DISSERTATION

PROTEOMIC AND BIOLOGICAL DIVERSITY OF CLOSELY RELATED CLINICAL ISOLATES OF *MYCOBACTERIUM TUBERCULOSIS*

Submitted by

Martha Carolina Mehaffy

Department of Microbiology, Immunology and Pathology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2009

COLORADO STATE UNIVERSITY

November 05, 2009

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY MARTHA CAROLINA MEHAFFY ENTITLED PROTEOMIC AND BIOLOGICAL DIVERSITY OF CLOSELY RELATED CLINICAL ISOLATES OF *MYCOBACTERIUM TUBERCULOSIS* BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate work

Angelo Izzo

John Belisle

Norm Curthoys

Advisor: Karen Dobos

Department Head: Edward Hoover

ABSTRACT OF DISSERTATION

PROTEOMIC AND BIOLOGICAL DIVERSITY OF CLOSELY RELATED CLINICAL ISOLATES OF *MYCOBACTERIUM TUBERCULOSIS*

Recent studies have indicated that some *Mycobacterium tuberculosis* (*Mtb*) strains may be more easily transmitted and successful in causing disease. While this phenotypic diversity exits and is well documented, protein variation between *Mtb* strains appears to be small. Historically, the analysis of virulence determinants in *Mtb* has focused on the comparison of virulent to avirulent *Mtb* strains as well as on the comparison of single gene knock-out mutants to *Mtb* reference strains. Biological and biochemical comparison of *Mtb* clinical isolates with different degrees of virulence, although more complex than those described above, might provide additional information regarding the natural phenotypic variability that exists in *Mtb*. The focus of this study was to determine if a group of closely related clinical isolates of *Mtb* (BE, C28 and H6) as well as the more distantly related *Mtb* CDC1551 present differences in both, virulence in the guinea pig model of TB infection and in the protein expression profiles of secreted and cytosolic fractions. The three closely related *Mtb* isolates comprise the S75 group, which has been reported to be very successful in causing disease. In contrast, *Mtb* CDC1551, although reported to be highly transmissible, is associated with only a few number of active TB cases. To test the hypothesis that closely related *Mtb* strains would have similar

virulence between each other, but higher virulence when compared to *Mtb* CDC1551, guinea pigs were infected with each *Mtb* strain and time-to-death analysis was performed. Proteomes of secreted and cytosolic fractions of these strains were also analyzed and compared by 2D-GE and isobaric tag labeling (iTRAQ) followed by tandem mass spectrometry in order to test whether protein levels of these strains would present statistically significant differences when compared to each other and when compared to *Mtb* CDC1551.

Unexpectedly, differences in both virulence and protein levels between the closely related *Mtb* clinical isolates were identified. Strains *Mtb* C28 and *Mtb* H6 appeared to be more virulent than *Mtb* BE based on analysis of survival curves of infected guinea pigs. Given that differences in virulence between these closely related strains were identified, the analysis of secreted and cytosolic proteomes was focused in the identification of differential proteins between *Mtb* BE and the more virulent strains *Mtb* C28 and *Mtb* H6 that could explain the differences observed in the survival of guinea pigs infected with these strains. Analysis of proteomes by 2D-GE allowed the identification of 13 spots presenting statistically significant differences between at least 2 of the four analyzed strains. However, most of the differences were between *Mtb* CDC1551 and the S75 group, indicating that 2D-GE might not be able to resolve subtle dif differences between closely related *Mtb* strains. A rigorous statistical analysis of iTRAQ data allowed the identification and quantification of 101 and 139 proteins in the secreted and cytosolic fraction respectively. Similar to 2D-GE, most of the differences were observed between *Mtb* CDC1551 and the S75 group. Several differences in protein levels were also observed between closely related strains, especially between *Mtb* BE and both, *Mtb* C28

and *Mtb* H6. Proteins involved in virulence, detoxification and adaptation were more abundant in the more virulent strains C28 and H6. Similarly, enzymes related to metabolic pathways (GltA2, SucC, Gnd1, Eno) and proteins involved in nutrient uptake and storage (BfrB, ViuB, TB15.3 and SseC2) were also present in higher quantity in these strains when compared to *Mtb* BE. In contrast, proteins related to cell wall and cell processes were more abundant in *Mtb* BE. These observations allowed us to construct hypotheses regarding the higher virulence of *Mtb* C28 and *Mtb* H6 in the guinea pig model, including an increased ability for adaptation of these *Mtb* strains to the challenging host environment, a better utilization of energy sources, as well as an advantage for nutrient uptake. Future studies are proposed and discussed to test the hypotheses and to validate our results in an 'in vivo' model.

In a separate study, and based on the fact that the most remarkable proteomic difference between *Mtb* CDC1551 and the other three *Mtb* clinical isolates was a considerable higher amount of an immunogenic protein called Cfp2, preliminary analysis of this protein secretion mechanisms, with an emphasis on its signal peptide, was performed. In addition, possible Cfp2 protein-protein interactions were identified by pull-down assays followed by Tandem MS/MS. As a first step to understand the processing of this protein for secretion, it was demonstrated that a substitution at position -3 of the cleavage site reduced the secretion of this protein. However, mutations in other positions (-1, +1, +2) did not decrease the secretion of this protein when compared to the wild type. Cfp2 co-eluting proteins were identified, including several proteins involved in fatty acid biosynthesis, indicating that Cfp2 might have a role in this cellular process.

In conclusion, the combined use of the guinea pig model and proteomic analysis demonstrated that closely related strains of *Mtb* not only presented differences in virulence, but they also presented several differences in the protein expression levels in both cytosol and secreted protein fractions. The pattern of expression of certain groups of proteins suggests a possible relation to the different degrees of virulence of each of the strains and provides a foundation for further validation of these proteins in the context of *Mtb* virulence.

Preliminary studies on Cfp2 generated protocols for Cfp2 native purification and recombinant production and provided the basis for further studies on *Mtb* signal peptides as well as the possible role of Cfp2 in fatty acid biosynthesis.

Martha Carolina Mehaffy
Microbiology, Immunology and Pathology
Colorado State University
Fort Collins, CO 80523
Fall 2009

ACKNOWLEDGMENTS

I would like to thank my graduate committee members, Dr. Karen Dobos, Dr. Angelo Izzo, Dr. Norm Curthoys and Dr. John Belisle for their helpful advice and constructive criticisms. I would like to give special thanks to my advisor and mentor, Dr. Karen Dobos, for all of her support and guidance during my time as a graduate student, both in and out of the workplace. I will never forget how Karen made me feel welcomed and at home in the first months following arrival after landing thousands of miles away from my country and family. Thank you, Karen. I would like to thank Dr. Curthoys for his advice and time. It really was an honor to have him on my committee. I want to thank Dr. Belisle for the financial support during my first year as a graduate student. I want to thank Dr. Izzo and members of his lab, JoLynn, Kim and Drew who performed the work with the guinea pigs. It was a wonderful help. I thank Dr. Ann Hess who developed the statistical procedure for the analysis of iTRAQ data. She was always very patient responding to dull questions, and receptive to any suggestions. I would also like to thank Dr. Jessica Prenni for her help with mass spectrometry of iTRAQ samples. Finally, I would like to thank my family. I give special thanks to my wonderful husband and best friend, John, for his tolerance and understanding throughout this process. He had the special ability to make me periodically return to reality and see the marvels of life. I thank my children, Nicolas and Emily, for their love and the magic moments in my life. And finally, I would like to thank my parents for their unconditional love and support.

To Nicolas and Emily

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Chapter I. Literature review: Tuberculosis and the tubercle bacillus

1.1. Introduction

It is estimated that one third of the world population is infected with *Mycobacterium tuberculosis* (*Mtb*), the causal agent of tuberculosis (TB). Furthermore, its global incidence is increasing approximately 1% per year with a projection of 8 million new cases each year [1]. In 2007 alone, the World Health Organization (WHO) estimated more than 9 million new cases of TB in addition to 1.16 million cases of relapsing TB [2]. Adding to this dramatic information is the number of deaths due to TB, which reaches the 2 million person mark every year. TB infection also accounted for 26% of all avoidable adult deaths worldwide at the end of last century [3]. For these reasons, TB is considered the second leading cause of infectious disease-related death after HIV/AIDS [4].

In addition to the impact that this disease has on people's lives, TB infection is also very relevant in terms of the global economy. TB cases occur predominantly in the economically most productive age group (15 to 49 years) and in low-income and emerging markets. India, Bangladesh, Indonesia, and Pakistan collectively are responsible for more than 50% of the global TB burden. Sub-Saharan Africa has the highest incidence rate of TB, where the number of cases reported annually (290 per

100,000) doubled in the early 90s [5, 6]. Consequently, the economic burden due to TB infection in the world is approximately 12 billion dollars per year [7].

In industrialized countries, where TB was at some point considered almost eradicated, the infection rates are starting to rise. Some of the reasons for this increase are the influx of migrants from countries with higher TB rates, the HIV co-infection and the appearance of drug resistant strains. Dissemination of HIV infection throughout the world has a direct association with increases in the incidence of *Mtb* as well as non-tuberculous mycobacteria (NTM) infections in many parts of the world [8]. It has been suggested that HIV co-infection accelerates the development, worsens the severity, and amplifies the transmission of TB [9].

For these reasons, basic studies leading to an increase of knowledge on TB, as well as practical studies focused on development of new vaccines, drugs and diagnostic tools are imperative for the control of this disease.

1.2. History of tuberculosis

Even though TB is a leading cause for sickness and death in our modern world, it is not, a new disease. In fact, TB is an ancient disease which has probably killed more people than any other disease in the history of humanity [10].

It has been hypothesized that the genus *Mycobacterium* dates from about 150 million years ago [11]. The early ancestor of *Mtb* is estimated to have appeared in Africa for the first time about 3 million years ago[12]. Modern *Mtb* seems to have originated about 20,000 years ago [13]. *Mtb* infections can only be documented from about 5,000 years ago in ancient Egypt, where Egyptian mummies and different forms of art have been

obtained showing physiological characteristics typical of bony tuberculosis, skeletal disease caused by *Mtb* [14]. The presence of tuberculosis in the ancient world has been further supported by the amplification of *Mtb* DNA from Egyptian [15] and Peruvian mummies [16]. Those studies also confirm the fact that *Mtb* was present in South America before the first European settlers arrived. Recent molecular studies suggest that the evolutionary expansion of human tuberculosis began on the Indian subcontinent, spreading to East Asia and then to other parts of the world [17].

One of the first written descriptions of tuberculosis was likely made by Hippocrates in classical Greece. He not only named the disease phthisis, but also described its clinical manifestations [10, 18, 19]. After this, written reports of tuberculosis in Europe during the middle ages are few. However, there is some archeological evidence that TB was present in Europe [10].

It is in modern Europe that well documented descriptions of tuberculosis began to appear, in addition to an increased understanding of its clinical and physiological manifestations. Aristotle was perhaps the first person to suggest that TB was contagious. Then in 1720, Benjamin Marten, an English physician, suspected the contagious nature of TB and advised that prolonged contact to a phthisic should be discouraged due to the risk for the healthy person to acquire the disease [20]. However, it was not until the 19^{th} century that the infectious nature of TB was first proved. Jean-Antoine Villemin, a French military surgeon inoculated a rabbit with material obtained from a tuberculous cavity. The rabbit presented signs of extensive tuberculosis when it was autopsied three months later [20]. In addition, Rene Theophile Hyacinthe Laennec, the inventor of the stethoscope, elucidated the pathogenesis of pulmonary and extrapulmonary tuberculosis in the early

1800s [21]. At this point, the tuberculosis epidemic in Europe and North America was at its highest peak, presenting death rates of 1,000/100,000 per year [22].

In 1882 the tubercle bacillus was identified and confirmed as the etiologic agent of TB by Robert Koch. He not only earned the Nobel Prize due to his efforts on TB, but also changed the way TB was seen and treated. Koch isolated a substance from the bacilli and called it Tuberculin [10], which later became the first diagnostic test for TB. In 1909 Clemens Freiherr von Pirquet published studies on the use of Tuberculin that was injected intracutaneously to detect latent TB [23]. After this, Florence Seibert developed the purified protein derivative (PPD) which led to controlled studies to understand the tuberculin reaction sizes [10]. It is remarkable that almost a hundred years later the general diagnosis of TB remains for the most part unchanged.

By the mid-19th century, the TB epidemic started to disappear in both North America and Europe [24]. To explain this decline some hypotheses have been suggested, including the improvement of living and nutritional conditions and herd immunity, but the reasons for this drop in tuberculosis rates remain unknown [25].

At the beginning of the 1900s, death rates due to TB were lower but the disease was still considered a major public health concern and efforts to develop a vaccine began to take place. In 1921, Albert Calmette and Camille Guérin developed a live-attenuated *M. bovis* vaccine [26]. Over the next seven years, over 100,000 children were vaccinated with the BCG (Bacille Calmette-Guérin) vaccine and its use was widely spread in Europe. By the mid 1900's, PPD testing and BCG vaccination was disseminated worldwide, comprising the first disease control program by the WHO [10]. Even though BCG vaccination has been shown to have a very wide range of efficacy [27-29], it is still the only vaccine

available for use in humans and with more than 120 million doses administered worldwide, and a key component of TB control in most countries [29].

Tuberculosis keeps challenging physicians and scientists around the world. Today the HIV co-infection and the development of multidrug (MDR-TB) and extensively-drug resistant TB (XDR-TB) make tuberculosis a major public health issue again. Recent estimates indicate that in 2003 there were 458,000 incident cases of MDR-TB globally, including new and retreatment cases [30]. Prevalent cases may be two or three times higher than incident cases, in addition a much larger number of MDR-TB latently infected individuals may exist [31].

The emergence of XDR-TB threatens to return TB to the pre-antibiotic era when fifty percent of patients died of the disease [32].

TB continues to be a major public health concern especially in developing countries, with the lowest gross national products producing countries in Africa, Asia and Latin America being at the highest risk. Therefore, it is imperative to continue scientific research on *Mtb* focused on improving diagnosis and treatment, including MDR-TB and XDR-TB. Basic research focused on new drug targets and virulence determinants, taking advantage of available clinical isolates, are an important part of more practical applications and should also be considered a priority on TB research. Otherwise, we will continue to lose thousands of lives to the hands of the "white plague".

1.3. The genus Mycobacterium

The genus *Mycobacterium* was first described by Lehmann and Neumann in 1896 when *Mtb* (bacterium tuberculosis) and *M. leprae* (bacterium leprae) were grouped together

[33]. Genus *Mycobacterium* belongs to the Mycobacteriaceae family, order Actinomycetales and class Actinomycetes. In the mid 90s, there were approximately 75 mycobacteria species [34]. This number has grown fast and almost doubled. By the mid 2000s more than 130 species had been identified [35]. Members of the genus share the following characteristics: 1. acid-alcohol fastness, 2. presence of mycolic acids, containing 60-90 carbons that are cleaved to C22-C26 fatty acid methyl esters by pyrolysis and 3. G-C content of 61-71% [34, 36].

Species within the genus *Mycobacterium* are divided into Tuberculous and nontuberculous mycobacteria (NTM). NTM comprise all mycobacteria species with exception of the *Mtb* complex (MTC) and *M. leprae* [37]. Species of the MTC are all slow growers and *M. leprae* is the only non-cultivable mycobacteria. In contrast, NTM species can be subdivided based on growth: rapidly or slowly growing [34]. NTMs that present slow growth are the *M. avium complex* (MAC), in addition to *M. ulcerans, M. marinum, M. microti,* and *M. kansassi*. Some of the rapid growers that are classified as pathogenic are *M. abcessus*, *M. chelonae* and *M. fortuitum* among others [34].

M. leprae is the causative agent of Hansen's disease, also known as leprosy. This mycobacterium is an obligate intracellular parasite. It has been postulated that its requirement of host macrophages in order to grow and survive is due to a minimal genome, which is not only reduced in size, but also in its number of functional genes [38, 39].

Pathogenic members of the NTM group cause pulmonary and nonpulmonary infections, especially in immunocompromised individuals, with the MAC being the most important in terms of clinical relevance [40]. Species in this complex are *M. avium*, *M.*

intracellulare and *M. scrofulaceum* among others. In addition, *M. avium* contains some subspecies such as *M. avium subsp avium* and *M. avium subs paratuberculosis* in addition to others.

In addition to MAC, some other NTM that cause disease in humans are *M. malmoense*, *M. xenopi*, *M. simiae*, *M. genavense*, *M. celatum*, *M. kansassi*, *M. asiaticum*, and *M. szulgai* [34, 37]. In addition to presenting a challenge for accurate diagnosis and treatment, direct interaction with HIV is also an issue. For example, *M. xenopi* was recently found to increase HIV replication in vitro [41]. MAC species and *M.xenopi* are the most frequent NTM that cause pulmonary disease [42] and is therefore important to be able to differentiate them from *Mtb* to initiate prompt treatment.

Finally, some mycobacteria species cause skin infections including *M. haemophilum*, *M. marinum* and *M. ulcerans*. The latter is the causative agent of Buruli ulcer disease which is characterized for ulcerative and progressive skin lesions and is mostly found in tropical countries [43].

TB is generally caused by *Mtb*; though other species within the MTC, such as *M. canettii*, *M. africanum*, and *M. bovis*, are also known to cause TB in humans [44, 45]. Recently, additional species have been integrated into the complex, including, *M. pinnipedii*, which causes tuberculosis in seals [46] and *M. caprae*, which is very close to *M. bovis*, but is sensitive to pyrazinamide and is isolated primarily from goats [47]. In addition to these species, *M. microtti*, known to cause disease in rodents, also belongs to the MTC [48].

M. canettii, is a very rare species and has been linked to tuberculosis cases only in western Africa [45]. Molecular analyses based on genetic polymorphisms in the *hsp65*

and *gyrB* genes suggest that *M. canettii* is, in evolutionary terms, much older than other species within the MTC [49, 50].

Mtb and *M. africanum* primarily infect humans. However, infections in animal, although rare, have also been shown [51, 52]. In contrast, *M. bovis* and *M. caprae* have a very broad range of hosts, causing disease in wild and domestic animals in addition to humans [50, 53, 54]. Tuberculosis caused by these mycobacteria is not common and prevalence rates vary between countries. In Latin America, it is estimated that *M. bovis* causes 0-2.5% of all TB cases [55]. In the United States, the prevalence seems to be around 1.4% [56]. However, appropriate media to isolate *M. bovis* is not routinely use in many setting, and therefore, the estimated number could be higher [55].

1.4. Structure of Mycobacterium tuberculosis

The cell envelope of mycobacteria is a complex structure that differs significantly from counterpart structures of both Gram-negative and Gram-positive bacteria. It consists of three major entities: the plasma membrane, the cell wall core and the outermost layer. This last one is sometimes referred as capsule, as in the case of pathogenic mycobacteria such as *Mtb* [57].

The cell wall core is composed of peptidoglycan (PG), a structure that is common to most eubacteria and is responsible for the size and shape of the cell. However, the surface of *Mtb* is rich in lipids and carbohydrates with an outermost capsular layer composed primarily of the polysaccharides arabinomannan (AM) and glucan [58]. In addition, muramic acid residues, which are a central component of other bacteria cell wall, are not present in mycobacteria. Instead, they are replaced by a complex polysaccharide, AG

(arabinogalactan) which is attached to PG through a unique linker unit and is acylated at its distal end to PG with mycolic acids [59, 60].

The entire complex is abbreviated as the mAGP (mycolylarabinogalactan–peptidoglycan) and is essential for viability in *Mtb* and other mycobacteria [61]. Moreover, mycobacteria unique cell wall structure is responsible for their particularly low permeability and resistance to common antibiotics [62], in addition to dehydration resistance and the ability to survive in the hostile macrophage environment [63, 64].

Mycolic acids form a large family of related lipids that are unique to mycobacteria and specific types of mycolic acids are usually related to particular mycobacteria species [60]. *Mtb* and *M.bovis* produce three different types of mycolic acids: alpha mycolates, methoxymycolates and ketomycolates [65-67]. In addition, pathogenic slow growing mycobacteria produce significant amounts of cyclopropanated mycolic acids which in *Mtb* has been recently associated with cell viability and drug resistance [68].

Similarly, the carbohydrate arabinan is an important component of mycobacteria structure. It is not only part of AG, but it is also part of another unique element of mycobacteria, Lipoarabinomannan (LAM). LAM is composed of a phosphotidil-myo-inositol mannan (PIM) core which is entirely embedded within the cell wall. However, the arabinan of LAM has been demonstrated to be exposed on the surface and is directly implicated in the immunopathogenesis of leprosy and tuberculosis [57].

It has been shown that virulent strains of *Mtb* as well as *M.bovis*-BCG show extensive mannose capping of the arabinan termini of LAM (Man-LAM), while opportunistic rapidly growing species, such as *M. chelonae* lack these mannose residues [69]. Mannose caps in *M. smegmatis* are replaced by inositol phosphate [70]. Man-LAM has a reduced

ability to stimulate monocyte chemotaxis and has been suggested as one of the mechanisms by which *Mtb* evades early innate immune responses [71]. Furthermore, the presence of mannose caps on LAM has also been implicated in the binding to human macrophage mannose receptors [72] and lung surfactant proteins [73], thereby affecting the adherence of whole bacterium to macrophage host. In addition, LAM seems to be directly involved in inhibition of phagosomal maturation [74, 75].

The main component of the *Mtb* capsule is glucan; nonetheless, other carbohydrates, especially arabinose and mannan are also present. In addition to carbohydrates, proteins are part of the capsule and some of them have been identified. Surprisingly, lipids are not a significant component of this outer envelope [76, 77].

Inside the cell wall skeleton is the plasma membrane, which is supposed to be very similar in structure and function to other bacterial plasma membranes [60], yet, only few studies have been performed to address the particularities of *Mtb* cell membrane. Rezwan et al (2007) analyzed the different subcellular fractions of *M. smegmatis*. In the fraction corresponding to the membrane, they found different classes of characteristic phospholipids of mycobacteria, such as phosphatidyl inositol mannosides (PIMs), phosphatidyl glycerol (PG) and phosphatidyl ethanolamine (PE). They also showed that this fraction contains mainly nonhydroxylated fatty acyl residues and only trace amounts of mycolates [78].

Mycobacteria and phylogenetically related bacteria of the suborder Corynebacterineae, unlike other gram-positives, contain porins or pore-forming proteins that facilitate the entry of small molecules through the permeability barrier composed of the cell wall mycolates and other lipids [57, 62]. In *M. smegmatis* the main porin has been

characterized and called MspA [79]. In *M. bovis* and *Mtb* a channel-forming protein called OmpA has been identified [80, 81]. Recently, an additional channel forming protein (Rv1698) that is likely involved in transport processes across the outer membrane of *Mtb* was recognized [82]. Additionally, more than 140 putative outer membrane proteins (OMPs) have been identified through bioinformatics analysis of *Mtb* secreted proteins, and two of them (Rv1695 and Rv1973) have been characterized [83]. For the most part, this subject remains open for study.

Finally, it has been recently suggested that *Mtb* possesses a pili-like structure composed of subunits encoded by the Rv3312A gene. Sera from TB patients react against Rv3312A protein which suggests that this structure is produced during human infection [84]. However, more studies are needed to confirm the production of this structure during in vivo infection as well as its significance for *Mtb* pathogenesis.

