	17.85/nss 18.29/nss	4.63/nss 4.51/nss	No statistically significant	[21, 24] [21, 24]
8	gi15609043	Rv1906c ^d	--	2	17%	15.5/12.5	5.01/4.68	↑BE vs ↓CDC1551, C28, H6	[23]
9	gi15609063 gi15609043	Rv1926c Rv1906c ^d	mp63 --	2 3	24% 26%	16.5/13.6 15.5/12.5	4.6/4.5 5.01/4.68	No statistically significant	[16, 17, 21, 23, 24] [23]
10	gi15609043	Rv1906c ^d	--	1	9.6%	15.5/12.5	5.01/4.68	Absent in CDC1551	[23]
11	gi15610554 gi15609513 gi15608492	Rv3418c Rv2376c Rv1352	groES mbt12, cfp2 --	10 2 1	97% 15% 11%	10.77/nss 16.63/12.05 12.85/9.98	4.34/nss 6.52/5.1 5.62/4.63	↑H6 vs ↓BE, C28	[16, 17, 21, 23, 24] [16, 17, 23] [23]
12	gi15610554 gi15611010	Rv3418c Rv3874	groES esxB, cfp10	7 4	67% 69%	10.77/nss 10.79/nss	4.34/nss 4.31/nss	No statistically significant	[16, 17, 21, 23, 24] [16, 17, 21, 23, 24, 68, 69]
13	gi15610015 gi15609341	Rv2878c Rv2204c	mpt53 --	2 3	20% 25%	18.38/14.6 12.54/nss	4.98/4.57 4.18/nss	↓CDC1551 vs ↑C28	[16, 17, 23] ---
14	gi15609513 gi15609168 gi15610015	Rv2376c Rv2031c Rv2878c	mbt12, cfp2 hspX, acr mpt53	3 2 2	23% 19% 24%	16.63/12.05 23.26/nss 18.38/14.6	6.52/5.1 5.45/nss 4.98/4.57	No statistically significant	[16, 17, 23] [16, 17, 21, 23] [16, 17, 21, 23, 24, 70]
15	gi15609513 gi15609168	Rv2376c Rv2031c	mbt12, cfp2 hspX, acr	4 3	24% 29%	16.63/12.05 23.26/nss	6.52/5.1 5.45/nss	No statistically significant	[16, 17, 23] [16, 17, 21, 23, 24, 70]
16	gi15609513 gi15610554	Rv2376c Rv3418c	mbt12, cfp2 groES	5 3	54% 28%	16.63/12.05 10.77/nss	6.52/5.1 4.34/nss	No statistically significant	[16, 17, 23] [16, 17, 21, 23, 24]