1.5. TB infection

One of the first documents describing the immunological course of the tuberculosis infection was by Wallgren (1948) who divided the infection into four different stages [85].

The first step on tuberculosis infection begins with the inhalation of droplets containing viable *Mtb*. Once the particles of 1 to 5 μm in diameter are inhaled and phagocytosed by resident alveolar macrophages (MΦ), *Mtb* can be killed or inhibited depending on the state of activation of these MΦs. More recent studies have shown that dendritic cells (DC) are an important part of the initial response to *Mtb,* traveling to the lymph nodes

and presenting antigens to T-cells, which is one of the first steps of the cell mediated immune response (CMI) [86].

Attachment of the mycobacteria occurs via recognition by pattern recognition receptors (PRRs), such as Toll-like receptors (TLR), on the surface of MΦ and DCs of distinct *Mtb* surface molecular structures known as pathogen associated molecular patterns (PAMPs). The resulting influx of neutrophils, MΦ , Natural Killer (NK) cells and other cells to the site of infection, as well as the production of cytokines and chemokines, serve as a stimulus for granuloma formation and act directly to limit the extent of mycobacterial replication at the early stage of infection [87]. NK cells are also a key component during the first stage of TB infection where they are probably the most important source of IFN-γ and therefore can activate macrophages before CMI is established and T-cells take over this role. NKs can also kill intracellular bacilli by mediating the apoptosis of infected cells and are involved in the direct killing of the bacilli. The exact mechanism of both apoptosis and direct killing remains to be clarified [88]. It has been suggested that the efficiency of these early innate responses may help to determine whether a latent infection is established and whether that infection ultimately will progress to active disease [89, 90].

After these initial steps, MΦ incapable of killing the mycobacteria are disrupted, causing an influx of blood monocytes and other inflammatory cells to the lung. These monocytes differentiate into MΦ, and bacilli are again ingested. Within 2 to 6 weeks after infection, a process called Delayed Type Hypersensitivity (DTH), also known as the tissue damage response, is established. DTH is a type of CMI characterized by destruction of infected, non-activated MΦ and nearby tissue causing necrosis. DTH seems to be required during

the first stages of the disease to control infection, however, it can also be the cause of excessive damage and immunopathology of TB, leading to excessive necrosis and cavity formation [91].

Final containment of the disease is mediated by the establishment of protective CMI, which, in contrast to DTH, requires a relatively large concentration of antigen and it is believed to induce long-term protection. Protective immunity is characterized by granuloma formation. *Mtb* infected DCs present antigens to CD4+ T-cells in the lymph nodes and in conjunction with interleukin (IL)-12 secretion, induce them to differentiate into Th1 cells. Then differentiated T-cells migrate to the site of infection where they contact MΦs and activate them in an interferon (IFN)-γ dependent manner. As more cells come to the site of infection, a general granuloma structure begins to form. Infected cells can fuse together forming giant cells while lymphocytes and activated MΦs surround them and other infected MΦs, preventing the escape of bacilli from the center. Tumor necrosis factor (TNF)-α has an important role during granuloma formation since it is involved in recruitment of new monocytes to the site of infection.

In general, CD4+ T-cells are believed to play a more important role in control of TB infection in comparison to CD8+ T-cells. However, both are present in the lymphocyte layer of the granuloma and both are responsible for IFN-γ production. CD8+ T-cells are believed to play a more active role during persistence while CD4+ T-cells are more involved during the acute infection [87, 92].

Infected MΦs are contained in the center of the granuloma, where they eventually die forming a solid caseous center. At this point the number of viable bacilli becomes stationary because more MΦs become activated, thus killing the bacteria. Based on the

important role that oxygen plays in *Mtb* replication, several authors have suggested that the environmental conditions in the caseous center may also contributed to the inhibition of *Mtb* growth [93-95]. On the other hand, in 10% of presumed immunocompetent individuals, the infection is not contained and continual bacillary replication results in disease symptoms and associated pathology, including tissue necrosis and cavitation [87, 96]. In this case, bacilli escape the edge of caseous necrosis, infecting and replicating in non-activated MΦs which results in host cell necrosis producing an enlargement of the caseous necrotic center. This, in turn, results in progression of the disease and clinical signs.

Finally, liquefaction of the caseous center occurs and bacilli are now able to multiply extracellularly and reach high numbers. Extended necrosis is produced leading to erosion of the bronchial wall and cavity formation. At this point, the bacilli have access to the airways and the patient becomes infectious, the bacillus becomes transmissible and the cycle of the disease can start again in a different person.

Although the immune status of the host plays an essential role on the outcome of tuberculosis infection, several factors from *Mtb* itself can interfere with some of the host anti-mycobacterial mechanisms and stimulate mycobacteria survival. Several mycobacterial components have been associated with, or are essential for complete virulence of *Mtb*. A list of genes required for growth has also been obtained by using a transposon site hybridization technique (TraSH) [97, 98].

Many of the genes associated with virulence or required for appropriate mycobacteria growth have been identified and characterized using single gene knockout mutants. The relationship with the host cell resulting from the expression of these genes or other cell

components is largely unknown. Some of these factors are extensively reviewed by Smith (2003) [99]. For a number of factors, their association with the host-immune response or mycobacteria physiology in the context of infection has been revealed and some of them are listed in Table 1-1.

Table 1-1. Virulence factors of *Mtb* and their association with the host immune response.

Virulence factor	Function within the bacillus	Association to host immune responses	Reference
ManLAM	Mannose-capped Lipoarabinomannan Cell envelope structure	Involved in uptake of the bacilli by MΦ , inhibition of phagosome maturation, inhibition of immunomodulatory responses	[100-102]
Sulfolipid 1(SL-1)	Anionic glycolipid, cell envelope structure	Inhibit granuloma formation and TNF-alpha production.	[103]
Phenolic glycolipid (PGL-Tb)	Cell envelope structure	Inhibition of inflammatory cytokines	[104]
Threhalose dimycolate mannoside (TDM)	Neutral glycolipid, cell envelope structure	Intracellular survival	[105]
19kDa antigen	Lipoprotein LpqH	Decrease antigen presentation to T cells	[106, 107]
SapM	Acid phosphatase	Hydrolyzes host phosphatidyl inositol 3-phosphate (PI3P) resulting in inhibition of phagolysosome fusion	[108]
PknG	Eukaryotic-like serine/threonine kinase	Inhibition of phagolysosome fusion	[109]
DosR/DosS -DosT	Two component system, gene regulation	Involve in nitric oxide and hypoxic stresses responses	[110, 111]
HspX	Heat shock protein	Involved in hypoxic responses, linked to the ability of the bacteria to arrest growth and become dormant	[112]

SodA	Superoxide dismutase	Responds to superoxide stresses, inhibit innate immune responses	[113]
Icl1 and Icl2	Isocitrate lyases, involved in the glyoxylate cycle	Intracellular survival (Involved in the ability of the mycobacteria to survive on fatty acids as the only carbon source)	[114]
Mycobactin, carboxymycobactin	Iron uptake	Nutrient uptake, Intracellular survival	[115, 116]
MbtB	Involved in synthesis of carboxymycobactin and mycobactin	Nutrient uptake, Intracellular survival	[99]
Esat-6/Cfp10	Secreted antigenic proteins	Inhibits TLR signaling in macrophages	[117-120]
Mce1	Macrophage cell entry protein	Involved in uptake of the bacilli by nonphagocitic cells and intracellular growth	[121, 122]
HbhA	Heparin-binding haemagglutinin	Adhesion to epithelial cells and dissemination	[123, 124]

1.6. Diagnosis and treatment of tuberculosis

One of the major issues for TB control is the lack of accurate diagnostic tests for both TB infection and drug resistance. Advances in this field will be discussed in the following sections.

1.6.1. Diagnosis of active Tuberculosis

In many parts of the world, especially in developing countries, active TB is diagnosed based on clinical presentation, in addition to Ziehl-Neelsen stains from sputum samples and in some cases X-rays [55, 125]. However, two major disadvantages of microscopy are the potential contamination of the specimen with environmental mycobacteria and the

high rate of false negatives which in some cases can be as much as 50% [126, 127]. This low sensitivity is also a major concern in patients with HIV/AIDS [128, 129].

In more advanced laboratories, cultures are performed. However, the time to obtain a positive culture can vary from more than 7 days to a month [126]. When liquid media containing specific markers such as BACTEC™ MGIT-960 is used, results can be obtained much faster than regular culture in solid media but contamination with fast growing bacteria is a major drawback for this method [126]. Some advantages of in vitro culture include the ability to test for drug resistance and to identify mycobacteria at the species level. Both of these important procedures require additional time, and in some cases by the time results are available, the patient has already initiated treatment [130], which in drug resistant cases is far from being appropriate.

Since the development of the polymerase chain reaction (PCR), several studies have been performed in an effort to develop a more accurate and rapid diagnostic test for tuberculosis [131-135]. Nucleic acid amplification tests (NAAT) provide higher sensitivity and faster results than traditional diagnostic tests. With the advances in point-mutation identification and development of multiplex PCR, these tests not only provide information about drug resistance status by targeting the genes involved in drug resistance mechanisms, but also differentiate between mycobacterial species in one single test [136-138]. Two commercial tests have been recently approved by the Federal Drug Association (FDA) [139] and are available, they are the Amplicor MTB test (Roche diagnostics System) and the Amplifier Mycobacterial Tuberculosis Direct test (Gen Probe, Inc San Diego).

NAATs have a disadvantage due to the requirement of specialized training, high expenses for reagents, and expensive equipment. This makes their suitability to be implemented in developing countries very low. In addition, certain studies have shown that the reproducibility of these methods is not as high as expected and the sensitivity on AFB- (Acid fast bacilli) and extra-pulmonary samples is low [139, 140].

1.6.2. Diagnosis of latent Tuberculosis

Identification of individuals that are at risk of developing TB is an important step for the control of TB. In addition, in immunocompromised patients, early diagnosis of latent TB, followed by appropriate treatment can increase their chances of survival [141]. Diagnosis of latent TB relays mainly on the Tuberculin Skin Test (TST) which is based on the delayed type hypersensitivity (DTH) reaction to the PPD.

TST is relatively cheap and easy to perform. However, due to the nature of PPD, which contains several hundred antigens, TST is not a reliable method for detecting latent TB in countries where BCG vaccination is performed or where environmental mycobacteria are endemic, since both can lead to false positive results [142, 143]. Unfortunately, these factors are present where TB is a major health problem; therefore testing to treat latent TB in these countries is very challenging. False negative results are also an important concern in HIV patients and children, and are directly related to their CD4+ T lymphocyte counts and potentially impaired T-cell responses respectively [144, 145].

In an effort to overcome these issues, other tests have been developed. The IFN-γ release assays (IGRAs) mimics a DTH response 'in vitro'. Peripheral blood mononuclear cells (PBMCs) contained in whole blood samples are exposed to mycobacterial proteins, then,

if T-cells have been sensitized due to TB infection, they release IFN-γ which is measured by enzyme-linked immunosorbent assay (ELISA) [127, 146]. The most accepted antigens to use in this assay are Cfp10 and Esat-6, which are *Mtb* specific and therefore absent in the *M.bovis* BCG strain and environmental mycobacteria, decreasing the number of false positives when compared to TST [147-149]. However, individuals and contacts infected with *M. africanum*, which accounts for almost half of all TB infections in West Africa [150], are less likely to respond to Esat-6 and other RD-1 (Region of Difference 1) antigens [151]. IGRAs also have the potential for false negatives in HIV/AIDS patients [152].

Nevertheless, due to its higher specificity in comparison to TST, three tests based on IFN-γ detection after sensitation of T-cells ex vivo are commercially available. QuantiFERON®-TB (Cellestis Limited, Carnegie, Australia) was approved for use in the United States in 2001 [153]. However, this test uses PPD as the antigen, therefore, nonspecific reactions can occur similar to that observed with TST. To improve, a second generation test, QuantiFERON®-TB-Gold was developed. This test uses antigens Esat-6 and Cfp-10 and was approved for use in USA in 2004 [154]. Another test, QuantiFERON®-TB-Gold 'in tube', minimizes sample handling and in addition to Esat-6 and Cfp10, antigen Tb7.7 (Rv2645) is added to the cocktail. This test presents a high specificity and also the advantage of being performed in a single patient visit with results available within 24h after the specimen is collected. Unfortunately, this test has not been approved for use in the United States. In contrast, an assay in which the number of IFN-γ producing cells can be quantified by enzyme-linked immunospot (ELISPOT) [143, 155] has been recently approved for use in the United States. The T SPOT.TB® assay (Oxford

Immunotec, Oxon, UK) has the same fundamentals as Quantiferon tests, using Esat-6 and Cfp10 as antigens; however, to increase the sensitivity, a plate is previously coated with anti-IFN-γ monoclonal antibodies, to avoid the problem of cytokine consumption during culture. In addition, it is possible to relate production of IFN-gamma to the number of sensitized T-cells [155].

Finally, detection of additional biomarkers in addition to IFN- γ has recently been shown to be useful to differentiate between active and latent TB [156] and might provide the basis for better and more accurate tests.

1.6.3. Treatment of Tuberculosis infection

Due to its unique cell wall, *Mtb* has an inherent resistance to several drugs, making this bacillus a difficult target for treatment. The streptomycin (SM) discovery in 1944 [157], followed by that of Isoniazid (INH) in 1954 [158], the rifammycins in 1957 [159] and Ethambutol (EMB) in 1962 [160] changed the history of TB. Drug treatment improved prognosis and increased survival rates of TB patients.

The current regimen for TB treatment which was designed in an effort to increase efficacy and prevent the emergence of resistant organisms, is comprised by 2 months of INH, Rifampin (RIF), EMB and Pyrazinamide (PZA), followed by 4 months of INH and RIF [20]. EMB can be discontinued if drug susceptibility tests show that the strain involved in infection is pansensitive [161]. The most common regimen in the United States consists of daily antibiotics administration for the first 2 weeks, followed by two or three times per week after that [161, 162]. However, the WHO recommends daily uptake during the entire duration of treatment [163].

Treatment of latent TB is controversial due to the low efficacy of antibiotics on non-replicative cells and the possibility of drug resistance appearance. Nonetheless, INH preventive therapy is recommended by the WHO and the Joint United Nations Program on HIV/AIDS (UNAIDS) [164] and some studies indicate that treatment with isoniazid for 6 months in HIV positive patients can decrease the incidence of active TB by 42 - 60% [165, 166].

Tuberculosis requires a long treatment period and secondary effects from flu-like symptoms to serious hepatotoxicity can occur. Consequently, patient compliance to treatment is poor which increases the risk of reactivation and/or generation of drug resistant TB.

MDR-TB strains are those that have resistance to both INH and RIF, while XDR-TB strains present additional resistance to any fluoroquinolones plus one of the second-line injectable drugs (amikacin, capreomycin, or kanamycin) [167].

Both MDR-TB and XDR-TB have been associated with very high morbidity and mortality, prolonged treatment to cure, and an increased risk of spreading drug-resistant isolates in the community [32, 168-170].

Treatment of drug resistant TB is even longer and requires the implementation of second-line drugs as such as streptomycin, capreomycin, kanamycin, amikacin, ethionamide, para-aminosalicylic acid, cycloserine, ciprofloxacin, ofloxacin, levofloxacin, moxifloxacin, gatifloxacin and clofazimine. Unfortunately, these drugs are more expensive, less efficacious and less tolerable [171, 172].

In an effort to minimize the appearance of *Mtb* resistant to drugs, Directly Observed Treatment – Short course (DOTS), has been implemented. The DOTS program includes

supervised administration of antibiotics by trained personnel and has been proven to be one of the most cost-effective global health interventions available today [173]. DOTS has been implemented in 184 countries [174] but its level of implementation vary as it is very demanding for patients and for health care staff [172].

Several new drugs that could be more effective are in different stages of clinical trials. If approved, they would not only improve TB treatment, but could have a major impact on MDR-TB and XDR-TB control [161].

Even though these interventions have had some success, the rates of MDR-TB and XDR-TB have increased within the last years [175]. Therefore new and better drugs and vaccines are needed to achieve complete control of TB.

1.7. Molecular epidemiology of TB

The development of molecular techniques that allow genotyping of clinical and laboratory isolates, including those that are currently circulating has bolstered TB research efforts. It has been suggested that genotyping of *Mtb* increases our knowledge of moving and expanding TB clones which will help define better control measures for the disease [176].

Current circulating strains fall into several major lineages, or clades. Filliol et al (2003) described 36 clades and 9 major superfamilies including *M. africanum*, Beijing, *M. bovis*, East-African-Indian (EAI), Central Asian (CAS), T, Haarlem, X, and Latin American and Mediterranean (LAM) [177]. However, most *Mtb* strains fall into six major clades, including Beijing, Haarlem, Africa, East African Indian (EAI) and Latin American and

Mediterranean (LAM) [178]. All of these strains are present in East Africa; however their global distribution varies [179].

Transposable and repetitive elements appear to be the main source of genomic diversity in *Mtb*. Therefore, several methods based on the study of these elements have been developed for typing of *Mtb* strains. These include: IS6110-based restriction fragment length polymorphism genotyping (RFLP) [180], spacer oligonucleotide typing (spoligotyping) [181], and more recently, mycobacterial interspersed repetitive units-variable number tandem repeats (MIRU VNTR) typing [182] and Single Nucleotide Polymorphisms (SNPs) [17].

IS6110-RFLP is the most widely used and accepted method of generating genotypes due to its reproducibility and relatively straightforward procedure and has been standardized as the 'gold standard' among molecular biology laboratories [183]. However, it has been shown that IS6110-based RFLP typing does not allow the discrimination between *Mtb* isolates with only one or two IS6110 copies [184]. Therefore, studies involving low-copy-number isolates use additional genotyping techniques to better describe relatedness among *Mtb* strains [185].

SNPs represent another source of genetic diversity and have been shown to be useful markers for evolutionary and epidemiological studies in clinical *Mtb* [176]. For example, SNPs at codon 463 of the *kat*G gene and at codon 95 of the *gyr*A gene are the basis for the classification *of Mtb* isolates into the three principal genotypic groups [13]. SNPs have been used to investigate the associations between these genotypic groups and the infectivity and pathogenicity of *Mtb* strains [186].

SNPs have also been shown to be species-specific. An SNP in the *nar*GHIJ promoter region appears to be specific for *Mtb* [187], one in *oxy*R is specific to *M. bovis* [188], and one in *mgt*C allows the discrimination between Haarlem vs non-Haarlem genotypes [176]. An SNP in *pho*P of *Mtb* H37Ra has been shown to be related to the avirulent phenotype of this strain in relation to *Mtb* H37Rv [189].

Molecular typing using the MIRU approach was used to study the evolutionary genetics of the MTC using clinical isolates from diverse geographic regions. This technique corroborated findings obtained by spolygotiping where *M. canettii* appeared to be the first branch to have diverged from a common MTC ancestor, followed by the *Mtb* East African Indian (EAI) clade. Then, a third branch separates a *M. africanum–M. bovis* clade, followed by a node separating *Mtb* Beijing versus non-Beijing. Among non-Beijing *Mtb* strains, branches such as the Latin American and Mediterranean (LAM), X and Haarlem clades diverged later [190].

In addition to the specific examples describe above, molecular genotyping has been used to study evolution and geographic distribution of TB strains [191, 192], as well as a tool to identify transmission chains [184], characterize isolates during an outbreak [193] and to discriminate between re-activation or re-infection episodes and relapse after treatment [194-196]. Thanks to DNA fingerprinting it has been recognized that most TB cases are due to ongoing and recent transmission [197, 198], instead of reactivation, as was thought previous to the introduction of this technique [199]. This knowledge itself has had an impact on TB control strategies.

1.8. Importance of clinical and laboratory strains in the study of TB

To date, most knowledge of host and pathogen factors involved in TB infection have been achieved through the study of *Mtb* infection in animal models and the use of *Mtb* mutants with impaired phenotypes for in vitro or in vivo growth.

However, the study of currently circulating *Mtb* clinical isolates has increased our understanding of how the pathogen interacts with the host, resulting in different disease outcomes and transmission patterns. Clinical isolates have been shown to have different survival and replication rates in whole blood cultures and human monocytes, showing a relationship between bacterial growth and innate host immunity [89, 200]. Moreover, using the mouse model, Dormans et al (2004) have shown differences in survival, lung pathology, bacterial load, and DTH response between nineteen *Mtb* strains comprising eleven major genotype families [201].

Study of immune responses to predominant strains that are responsible for major outbreaks is important in order to determine the mechanism by which these strains are more successful in causing disease. Studies have shown a link between hypervirulence of *Mtb* strains and lack of development of Th1 immunity in murine models. As a result, depressed Th1 and elevated Th2 immune responses are associated with the virulence of some mycobacterial strains [104, 193, 202-204].

Lipids from different *Mtb* clinical isolates have been studied in the context of stimulation of innate immunity. Interruption of Sulpholipid 1 (SL-1) synthesis attenuates *Mtb* infection in mice and is associated with reduced production of cytokines such, as IL-10, which is involved in the regulation of the Th1 response [205].

Similarly, LAM from avirulent *Mtb* H37Ra, but not from virulent Mtb Erdman strain, induces the production of TNF-α in murine bone marrow-derived macrophages [206] suggesting that the host immune response is dependent on the strain virulence and emphasizes the importance of cell wall lipids in the biology of the bacilli.

Some of the *Mtb* strains (laboratory and clinical) and the immunological studies associated with them are described below.

1.8.1. H37Rv

Mtb H37Rv is the most commonly studied laboratory strain. It was first isolated in 1905 and showed high virulence in the guinea pig model. Since that time, this strain has been continually passaged. In 1934, strain H37 was dissociated into "virulent" (Rv) and "avirulent" (Ra) strains [207].

Although *Mtb* H37Rv has maintained virulence in animal models, its virulence in humans is unknown and due to the continuous passages, there have been concerns as to whether this may have led to attenuation of virulence [208, 209]. In this regard, some studies have shown that *Mtb* H37Rv has significantly less growth in human MΦs than clinical isolates with minimal in vitro passages [89, 208]. Recent studies have also shown that there is considerable heterogeneity in *Mtb* H37Rv clones, which results from subculturing and seems to be associated with mutations that affect virulence indicators such as neutral red and PDIM (phthiocerol dimycocerosate) [210].

Studies using murine MΦs infected with *Mtb* H37Ra vs *Mtb* H37Rv have shown marked differences in the immune response to these strains. The growth of *Mtb* H37Ra is controlled in the lungs of immunocompetent mice, whereas *Mtb* H37Rv grows

extensively at this site [211]. It seems that MΦs infected with the avirulent strain *Mtb* H37Ra are able to control bacteria growth by the induction of Th1-type cytokines and chemokines. On the other hand, *Mtb* H37Rv induces an upregulation of Th2-type cytokines, including interleukin-5 (IL-5), IL-10 and IL-13, resulting in the inability of macrophages to control the growth of this strain [212]. More recently, the potential role of IL-23 in TB infections was suggested after comparison of alveolar MΦs infected with *Mtb* H37Rv or *Mtb* H37Ra. MΦs infected with *Mtb* H37Rv produced higher quantities of IL-23 than did MΦs infected with the avirulent strain [213].

Differences inherent to the bacilli have been found between *Mtb* H37Ra and *Mtb* H37Rv that might explain some of their biological behavior. *Mtb* H37Ra lacks the expression of virulence-associated lipids such as sulfolipid (SL), diacyltrehaloses (DAT), and polyacyltrehaloses (PAT) along with several genes implicated in virulence [214, 215]. Some of these differences seem to be associated to a point mutation in *Mtb* H37Ra *phoP* gene, which is the response regulator of the two component system PhoPR [189, 216, 217].

Comparison of *Mtb* H37Rv to the *Mtb* K-strain (most prevalent strain in Korea) has shown marked differences that could be related to a less virulent phenotype in *Mtb* H37Rv. Apparently, *Mtb* H37Rv induces higher levels of protective cytokines (TNF-α, IL-6, IL-12p40) and apoptotic markers than *Mtb* K-strain, resulting in diminished necrotic cell death [218]. An increase production of protective cytokines (TNF-α, IFN-γ) by *Mtb* H37Rv was also observed when this strain was compared to clinical isolates belonging to the Central Asian Strain 1 (CAS1) and Beijing families [219]. It seems then,

that *Mtb* H37Rv is fully virulent when compared to completely avirulent *Mtb* strains; however, it presents a lower virulence in relation to strains freshly isolated from patients.

1.8.2. CDC1551

Mtb CDC1551 was isolated from an outbreak on the Kentucky-Tennessee border between 1994 and 1996. It was initially described as hypervirulent in humans due to high rates of PPD conversion among individuals exposed to patients with active disease. Supporting this idea, initial analysis of *Mtb* CDC1551 showed 100-fold higher numbers of bacilli compared to the numbers of bacilli isolated from the lungs of mice infected with the *Mtb* Erdman laboratory strain [193]. However, subsequent studies using the mouse and rabbit models indicated that *Mtb* CDC1551 is, in fact, less virulent than other clinical and lab *Mtb* strains, showing prolonged survival in mice after infection [220] and slower growth in the rabbit lung [221, 222]. During mixed mouse infections with the hypervirulent *Mtb* strain HN878, *Mtb* CDC1551 grew more slowly and the recovery post-infection was poor [223].

Immunologic aspects of *Mtb* CDC1551 have also been studied. In the mouse, *Mtb* CDC1551 induces granulomatous differentiation in the lungs at an earlier time point than *Mtb* HN878, *Mtb* H37Rv and *Mtb* Erdman strains. It also induces relatively high levels of Th1-associated cytokines in infected mouse lungs and in human blood monocytes in vitro.

Since Th1 immunity is critical for controlling TB infection, the strong induction of this host response by *Mtb* CDC1551 may serve to protect the host from disease progression

[220, 224]. Thus, in the mouse model, this clinical isolate has been categorized as both hypovirulent and hyperimmunogenic [223].

This differential response was shown to be directly attributable to the extractable polar lipid constituents of CDC1551 cell wall indicating that the lipid-associated immunogenicity of CDC1551 may contribute to its attenuated virulence [224].

1.8.3. HN878

Isolate *Mtb* HN878 was the causative agent of an outbreak in Texas between 1995 and 1998. This strain is associated with an unusually high proportion of active cases of disease and a high frequency of extrapulmonary disease [13]. *Mtb* HN878 has been shown to cause rapid progression to death in mice in comparison with other clinical isolates and standard laboratory strains *Mtb* H37Rv and *Mtb* Erdman [203, 220].

Moreover, this strain belongs to the W-Beijing family of *Mtb* isolates, which have been associated with outbreaks throughout the world and with clusters of drug-resistant TB in the United States [222, 225]. *Mtb* strains from the W-Beijing family have also been associated with multidrug resistance in extrapulmonary TB [226], and studies in rabbits suggest that these strains are more likely to leave the lungs than other less virulent strains [222].

As mentioned above, this strain, in contrast to *Mtb* CDC1551 induces a relatively weak Th-1 associated cytokine response in the lungs of infected mice [203, 220]. In vitro infection of monocytes with *Mtb* HN878 and other W-Beijing isolates is characterized by Th-2 polarized immunity as seen by high induction of IL-4 and IL-13 [224]. Therefore, *Mtb* HN878 and other related W-Beijing strains are considered to be both hypervirulent

and hypoimmunogenic in mice and human monocytes [223]. It is hypothesized that their failure to induce a strong protective Th-1 response would ultimately result in delayed or impaired protective immunity, leading to more severe disease which in turn would be responsible for their high virulence [223, 224].

The immunosuppressive effect of *Mtb* HN878 and other Beijing strains is due to the production of a characteristic phenolic glycolipid (PGL-T). Production of this glycolipid is associated with an intact pks1-15 gene. In contrast, *Mtb* H37Rv and *Mtb* CDC1551 have a 7 bp deletion resulting in a frame shift on pks1-15, which results in an inability to produce PGL-T. Supporting the role of PGL-T in virulence, Reed and co-workers showed that the deletion of the gene involved in the production of this lipid in the *Mtb* Beijing strains results in a less virulent phenotype [104].

1.8.4. Other strains

C-strain: This strain is one of the most widespread *Mtb* strains in New York. It has been suggested that the widespread dissemination of the *Mtb* C strain within the city might be related to an intrinsic resistance to Reactive Nitrogen Intermediates (RNI) [227].

The importance of radical oxygen species (ROS), reactive nitrogen intermediates (RNI) and nitric oxide (NO) in tuberculosis has been cause for debate throughout the years. Although some studies suggest that these molecules play a key role in the control of the infection in 'in vitro' MΦs [228] or in guinea pigs [229], their significance in humans might not be as relevant [230, 231]. However, the fact that the *Mtb* C-strain has evolved to be resistant to RNI suggests that oxidative response by MΦs to TB infection could also present some relevance in humans.

K-strain: This *Mtb* strain was isolated from an epidemic in high school students and was also found to be the most prevalent clinical isolate in Korea [232]. As mentioned above, it has been reported that this isolate induced significantly higher levels of necrotic cell death than *Mtb* H37Rv, suggesting that highly virulent strains inhibit cellular apoptosis as a host defense mechanism and instead induce necrosis in MΦs [218].

1.9. Conclusion

Tuberculosis is not a new disease. Control of TB has come a long way, and without a doubt, historical advances, from the ancient description of the disease to identification of the tubercle bacilli, in addition to establishment of the TST and development of the BCG vaccine, have changed the history of humanity and saved millions of lives. However, there are still millions more to save.

Modern studies on TB have lead to an increased understanding of the tubercle bacilli and the steps involved in the progression of the infection; new diagnostic tools have been developed and more are currently under study, in addition to new therapeutic drugs that could improve treatment of both susceptible and drug-resistant TB. In a similar way, understanding of the host immune response and the factors associated with it, as well as the pathogenic determinants inherent to the bacilli has increased in recent years.

TB still presents a challenge to community leaders and scientist around the world. Even though new tools for diagnosis have been developed, microscopic analysis of sputum smears remains the only possibility in many settings. Also, in spite of the attention that MDR-TB and XDR-TB have received, their numbers continue to rise and they are a constant threat to the global control of tuberculosis. Consequently, new advances in

diagnostics tools, vaccine development, increase of new and more effective drugs and drug regimens are all essential.

Studies that increase our understanding on the physiology of *Mtb*, including its interactions with the host and the immune responses elicited by this bacillus, with a special focus on virulent determinants associated to disease outcome, are imperative and should continue.

In this regard, the study of *Mtb* clinical isolates is very relevant. Currently circulating *Mtb* strains present different patterns of transmission and virulence, as well as interactions with the host, resulting in different degrees of the immune response. These studies are not only important to understand the association between host-pathogen interactions and disease outcome, but they also complement the development of new vaccines and diagnostic tools.

2. Chapter II. Literature review: The tubercle bacilli seen through proteomics

2.1. Introduction

In contrast to many other infectious diseases, TB is unique in the sense that it presents high variations in disease outcome from person to person. While the majority of people infected with *Mtb* will not develop the disease, about 10% will develop active tuberculosis after a latent interval that can vary from weeks to decades [233]. In addition, there is a whole spectrum of clinical presentations, from subclinical to rapidly fatal. Although pulmonary tuberculosis is the most common form of the disease, almost any organ can be involved.

Pathogenesis of TB and the factors involved in this great variation are only partially understood. However, it is certain that it depends on a complex interplay between environmental, bacterial and host characteristics [201]. To date, several of the determinants affecting the outcome of the infection have been identified including environmental and social factors, host genetics and immune status and factors inherent to the bacilli. Environmental and social factors that affect the susceptibility to tuberculosis include poverty, malnutrition, stress, overcrowding and exposure to environmental mycobacteria [234, 235]. Host immune status is often linked to social conditions like

poverty and malnutrition. The HIV epidemic also plays an important role in susceptibility to *Mtb* [236]. Finally, some genetic markers, like the NRAMP gene have been associated to an inclination to develop tuberculosis [237]. Mycobacterial components have also been studied in the context of virulence and infection; and some of these are listed in Table 1-1 in the previous chapter. However, the role and function of many other cellular components from *Mtb* remain to be studied. In particular, the identification of proteins that could impact the disease outcome is important, not only to increase our understanding of the biology of *Mtb*, but also to complement the design of new diagnostic tools with a special emphasis on disease prognosis. Studies of clinical isolates play an important role in this regard.

Mtb clinical isolates usually show differences in their ability to cause disease, transmission rates, survival in different animal models, and in vitro and in vivo growth [89, 104, 201, 203, 220, 238]. Therefore, they represent an excellent source for studying mycobacterial determinants that have an impact on disease outcome and transmission. In this chapter, several proteomic studies, with a special attention to those performed on *Mtb* strains are discussed.

2.2. The pre-proteomic era

2.2.1. *Mtb* genome and related studies

Mtb is considered to present very low genetic diversity [239]. Nonetheless, the sequencing of the entire genome of *Mtb* H37Rv [240] opened a new era in mycobacteriology permitting the study of entire genomes, which in turn has allowed the

identification of genetic and genomic differences between mycobacteria species and strains.

Genomic studies of clinical and laboratory *Mtb* strains are focused mainly on two different aspects of the genome. One is the identification and characterization of regions of difference (RD) which are present in virulent strains of *Mtb* and *M. bovis* but not in the *M.bovis* BCG strain [241]. Fourteen regions of difference (RD1–14), ranging in size from 2 to 12.7Kb, that were absent from *M.bovis* BCG Pasteur relative to *M.tb* H37Rv were identified by differential hybridization arrays [242, 243].

The second aspect of genomic studies involves the identification of insertions/deletions, including single genes present in RDs, that can be associated with certain biological/immunological properties (i.e. deletion of specific genes that increase or decrease virulence) [244].

Regions of deletion are also useful to discriminate between *Mtb* and other species within the MTC. Six regions, (RvD) 1–5 and TbD1, that are absent in *Mtb* H37Rv but present in other members of the MTC have been identified by comparative genomics approaches employing pulsed-field gel electrophoresis techniques and in silico analysis [243, 245]. TbD1 can also be used to divide *Mtb* into ancestral and ''modern'' strains, including representatives of major epidemics like the Beijing, Haarlem, and African *Mtb* clusters. [246].

Whole genome comparisons within the MTC revealed around 100 genes present in *Mtb* that are absent in *M. bovis* and an additional 30 genes that are absent in the *M. bovis* BCG strain, suggesting that the years of in vitro passage led to the selective loss of these genes [242, 243].

This has important implications. Differential genes between *Mtb* and *M. bovis* could be associated with biological differences of these two pathogens, including host tropism and disease presentation. In the same way, differential genes between the two pathogenic species and *M. bovis* BCG have been the basis for virulence studies as well as important developments in the vaccine and diagnostics fields [3, 117, 247]. Regions of deletion between strains of *Mtb* have provided clues in regard to its virulence. For instance, Lazzarini et al (2007) identified a genomic segment (RD^{RIO}) that is deleted in a prevalent group of strains in Brazil [248]. Ten genes, including two encoding immunogenic PPE proteins, are missing as a consequence of this deletion and the authors hypothesized that they might have a positive effect in transmission and virulence of these strains.

The completion of the genome of the clinical isolate *Mtb* CDC1551 has allowed the comparison of entire genomes between *Mtb* strains [249]. The *Mtb* CDC1551 genome contains 49 insertions and 37 deletions when compared to *Mtb* H37Rv genome. Almost half of the insertions/deletions involve genes encoding PPE or PE_PGRS family proteins which have been noted to be very polymorphic by other authors [250].

Even though gene deletions are generally considered deleterious, they are not uncommon, and other clinical isolates of *Mtb* have also shown to have several genes missing from their chromosomes [251-253]. In some cases, gene deletions have been shown to confer advantages to the bacilli. That is the case of a common Asian *Mtb* lineage which has a deletion affecting the gene Rv1519 [244]. Strains bearing this deletion induced a less protective and more anti-inflammatory reaction in monocyte-derived macrophages, resulting in a harmful immune phenotype responsible for the persistence and high transmission potential of these strains. Several other deletions present in clinical isolates

include genes that are possibly involved in pathogenicity or latency [253], including phospholipase-C genes [254] and polyketide synthase gene (pks5). This latter has been shown to encode the toxin responsible for cutaneous lesions during *Mycobacterium ulcerans* infection [255].

The analysis of the correlation between deletions and disease presentation suggests that the likelihood that a clone of *Mtb* will cause pulmonary cavitation decreases as the amount of genomic deletion increases, indicating that variations in gene content can be associated with outcome of the disease [253].

2.2.2. Gene expression studies in *Mtb*

Gene expression studies are based in different and complementing techniques; the most widely used are mRNA microarrays and quantitative real time PCR. Other approaches are based on subtractive hybridization and differential display.

Microarrays offer the advantage of global analysis of gene expression under different environmental conditions, including those mimicking the host environment as well as drug treatment, providing information regarding virulence, adaptation, intracellular survival and drug resistant phenotypes. Analysis of *Mtb* gene expression upon treatment with antimycobacterial drugs can also provide insights on drug resistance mechanisms. On this matter, expression profiles after treatment with isoniazid, ethambutol, ethionamide and triclosan have been obtained [256-258]. These studies will not be discussed here. Rather, studies concerning the identification of genes involved in virulence and pathogenicity by global expression analysis of *Mtb* strains with different

virulent phenotypes or under environmental conditions that mimic different aspects of natural infection will be the focus of this review.

Many studies have compared the gene expression profiles of *Mtb* H37Rv to those of *Mtb* H37Ra. This analysis has led to the identification of genes with potential involvement in virulence. These include the two component system devR-devS (a.k.a. dosR-dosS) [259, 260], Rv2770c, a member of the PPE protein family, Rv1345, a probable polyketide synthase, and Rv0288 which is a member of the ESAT-6 family [261]. Other genes with differential expression between *Mtb* H37Rv and *Mtb* H37Ra include some with involvement in cell wall and cell processes as well as in intermediary metabolism and respiration. Genes involved in lipid metabolism have been found to have different expression levels between strains, which in turn could account for the differences observed in cord formation between *Mtb* H37Rv and *Mtb* H37Ra [189, 214, 262].

Interestingly, the dosR regulon is upregulated in *Mtb* H37Rv under dormancy conditions [263, 264]. However, is has been shown to be upregulated even higher in the avirulent strain *Mtb* H37Ra, leading to the speculation that attenuation of this strain is caused by hyperdormancy [189]. Puzzling, the dosR regulon was also shown to be constitutively upregulated in hypervirulent *Mtb* strains of the Beijing lineage [265]. This regulon comprises a set of 48 genes that are regulated in response to hypoxia or treatment with NO [263, 266]. Therefore, it has been hypothesized that the overexpression of these genes in the *Mtb* Beijing strains would confer them a significant advantage since they would be "preadapted" to the environmental conditions that are thought to exist during in vivo infection [265]. More studies are needed to clarify the specific role that overexpression of this regulon plays in virulence.

Gene expression of *Mtb* H37Rv and its avirulent counterpart has also been evaluated in the context of their interaction with murine macrophages. In this case, the frd operon (encoding the fumarate reductase enzyme complex) is upregulated at early points post-infection in *Mtb* H37Rv and at late points post-infection in *Mtb* H37Ra [267]. The importance of this type of information was highlighted by the authors in the same study by treating intracellular mycobacteria with an inhibitor of fumarate reductase, showing a significant reduction of bacterial growth.

Similarly, studies focused on *Mtb* H37Rv and *M. bovis* differential gene expression provide interesting data that could help understand the differences in host tropism as well as virulence levels presented by these two strains [268].

In addition to gene expression analyses of well characterized *Mtb* laboratory strains, studies on clinical isolates are also very valuable. Recently, real time PCR was used to assess the expression levels of a group of genes in three hypervirulent *Mtb* strains isolated from tuberculosis meningitis patients and three hypovirulent *Mtb* clinical isolates. Using this approach in conjunction with ex vivo macrophage infections, the authors found that expression of *fad*E28 is associated with virulent phenotypes in *Mtb* [269]. *fad*E28 encodes acyl-coenzyme A (acyl-CoA) dehydrogenase, which participates in the first step of β-oxidation [270] and its inhibition was previously shown to be related to attenuation of the mycobacteria [271].

Other studies have shown that variable gene expression between clinical isolates is significant. In particular, genes associated with T-cell antigens, PE and PPE families, and lipid metabolism all of which are thought to be important in host–pathogen interactions, present high levels of variation [272, 273]. Interestingly, PE/PPE family of genes also

presents high variability at the genomic level [250], suggesting that differential expression of genes could be associated with variation in the genome. One particular case is the gene *ppe*44 which presents nucleotide substitutions and also significant higher expression levels in Beijing *Mtb* isolates compared to non-Beijing *Mtb* strains [274].

Analysis of gene expression of clinical isolates is therefore important since genetic variation in circulating strains might impact virulence, transmission, treatment, vaccines and diagnostics.

2.3. Proteomic tools in the study of infectious diseases

Proteomics allows for high throughput profiling of the expressed proteins at the cellular and subcellular levels. It complements genomics and classic gene expression studies in showing which proteins are really expressed, taking into account mRNA stability, as well as post-translational modifications and protein degradation.

Moreover, proteomics offers insights into subcellular protein localization and functional status of a cell in response to environmental stimuli [275, 276]. In addition, it can be used to measure qualitative or quantitative differences in protein expression after exposure to different environments or between different organisms or cells. This information can then be used for bioinformatics modeling of cellular processes, including pathways, cell growth and cell metabolism [277].

In contrast to the genome, the proteome is not static but highly dynamic [278], and it represents a more realistic view of the intricate status of a living cell.

During the past decade, proteomic studies have been dominated by two technologies: 2-D gel electrophoresis to separate proteins, and mass spectrometry to identify these proteins.

Although mass spectrometry continuous to be a central tool in proteomics, other technologies have been developed to either improve, complement or replace 2D gel electrophoresis.

A brief description of some of the most common proteomic techniques, including the central role of mass spectrometry, is the focus of this section. Special attention will be given to 2D-GE and iTRAQ techniques due to their relevance towards the studies described in the following chapters.

2.3.1. Mass spectrometry

Mass spectrometry (MS) and tandem mass spectrometry (MS/MS) are the center of all proteomic studies. These techniques allow for identification of complex molecules in a fairly pure form or in a composite mix. There are several different types of mass spectrometry, however, only those used for protein and peptide identification will be described here.

Tandem MS/MS is generally the preferred method used to identify proteins by matching the measured fragment ion spectra of peptides with theoretical spectra calculated from known DNA or protein sequences [279]. Tandem mass spectrometry is based on the retention of precursor ions (i.e. peptides) and their fragmentation and ionization resulting in obtaining information at the amino acid level.

The major advances in mass spectrometry that have had an impact in proteomic research have been electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) [280-282]. In ESI, highly charged droplets dispersed from a capillary in an electric field are evaporated and the resulting ions are attracted to the inlet of the mass

spectrometer where they can be detected. In MALDI, sample molecules are mixed with an excess of solid or liquid matrix containing a highly UV-absorbing substance that will absorb the energy of a UV laser shot and then transfer it to the sample for desorption and ionization [283].

Quadrupole ion traps and triple-stage quadrupole instruments are among the most used instruments for protein analysis due to their power for structural analysis through tandem MS/MS [284]. The operation of quadrupoles and quadrupole- ion trap instruments is based on the motion of ions in oscillating (radio-frequency, RF) electric fields. The basis for the mass analysis is that in an RF field an ion experiences a force proportional to its distance from the center of the field [284].

Two of these instruments will be briefly described below.

2.3.1.1. LTQ Tandem Mass spectrometer

The linear trap quadrupole (LTQ) is one of the most commonly mass spectrometers utilized for proteomic analysis. LTQ, in comparison to previous ion trap instruments has 15 times higher ion capacity, three times faster scan rate and better detection and trapping efficacy [285]. In these instruments ions are injected from a external ESI source and are then trapped for a given period of time prior to their scanning in a mass-dependent manner [284]. Most of these instruments use collision-induced dissociation (CID) for the generation of fragment ions, however, this type of fragmentation is associated with a decreased stability of fragment ions that are less than 30% of the m/z of the precursor peptide, also known as the one-third rule [286]. Unfortunately, this presents a limitation

for the use of LTQ instrument in analysis of iTRAQ labeled peptides due to unreliable MS/MS detection of the low m/z values of the tags (i.e. 114 – 117) [287].

The introduction of a new dissociation technique called pulsed-Q dissociation (PQD) [288] was shown to overcome the issue of the one-third rule, and iTRAQ analysis in a LTQ is now possible [287, 289]. However, the collision energy (CE) for the PQD needs to be carefully standardized for each experiment and even regular procedures of cleaning and re-tuning can have an impact on previous "standardized" CE [287].

2.3.1.2. Q-TOF Tandem Mass spectrometer

Quadrupole-Time-Of-Flight can be described as a triple quadrupole with the last quadrupole replaced by a Time-of-flight (TOF) analyzer [290]. Q-TOF instruments present a high mass accuracy and resolution for analysis of proteins and peptides in both MS and MS/MS modes [290, 291].

For Tandem MS/MS, Quadrupole 1 (Q1) functions as a filter based on mass transmitting only the parent ions of interest, which are subsequently accelerated before entering the collision cell Q2. In the Q2, parent ions are subjected to collision induced dissociation (CID) after the first few collisions with neutral gas molecules (usually argon or nitrogen). The resulting fragment ions (in addition to the remaining parent ions) are collisionally cooled and focused. After leaving the collision cell, ions are re-accelerated and focused into a parallel beam that enters the ion modulator of the TOF analyzer. Then a pulsed electric field is applied at a high frequency pushing ions into the accelerating column where they acquire their final high energy. From the accelerating column, ions arrive in the field-free space where TOF mass separation and posterior detection occurs [290].

The stability of ions generated by CID in a Q-TOF is much greater than in a LTQ instrument; therefore, the one-third-rule does not apply here. In addition, its capability to amplify low mass fragments in MS/MS mode improves the signal strength of iTRAQ reporter ions [287, 292]. Consequently, the Q-TOF mass spectrometer is the gold-standard in terms of quantitative accuracy for iTRAQ labeled peptides [293, 294].