\$watermark-text

\$watermark-text

\$watermark-text

Spot #	NCBI Accession number	Protein	Gene name	Number of unique peptides	Sequence coverage	pMr/mMr (kDa) ^a	pIP/mIP ^b	Comparison between strains ^c (p-value <0.05)	Reference
17	gi 15609513	Rv2376c	mbt12, cfp2	4	34%	16.63/12.05	6.52/5.1	Only present in CDC1551	[16, 17, 23]
18	gi 15609513 gi 15610452	Rv2376c Rv3616c	mbt12, cfp2 ---	2 2	16% 6.9%	16.63/12.05 39.88/nss	6.52/5.1 5.01/nss	No statistically significant	[16, 17, 23]
19	gi 15609513 gi 15610554	Rv2376c Rv3418c	mbt12, cfp2 groES	3 2	23% 16%	16.63/12.05 10.77/nss	6.52/5.1 4.34/nss	↑CDC1551 vs ↓BE, C28	[16, 17, 23] [16, 17, 21, 23, 24]
20	gi 15610554 gi 15609168	Rv3418c Rv2031c	groES hspX	11 4	97% 33%	10.77/nss 23.26/nss	4.34/nss 4.31/nss	No statistically significant	[16, 17, 21, 23, 24] [16, 17, 21, 23, 24, 68, 69]
21	gi 15608947 gi 57117165	Rv1810 ^d Rv3875	-- EsxA, Esat-6	2 2	27% 33%	12.14/8.42 9.93/nss	6.06/4.5 4.19/nss	↑CDC1551 vs ↓C28, H6	[23] [16, 17, 21, 23, 24, 68, 69]
22	gi 15610784	Rv3648c	cspA	3	37%	7.37/nss	4.95/nss	↓CDC1551 vs ↑BE, C28	[16, 17, 21, 24]
23	gi 15609357 gi 15610384	Rv2220 Rv3248	glnA1 sahH	14 6	32% 13%	53.53/nss 54.32/nss	4.84/nss 4.84/nss	No statistically significant	[16, 17, 23] [16, 17, 23]
24	gi 15608214	Rv1074c	fadA3	18	57%	42.65/nss	4.66/nss	No statistically significant	[21, 23, 24]
25 ^e	gi 15609023	Rv1886c	fbpB, Ag85B	5	22%	34.58/30.5	5.71/4.87	Absent in CDC1551	[16, 17, 21, 23, 24]
26	gi 15607456	Rv0315	--	3	13%	32.18/28.8	4.7/4.79	Only present in C28	[23]
27	gi 15609513 gi 15609168	Rv2376c Rv2031c	mbt12, cfp2 hspX, acr	3 5	23% 36.8%	16.63/12.05 23.26/nss	6.52/5.1 5.45/nss	Absent in H6	[16, 17, 23] [16, 17, 21, 23, 24, 70]
28	gi 15609117	Rv1980c	mpt64	9	41%	24.82/22.43	4.59/4.6	Present only in H6	[16, 17, 21, 23, 24]

^a pMr= pre-protein molecular weight, mMr= mature protein molecular weight.

^b pIP= pre-protein isoelectric point, mIP= mature protein isoelectric point, nss= no signal sequence.

^c Individual spots were quantified by densitometry and differences between strains for each spot were evaluated by the t-student test.

^d Signal sequence experimentally confirmed by Malen et al [23].

^e this spot was identified as FbpB (Rv1886c) by MS/MS, however, western blot analysis indicated that this spot corresponds to FbpA(Rv3804c) (Data not shown).

Table 2

Cytosolic proteins identified by Tandem MS/MS, quantified by the iTRAQ approach and with significantly different protein levels between strains ($p < 0.05$).

Cytosolic proteins with significantly different levels between strains (p-value < 0.05)									
Ratio against BE (BE= 1.0)									
NCBI Accession number	Rvnumber	Gene	CDC1551	C28	H6	Functional category	Number of unique Peptides ^c	% Sequence coverage	References
15607151	Rv0009	ppiA	0.87	1.20	1.23	2	4	27.42	[13, 21, 23–25, 44]
15607162	Rv0020c	TB39.8	1.20	0.52	0.65	9	2	10.42	[16, 17, 21, 25]
15607195	Rv0053	rpsF	0.89	0.81	0.99	2	4	43.75	[21, 24, 25]
15607196	Rv0054	ssb	1.25	1.55	1.17	10	2	16.77	
15607391	Rv0250c	--	0.87	0.96	1.02	10	2	37.11	
15607411 ^a	Rv0270	fadD2	1.34	1.12	0.95	1	1	1.75	[17, 25]
15607491	Rv0350	dnaK	1.18	1.40	1.16	0	7	22.56	[17, 21, 23–25]
57116727	Rv0379	sedE2	0.79	0.65	0.78	3	4	60.56	
15607706	Rv0566c	--	1.49	0.94	1.12	10	4	28.83	[21, 24, 25]
15607776	Rv0636	hadB	0.74	1.46	1.03	7	1	8.45	[17, 25]
15607780	Rv0640	rplK	0.96	1.26	1.10	2	2	30.99	[25]
15607792	Rv0652	rplL	1.03	1.50	1.26	2	6	44.62	[17, 21, 25]
15607823	Rv0683	rpsG	0.55	0.82	0.97	2	3	30.13	[25]
15607825	Rv0685	tuf	1.51	1.55	1.13	2	5	22.73	[16, 17, 21, 24, 25]
15607841	Rv0701	rplC	1.00	0.71	1.35	2	3	17.51	[25]
15607843 ^a	Rv0703	rplW	0.89	1.00	1.20	2	1	11.00	[25]
15607850	Rv0710	rpsQ	0.61	0.87	0.98	2	3	17.65	
15607858	Rv0718	rpsH	0.77	0.82	0.99	2	5	42.42	
15607859	Rv0719	rplF	0.84	1.05	1.29	2	3	21.23	[25]
15607873	Rv0733	adk	1.36	1.27	1.16	7	9	65.75	[17, 21, 24]
15607954	Rv0814c	sseC2	1.37	1.72	1.28	7	4	56.00	
15607994	Rv0854	--	1.31	0.94	0.97	10	2	14.29	
15608003	Rv0863	--	1.36	1.14	0.97	10	2	56.99	
15608220	Rv1080c	greA	0.86	0.85	0.94	2	4	29.27	[21, 24, 25]