2.3.2. Gel based proteomic techniques – 2D-GE

2D gel electrophoresis (2D-GE) is based on the separation of proteins according to their isoelectric point by isoelectric focusing in a pH gradient and by mass in a SDS-PAGE. After separation, proteins are usually visualized by staining with silver or coomasie blue. Protein spots are excised from the gel and subjected to trypsin digestion followed by mass spectrometry analysis. This technique allows the separation and further characterization of proteins in one single gel, and allows the creation of maps representing the proteomes of whole cells or subcellular fractions. These maps are then analyzed using specialized imaging software, allowing the comparison of different gels representing different samples or conditions.

Major disadvantages of this technique are as follows: i). the low reproducibility between gels [295], which makes the comparison of protein levels between different samples very challenging and sometimes subjective, ii). the potential for several proteins to be present in the same spot and iii). the possibility of single proteins to appear as multiple spots, making it difficult to draw conclusions based on relative abundance of these proteins. The image analysis is time consuming, especially when more than two conditions (stimuli, strains, patients, etc) are being evaluated. In addition, the analysis of certain classes of

proteins is difficult due to the broad range of physicochemical properties of proteins [296].

Labeling of protein samples with fluorescent tags can improve not only the sensitivity of 2D gels but also their reproducibility. This technique (a.k.a. DIGE) allows for the separation and analysis of up to three different samples in one single gel, which facilitates the comparison between maps. Reproducibility between gels, when more than three samples are analyzed, and between technical replicates, is still a major concern. In addition, since the electrophoresis procedure is basically the same as for conventional 2D gels, the problems in resolution and quantification of certain proteins remain. This technique also requires the use of specialized and expensive imaging-acquiring equipment.

Even though these limitations are important considerations for experimental design, 2D-GE offers the advantage of following qualitative differences between samples (i.e. absolute absence or presence of spots). In addition, when post-translational modifications are studied in the context of cellular physiology, 2D-GE allows easy visualization of different isoforms for a particular protein. Classical 2D-GE with coomasie blue or silver staining is much more accessible technique in laboratories with limited resources.

2.3.3. Gel free techniques

In contrast to 2D-GE, gel free techniques do not perform separation of the proteins prior to mass spectrometry analysis and therefore are sometimes called "shotgun" strategies. Proteins are usually subjected to digestion by proteases and peptides are separated and analyzed by liquid chromatography and mass spectrometry followed by identification

using bioinformatics. Some of the advantages of gel free techniques include better sensitivity and reproducibility, as well as higher throughput. Additionally, the amount of sample needed to obtain high quality results is typically smaller than for 2D-GE.

2.3.3.1. Isobaric tags for relative and absolute quantification (iTRAQ)

As can be deduced from its name, iTRAQ analysis uses tags with identical mass for peptide labeling. Up to eight different samples can be labeled and analyzed in a single experiment. Therefore, the popularity if this technique has grown in recent years.

Each label is composed of a peptide reactive group (NHS ester) and an isobaric tag of 145 Da that consists of a balancer group (carbonyl) and a reporter group (based on N-methylpiperazine) [294]. Between the balancer and the reporter group is a fragmentation site (Figure 2-1). The peptide reactive group attaches specifically to free primary amino groups (i.e. N-termini and amino groups of lysine residues), although labeling of tyrosines has also been reported [294]. Labeled peptides are then analyzed by Tandem MS/MS using techniques that allow the resolution of low m/z ions. Finally, the peak intensity corresponding to each iTRAQ label (114, 115, 116, 117 m/z in a fourplex) are used to perform quantification of each peptide within each sample.

One of the major advantages of this technique is the ability of multiplexing, making it possible to analyze several samples in a single experiment, reducing issues of variability related to differences in sample preparation. Disadvantages of this technique include the lack of labeling if the primary amino group is modified, for example by acetylation or by the formation of pyroglutamic acid from N-terminal glutamine or glutamic acid [278]. In addition, a particular proteolytic peptide may be derived from different proteins resulting

in significantly different quantities of that protein [296]. This can be overcome with the automatic or manual removal of peptides associated with more than one identified protein.

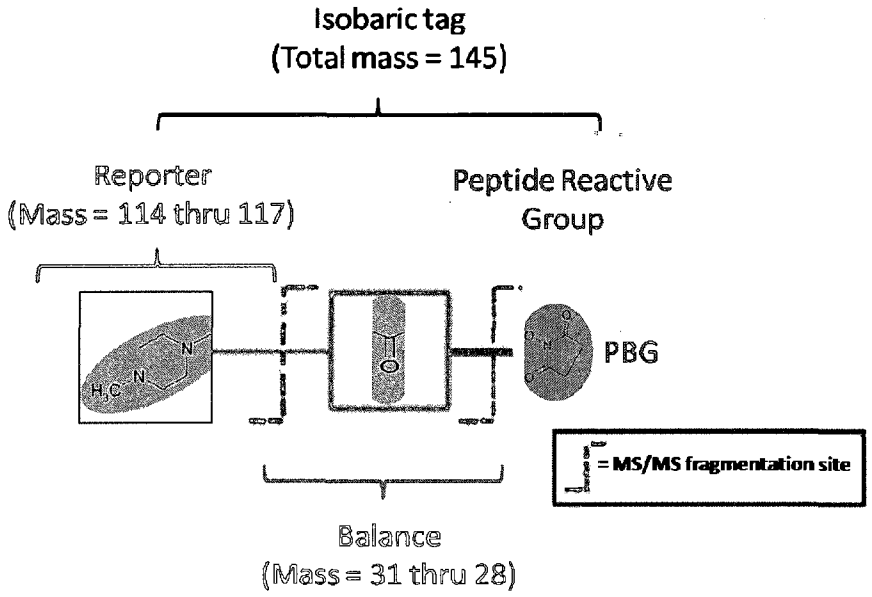


Figure 2-1. Scheme of the iTRAQ reagent.
Modified from [278, 294] and the iTRAQ user manual (Applied Biosystems).

Several software programs have been developed for the analysis of iTRAQ data including i-Tracker [297], MassTRAQ [298], Quant [278], iTRAQPak [299], Pro-Quant™ (Applied Biosystems (ABI), Darmstadt, Germany), ProteinPilot™ (Applied Biosystems (ABI), Darmstadt, Germany) and Mascot™ (Matrix Science, London, UK). The approach for the analysis of data is significantly different in each of these programs, therefore and special consideration for appropriate data retrieval and interpretation using these programs must be given (i.e. peptide vs protein quantification, statistical analysis, normalization, etc). Analysis of iTRAQ data to determine differential expression of proteins is very variable between studies. Several studies use an arbitrary cutoff of log values or fold-ratios to consider a particular protein significantly up or down-regulated

[300, 301]. In other more rigorous studies, a statistical analysis, usually an ANOVA, is performed [302-304].

In the mycobacteria field, iTRAQ analysis of proteomes has rarely been used. Radosevich and co-workers performed a quantitative analysis of membrane and cytosolic proteomes of two *M. avium* subsp. p*aratuberculosis* isolates [300]. In this study, proteins that had a change of 2 or more standard deviations in their ratio value were considered to be differentially expressed between strains, but no statistical tests were used to validate the results and no normalization was applied to the raw data. Additionally, it is not clear if replicates were performed. Even though these considerations are very important with regard to the results presented in the study, immunoblot analysis of three proteins confirmed the findings obtained for these proteins by iTRAQ.

Another study used iTRAQ and 2D-GE for the comparison of proteomes between two strains of *M. ulcerans* (i.e. wild type vs mycolactone-deficient mutant) [305]. In this study, 123 proteins were differentially expressed between the mutant and the wild type. Unfortunately, iTRAQ data analysis, including statistical analysis or cutoff values were not described, making it difficult to perform a critical review of the paper.

Biomarker discovery using iTRAQ analysis of serum proteins from cattle infected with *M. bovis* or *M. paratuberculosis* has been recently published [301]. From this approach, several proteins were identified as differentially expressed in control vs diseased animals and also in *M. bovis* vs *M. paratuberculosis* infected animals. Even though the level of variability for most of the differential proteins identified in the study were only 0.5 or -0.5 log, the authors validated the use of vitamin DBP as a biomarker for mycobacteria infections in cattle by using chemiluminiscent dot-blots.

Host responses have also been evaluated by labeling host proteins with iTRAQ tags after exposure to *Mtb* lipids [306]. Interestingly, MΦ receptors and proteins involved in antigen presentation were some of the 166 proteins differentially expressed after *Mtb* lipid exposure. The authors compared iTRAQ labeling to SILAC and found a good correlation between both methods, although iTRAQ seems to be able to detect smaller changes in protein levels than SILAC.

2.3.3.2. Other gel-free techniques

Other 'shotgun' approaches include the labeling of peptides using isotope-coded affinity tags (iCAT) or radioabeled aminoacids. Label free approaches are also possible thanks to the development and increased power of programs and algorithms that facilitate the analysis of mass spectrometry data. Some of these techniques are described in Table 2-1.

Table 2-1. Brief summary of some gel-free proteomic techniques

Technique	Brief description	Advantages	Drawbacks	References
Isotope-coded affinity tag (iCAT)	Proteins are labeled at cysteine residues using two isotopic tags (i.e. light and heavy).	Samples are combined prior to trypsin digestion and tandem MS/MS, reducing technical variability.	Only cysteine-containing proteins are labeled. No more than 2 samples can be analyzed at a time.	[307, 308]
Stable isotope labeling of amino acids in culture (SILAC).	Isotopically distinct amino acids are incorporated during during culture.	Identification of de novo synthesized proteins.	Radiolabeling requires especial infrastructure.	[297]
Label free approaches	Samples are processed separately and the number of spectral counts for each protein in each sample is obtained by specialized software.	Involves very small number of steps, minimizing variability due to the inherent manipulation of the sample	Requires complex calculations and normalization. No multiplexing of samples, increasing technical variability	[309]

None of the techniques described above are perfect. However, their use and development in conjunction with almost day to day improvements of mass spectrometry and bioinformatics software have made proteomics one of the most exciting fields in the study of infectious and non infectious disease.

Proteomics provides the opportunity to determine whether or not proteins exist in multiple species, and to examine their extent of post-translational modifications. Furthermore, proteomics may be used to examine the cellular and subcellular distribution of proteins and their relative concentrations, and to identify the proteins that have not been predicted by genome analysis, increasing the accuracy of genome annotation [310].

2.4. Proteomic studies on Mtb

Most of the studies that have been done on *Mtb* to assess gene expression variability have been performed at the genomic and mRNA level. However, in recent years, the global analysis of proteins or proteomics has become a useful development in the study of mycobacterial physiology in the context of virulence, pathogenesis and environmental adaptation.

Some of the advantages of using proteomics were described in the previous section. In addition, proteomic analysis in *Mtb* provides clues in relation to pathogenicity and prevalence of strains. Moreover, identification of protein expression differences between strains complements the development of vaccines, serodiagnostic tests and choice of drug targets [311].

For many years, proteomics studies on TB have been focused on the characterization of proteomes of different subcellular fractions of *Mtb*. Due to their importance in terms of antigenic properties [312], with implications in the vaccine and diagnostic fields, secreted proteins are one of the most studied fractions [249, 275, 276, 313-318] and include studies on the Ag85 complex, Mpt63, Mpt64, GroES, SodA, DnaK and members of the Esat6 family; the most abundant proteins in the secreted fraction of *Mtb* H37Rv [313, 315]. Cytosolic and cell wall proteomes [276, 315, 319] as well as whole cell lysates [320] have also been studied, although to a lesser extent. These original studies have opened the field of proteomics in TB research and have provided useful annotation of proteins and 2D-GE maps. Some studies have also identified open reading frames that had not been predicted by genomics [321].

Proteomic approaches have also been used to identify antigens that are recognized by most TB patients' sera and that could be potentially used as biomarkers for different stages of the disease (i.e. active vs latent; cavitary vs non-cavitary) [316, 322, 323].

Recently, an effort to create an *Mtb* glycoprotein database was possible in part by proteomic studies [324]. Glycoproteins are considered to be involved in host-pathogen interactions [325], and although some of the *Mtb* glycoproteins have been identified [326, 327], the recognition of these proteins at a global level is one of the first steps to completely understand their role in the physiology of the mycobacteria.

The identification of proteins involved in pathogenesis and virulence by studying the *Mtb* proteomes under different environmental conditions, as well as by comparison of clinical and laboratory strains, have been performed and will be the focus of this section.

2.4.1. Analysis of proteins expressed under conditions mimicking the host environment

Proteins that are expressed by *Mtb* in vivo present a major likelihood of being recognized by the host, and therefore their recognition is important in the context of vaccine and immunodiagnostic development. Identification of TB proteins in 'in vivo' infections is problematic due to the low ratio of bacillus to host cell proteins. Thus, most of these studies have been performed 'in vitro' by using different conditions that resemble the host environment, including starvation and growth on low oxygen.

Proteins such as the 45kDa antigen (ModD), Mpt64 and Tig, a protein similar to the trigger factor, are less abundant during starvation suggesting that their expression during natural infection might be low. Two hypothetical proteins, Rv2557 and Rv2558, in addition to HspX, were found to be induced under nutrient starvation conditions [328]. Proteomic approaches have also shown HspX to be predominantly present during hypoxic conditions [329, 330], in aging cultures [331], in standing versus shaking cultures [332] and during intracellular growth [333], all of which supports its proposed role during latency.

Levels of L-alanine dehydrogenase (Ald), HspX, GroEL2, probable fructose-biphosphate aldolase (Fba), bacterioferritin (BfrB) and elongation factor EF-Tu among others have been shown to be higher under low oxygen conditions [330, 334]. EF-Tu has also been shown to be more abundant during intracellular growth [333], suggesting that these proteins are being expressed during infection. These proteins have also been identified as differential between *Mtb* H37Rv and *M. bovis* BCG [275, 314, 333], emphasizing their potential role in pathogenesis of *Mtb*.

Using the iCAT approach, Cho and co-workers (2006) identified several other proteins showing differential levels between different stages of non-replicating persistent *Mtb*. A great majority of the differential proteins belong to the small molecule metabolism category, especially proteins involved in energy metabolism and degradation [335].

It has been suggested that one of the stresses that mycobacteria encounter during natural infection are radical oxygen species (ROS) and radical nitrogen intermediates (RNI), such as superoxide and nitric oxide [228, 229]. Therefore, protein levels after exposure to these radicals have also been evaluated. As expected, DnaK and other heat shock proteins were induced after exposure to menadione, a compound that increases intracellular superoxide levels [336].

However, overexpression of heat shock proteins has been shown to reduce survival of the mycobacteria during the chronic phase of infection. Particularly, a mutant overexpressing DnaK showed a reduction of bacterial load in the spleens of infected mice in a similar manner to what is obtained after immunization with BCG [337]. It seems then, that a balance between expression of these types of proteins at different points during the infection must occur in order to be beneficial for the mycobacteria.

Keeping up with the description of protein expression under stimuli that mimic the MΦ environment, it has been shown that several antigens and enzymes (Lsr2, Hsp16.3, oxidoreductase, PPIase, Aconitase, Fur) present differential levels under high or low iron conditions [338], which are believed to be important during natural infection [116].

Studies of in vivo expression of mycobacterial proteins are considerably scarce; however they are very relevant since they represent a more realistic view of what occurs during natural infection. Metabolic labeling of *Mtb* and infection of THP-1 cells showed the

differential expression of 44 proteins, from which 16 were more abundant and 28 were repressed after infection [339]. In another study, proteins induced upon phagocytosis of *M. bovis* BCG by THP-1 cells included HspX, the chaperones GroEL1 and 2 and Elongation factor EF-Tu [333].

More recently, Mattow and co-workers (2006) identified eleven proteins that were unique to intracellularly growing *Mtb,* including several enzymes involved in cell metabolism such as cystathionine (beta)-synthase (CysM2), GMP synthase (GuaA), malate dehydrogenase (Mdh) and phosphoglycerate mutase I (Gpm) among others. A hypothetical protein, Rv1130 was also detected in eight spots unique to intracellular mycobacteria [340].

Analysis of *Mtb* proteome in response to drug treatment and the comparison of proteomes between drug-resistant and drug-susceptible strains are important in order to elucidate potential drug resistant mechanisms and novel drug targets. Most studies on this matter have been performed following exposure to, or with cells resistant to, isoniazid (INH). This has led to the identification of five proteins that are upregulated in INH-resistant strains (Rv1446c, Rv3028c, Rv0491, Rv2971, and Rv2145) [341]. Most of these are membrane proteins, which are known to have potential as diagnostic and therapeutic targets. Some of the proteins that are differentially expressed after INH exposure corresponded to proteins from the *kas* operon, proteins involved in fatty acid metabolism (FadD26 and DesA2), and proteins involved in cell wall and cell processes (DdlA, Wag31) [309, 342].

A recent study using a shotgun label-free proteomic technique assessed the differences of proteomes after exposure of *M.smegmatis* to INH and EMB. Using this approach, the

authors were able to identify more than 400 proteins with differential expression between control and drug-treated mycobacteria [343].

2.4.2. Comparison of virulent and avirulent mycobacteria strains

In the search for determinants of virulence, novel vaccine candidates and better diagnostic targets, virulent and avirulent strains have been compared at the protein level. Proteomic analysis by 2D-GE of *Mtb* H37Rv and *M. bovis* BCG allowed the identification of several proteins unique to each of these species [275, 314]. *Mtb* specific proteins included L-alanine dehydrogenase (40 kDa antigen, ald, Rv2780), isopropyl malate synthase (Rv3710), nicotinate-nucleotide pyrophosphatase (Rv1596), Mpt64 (Rv1980c) and two conserved hypotheticals (Rv2449c and Rv0036c) [275]. Esat-6, Cfp10 and other Esat-6 like proteins were also identified only in *Mtb* [314]. Remarkably, other proteins that are not predicted to be absent in the *M. bovis* BCG genome were identified only in *Mtb* and include the chaperone GroES, elongation factor EF-Tu (Tuf), Rv3269 and seven hypothetical proteins (Rv0020c, Rv1684, Rv1893, Rv3046c, Rv3881c, Rv1198 and Rv1793).

Proteomes of *M.tb* H37Rv and *M. bovis* BCG have also been compared using a complementary approach of 2D-GE and ICAT [320]. This study confirmed findings from previous genomic analyses in which two proteins (Rv0223 and Rv1513) were found to be unique to *Mtb* H37Rv. In addition, Rv0570 was also found to be present in *Mtb* H37Rv, and not in *M. bovis* BCG. These three proteins represent potential vaccine candidates but they have not been characterized in the context of the immune response.

Similarly, proteomes of *Mtb* H37Rv and *Mtb* H37Ra have been compared, resulting in the identification of proteins unique to *Mtb* H37Rv but absent in *Mtb* H37Ra. Eleven

protein spots that were present in both strains, but upregulated in *Mtb* H37Rv, as well as three protein spots downregulated in this strain, were identified [344].

Proteins unique to or upregulated in *Mtb* H37Rv in comparison to non virulent strains are potentially involved in virulence and pathogenesis mechanisms. In addition, proteins that are only present in *Mtb* but absent in *M. bovis* BCG strains, such as Cfp10 and Esat-6, have been exploited as better and more specific diagnostic test for TB infection [148, 149].

2.4.3. Comparison of *Mtb* clinical isolates

The identification of antigens and other immunogenic components that have a fairly similar expression in different clinical isolates of *Mtb* is a key component of the development of new and better vaccines and diagnostic tests for TB. Consequently, proteomics studies on clinical and laboratory strains of *Mtb* represent a good approach to identify potential antigens that are well represented in different *Mtb* isolates and also allow the identification of the most abundant proteins expressed by this bacillus. This is one of the first steps to identify immunogenic proteins that are recognized for the majority of people infected with TB and can be selected for the development of vaccines and diagnostic tests.

Differences observed between *Mtb* clinical isolates could be used for identification of specific strains during an outbreak. For example, proteomic analysis of distinct *Mtb* genotypes led to the identification of Rv0927c, which is only present in non W-Beijing *Mtb* strains and could be used as a marker to differentiate between W-Beijing and non W-Beijing strains [345].

Identification of proteins expressed at different levels by *Mtb* clinical isolates representing diverse phenotypes in natural human infections represents a valid approach to identify proteins that are potentially involved in the variability observed in virulence, disease outcome and transmission. Studies involving comparative proteomic analysis of *Mtb* strains will be described below. A summary of findings obtained by this approach are listed in Table 2-2.

Table 2-2. Summary of differential proteins identified by comparative proteomic studies in *Mtb* clinical isolates.

Mtb Strains	Subcellular fraction	Technique	Differential proteins	References
K-strain, H37Rv, CDC1551	Secreted proteins	2-DE	↑K-strain: Rv0652, Rv1636, Rv2818c, Rv3369, Rv3865, Rv0566c, MT3304, Rv3160 ↑K-strain, CDC1551: Cfp10, Rv0560c, Rv3648c	[346]
K-strain, H37Rv	Proteins recovered upon U-937 cells infection	2-DE	↑All strains: Rv3521, FadE8, Rv2558, AtpG, AmiA and SufC ↑K-strain: GlgP and LinB ↑H37Rv: Rv3871	[347]
W-Beijing and Non-W-Beijing	Whole cell lysate	2-DE	Rv0927c	[345]
CDC1551, H37Rv	Secreted proteins	2-DE	Unique to CDC1551: Rv0927c and 4 unidentified spots Unique to H37Rv: HisA ↑H37Rv: AhpC	[249]
Erdman, H37Rv	Whole cell lysates, secreted proteins	2-DE	Unique to H37Rv: Ald, SppA, Rv2296 and two non-identified spots Unique to Erdman: Rv3213, Rv2641 ↓H37Rv: ArgD, Rv1284, ArgB, ArgC	[275]

Beijing strains, Family 23 strains, H37Rv	Whole cell lysates, Secreted proteins	2-DE, Western blot	↑Beijing and family 23: HspX, Rv2005c, Rv2744c ↑H37Rv: Ag85A ↓Beijing: PstS1	[311]
H37Rv, 2 distinct clinical isolates	Whole cell lysates	Western blot	No significant differences between strains, but variation related to growth stage	[349]

Comparison of the secreted proteomes of *Mtb* H37Rv and *Mtb* CDC1551 have shown that only 13 proteins, from over 1500 present differential levels between these strains, most of which are due to absolute absence/presence of specific proteins rather than differences in quantity [249]. This study also showed that the transcriptional regulator MoxR has a mobility shift between these strains, which is probably due to post-translational modifications. The comparison of *Mtb* H37Rv and *Mtb* Erdman strains revealed 18 differential spots of which 8 were due to absolute presence or absence in one of the strains compared to the other [275].

Clinical isolates *Mtb* CDC1551 and *Mtb* K-strain have also been compared to *Mtb* H37Rv [346]. Eight proteins that were more abundant in the *Mtb* K-strain and three proteins that were more abundant in both *Mtb* K-strain and *Mtb* CDC1551 compared to *Mtb* H37Rv were identified. These differential proteins were produced by recombination technology in *E.coli* and used to test TB patient serum in an ELISA format. Two proteins, Cfp10, which is now used for diagnosis, and Rv3369 showed potential for diagnostic use. Based on their results, however, Rv3369 should be used with caution because its expression in *Mtb* CDC1551 and *Mtb* H37Rv was lower than in the *Mtb* K-strain, suggesting that variability in patients' immune response towards this antigen could also occur. Rv0560c was found to be more abundant in both clinical isolates in comparison to

Mtb H37Rv. This protein is a possible methyltransferase and has been shown to be present in high levels during anaerobic conditions [330], suggesting a potential role during intracellular growth.

Comparison of proteomes of the *Mtb* K-strain and other *Mtb* clinical isolates, in addition to *Mtb* H37Rv has been done after infection of U-937 cells [347], providing additional information regarding the differences between these strains in the context of the interactions with the host. Six proteins were identified as overexpressed upon infection by all strains analyzed in this study (Rv3521, FadE8, Rv2558, AtpG, AmiA and SufC) [347]. Therefore, these proteins are probably associated with intracellular survival, virulence and adaptation, and are good targets for further studies on their physiologic roles. Two proteins (GlgP and LinB) were present in higher levels only in the *Mtb* K-strain, while FtsK was more abundant in *Mtb* H37Rv. GlgP and LinB represent proteins with important cellular functions. Specifically, GlgP has been associated to MΦ migration inhibition [348] and therefore could be advantageous to the mycobacteria during human infection. Consequently, higher expression of these proteins by the *Mtb* K-strain has been hypothesized to be related to the high prevalence of this strain in Korea [347].

Proteomes of strains belonging to the Beijing and Family 23 (F23) *Mtb* lineages have been analyzed and compared to *Mtb* H37Rv [311]. In this study, eight protein spots that were more abundant in the whole cell lysates of the *Mtb* clinical isolates (Beijing and F23) in relation to *Mtb* H37Rv were identified. Interestingly, seven of these were identified as α-crystallin. The remaining spot contained two proteins, Rv2005c and the 35kDa antigen Rv2744c. Hypothetical protein Rv2005c, a protein showing similarity to

universal stress proteins (USP) has also been shown to be expressed specifically in anaerobic conditions, suggesting that it is expressed during natural infection and could play a role in virulence and survival of the mycobacteria [330]. Three spots identified as Ag85A had decreased intensity in the CFPs of both *Mtb* clinical isolates compared to *Mtb* H37Rv. In addition, PstS1 was found to be the most marked difference between Beijing and F23 *Mtb* strains, being less abundant in the former. Lower expression of this protein in the *Mtb* Beijing strains was correlated with decrease recognition of this protein by sera from *Mtb* Beijing-infected patients [311, 316].

Results from these studies show that protein variation between distantly related strains is little. However, the few differences observed could have an impact in disease outcome and also provide valuable information that can be used for diagnostic purposes and impact vaccine development.

2.5. Conclusion

The study of protein expression on a global basis has been shown to provide clues to understanding the physiology of *Mtb* by identifying proteins expressed under conditions mimicking the host cell environment and other stimuli. In addition, the identification of the secreted proteome and related antigens has been useful for designing diagnostic tests and new vaccines.

The comparison of laboratory and clinical isolates with different degrees of virulence at the protein level has also been shown to be important not only to identify virulent determinants, but also to provide useful information that can be used for the development of serodiagnostic tests. That said, protein variation observed between clinical isolates

seems to be relatively small. This fact could be due to the inherent low genomic and genetic variability in *Mtb* or could be linked to the use of proteomic techniques that do not allow the identification of proteins with relatively small differences in expression.

The use of new proteomic techniques with higher sensitivity and reproducibility such as some shotgun techniques could have an important positive impact on the study of distantly and closely related clinical isolates at the protein level. This in turn could have an impact in our understanding of virulence and transmission.

3. Chapter III. Biological diversity of closely related clinical isolates of *Mycobacterium tuberculosis*

3.1. Introduction

The global incidence of tuberculosis (TB) is increasing at approximately 1% per year, with an estimated more than 9 million new cases each year (WHO, 2009). More than 4 million deaths were reported in 2007 due to infection with *Mtb*, the causal agent of TB [2]. Unfortunately, centuries after the first description of tuberculosis infection, there is still a major need for more effective vaccines, diagnostic tests and drug treatments, which are crucial for the control of this disease [311].

Mtb presents a very limited genetic variation. However, there is a high degree of phenotypic variability among isolates, including differences in clinical outcome and epidemiological behavior. Both host and bacterial factors play an important role in this variability [104, 350, 351]. Most of this variability has only been demostrated in distantly related *Mtb* strains with defined characteristics of hyper or hypovirulence [89, 200, 218, 222, 223]. It can be deduced from these studies that biological diversity in closely related *Mtb* strains is assumed to be small, but is in general, unknown.

The S75 group of *Mtb* strains was identified in an epidemiologic study of tuberculosis clinical isolates from New Jersey. In that study, 22% (381) of strains were determined to have a low number of copies of the insertion fragment IS6110. Fifty six pansusceptible

strains were further grouped on the basis of their unique spoligotype, CDC0075 (octal code, 777776407760601) and this cluster was nominated S75. This group comprised 3 strains, each with a distinct IS6110 pattern. Strains were arbitrarily labeled BE (n=41 isolates), H6 (n=13 isolates), and C28 (n=2 isolates), corresponding to 1, 2, and 3 IS6110 copies, respectively [352]. Finally, it was noted that in contrast to other isolates with a single or small IS6110 copy number which are consistently members of genetic group 1 [13], the S75 cluster belongs to principal genetic group 2, Lineage IV.

In order to determine if closely related *Mtb* clinical isolates present phenotypic differences in the context of virulence and pathogenicity, the three *Mtb* strains comprising the S75 group, in addition to the clinical isolate *Mtb* CDC1551, were studied using the guinea pig model of TB infection. This animal model is supposed to resemble human's immunology and pathology to a great extent and provides the convenience of working with small animals [104, 203, 221, 223]. The guinea pigs inherent susceptibility to infection with *Mtb* may highlight subtle differences in the biological behavior of genetically very closely related *Mtb* strains.

3.2. Hypotheses

Mtb strains BE, C28 and H6 are very closely related. Based on molecular data, it has been proposed that the stepwise acquisition of IS6110 from *Mtb* BE to *Mtb* H6 to *Mtb* C28 resulted from two single genetic events with a probable evolutionary time frame of only one decade [352]. The variation at the genomic level of these strains is small, explained by their low copy number of the insertion fragment IS6110, which has been associated with major genomic variations in clinical isolates [252].

Based on this information, we hypothesize that closely related *Mtb* strains comprising the S75 group will not present significant differences in their virulence as measured by time to death of infected guinea pigs.

Mtb CDC1551, which has four copies of IS6110, also belongs to principal genetic group 2 but to a different lineage (V) and it is genetically more distantly related to the S75 group. *Mtb* CDC1551 has been catalogued as very transmissible, yet hypovirulent [220, 223, 353]. Therefore, we hypothesize that this strain is less virulent than members of the S75 group, which in contrast seem to be very successful in causing disease [352]. We hypothesize that the possible lower virulence of *Mtb* CDC1551 will be demonstrated by a significantly longer survival time of guinea pigs infected with this *Mtb* strain in relation to the S75 group of *Mtb* strains.

3.3. Materials and methods

3.3.1. *Mtb* strains and culture conditions

Cultures for guinea pig infection were prepared as follows. One ml glycerol stock of each strain was inoculated in 50ml of Proskauer & Beck media (5g KH2PO4, 5g asparagine, 0.6g MgSO4.7H2O, 2.5g Magnesium citrate, 20ml glycerol, 1L water, pH 7.8). Cultures were incubated without agitation at 37°C for approximately 3 months. Then, the pellicle or film formed by virulent cells on the surface of the media was transferred to 50 ml of 7H9/OADC/0.05% Tween and incubated for 10-15 days in agitation at 37°C. Three days before infection, each culture was transferred to new 50ml of media to a final concentration equal to an O.D600 of 0.1. This results in an O.D600 of 0.2 by the time of

infection. Finally, 25ml of this culture were harvested and resuspended in 10ml of sterile PBS. This suspension was used directly to perform the aerosol infection of guinea pigs.

3.3.2. Low dose aerosol infection of Guinea Pig

Out-bred female Hartley guinea pigs weighing 450-500 grams were purchased from Charles River Laboratories (Wilmington, MA). Guinea pigs were maintained under ABSL-3 barrier conditions in isolator cages (Thoren, Hazleton PA). Bacterial suspensions describe above were used to infect 18 guinea pigs (Male Hartley) per strain by the low aerosol method with a Madison Chamber [354]. In addition, a small aliquot of the suspension was plated in 7H11 media to estimate the total number of cells used in each infection.

After infection, individuals in each group were weighted daily and checked for signs of disease. The Karnofsky scale for pain and distress was used to evaluate the well being of each guinea pig. Guinea pigs with a score of 8 or more were euthanized. In addition, five guinea pigs per group were sacrificed at days 30 and 60 post infection (p.i.) to obtain measurements of total colony forming units (CFUs) in the lung and spleen. CFUs were determined by plating organ homogenates onto nutrient 7H11 agar supplemented with OADC. Colonies were enumerated after 21-days incubation at 37°C. The remaining guinea pigs were followed for survival. All experimental procedures were approved by the Colorado State University Institutional Animal Care and Use Committee.

3.3.3. Statistical analysis

Guinea pig survival was plotted using the Kaplan-Meier method [355], and differences between curves were analyzed using the log-rank test [356]. Pathology scores were analyzed by the Kruskal-Wallis test using SAS. Differences in the CFU counts at day 30 and 60 p.i. were evaluated by the t-student test.

3.3.4. Lipid extraction

Lipids were extracted from lyophilized cells 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.

3.3.5. Thin layer chromatography (TLC)

3.3.5.1. 1D TLC

Lipids in each sample (100μg) were fractionated in a 10 by 10cm aluminum-backed TLC plate (EM Science) using a TLC tank (Kontes) as describe elsewhere [357, 358]. Different solvent systems were used, including: Chloroform: methanol: water (65:25:4 vol:vol:vol), Chloroform: methanol: water (60:20:4 vol:vol:vol), Chloroform: methanol: water (60:30:6 vol:vol:vol) and Cyclohexane: ethylacetate (3:2 vol:vol) and Toluene: acetone (95:5 vol:vol).

After separation lipids were visualized with either Charring spray (1% cupric sulfate, w/v, 8% phosphoric acid, v/v) for total lipids or α-Naphtol spray (5g of α-Naphtol, 25ml of sulfuric acid in 350ml of ethanol) for glycolipids.

3.3.5.2. 2D TLC

Two dimensional TLC was performed as previously described elsewhere [358]. Two different solvents systems were used: 1). 1D: Chloroform:methanol:water (60:30:6 vol:vol:vol), 2D: Chloroform:acetic acid:methanol:water (40:25:3:6 vol:vol:vol:vol) and 2). 1D: Petroleum ether:acetone (92:8 vol:vol) three times, 2D: Toluene:acetone (95:5 vol:vol). After separation, lipids were visualized as described above.

3.4. Results

3.4.1. Assessment of S75 group virulence in the Guinea Pig model

Guinea pigs were infected with each *Mtb* strain (BE, C28, H6, CDC1551) by the low dose aerosol method with a suspension corresponding to 6 x 10^6 CFU/ml in the nebulizer chamber. Each strain was previously grown as a pellicle which selects for organisms that proliferate and release increased amounts of TDM from their surfaces, thus selecting for virulent bacilli [359]. This, in addition to the natural susceptibility of guinea pigs to TB infection allows the identification of biological differences between TB strains that are closely related to each other.

Analysis of survival rates was performed by the Kaplan Meier method and the long rank test. Analysis of survival curves of guinea pigs infected with the four different *Mtb* strains revealed that strain BE infected animals had a significantly higher survival rate when the survival curve of this *Mtb* strain was compared to survival curves of guinea pigs infected with *Mtb* C28 and *Mtb* H6 (p-value = 0.019 and 0.008, respectively) (Figure 3-1). The survival curve from guinea pigs infected with *Mtb* CDC1551 was not

significantly different compared to those infected with *Mtb* C28 and *Mtb* H6 (p-values = 0.930 and 0.975, respectively). However, when the long rank test was applied to compare survival curves of *Mtb* CDC1551 vs *Mtb* BE, the difference was not statistical significant (p-value = 0.063) (Figure 3-1). This was unexpected because the median survival time (MST) for guinea pigs infected with *Mtb* CDC1551 (36 days) was smaller than MST of *Mtb* C28 (43 days), *Mtb* H6 (43 days) and *Mtb* BE (50 days). However, the maximum day of survival for guinea pigs infected with *Mtb* CDC1551 was 68, whereas the number was 57 and 62 for animals infected with *Mtb* C28 and *Mtb* H6, respectively. Maximum day of survival for guinea pigs infected with *Mtb* BE was 71. The differences between early and late time points in the survival curves of these *Mtb* strains may explain the lack of statistical significance between survival curves of *Mtb* CDC1551 and *Mtb* BE because the long rank test analyzes the complete shape of the survival curve and gives equal weight to all time points.

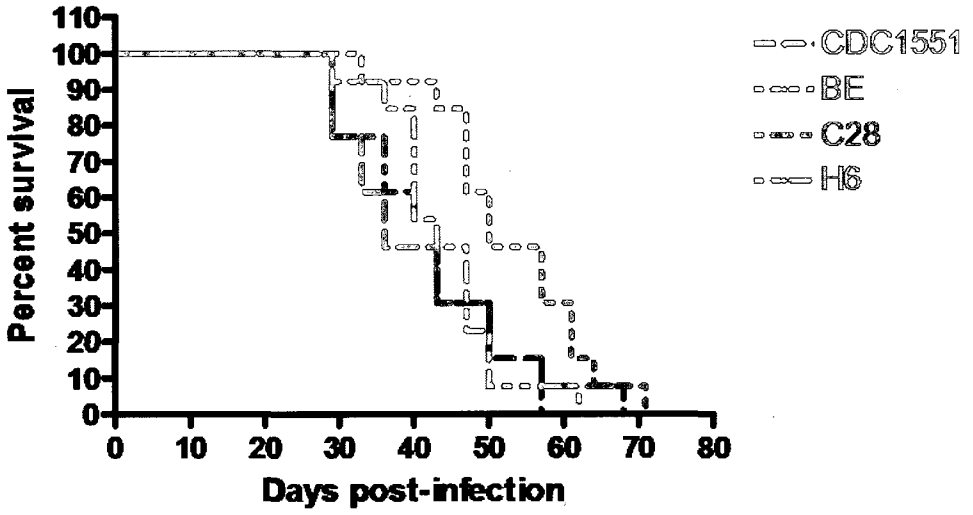
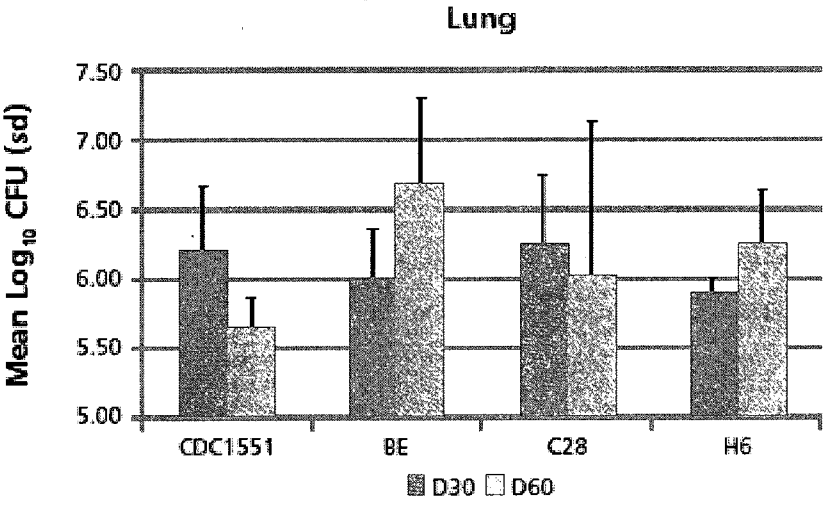


Figure 3-1. Kaplan-Meier plot of the survival of guinea pigs infected with *Mtb strains* CDC1551, BE, C28 and H6.

To avoid any bias, guinea pigs for each group (30 days p.i.; 60 days p.i.; and survival) were pre-selected during the experimental design before the low-dose aerosol infection was performed. The majority of the guinea pigs that were pre-selected to be euthanized at day 60 p.i. died before reaching this day and were included in the survival analysis. When CFU counts at day 30 and day 60 p.i in lung and spleen were evaluated, no statistical differences were observed between any of the *Mtb* strains. However, there was a trend for the CFUs of the BE strain to be higher than other strains at day 60 p.i. in both lung and spleen. There was also a trend for the CFUs of BE at day 60 p.i. to be higher than CFUs of this strain at day 30 p.i (Figure 3-2).

In an effort to determine if differences in survival rates were due to greater tissue damage produced by strains C28 and H6, pathological examination of lungs, spleen and liver at days 30 and 60 p.i. were performed. No significant differences were observed between strains at day 30 p.i. in any of the organs (Data not shown). Statistical differences between CDC1551 and BE were observed when total pathological score for the lungs was evaluated at day 60 p.i. ($p<0.05$) (Figure 3-3). No significant differences between any of the strains were observed in the pathology of spleen and liver. When each individual pathological parameter in each organ was statistically evaluated, some significant differences ($p<0.05$) were observed between at least two of the four strains for some of the parameters, however, no pattern that could be related to differences found in virulence was determined (data not shown). However, a general pattern in which BE shows diminished lung involvement, secondary lesions, mineralization and fibrosis at day 30 p.i. was observed (Figure 3-4).

A



B

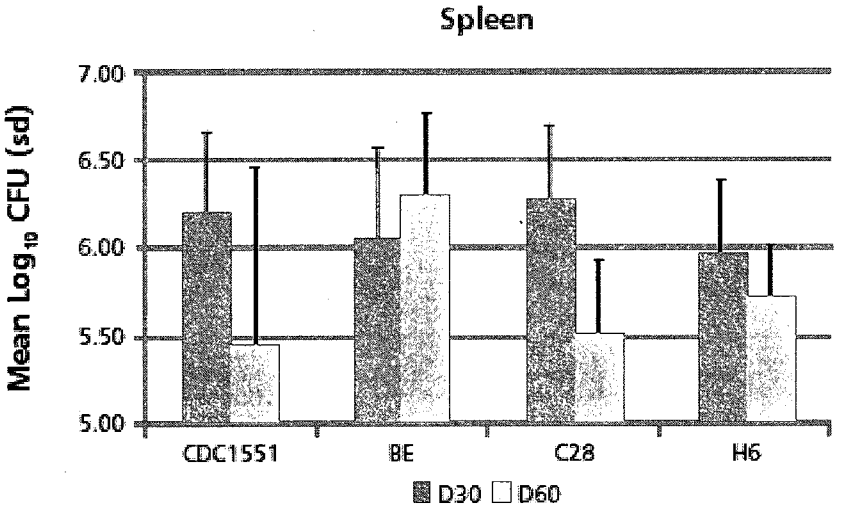
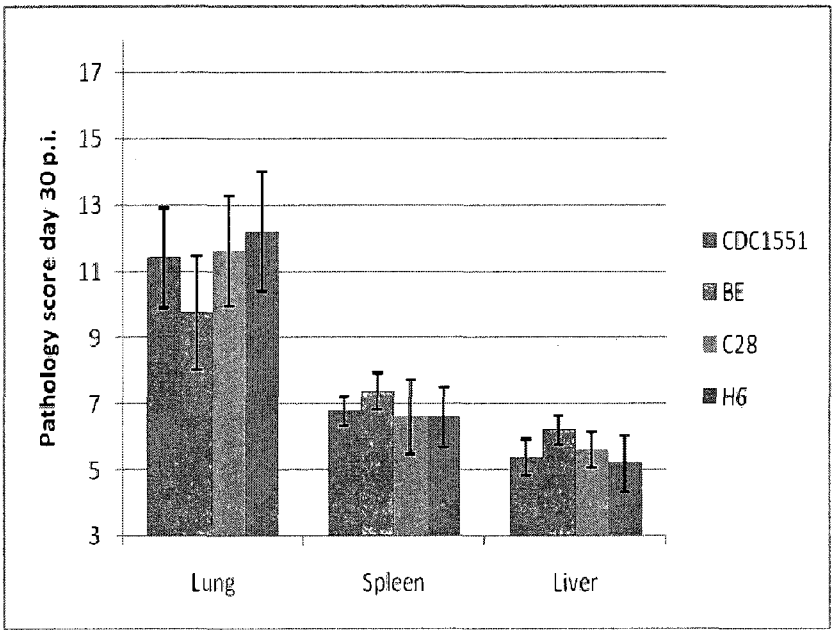


Figure 3-2. Colony forming units in lung (A) and spleen (B) of guinea pigs infected with each *Mtb* strain at days 30 and 60 post-infection.
Initial inoculum 6 x 10^6.

A



B

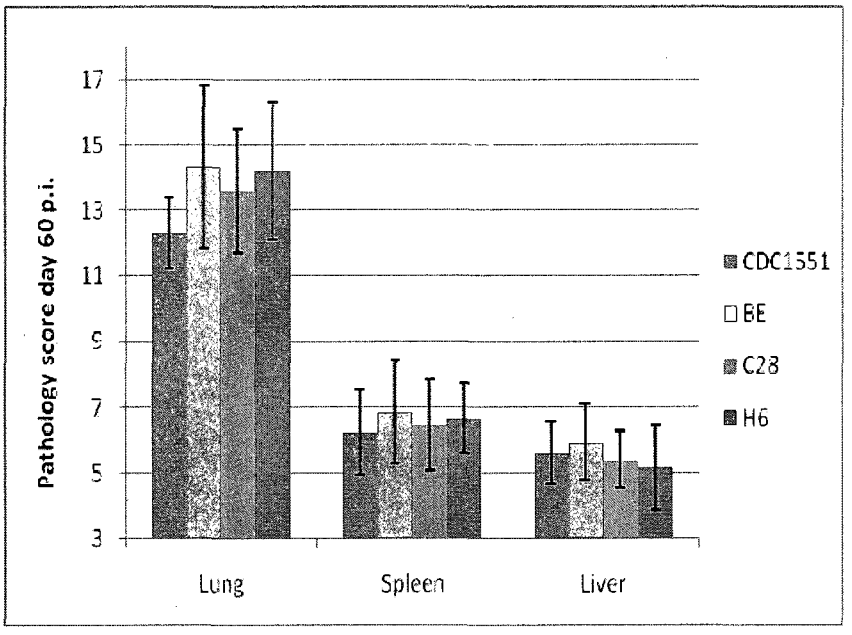
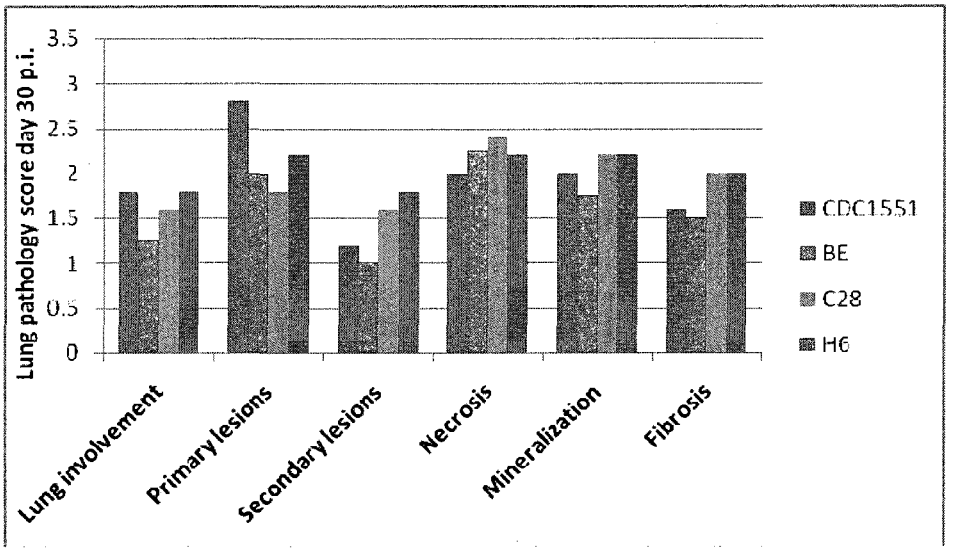


Figure 3-3. Pathological assessment of lung, spleen and liver after infection with each *Mtb* strain.
A: day 30 p.i. and **B:** day 60 p.i.