\$watermark-text

\$watermark-text

\$watermark-text

Cytosolic proteins with significantly different levels between strains (p-value <0.05)									
Ratio against BE (BE= 1.0)									
NCBI Accession number	Rvnumber	Gene	CDC1551	C28	H6	Functional category	Number of unique Peptides ^c	% Sequence coverage	References
57116835	Rv1159A	phhB	1.55	1.20	1.15	10	1	12.77	[71]
15608448	Rv1308	atpA	0.86	1.22	0.89	7	1	2.00	[17, 25]
15608475	Rv1335	cfp10A	1.25	1.31	1.05	10	2	40.86	[20]
15608704 ^a	Rv1566c	inv	1.76	0.87	0.90	0	1	6.96	
15608768	Rv1630	rpsA	0.42	1.37	1.08	2	2	4.99	[25]
15608774	Rv1636	TB15.3	1.15	1.50	1.30	10	3	29.45	[16, 17, 21, 24]
15608779	Rv1641	infC	1.17	0.74	0.72	2	2	15.92	[25]
15608822 ^a	Rv1684	--	1.32	1.15	0.83	10	1	18.67	
15608876	Rv1738	--	0.97	1.51	1.23	10	2	23.40	[25]
15608964	Rv1827	cfp17	0.90	0.84	0.89	10	7	67.28	[13, 17, 23]
15609069	Rv1932	tpX	1.10	1.26	1.19	0	4	39.05	[23]
15609117	Rv1980c	mpt64	1.64	1.26	1.30	3	6	35.53	[17, 21, 23, 24]
15609168	Rv2031c	hspX	0.73	1.05	1.08	0	5	43.75	[16, 17, 21, 24, 25, 42, 44, 70]
15609192	Rv2055c	RpsR2	0.91	0.74	1.02	2	2	26.14	
15609231 ^a	Rv2094c	tatA	1.17	0.80	1.05	3	1	20.48	
15609248 ^a	Rv2111c	pup	4.04	0.83	0.76	10	1	31.25	
15609282	Rv2145c	wag31	1.62	1.19	0.98	3	5	24.62	[21, 24]
15609322	Rv2185c	TB16.3	1.23	1.23	1.08	10	5	45.83	[16, 17, 21]
15609352 ^a	Rv2215	dlaT	0.84	0.79	0.65	7	1	3.61	[17, 25]
15609381	Rv2244	acpP	1.16	0.86	0.93	1	7	56.52	[17, 25]
15609439	Rv2302	--	1.00	0.56	0.62	10	3	60.00	
15609513 ^a	Rv2376c	cfp2	2.24	0.46	0.52	3	1	8.33	[17, 23]
15609549	Rv2412	rpsT	0.64	0.74	0.92	2	2	23.26	[25]
15609579 ^a	Rv2442c	rplU	0.83	0.98	1.19	2	1	18.27	
15609599	Rv2462c	tig	1.35	1.26	1.11	3	5	16.31	[16, 17, 21, 25]
15609657 ^a	Rv2520c	--	1.26	0.88	1.06	3	1	17.33	
15609671	Rv2534c	efp	1.45	1.22	0.99	2	1	14.97	[16, 17, 21, 24]