A



B

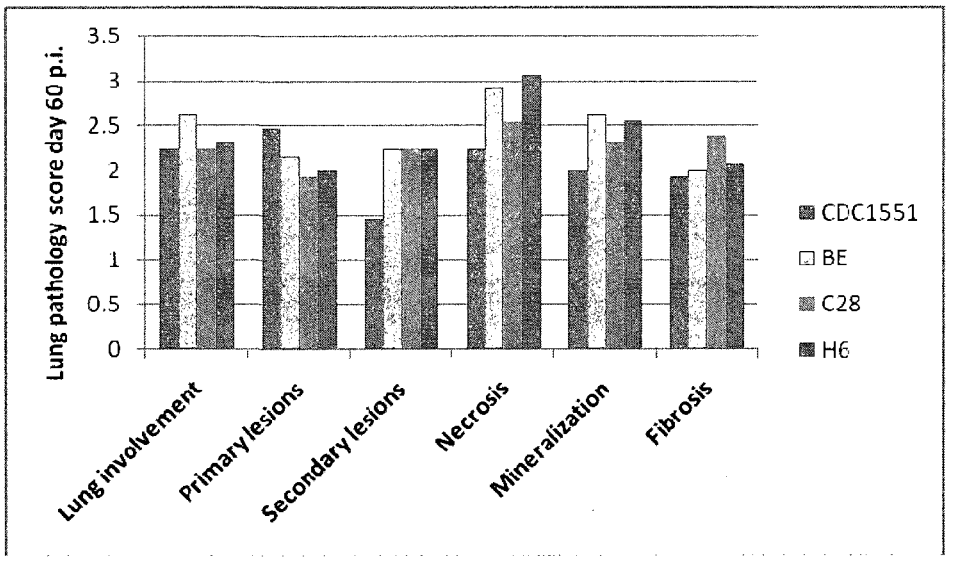


Figure 3-4. Pathological assessment of lung itemized by each evaluated parameter.
A: day 30 p.i., **B:** day 60 p.i.

3.4.2. Lipid analysis

No differences were observed between the lipid component of the strains by 1D TLC (Figure 3-5) or 2D TLC (Figure 3-6), at least with the pool of solvent systems used in this study.

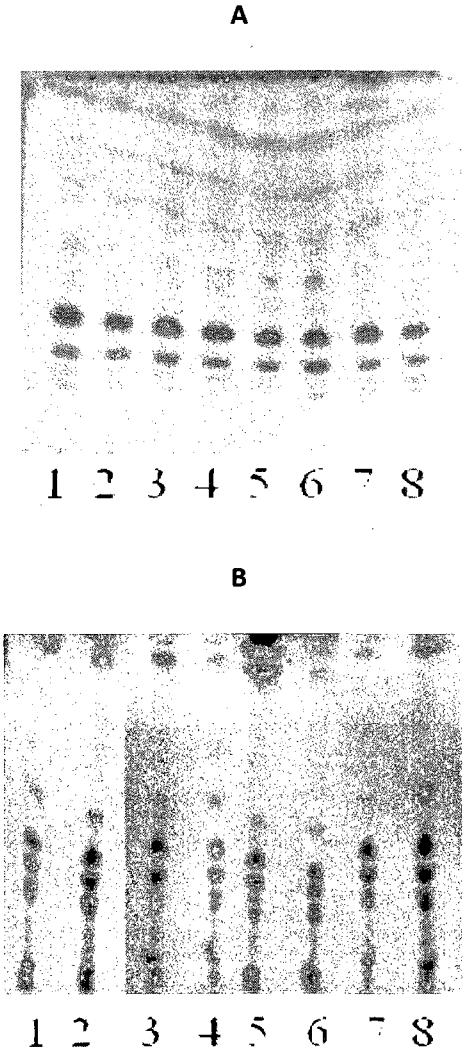


Figure 3-5. Example of similarities in lipid component between *Mtb* strains by thin layer chromatography.
A. 2:1 fraction (Total lipids) and **B.** 10:10:3 fraction (LAM, PIMs). **Lanes 1, 2**: CDC1551; **Lanes 3,4**: BE; **Lanes 5,6**: C28; **Lanes 7, 8**: H6. Solvent system: Chloroform:Methanol:Water (65:25:4 vol:vol:vol)

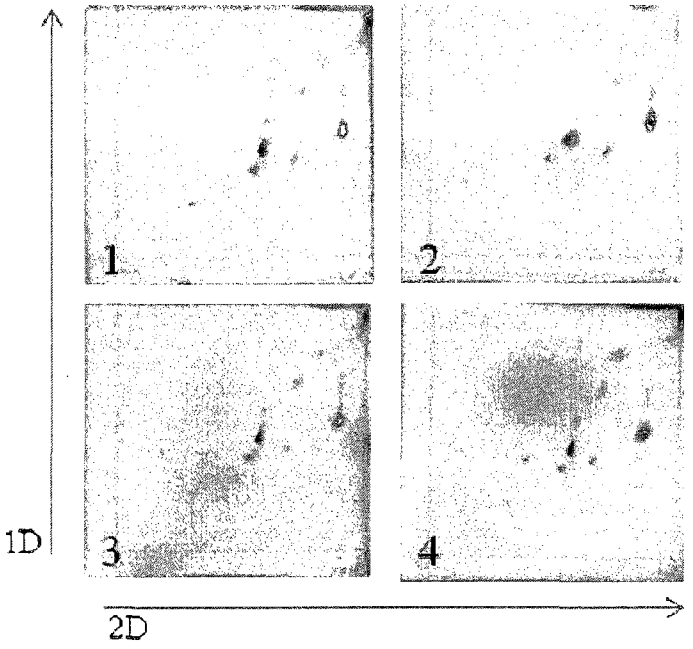


Figure 3-6. Example of similarities in lipid component between *Mtb* strains by 2D-TLC of total lipids (2:1 fraction).
1: CDC1551; **2:** BE; **3:** C28; **4:** H6. Solvent system: 1D=Chloroform:Methanol:Water (65:25:4 vol:vol:vol); 2D= Chloroform:acetic acid: methanol:water(40:25:3:6 vol:vol:vol:vol)

3.5. Discussion

The S75 group was identified in an epidemiologic study of clinical isolates from a TB cluster in New Jersey [352]. This group of strains was identified based on its unique spoligotype (CDC0075,octal code, 777776407760601) and fifty six pansusceptible strains were found to belong to this group. In addition, strains from this group have a very low copy number of the insertion fragment IS6110 and were further arbitrarily classified as BE, H6 and C28 corresponding to 1, 2, and 3 IS6110 copies, respectively. It was found that of these strains *Mtb* BE was the most common strain, identified in 73.2% of the isolates, followed by *Mtb* H6 and *Mtb* C28 (23.2% and 3.6% respectively).

Several animal models can be used to study and measure the virulence of different *Mtb* strains [104, 203, 221, 223]. In general, the mouse model is the most widely used not

only because it is generally a low cost model, but also because it offers several advantages including a significant number of existing reagents, extensive literature on mouse immunology and the availability of Knock-out animals for most of the key genes involved in the immune response to pathogens [360]. Mice as well as rabbits have an inherent tolerance to TB infection, and is a reason why these two animal models have been postulated to mimic the human infection in immunocompetent individuals [91, 361]. On the other hand, it has been proposed that humans are more susceptible to TB infection than mice and rabbits [221], which has led to some controversy about the extent to which mice mimic tuberculosis infection in humans. Therefore other animal models, such as the guinea pig, have been used in the study of TB.

Guinea pigs, in contrast to mice and rabbits, are believed to be very susceptible to TB infection [360, 361]. This is supported by the fact that while guinea pigs generally succumb to the disease, mice develop a chronic infection that unlike humans is characterized by a large number of bacilli in lungs and spleen [360, 362]. In addition, induction of DTH in mice requires a higher dose of PPD compared to humans or guinea pigs [363]. Guinea pigs develop not only primary granulomas, but also intragranulomatous necrosis and caseation as seen in humans [364-366]. Intragranulomatous necrosis in TB is the origin of tissue destruction and extracellular bacilli and is the source of latent mycobacteria [93]. Additionally, its progression and liquefaction is the origin of cavitation, the primary source of infecting bacilli [367].

Thus, the guinea pig model offers several advantages over other animal models. It resembles to a great extent human's immunology and pathology and provides the convenience of working with small animals.

The utility of the guinea pig model to differentiate between different degrees of virulence of closely related *Mtb* strains was demonstrated in the present study. Unexpectedly, and in contrast to our initial hypothesis, when the virulence of the S75 group of *Mtb* strains was evaluated in this model using a low dose aerosol infection, it was determined that strains *Mtb* C28 and *Mtb* H6 were more virulent than *Mtb* BE in terms of survival rate of infected guinea pigs (Figure 3-1). Differences between *Mtb* BE and *Mtb* CDC1551 were not statistically significant, however, at five weeks post infection, only 50% of GP infected with *Mtb* CDC1551 were alive, compared to 90% in the BE group, suggesting that *Mtb* CDC1551, together with C28 and H6 might be more successful in the progression of disease.

Virulence of these strains, as measured by the survival time of infected guinea pigs was estimated using the Kaplan-Meier method [355] and the log rank test [356]. In contrast to other statistical analyses in which the parameters to be measured to test a hypothesis are discrete numbers (i.e. means, standard error, etc), analysis of survival rates are based in the entire shape of a survival curve, also referred to as the survival function. The Kaplan Meier method was initially developed to allow the inclusion of data obtained from subjects that did not reach the end of the study (censoring) due to unexpected reasons [355]. However, even if censoring is not an issue, the Kaplan-Meier estimator allows the calculation of the probability of survival between time 0 and each of the time points where an event occurred, in this case, death. This at the end will be representative of the entire survival curve. The intrinsic susceptibility of guinea pigs to *Mtb* infection allows the assessment of differences in survival curves between different virulent *Mtb* strains. In some cases, for example during vaccine efficacy studies, survival rates at specific time

points can be evaluated by different statistic methods such as the Fisher exact test and the Chi-square test [368-370]. However, even in these types of studies, survival time is generally estimated using the Kaplan-Meier method and the log rank test [368, 370-372]. To determine if the survival function for each strain was significantly different from each of the other strains, the log rank test was performed. This test calculates the variance for the Kaplan-Meier estimator and determines the probability that the differences in survival function of two groups are due to chance. The long rank test calculates this probability at each time point and then the values are added, so that the final calculation is representative of the entire curve. In this study, survival functions for *Mtb* C28, *Mtb* H6 and *Mtb* CDC1551 were not statistically significant. However, survival functions were significantly different only between *Mtb* C28 and *Mtb* H6 versus *Mtb* BE, but not between *Mtb* BE and *Mtb* CDC1551. This discrepancy is explained by the fact that the log rank test gives equal weight to all time points, and therefore, the differences in the maximum survival time between *Mtb* CDC1551 (68 days p.i.) and *Mtb* C28 (57 days p.i.) and *Mtb* H6 (62 days p.i.) influence the final statistical significance of each of the survival curves of these strains when compared to *Mtb* BE (maximum survival time: 71 days p.i.). Differences in two or more survival functions can be estimated using the Gehan-Breslow-Wilcoxon test, which in contrast to the log-rank test, gives more weight to deaths at early time points; however this test requires one group to have a consistently higher risk than the other. Since our initial hypothesis was that no differences in survival times of guinea pigs infected with closely related strains would exist, the study design used the log-rank test instead of the Gehan-Breslow-Wilcoxon test. Similar studies in *Mtb* and the guinea pig model also use the log-rank test and Kaplan-Meier estimator to

evaluate differences in virulence between strains [373, 374]. Additionally, studies to assess protection after vaccination and subsequent challenge with *Mtb*, which theoretically should present differences in risk between groups, are usually also based in the log-rank test to estimate statistic significance between groups [370, 371, 375]. Another way to compare survival rates between groups is by using the median survival time. Median survival time is useful in long-term survival studies or when one or all of the groups in the study do not reach 100% mortality. In the mycobacteria field, median survival time is commonly applied in virulence studies involving the mouse model of TB infection [376-379]. However, time to death outcomes are most appropriately analyzed by survival function which, as mentioned above are based on the entire shape of the survival curve [380].

Finally, confidence in the results obtained in this study is highlighted by the number of guinea pigs used to determine the survival function. Similar studies using the guinea pig model of infection reach conclusions using a standard number of 5 - 10 guinea pigs per group [371, 373, 374]. In the present study, the survival function for each of the strains analyzed was calculated using data obtained from 13 animals per strain.

As mentioned previously, each animal model of TB infection has unique advantages and drawbacks. In this study the guinea pig model was chosen over the mouse model based on the high susceptibility of guinea pigs to TB infection, desiring to highlight subtle differences that closely related strains of *Mtb* may present in terms of virulence and pathogenicity.

Even though the mouse and guinea pig models of TB infection present marked differences, results presented in this study, in which *Mtb* BE seems to present a lower

virulence than *Mtb* C28 and *Mtb* H6, have been validated using the mouse model (B. Kreiswirth, personal communication, unpublished results).

Several studies have shown that more virulent strains grow faster than less virulent strains, and in some cases, this difference in growth has also been suggested to be associated with higher transmission rates [223, 381].

Concordantly, it was expected that a higher number of CFUs would be found in the lungs and spleens of guinea pigs infected with strains *Mtb* C28 and *Mtb* H6, correlating to a decrease in survival rates. Differences observed between survival rates between strains were not associated with CFU in either lung or spleen at days 30 and 60 p.i. Evaluation of the lung pathology caused by these strains showed a trend in which pathological scores for guinea pigs infected with *Mtb* BE are lower in terms of lung involvement, secondary lesions, mineralization and fibrosis at day 30 p.i. when compared to *Mtb* C28 and *Mtb* H6. Interestingly, the lung pathology caused by *Mtb* BE seems to increase at day 60 p.i., while the pathology caused by the other strains remains stable (Figure 3-4). Miscorrelation between survival rates and CFUs has been shown previously. Kaushal and colleagues (2002) demonstrated longer survival rates of mice intravenously infected with an *Mtb sig*H mutant versus those infected with the wild type [377]. The CFU from infected organs between the two groups was not significantly different. Rather, the *Mtb sig*H mutant showed a reduction in immunopathology in the lung and spleen of infected mice. This was found to be associated with a reduced recruitment of T cells to the site of the infection. The authors hypothesized that the failure to recruit T cells in the mice infected with the *Mtb sig*H mutant was due to a lack of virulence factors that induce pro-inflammatory responses, resulting in longer survival rates [377]. This previous study,

combined with our observation that guinea pigs infected with *Mtb* BE demonstrate higher survival rates compared to *Mtb* C28, *Mtb* H6, and to a some extent *Mtb* CDC1551 infected guinea pigs, suggests that specific molecular determinants play a role in how *Mtb* interacts with the host, resulting in the higher virulence of *Mtb* C28 and *Mtb* H6.

Therefore, in an effort to determine if some molecular components, specifically lipids, could be related to the differences observed between these *Mtb* strains in the guinea pig model, lipid profiles of S75 members and *Mtb* CDC1551 strains were studied.

No differences in the lipid profiles of any of the strains, including *Mtb* CDC1551, were observed (Figure 3-5, Figure 3-6). This was surprising given that mycobacterial lipids have been implicated as virulence factors in TB and also play a role in host–pathogen interactions [76, 104, 382, 383]. In addition, extractable polar lipids from *Mtb* CDC1551 cell wall have been associated with the ability of this strain to induce high level of Th-1 cytokines in infected mouse lungs and in cultured human monocytes [220, 224] and variability in lipid components have been also demonstrated in different *Mtb* strains [104, 220, 265].

Since differences in lipid profiles could not be identified between strains, we explored how proteins might be related to the increased virulence of *Mtb* C28 and *Mtb* H6 in comparison to *Mtb* BE, and will be the subject of study in the next chapter.

3.6. Conclusions

Identification of *Mtb* clinical isolates with different degrees of virulence and transmission represents an important step towards the characterization and identification of

determinants of virulence. In this study, different degrees of virulence were observed between closely related *Mtb* strains.

Guinea pigs infected with *Mtb* strains C28 and H6 had a lower survival rate than guinea pigs infected with *Mtb* strain BE, indicating that while all three strains were virulent, *Mtb* strains C28 and H6 possessed a higher virulence phenotype based on guinea pig mortality.

Based on previous epidemiological studies [352], *Mtb* BE is the most prevalent strain in the New Jersey cluster from which this group of strains were identified. This conflicts with its reduced virulence in the guinea pig model, as measured by survival rates. It is possible that the reduced morbidity and mortality in the guinea pig model may be correlated to a milder disease in humans, leading to delays in seeking medical help and retention of infectious persons among contacts, with subsequent increased likelihood of spreading disease. Continued studies in epidemiology and the host response to infection with each of these strains are required to explore this hypothesis. Particularly, it would be interesting to compare the symptomatology of TB patients infected with each of these strains and the estimated time from onset of symptoms to diagnosis and treatment. If our hypothesis is correct, we would expect to find that patients infected with *Mtb* strain BE present milder symptoms and the time before seeking medical help would be longer than for patients infected with *Mtb* strains C28 and H6.

The presence of at least one known risk factor to acquire and develop TB was present in 66% of the patients from where the S75 strains were initially identified [352]. However there was no information regarding the number of patients with these factors in relation to each of the three *Mtb* strains, possibly due to the small number of patients infected with

strain *Mtb* C28. Future epidemiological studies to estimate a possible correlation of TB infection for each of the S75 *Mtb* strains with the presence of factors affecting an appropriate immune response (i.e HIV coinfection, malnutrition) would be worth of study. Finally, possible differences in transmission of these *Mtb* strains should be evaluated by performing a follow-up of PPD conversion rates for household contacts of the S75-infected patients, an indirect measurement of the transmission and infectivity of these strains. This type of study might be difficult to perform in the U.S. due to the low rates of TB infection, and perhaps only feasible in the case of an outbreak involving this group of strains. It is unknown is this group of *Mtb* strains has a global distribution, but transmission studies involving household contacts in a setting with a high transmission rate of TB would be very informative.

In the present study, differences in the degree of virulence of each *Mtb* strain did not correlate with the number of CFUs in either lung or spleen, suggesting that inherent molecular components produced by each strain could be responsible for the differences observed in the guinea pig model, and may be related to specific host-pathogen interactions. Similar lipid profiles were observed in each strain, indicating that the differences observed in virulence are probably not related to a particular presence or absence of certain lipids in the more virulent strains *Mtb* C28 and *Mtb* H6. Further studies utilizing more sensitive techniques to evaluate the lipid component of each strain, such as LC-MS-MS profiling, are possible. Other possibilities include the analysis of protein levels between strains in the context of the differences observed in the guinea pig model. This will be the center of study in the next chapter of this dissertation.

4. Chapter IV. Closely related clinical isolates of *Mycobacterium tuberculosis* with different degrees of virulence present protein expression variability

A modified version of this chapter has been submitted for publication in the Molecular and Cellular Proteomics journal with the title: **Descriptive proteomic analysis shows protein variability between closely related clinical isolates of *Mycobacterium tuberculosis*.** Carolina Mehaffy, Ann Hess, Jessica Prenni, Barun Mathema, Barry Kreiswirth, Karen Dobos.

4.1. Introduction

Evaluation of protein expression in both qualitative and quantitative terms offers important information on different aspects of cell physiology and biology with implications in medicine, infectious diseases and cell development.

Throughout 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.

2D gel electrophoresis (2D-GE) 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 2D-GE, 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 [294] and more recently, 8 [384] different samples in a single mix. iTRAQ labeling has been successfully used for the analysis and identification of cancer biomarkers [385, 386], virulent factors and other molecular determinants, as well as host responses to pathogenic bacteria [387-389].

Several studies have been performed in order to characterize the proteome of *Mtb*. Most of these studies have been done using laboratory reference strains of *Mtb*. The *Mtb* secreted proteome has been studied thoroughly by both 2D-GE [249, 275, 276, 314, 315, 317, 318] and shotgun strategies [313] . This last approach when complemented with 2D-GE provided the most complete analysis of *Mtb* proteins, from which 257 proteins were identified, therefore increasing the annotation of exported proteins in *Mtb*. Cytosolic and cell wall proteomes [315, 319] as well as whole cell lysates [276, 320] and membranes [319, 390] have also been studied, although to a lesser extent. Similarly 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 [319].

In addition to the broad characterization of *Mtb* proteomes, the identification of protein expression differences between *Mtb* strains is also valuable and can complement the development of vaccines, serodiagnostic tests and choice of drug targets [311].

Proteomic analysis with special focus on *Mtb* clinical isolates may provide clues to pathogenicity and prevalence of *Mtb* strains. The comparison of *Mtb* clinical isolates with different degrees of virulence may offer a newer perspective in the identification of virulent determinants than more classical studies in which completely avirulent and fully virulent strains are compared. The analysis of proteomes of clinical isolates presenting different levels in their virulence and pathogencity, although complex, provide a more realistic insight to understand differences in disease outcomes encountered during natural TB infection.

Despite their genetic relatedness, we demonstrated that closely related *Mtb* strains (S75 group) present significant differences in their virulence in the guinea pig model of TB infection. To evaluate if these differences are reflected in variability in the proteome of these strains, we used both 2D-GE 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 [352]. Three of these strains belong to a previously characterized cluster called S75 [352], and the remaining is the well studied CDC1551 strain.

4.2. Hypotheses

Proteomic studies on distantly related strains indicate that though phenotypic variation at the proteins level exists, it is very low. From these studies, it can be deduced that the protein variation between *Mtb* isolates is small [249, 311]. This, in addition to the relatively small genetic variability of *Mtb* [13, 239], suggests that the closely related *Mtb* strains BE, C28 and H6 may not present major phenotypic differences in terms of their

protein profiles. However, findings related to the virulence of these isolates indicate that these strains present different degrees of virulence in the guinea pig model. Therefore, we hypothesized that even though the number of differential proteins between the closely related strains could be small these differences might reflect the phenotypic biological variability that these *Mtb* strains demonstrate in the guinea pig model. Specifically, it is expected that proteins involved in virulence and survival, as well as host-pathogen interactions, present higher levels in the more virulent strains *Mtb* C28 and *Mtb* H6 when compared to *Mtb* BE. It is also expected that these few differences will be statistically significant.