\$watermark-text

\$watermark-text

\$watermark-text

Cytosolic proteins with significantly different levels between strains (p-value <0.05)

Ratio against BE (BE= 1.0)

NCBI Accession number	Rvnumber	Gene	CDC1551	C28	H6	Functional category	Number of unique Peptides ^c	% Sequence coverage	References
57117019 ^a	Rv2744c	30kd_ag	0.68	1.23	1.54	10	2	9.63	[25]
15610015 ^b	Rv2878c	mpt53	1.39	1.51	0.92	3	1	7.55	[16, 17, 23]
15610026 ^a	Rv2889c	tsf	1.27	0.78	0.63	2	1	4.06	[21, 24, 25]
15610082	Rv2945c	lppX	1.14	0.97	0.96	3	3	28.76	[23, 25]
15610166	Rv3029c	fixA	1.22	1.50	1.21	7	5	22.56	[17, 21, 24, 25]
15610183	Rv3046c	--	0.67	0.76	0.85	10	2	32.26	
15610554	Rv3418c	groES	0.73	0.59	0.75	0	8	90.00	[17, 21, 23-25]
15610733	Rv3597c	isr2	1.35	0.95	1.01	10	2	13.39	[17, 25]
15610784 ^a	Rv3648c	cspA	1.58	1.74	1.40	0	2	49.25	[17, 21, 24, 25]
15610852	Rv3716c	--	0.91	1.42	1.21	10	3	41.35	[21, 24]
15610977	Rv3841	bfrB	2.01	2.37	1.64	7	3	19.34	[17, 23, 25]
15611015 ^a	Rv3879c	--	1.07	0.71	0.86	10	1	2.20	[25]

All proteins were identified with at least 95% probability at the protein levels, except for

^aProteins identified with at least 81 % probability at the protein level (n=30); and^bProteins identified with at least 79% probability at the protein level (n=2).^cAll unique peptides were identified with at least 95% probability

Table 3

CFP proteins identified by MS/MS, quantified by the iTRAQ approach and with significantly different levels between strains (p-value <0.05).

CFP proteins with differential levels between strains (p-value <0.05)										
NCBI Accession number	Rvnumber	Gene	Ratio against BE (BE= 1.0)				Functional category	Number of unique peptides ^c	% Sequence coverage	References
			CDC1551	C28	H6					
15607504	Rv0363c	fb ^a	0.74	1.32	1.16		7	1	4.65	[17, 21, 23]
15607539 ^a	Rv0398c	--	1.48	0.85	0.76		3	1	5.16	[23]
15607572	Rv0431	--	0.84	0.38	0.40		3	2	40.85	
15607581	Rv0440	groEL1	1.14	1.32	1.09		0	5	12.96	[17, 21, 24]
15607641	Rv0500	proC	0.61	0.96	1.02		7	2	9.83	
15607699 ^a	Rv0559c	--	0.95	0.59	0.67		3	2	38.39	[21, 23, 24]
15607709	Rv0569	--	0.64	0.55	0.69		10	3	51.14	[21, 24]
15607723 ^b	Rv0583c	LpqN	1.25	0.76	0.77		3	2	15.79	[23]
15607938 ^a	Rv0798c	cfp29	0.88	1.49	1.40		0	1	5.28	[17]
15608024 ^a	Rv0884c	serC	0.50	0.50	0.67		7	2	9.84	[23]
15608036	Rv0896	glrA2	1.61	2.71	2.14		7	9	32.64	[25]
57116801	Rv0934	pstS1	0.86	0.65	0.76		3	8	45.72	[23]
15608091	Rv0951	sucC	0.68	1.53	1.69		7	1	2.58	[25]
15608123	Rv0983	pepD	0.96	0.55	0.57		7	3	6.90	
15608163	Rv1023	eno	1.03	1.48	1.32		7	1	3.50	
15608219	Rv1079	metB	1.52	1.11	1.08		7	3	12.37	
15608492 ^a	Rv1352	--	0.57	0.49	0.60		10	1	11.38	[23]
15608531	Rv1392	metK	0.69	1.80	0.98		7	4	14.89	[23]
15608586 ^a	Rv1448c	tal	0.16	1.36	1.97		7	1	2.41	[23]
57116925	Rv1844c	gndI	2.06	1.67	2.09		7	2	7.01	
57116926	Rv1860	modD	1.46	1.21	0.99		3	6	31.69	[23]
15609043 ^a	Rv1906c	--	0.68	0.38	0.41		10	1	9.62	[23]
15609069	Rv1932	tpx	0.73	1.10	1.08		0	1	13.33	[23]
15609168	Rv2031c	hspX	0.92	2.43	3.87		0	4	40.28	[21]

\$watermark-text

\$watermark-text

\$watermark-text

CFP proteins with differential levels between strains (p-value < 0.05)									
Ratio against BE (BE= 1.0)									
NCBI Accession number	Rvnumber	Gene	CDC1551	C28	H6	Functional category	Number of unique peptides ^c	% Sequence coverage	References
15609247	Rv2110c	prcB	0.68	0.64	0.66	7	2	12.37	[23]
15609341	Rv2204c	--	0.61	1.23	1.11	10	4	45.76	
15609357	Rv2220	glnA1	0.64	0.66	0.78	7	10	36.40	[23]
15609482	Rv2345		1.30	0.65	0.82	3	3	8.64	
15609513	Rv2376c	cfp2	2.16	1.27	1.12	3	4	56.55	[17, 23]
57116994	Rv2467	pepN	0.98	1.54	1.16	7	4	4.99	
15609630	Rv2493	--	1.51	1.53	1.75	0	2	34.25	
15609858	Rv2721c	--	1.24	1.62	1.07	3	1	5.44	[23]
15610010 ^a	Rv2873	mp83	2.61	1.50	1.10	3	1	19.69	[23]
15610032	Rv2895c	viuB	1.29	1.45	1.66	7	1	7.77	[25]
15610143	Rv3006	lppZ	0.79	0.54	0.62	3	5	26.20	[23]
15610212	Rv3075c	--	1.04	1.69	1.69	10	1	4.89	[21]
15610403	Rv3267	--	0.68	0.55	0.48	10	2	7.83	[23]
57117088 ^a	Rv3312A	--	0.39	0.24	0.45	3	1	26.92	
15610490 ^a	Rv3354	--	0.66	0.49	0.47	10	1	20.93	[23]
15610525	Rv3389c	--	1.31	1.63	1.01	7	2	12.76	[21]
15610639 ^a	Rv3503c	fdxD	0.94	0.69	0.79	7	1	33.33	
15610723	Rv3587c	--	0.87	0.51	0.37	3	4	21.97	[23]
57117145	Rv3682	ponA2	0.82	0.35	0.52	3	2	4.57	[23]
15611010	Rv3874	esxB	1.23	1.74	0.90	3	3	49.47	[21, 23]

All proteins were identified with at least 95% probability at the protein levels, except for

^aProteins identified with at least 84% probability at the protein level (n=18); and

^bProteins identified with at least 79% probability at the protein level (n=2).

^cAll unique peptides were identified with at least 95% probability.