Findings related to the survival rates and pathology induced by these strains should provide important information leading to hypotheses generation regarding the role of differential proteins in the context of virulence and disease outcome. This in turn would provide a precedent for future studies with a focus on functional and biological characterization of selected differential proteins. It is important to note that some differences in protein expression may only occur in vivo and not in vitro. Similarly, some of the variability in protein expression may be due to intrinsic characteristics of in vitro growth, and represents the first pre-validation step towards the identification of molecular components with a possible involvement in different levels of virulence.

The survival curve of guinea pigs infected with strain *Mtb* CDC1551 was similar to those of *Mtb* C28 and *Mtb* H6, but it was not significantly different from *Mtb* BE. In accordance, the focus of this study will be to evaluate the differential protein expression in the closely related strains BE, C28 and H6, in an effort to identify proteins with a potential involvement in the different degrees of virulence of these strains. However,

comparison of protein levels between the closely related strains and the more distantly related *Mtb* CDC1551 is interesting and may reflect some phenotypic variation between distinct *Mtb* lineages (for example IV vs V). If closely related *Mtb* strains BE, C28 and H6 are taken as a group and compared to *Mtb* CDC1551, it is expected that the number of proteins presenting significantly different levels will be higher than the number of differential proteins when the closely related strains are compared to each other.

4.3. Materials and Methods

4.3.1. *Mtb* strains and culture conditions

Glycerol stocks of *Mtb* clinical isolates BE, C28 and H6, as well as *Mtb* 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. Then, cells were washed twice with sterile PBS (Invitrogen) and inoculated in 1L of GAS media [391]. Cultures were then incubated at 37°C in agitation for 4 weeks. All cultures were prepared in triplicate.

4.3.2. 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 [326]. Briefly, 1L culture filtrates were concentrated to approximately 25ml of volume, using a 10KDa MWCO membrane (Millipore). Then, the filtrate was further concentrated to approximately 300µl using an Amicon Ultracell- 15 with a 10KDa MWCO by centrifugation at 3000 rpm,

4°C. After this, proteins were subjected to buffer exchange three times using 10mM Ammonium bicarbonate. Proteins were kept at 4°C until quantification by the BCA assay (Pierce) and then at -20°C for longer storage.

4.3.3. 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 100,000 x g 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.

4.3.4. 2D gel electrophoresis (2D-GE)

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 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. 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. 2D 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.

4.3.5. In gel digestion of proteins

Spots from 2D gels were excised and destained in 60% acetonitrile, 0.2M Ammonium bicarbonate at 37°C for 30min, three times. Gel pieces were covered with 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.

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

100µg of CFP, and 25ug of cytosolic proteins of each strain were labeled using the iTRAQ system (Applied Biosystems) following manufacturer's instructions. Briefly, samples were resuspended in 20µl of dissolution buffer (0.5M triethylammonium bicarbonate) followed by denaturation and reduction with 1µl of denaturing solution (2% SDS) and 2µl of reducing reagent (50mM tri-(2-carboxyethyl) phosphine (TCEP)) respectively, at 60°C for 1hr. Cysteines were blocked with blocking reagent (200mM methyl methanethiosulfonate (MMTS) in isopropanol) for 10min at RT. Then, proteins were digested overnight at 37°C with 10µg of trypsin. Samples were then labeled with iTRAQ reagents at RT for 1h. 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.

4.3.7. Mass Spectrometry

4.3.7.1. In gel digested proteins

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 x 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 3.0 software (Thermo Scientific) with an intensity threshold of 5,000 and 1 scan/group. Spectra were subjected to database interrogation against the *M. tuberculosis* genome (GenBank) using the Sequest algorithm. All searches were performed assuming trypsin digestion, with a 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 [392]. Protein identifications were only accepted if they could

be established at greater than 90% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [392].

4.3.7.2. iTRAQ labeled proteins

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, 75um x 43 mm analytical column, 5um, 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 Database search engine (Matrix Science) 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 29 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 [393].

4.3.8. Statistical analysis

4.3.8.1. 2D-GE

Normalized density of each spot of biological replicates for each strain was analyzed by the Student's t-test assuming equal variances.

4.3.8.2. 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 $\log_2$ 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. Model fitting was performed using the REML estimation method [394]. 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. Protein level estimates, standard errors and p-values were calculated for each comparison using least squares means [394].

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. Comparisons between each strain for each protein were performed and 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.

4.3.9. Western blot

Antibodies for western blot were obtained through the TB vaccine testing and research materials contract at Colorado State University with the exception of Cfp2 and Mpt64 polyclonal serum.

Serum against these two proteins was obtained after injection of Balb/c mice with 50 μg of recombinant Cfp2 or recombinant Mpt64 using Incomplete Freud's adjuvant (Sigma).

4.3.9.1. Chromogenic western blot

5ug of protein for each sample was resolved by SDS-PAGE and electro-blotted to nitrocellulose membrane. Membranes were probed with primary antibodies for 1 hr. Secondary antibody (Alkaline-phosphate conjugated Anti-mouse IgG or Anti-rabbit IgG; Sigma), was added and membranes incubated for 1 hr. Membranes were developed using 5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (Sigma).

4.3.9.2. Fluorogenic western blot

2.5ug or 5ug of each sample were resolved by SDS-PAGE as above and electro-blotted to a PVDF membrane (Hybond LFP, GE Healthcare). Western blot procedure was performed as recommended by the manufacturer. Secondary antibodies were Cy3 conjugated anti-mouse IgG and Cy5 conjugated anti-rabbit IgG. Blots were scanned in a Typhoon scanner (GE Healthcare) and bands were analyzed using ImageQuant™ TL software.

4.4. Results

4.4.1. 2D-GE of Culture Filtrate Proteins

2D gels for each biological replicate were analyzed and composite proteome maps representing each strain were created.

Approximately 120 spots were visualized in each sample. Twenty eight spots that presented some extent of variability between strains were selected for identification by MS/MS (Figure 4-1 and Table 4-1). Identification of the protein was achieved for the majority of the spots and the identity was the same regardless of the strain from which spot was originated, with one exception. Spot # 9 was identified as Rv1926c in strain *Mtb* 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 *Mtb* CDC1551 contains in fact both proteins, but Rv1926c is present in higher quantities. Similarly, Rv1926c could be

present in spot #9 of *Mtb* strains BE, C28 and H6 but in quantities that are not detectable by LTQ-MS/MS.

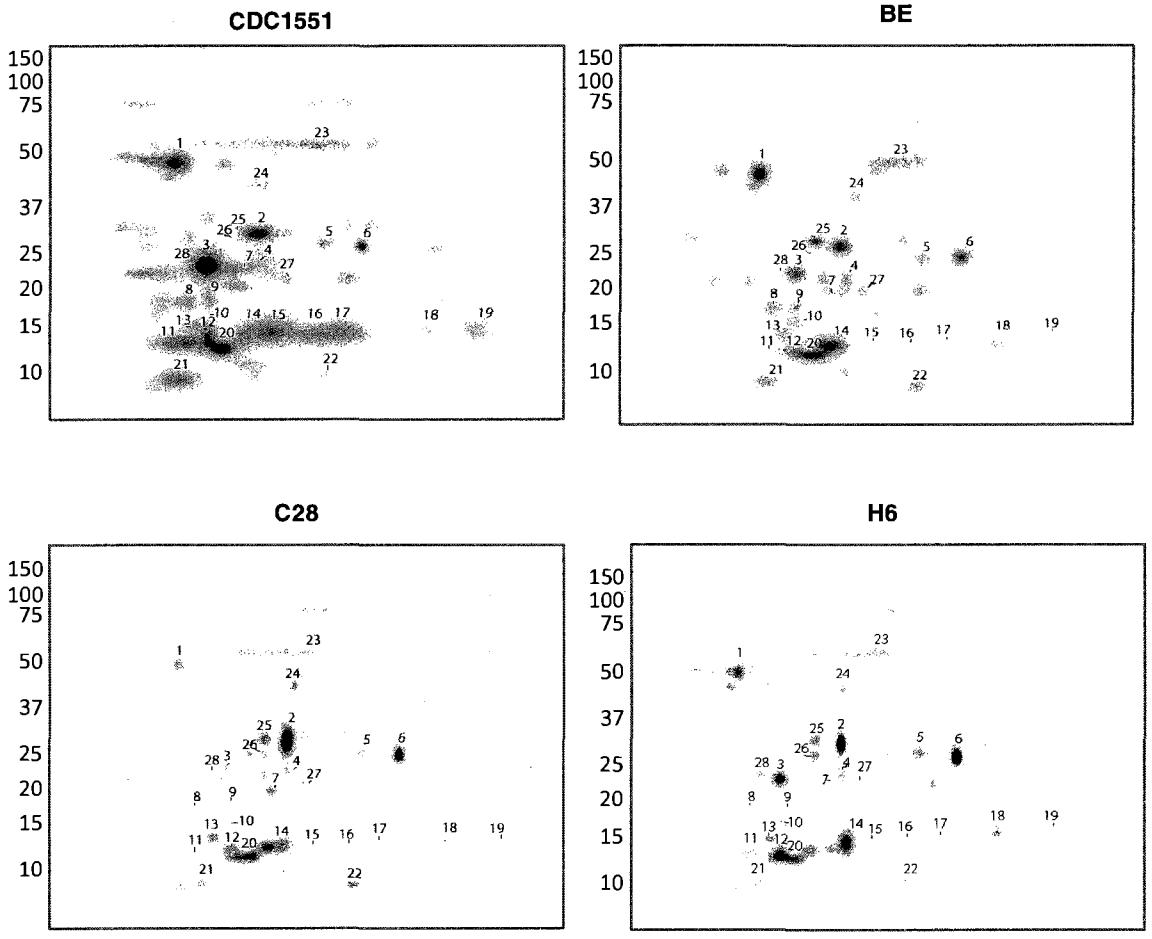


Figure 4-1. Composite 2D-GE maps from the secreted fraction of each analyzed *Mtb* strain. Composite maps were obtained from three biological replicates for each strain. Numbers correspond to identified proteins in Table 3-1. pH range is 4-7.

Fifteen of the spots were identified as more than one protein. In addition, several proteins such as SapM, Rv1906c, FbpC1, Cfp2, GroES, HspX, Cfp10 and Mpt64 were present in more than one spot, probably corresponding to specific isoforms of these proteins. In this regard, other authors have previously reported the presence of some proteins such as HspX and GroES, in multiple spots [249, 276, 311, 317].

In the present study, the protein with the most isoforms was the low molecular antigen Cfp2, also called Mbt12, which was present in 7 different spots, followed by HspX, GroES, and Cfp10 which were indentified in 3 spots each. Interestingly, the only spot identified as Cfp2 (Spot # 17) that did not contain any other proteins was only found in *Mtb* CDC1551. Puzzled by this finding, native Cfp2 was purified and resolved in a 2D-GE (Chapter V, Figure 5-8). Different spots were subjected to N-terminal sequencing and it was possible to determine that the Cfp2 isoform 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 *Mtb* CDC1551 lacks the first two amino acids (DP) from the mature protein and therefore the N-terminal sequence for that isoform 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 proteins presented significant differences between at least two of the four strains (Table 4-1).

Cfp2, Rv1810, Ag85B and a spot containing Mpt64 and LprA presented higher abundance in *Mtb* CDC1551 when compared to the closely related strains BE, C28 and H6. Meanwhile, Ag85A was absent from *Mtb* CDC1551 CFP. The lower abundance of Ag85A in the secreted fraction of *Mtb* CDC1551 was confirmed by western blot analysis of 2D-gels (data not shown).

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

Table 4-1. Proteins identified by Tandem MS/MS of selected 2D-GE spots and differences observed between *Mtb* strains.

Spot #	Protein	Gene name	pMr/mMr	pIP/mIP	Comparison between strains (p-value <0.05)	Reference
1	Rv1860	*apa, modD*	32.7/28.8	4.7/4.5	No statistically significant	[249, 276, 313, 315]
	Rv0462	*lpdC*	49.2/nss	5.7/nss		[275, 276, 313, 315, 395]
2	Rv1886c	*Ag85B, fbpB*	34.6/30.5	5.7/4.9	↑CDC1551 vs ↓BE, H6	[275, 276, 313, 315, 395]
3	Rv1980c	*mpt64*	24.8/22.4	4.6/4.6	↑CDC1551 vs ↓BE, C28, H6	[275, 276, 313, 315, 395]
	Rv1270c	*lprA*	24.9/21.6	5.2/5.0		[313]
4	Rv3705c*	--	22.4/20.0	4.7/4.9	No statistically significant	[313]
5	Rv3310	*sapM*	31.8/27.1	6.6/5.4	↓CDC1551 vs ↑BE, C28	[313]
	Rv3803c*	*fbpC1, fbpD*	31.1/27.8	6.6/5.5		[275, 276, 313, 315, 395]
	Rv0925c	--	26.9/nss	5.5/nss		---
6	Rv3803c*	*fbpC1, fbpD*	31.1/27.8	6.6/5.5	No statistically significant	[275, 276, 313, 315, 395]
	Rv3310	*sapM*	31.8/27.1	6.6/5.4		[313]
7	Rv1080c	*greA*	17.9/nss	4.6/nss	No statistically significant	[275, 395]
	Rv3628	*ppa*	18.3/nss	4.5/nss		[275, 395]
8	Rv1906c*	--	15.5/12.5	5.0/4.7	↑BE vs ↓CDC1551, C28, H6	[313]
9	Rv1926c	*mpt63*	16.5/13.6	4.6/4.5	No statistically significant	[275, 276, 313, 315, 395]
	Rv1906c*	--	15.5/12.5	5.0/4.7		[313]
10	Rv1906c*	--	15.5/12.5	5.0/4.7	Absent in CDC1551	[313]
11	Rv3418c	*groES*	10.8/nss	4.3/nss	↑H6 vs ↓BE, C28	[275, 276, 313, 315, 395]
12	Rv3418c	*groES*	10.8/nss	4.3/nss	No statistically significant	[275, 276, 313, 315, 395]
	Rv3874	*esxB, cfp10*	10.8/nss	4.3/nss		[275, 276, 313, 315, 395-397]
13	Rv2878c	*mpt53*	18.4/14.6	5.0/4.6	↓CDC1551 vs ↑C28	[276, 313, 315]
	Rv2204c	--	12.5/nss	4.2/nss		----
14	Rv2376c	*mbt12, cfp2*	16.6/12.1	6.5/5.1	No statistically significant	[276, 313, 315]
	Rv3914	*trxC*	12.5/nss	4.8/nss		[275, 276, 313, 315]
	Rv2031c	*hspX, acr*	23.3/nss	5.5/nss		[275, 276, 313, 315, 332, 395]
15	Rv2376c	*mbt12, cfp2*	16.6/12.1	6.5/5.1	No statistically significant	[276, 313, 315]
	Rv2031c	*hspX, acr*	23.3/nss	5.5/nss		[275, 276, 313, 315, 332, 395]

16	Rv2376c Rv3418c	*mbt12, cfp2* *groES*	16.6/12.1 10.8/nss	6.5/5.1 4.3/nss	No statistically significant	[276, 313, 315] [275, 276, 313, 315, 395]
17	Rv2376c	*mbt12, cfp2*	16.6/12.1	6.5/5.1	Only present in CDC1551	[276, 313, 315]
18	Rv2376c Rv0455c*	*mbt12, cfp2* --	16.6/12.1 16.6/13.7	6.5/5.1 6.9/5.6	No statistically significant	[276, 313, 315] [313]
19	Rv2376c Rv3418c Rv3874	*mbt12, cfp2* *groES* *esxB, cfp10*	16.6/12.1 10.8/nss 10.8/nss	6.5/5.1 4.4/nss 4.3/nss	↑CDC1551 vs ↓BE, C28	[276, 313, 315] [275, 276, 313, 315, 395] [275, 276, 313, 315, 395-397]
20	Rv3418c Rv3874 Rv1352*	*groES* *esxB, cfp10* --	10.8/nss 10.8/nss 12.9/10.0	4.3/nss 4.3/nss 5.6/4.6	No statistically significant	[275, 276, 313, 315, 395] [275, 276, 313, 315, 395-397] [313]
21	Rv1810*	--	12.1/8.4	6.1/4.5	↑CDC1551 vs ↓C28, H6	[313]
22	Rv3648c	*cspA*	7.4/nss	5.0/nss	↓CDC1551 vs ↑BE, C28	[275, 276, 315, 395]
23	Rv2220	*glnA1*	53.5/nss	4.8/nss	No statistically significant	[276, 313, 315]
24	Rv1074c	*fadA3*	42.7/nss	4.7/nss	No statistically significant	[275, 313, 395]
25	Rv3804c*	*fbpA*	35.7/31.7	6.5/5.3	Absent in CDC1551	[275, 276, 313, 315, 395]
26	Rv0315	--	32.2/28.8	4.7/4.8	Only present in C28	[313]
27	Rv2376c Rv2031c	*mbt12, cfp2* *hspX, acr*	16.6/12.1 23.3/nss	6.5/5.1 5.5/nss	Absent in H6	[276, 313, 315] [275, 276, 313, 315, 332, 395]
28	Rv1980c	*mpt64*	24.8/22.4	4.6/4.6	Absent in BE and C28	[275, 276, 313, 315, 395]

pMr= preprotein molecular weight, mMr= mature molecular weight, pIP= preprotein isoelectric point, mIP= mature isoelectric point, nss= no signal sequence. * Signal sequence experimentally confirmed by Malen and colleagues [313]

4.4.2. Statistical analysis of iTRAQ data

Since *Mtb* BE presented the lowest virulence in the guinea pig model, this strain was chosen as the reference strain for iTRAQ analysis in order to highlight the differences between this strain and the two closely related *Mtb* isolates, C28 and H6, which presented a higher virulence in the guinea pig model. 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.

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. Ratios for each peptide in each identified protein were Log_2 transformed 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. Normalization of iTRAQ data is critical because small differences between samples in the actual amount of protein that is subjected to labeling may result in quantification values that are not representative of each sample. Normalization is also used to correct for small differences due to labeling efficiency and technical variation, as well as run-to-run differences resulting from variation in the mass spectrometry process.

After log transformation and normalization, a mixed model was fit to the data using SAS proc mixed. Model fitting was performed using the Restricted Maximum Likelihood (REML) estimation method [394]. This estimation method is a way to generate variances

and covariances that can then be used to generate p-values and test for differences in each set of values (i.e. protein levels between strains). For a more extensive explanation of the mixed model used in this study please refer to the discussion section.

Because of the complexity of the data with multiple peptides representing a single protein and each of these peptides represented by multiple spectra; in addition to three biological replicates and three technical replicates, it is not feasible to calculate a sample based estimate of standard deviation by protein and strain. However, model-based standard error values for each comparison were calculated and included in Table 4-2 and Table 4-3. For review purposes, only SE for comparison between BE and the other three *Mtb* strains are included. However, these values were also calculated for the other three comparisons (CDC1551 vs C28, CDC1551 vs H6 and C28 vs H6). It is important to note that all modeling and analysis was done using $\log_2$ ratios relative to BE and the standard errors (SE) were also calculated on this scale.

The fact that the standard errors are the same for many of the strain comparisons for a given protein reflects the fact that the model is pooling information across strains (in the form of ratios). For this reason, it is not practical to include these standard errors as errors bars in the graphs of differential proteins. This impracticality would be reflected if the data were presented in graphs corresponding to ratio values, instead of $\log_2$ ratios. If this was the case, the bar corresponding to *Mtb* BE (1.0) would have a standard error equal to zero, giving the false impression that *Mtb* BE did not present any variation. Instead, variation presented in *Mtb* BE due to differences in labeling and protein levels is incorporated into the SE for each comparison (CDC1551 vs BE, C28 vs BE, H6 vs BE), including comparisons between the other three strains (CDC1551 vs C28, CDC1551 vs

H6 and C28 vs H6) because these comparisons are also based on the ratio calculated against *Mtb* BE. Even though these are important considerations, standard errors were included in graphic representation of the data to give a general idea of the total variation for the estimated ratio value of each protein.

After accounting for fixed effects (strain, peptide and strain-peptide interaction), the estimated proportion of variation due to biological and technical variability was calculated from the mixed model for each protein in each iTRAQ experiment (cytosol and CFP). For both experiments, biological variability was higher than technical variability with averages of 39% and 29% for biological variability for CFP and cytosol, respectively and averages of 24% (CFP) and 20% (Cytosol) for technical variability. This variability is similar to what has been reported in other iTRAQ studies, in which biological variability is close to 30% and technical variability is between 10 and 25% [398].

4.4.3. iTRAQ - Cytosol

195 proteins were identified in the cytosol (Cyt) fraction. 68 (34.87%) of them have not been described previously in any other subcellular fraction of *Mtb*, including the cytosol (Table 4-2 and Table 4-4). 139 proteins of this fraction were retained after data filtering and were used for quantification and statistical analysis (Table 4-2). Finally, 63 of these proteins (45.32%) presented statistically significant variations between at least two of the four strains.

Table 4-2. Cytosolic proteins identified by Tandem MS/MS and quantified by the iTRAQ approach.

Cytosolic proteins with no significant differences in protein expression between strains										
		Log_2 Ratio against BE (BE= 0.0)								
		CDC1551		C28		H6				
Rvnumber	gene	Log_2 ratio	SE	Log_2 ratio	SE	Log_2 ratio	SE	Functional category	Number of unique peptides	References
Rv0056	*rplI*	0.090	0.161	0.241	0.161	0.324	0.161	2	4	[276, 319]
Rv0125	*pepA*	-0.256	0.236	0.129	0.236	-0.208	0.236	7	2	[313]
Rv0175	--	-1.152	0.511	-1.026	0.348	-0.721	0.348	3	1	[319]
Rv0178	--	0.749	0.513	-0.954	0.500	-0.675	0.500	3	1	--
Rv0206c	*mmpL3*	0.100	0.248	-0.060	0.248	-0.259	0.248	3	2	[319]
Rv0285	*PE5*	0.614	0.393	-0.617	0.393	-0.940	0.393	6	1	[313]
Rv0287	*esxG*	-0.307	0.432	0.562	0.432	-0.167	0.432	3	2	[276, 313, 315]
Rv0351	*grpE*	-0.358	0.203	0.262	0.203	0.135	0.203	0	2	[276]
Rv0440	*groEL2*	-0.230	0.140	0.127	0.140	-0.170	0.140	0	11	[275, 315, 319, 395]
Rv0455c	--	-0.180	0.150	0.018	0.150	-0.035	0.150	10	1	[313]
Rv0475	*hbhA*	0.016	0.193	0.236	0.193	0.126	0.193	2	7	[319, 338]
Rv0559c	--	0.390	0.201	-0.188	0.201	-0.271	0.201	3	2	[313]
Rv0569	--	0.254	0.222	-0.488	0.220	-0.487	0.220	10	4	[275, 319, 395]
Rv0639	*nusG*	-0.116	0.262	0.494	0.262	0.227	0.262	2	2	[275, 395]
Rv0641	*rplA*	0.222	0.218	0.348	0.218	0.072	0.218	2	3	[276, 319]
Rv0702	*rplD*	-0.247	0.164	-0.191	0.161	0.227	0.161	2	3	[319]
Rv0704	*rplB*	-0.321	0.412	-0.564	0.412	-0.119	0.412	2	2	[319]

Rv0706	*rplV*	-0.422	0.225	-0.141	0.225	0.406	0.225	2	1	[319]
Rv0709	*rpmC*	-0.431	0.192	-0.207	0.192	-0.266	0.192	2	3	[276]
Rv0720	*rplR*	-0.683	0.229	-0.061	0.229	0.079	0.229	2	3	--
Rv0722	*rpmD*	0.632	0.364	-0.203	0.364	-0.013	0.364	2	1	--
Rv0723	*rplO*	-0.216	0.188	-0.056	0.176	-0.158	0.176	2	3	--
Rv0787A	--	-0.219	0.304	-0.391	0.304	-0.315	0.304	10	1	[321]
Rv0810c	--	0.233	0.248	0.191	0.248	0.037	0.248	10	1	--
Rv0871	*cspB*	-0.058	0.132	0.459	0.132	0.356	0.132	0	2	[276]
Rv0883c	--	-0.533	0.376	0.499	0.307	0.095	0.307	10	1	--
Rv1174c	*TB8.4*	-0.116	0.338	-0.772	0.338	-0.641	0.338	3	2	[313, 319]
Rv1196	*PPE18*	0.159	0.303	0.029	0.303	0.177	0.303	6	1	--
Rv1211	--	-0.095	0.244	-0.422	0.244	-0.109	0.244	10	2	[276]
Rv1223	*htrA*	0.364	0.270	-0.083	0.270	-0.059	0.270	7	3	[319]
Rv1270c	*lprA*	-0.291	0.214	-0.210	0.214	-0.055	0.214	3	2	[313, 319]
Rv1310	*atpD*	0.132	0.321	0.219	0.321	-0.124	0.321	7	1	[313, 319]
Rv1311	*atpC*	1.263	0.776	1.629	0.776	1.111	0.776	7	1	--
Rv1388	*mihF*	0.045	0.150	-0.108	0.150	-0.037	0.150	2	7	[319]
Rv1558	--	0.020	0.095	0.060	0.095	0.141	0.095	10	4	[275, 395]
Rv1626	--	-0.470	0.282	0.612	0.282	0.296	0.282	9	1	[275, 319]
Rv1638A	--	-0.206	0.262	-0.247	0.262	0.112	0.262	10	2	--
Rv1643	*rplT*	0.296	0.155	-0.002	0.155	0.264	0.155	2	1	--
Rv1860	*modD*	0.031	0.237	-0.416	0.237	-0.374	0.237	3	2	[313]
Rv1871c	--	-0.489	0.204	-0.014	0.194	-0.028	0.194	10	2	--
Rv1926c	*mpt63*	0.246	0.228	-0.407	0.228	-0.250	0.228	3	4	[275, 276, 313, 395]

Rv2493	--	-0.102	0.153	0.046	0.153	-0.155	0.153	0	5	--
Rv2536	--	-0.038	0.287	-0.719	0.287	-0.550	0.287	3	1	--
Rv2626c	--	0.033	0.152	-0.225	0.152	-0.156	0.152	10	2	[275, 276, 315, 319]
Rv2632c	--	-0.189	0.191	0.448	0.191	0.093	0.191	10	1	--
Rv2711	*ideR*	0.422	0.476	0.537	0.476	0.335	0.476	9	3	--
Rv2785c	*rpsO*	0.335	0.442	-0.163	0.442	0.201	0.442	2	2	[319]
Rv2882c	*frr*	-0.084	0.084	-0.073	0.084	0.011	0.084	2	4	[275, 276, 319, 395]
Rv2890c	*rpsB*	-0.096	0.365	0.342	0.365	0.317	0.365	2	1	[276, 319]
Rv2904c	*rplS*	0.396	0.679	0.361	0.679	0.332	0.679	2	1	--
Rv2909c	*rpsP*	0.000	0.247	-0.286	0.247	-0.278	0.247	2	1	--
Rv2986c	*hupB*	0.029	0.234	0.028	0.234	-0.143	0.234	2	3	[319]
Rv3028c	*fixB*	0.539	0.430	0.736	0.430	0.101	0.430	7	4	[275, 276, 319, 395]
Rv3133c	*devR*	-0.749	0.303	-0.480	0.303	-0.340	0.303	9	1	[275, 319, 395]
Rv3208A	*TB9.4*	0.280	0.205	0.507	0.205	0.159	0.205	10	5	[275, 276]
Rv3417c	*groEL1*	-0.044	0.420	-0.031	0.420	-0.092	0.420	0	1	[319, 333]
Rv3442c	*rpsI*	0.007	0.250	0.099	0.250	0.132	0.250	2	4	--
Rv3443c	*rplM*	-0.588	0.271	-0.265	0.271	-0.006	0.271	2	2	[319]
Rv3456c	*rplQ*	0.455	0.240	0.210	0.240	0.209	0.240	2	1	[319]
Rv3458c	*rpsD*	-1.301	0.799	-0.695	0.799	-0.443	0.799	2	1	[319]
Rv3460c	*rpsM*	-0.444	0.321	0.111	0.321	0.373	0.321	2	3	--
Rv3462c	*infA*	0.433	0.195	0.212	0.195	0.088	0.195	2	2	[319]
Rv3503c	*fdxD*	0.046	0.392	-0.168	0.392	-0.099	0.392	7	1	--
Rv3583c	--	ND	ND	-0.137	0.386	-0.084	0.386	9	2	--
Rv3678A	--	0.287	0.145	-0.167	0.145	0.056	0.145	10	1	--
Rv3688c	--	0.611	0.309	-0.182	0.309	-0.025	0.309	10	2	--

Rv3763	*lpqH*	-1.508	0.436	-0.423	0.436	-0.216	0.436	3	1	[276, 315]
Rv3780	--	0.144	0.303	0.470	0.303	0.335	0.303	10	2	--
Rv3788	--	0.119	0.206	0.107	0.206	0.057	0.206	10	1	[275]
Rv3849	--	-0.298	0.262	0.361	0.223	0.157	0.223	10	1	--
Rv3853	*menG*	-0.341	0.293	-0.553	0.293	0.113	0.293	9	1	--
Rv3867	--	0.817	0.368	0.265	0.368	0.197	0.368	10	2	[275, 313]
Rv3874	*esxB*	-0.680	0.326	-0.691	0.326	-0.507	0.326	3	5	[275, 313, 319]
Rv3875	*esxA*	-0.587	0.430	-0.614	0.430	-0.908	0.430	3	3	[247, 276, 313, 396, 397]
Rv3914	*trxC*	0.117	0.110	-0.020	0.110	-0.020	0.110	7	3	[275, 276, 313, 319]
Rv3920c	--	0.745	0.308	0.794	0.308	0.588	0.308	10	1	[319]

Cytosolic proteins with differential abundance between at least two of the four Mtb strains

Log2 ratio against BE (BE=0.0)

		CDC1551		**C28**		**H6**				
Rvnumber	**gene**	**Log_2 ratio**	**SE**	**Log_2 ratio**	**SE**	**Log_2 ratio**	**SE**	**Functional category**	**Number of unqiue peptides**	**References**
Rv0009	*ppiA*	-0.197	0.086	0.265	0.086	0.299	0.086	2	4	[275, 313, 319, 338, 395, 399]
Rv0020c	*TB39.8*	0.269	0.134	-0.953	0.134	-0.611	0.134	9	2	[275, 276, 315, 319]
Rv0053	*rpsF*	-0.172	0.060	-0.307	0.060	-0.013	0.060	2	4	[275, 319, 395]
Rv0054	*ssb*	0.325	0.187	0.633	0.187	0.223	0.187	10	2	--
Rv0250c	--	-0.205	0.079	-0.053	0.079	0.032	0.079	10	2	--
Rv0270	*fadD2*	0.420	0.110	0.160	0.110	-0.070	0.110	1	1	[276, 319]
Rv0350	*dnaK*	0.240	0.084	0.484	0.083	0.210	0.083	0	7	[275, 276, 313, 319, 395]
Rv0379	*secE2*	-0.346	0.077	-0.629	0.077	-0.354	0.077	3	4	--
Rv0566c	--	0.571	0.159	-0.090	0.159	0.166	0.159	10	4	[275, 319, 395]

Rv0636	*hadB*	-0.439	0.280	0.550	0.280	0.037	0.280	7	1	[276, 319]
Rv0640	*rplK*	-0.056	0.192	0.334	0.141	0.136	0.141	2	2	[319]
Rv0652	*rplL*	0.046	0.199	0.584	0.199	0.330	0.199	2	6	[275, 276, 319]
Rv0683	*rpsG*	-0.873	0.125	-0.284	0.125	-0.042	0.125	2	3	[319]
Rv0685	*tuf*	0.597	0.192	0.635	0.192	0.172	0.192	2	5	[275, 276, 315, 319, 395]
Rv0701	*rplC*	ND	ND	-0.495	0.325	0.436	0.325	2	3	[319]
Rv0703	*rplW*	-0.167	0.098	-0.006	0.098	0.257	0.098	2	1	[319]
Rv0710	*rpsQ*	-0.710	0.128	-0.208	0.128	-0.025	0.128	2	3	--
Rv0718	*rpsH*	-0.372	0.138	-0.290	0.139	-0.008	0.139	2	5	--
Rv0719	*rplF*	-0.255	0.179	0.066	0.174	0.371	0.174	2	3	[319]
Rv0733	*adk*	0.438	0.120	0.348	0.119	0.220	0.119	7	9	[275, 276, 395]
Rv0814c	*sseC2*	0.453	0.149	0.778	0.149	0.353	0.149	7	4	--
Rv0854	--	0.386	0.090	-0.097	0.090	-0.045	0.090	10	2	--
Rv0863	--	0.442	0.141	0.193	0.140	-0.051	0.140	10	2	--
Rv1080c	*greA*	-0.216	0.097	-0.236	0.097	-0.093	0.097	2	4	[275, 319, 395]
Rv1159A	*phhB*	0.628	0.190	0.262	0.190	0.208	0.190	10	1	[321]
Rv1308	*atpA*	-0.218	0.136	0.290	0.136	-0.161	0.136	7	1	[276, 319]
Rv1335	*Cfp10A*	0.324	0.133	0.391	0.129	0.073	0.129	10	2	[400]
Rv1566c	*inv*	0.815	0.203	-0.197	0.203	-0.146	0.203	0	1	--
Rv1630	*rpsA*	-1.244	0.334	0.450	0.334	0.106	0.334	2	2	[319]
Rv1636	*TB15.3*	0.207	0.178	0.585	0.178	0.374	0.178	10	3	[275, 276, 315, 395]
Rv1641	*infC*	0.231	0.215	-0.439	0.215	-0.478	0.215	2	2	[319]
Rv1684	--	0.402	0.198	0.201	0.198	-0.264	0.198	10	1	--
Rv1738	--	-0.044	0.236	0.591	0.236	0.299	0.236	10	2	[319]

Rv1827	*cfp17*	-0.145	0.076	-0.249	0.076	-0.165	0.076	10	7	[276, 313, 399]
Rv1932	*tpX*	0.142	0.152	0.332	0.152	0.249	0.152	0	4	[313]
Rv1980c	*mpt64*	0.711	0.206	0.328	0.206	0.383	0.206	3	6	[275, 276, 313, 395]
Rv2031c	*hspX*	-0.456	0.177	0.066	0.177	0.115	0.177	0	5	[275, 276, 315, 319, 332, 334, 338, 395]
Rv2055c	*RpsR2*	-0.130	0.144	-0.430	0.144	0.028	0.144	2	2	--
Rv2094c	*tatA*	0.228	0.158	-0.322	0.158	0.072	0.158	3	1	--
Rv2111c	*pup*	2.015	0.386	-0.266	0.386	-0.394	0.386	10	1	--
Rv2145c	*wag31*	0.700	0.093	0.250	0.087	-0.034	0.087	3	5	[275, 395]
Rv2185c	*TB16.3*	0.303	0.117	0.301	0.117	0.110	0.117	10	5	[275, 276, 315]
Rv2215	*dlaT*	-0.259	0.259	-0.342	0.259	-0.615	0.259	7	1	[276, 319]
Rv2244	*acpP*	0.219	0.058	-0.212	0.058	-0.112	0.058	1	7	[276, 319]
Rv2302	--	ND	ND	-0.842	0.167	-0.687	0.167	10	3	--
Rv2376c	*cfp2*	1.164	0.230	-1.111	0.230	-0.951	0.230	3	1	[276, 313]
Rv2412	*rpsT*	-0.652	0.111	-0.440	0.111	-0.127	0.111	2	2	[319]
Rv2442c	*rplU*	-0.272	0.160	-0.035	0.160	0.254	0.160	2	1	--
Rv2462c	*tig*	0.428	0.108	0.336	0.105	0.152	0.105	3	5	[275, 276, 315, 319]
Rv2520c	--	0.338	0.231	-0.176	0.231	0.085	0.231	3	1	--
Rv2534c	*efp*	0.536	0.216	0.283	0.216	-0.012	0.216	2	1	[275, 276, 315, 395]
Rv2744c	*30kd_ag*	-0.562	0.186	0.303	0.186	0.625	0.186	10	2	[319]
Rv2878c	*mpt53*	0.472	0.258	0.593	0.258	-0.124	0.258	3	1	[276, 313, 315]
Rv2889c	*tsf*	0.345	0.276	-0.349	0.276	-0.668	0.276	2	1	[275, 319, 395]
Rv2945c	*lppX*	0.184	0.099	-0.041	0.099	-0.056	0.099	3	3	[313, 319]
Rv3029c	*fixA*	0.292	0.117	0.583	0.115	0.273	0.115	7	5	[275, 276, 319, 395]
Rv3046c	--	-0.571	0.131	-0.403	0.130	-0.236	0.130	10	2	--

Rv3418c	*groES*	-0.453	0.131	-0.750	0.131	-0.414	0.131	0	8	[275, 276, 313, 319, 395]
Rv3597c	*isr2*	0.435	0.105	-0.080	0.105	0.011	0.105	10	2	[276, 319]
Rv3648c	*cspA*	0.657	0.277	0.799	0.277	0.488	0.277	0	2	[275, 276, 319, 395]
Rv3716c	--	-0.143	0.100	0.504	0.100	0.275	0.100	10	3	[275, 395]
Rv3841	*bfrB*	1.006	0.421	1.246	0.421	0.715	0.421	7	3	[276, 313, 319]
Rv3879c	--	0.096	0.215	-0.486	0.215	-0.226	0.215	10	1	[319]

Note: SE (Standard Error). Standard error are shown only for Log_2 ratio values of strain comparisons directly against BE. ND: Peak ion corresponding to the iTRAQ tag for the strain/protein was not detected by Mascot program.

4.4.4. iTRAQ-CFP

The global analysis by iTRAQ allowed the identification of 146 proteins in the culture filtrate fraction (CFP) from which 53 (36.3%) had not been characterized previously by other CFP proteomic studies (Table 4-3 and Table 4-4). Protein quantification and statistical analysis was performed in 101 proteins that were retained after data filtering (Table 4-3). 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. 28 proteins of the quantified proteins were found in both fractions, corresponding to 27.77% and 20.14% of the CFP and cytosol respectively.

4.4.5. Validation of iTRAQ approach by western blot

To confirm that the iTRAQ results were representing real differences between protein levels, equal aliquots of each biological replicate for cytosol or CFP were pooled and resolved in a SDS-PAGE followed by western blot against several proteins (i.e. Mpt64, HspX, ModD, HbhA, Cfp10, GroEL2 and Cfp2). In general, western blot corroborated the general trend of relative protein levels observed by iTRAQ (Figure 4-2, Table 4-2, Table 4-3). Some of the proteins that did not present significant differences by iTRAQ (i.e.Cfp10, GroEL2) showed different patterns by western blot. This was due to high variability between biological replicates and therefore they were not considered to be significant. iTRAQ analysis seems to be more sensitive to detect low protein levels. For example, HspX and Cfp2 were not detectable by western blot of *Mtb* CDC1551 cytosolic proteins or CFP of *Mtb* H6 respectively (Figure 4-2). However, lower levels of these proteins in the respective fractions were detected by iTRAQ (Table 4-2, Table 4-3).

Table 4-3. CFP proteins identified by MS/MS and quantified by the iTRAQ approach.

CFP proteins with no significant differences in protein expression between strains										
		Log_2 Ratio against BE (BE= 0.0)								
		CDC1551		C28		H6				
Rvnumber	gene	Log_2 ratio	SE	Log_2 ratio	SE	Log_2 ratio	SE	Functional category	Number of unique peptides	References
Rv0066c	*icd2*	-0.511	0.284	0.335	0.284	-0.102	0.284	7	14	[313]
Rv0211	*pckA*	0.239	0.342	0.354	0.342	0.150	0.342	7	3	[276, 313, 338]
Rv0234c	*gabD1*	0.013	0.153	0.085	0.153	0.057	0.153	7	6	[276]
Rv0250c	--	-0.012	0.277	0.199	0.277	0.570	0.277	10	2	
Rv0315	--	0.671	0.394	-0.250	0.394	-0.264	0.394	7	2	[313]
Rv0350	*dnaK*	-0.204	0.211	-0.365	0.211	-0.252	0.211	0	2	[275, 276, 313, 395]
Rv0390	--	0.078	0.362	0.190	0.362	-0.052	0.362	10	2	
Rv0462	*lpd*	-0.297	0.264	-0.623	0.264	-0.498	0.264	7	11	[275, 276, 313, 395]
Rv0468	*fadB2*	-0.009	0.180	0.050	0.180	0.001	0.180	1	5	[276]
Rv0577	--	0.011	0.232	0.019	0.232	0.124	0.232	10	2	[276]
Rv0652	*rplL*	0.623	0.483	1.178	0.483	0.960	0.483	2	6	[275, 276]
Rv0677c	--	0.519	0.725	0.085	0.725	-0.341	0.725	3	1	[275, 276, 313]
Rv0733	*adk*	0.023	0.556	0.435	0.556	0.920	0.556	7	3	[275, 276, 395]
Rv0932c	*pstS2*	0.435	0.203	0.016	0.203	-0.061	0.203	3	2	[313]
Rv0946c	*pgi*	0.513	0.454	-0.025	0.454	-0.201	0.454	7	3	
Rv0952	*sucD*	-0.445	0.719	0.607	0.719	0.539	0.719	7	1	[401]
Rv1001	*arcA*	-0.216	0.371	-0.214	0.371	-0.180	0.371	7	3	
Rv1017c	*prsA*	0.578	0.635	1.102	0.635	0.698	0.635	7	4	[276]

Rv1071c	*echA9*	0.483	0.242	0.511	0.242	0.149	0.242	1	2	[319]
Rv1074c	*fadA3*	0.081	0.191	0.023	0.191	0.041	0.191	1	9	[275, 313, 395]
Rv1080c	*greA*	0.508	0.209	0.258	0.209	0.147	0.209	2	4	[275, 395]
Rv1122	*gnd2*	0.367	0.161	0.262	0.161	0.224	0.161	7	3	[319]
Rv1177	*fdxC*	-0.378	0.280	0.271	0.280	-0.118	0.280	7	3	
Rv1211	--	0.839	0.669	0.180	0.669	0.573	0.669	10	1	[276]
Rv1271c	--	-0.085	0.337	-0.474	0.337	-0.488	0.337	3	2	
Rv1284	--	0.318	0.132	-0.029	0.132	0.007	0.132	7	2	[276]
Rv1388	*mihF*	1.273	1.089	0.569	1.089	0.506	1.089	2	2	[319]
Rv1477	--	-0.873	0.756	-1.075	0.756	-0.839	0.756	0	2	[313]
Rv1810	--	0.677	0.180	0.253	0.180	0.461	0.180	10	2	[313]
Rv1827	--	0.301	0.126	0.006	0.126	0.069	0.126	10	7	[276, 313, 399]
Rv1837c	*glcB*	0.016	0.550	0.831	0.550	0.247	0.550	7	13	[276, 313, 332]
Rv1876	*bfrA*	-0.395	0.247	0.018	0.247	0.025	0.247	7	4	[313]
Rv1887	--	-0.344	0.781	-0.394	0.781	-0.228	0.781	10	1	[313]
Rv1908c	*katG*	-0.683	0.390	-0.113	0.390	0.178	0.390	0	14	[276, 313, 332]
Rv1926c	*mpt63*	-0.051	0.379	-0.442	0.379	0.189	0.379	3	3	[275, 276, 313, 395]
Rv1980c	*mpt64*	0.506	0.406	0.586	0.406	0.925	0.406	3	4	[275, 276, 313, 395]
Rv2007c	*fdxA*	-0.241	0.258	-0.381	0.258	-0.478	0.258	7	2	
Rv2109c	*prcA*	0.162	0.210	0.091	0.210	0.290	0.210	7	1	[276]
Rv2244	*acpP*	1.134	0.750	0.861	0.750	0.927	0.750	1	8	[276]
Rv2290	*lppO*	-0.225	0.468	-0.476	0.468	-0.679	0.468	3	3	
Rv2391	*nirA*	-0.693	0.578	-0.877	0.578	-0.725	0.578	7	4	
Rv2799	--	0.107	0.445	-0.423	0.445	-0.404	0.445	3	2	[313]
Rv2971	--	-0.094	0.189	-0.040	0.189	-0.221	0.189	7	1	[276, 401]

Rv3029c	*fixA*	0.112	0.346	-0.036	0.346	0.141	0.346	7	2	[275, 276, 395]
Rv3169	--	0.322	0.229	0.157	0.229	-0.255	0.229	10	2	
Rv3248c	*sahH*	-0.509	0.372	0.577	0.372	0.038	0.372	7	5	[276, 313]
Rv3413c	--	0.279	0.490	-0.373	0.490	-0.621	0.490	10	2	[313]
Rv3418c	*groES*	0.292	0.636	0.589	0.636	0.655	0.636	0	6	[275, 276, 313, 395]
Rv3616c	--	-0.116	1.210	-1.809	1.210	-1.128	1.210	10	3	
Rv3628	--	0.493	0.302	0.372	0.302	0.068	0.302	7	3	[275, 395]
Rv3648c	*cspA*	-0.178	0.179	0.319	0.179	-0.082	0.179	0	1	[275, 276, 395]
Rv3716c	--	-0.463	1.010	-0.345	1.010	-0.500	1.010	10	1	[275, 395]
Rv3841	*bfrB*	0.475	0.388	-0.088	0.388	0.627	0.388	7	4	[276, 313]
Rv3846	*sodA*	-0.254	0.215	-0.536	0.215	-0.264	0.215	0	2	[313]
Rv3853	*menG*	-0.399	0.245	-0.311	0.245	-0.209	0.245	9	2	
Rv3875	*esxA*	0.083	0.291	0.341	0.291	-0.091	0.291	3	1	[247, 276, 313, 396, 397]
Rv3914	*trxC*	0.054	0.305	-0.183	0.305	0.142	0.305	7	2	[275, 276, 313]

CFP proteins with differential abundace between strains (p-value >0.05)

		Log2 ratio against BE (BE=0.0)								
		CDC1551		**C28**		**H6**				
Rvnumber	gene	Log2 ratio	SE	Log2 ratio	SE	Log2 ratio	SE	Functional category	Number of unique peptides	References
Rv0363c	*fba*	-0.430	0.252	0.396	0.252	0.216	0.252	7	1	[275, 276, 313]
Rv0398c	--	0.568	0.227	-0.238	0.227	-0.395	0.227	3	1	[313]
Rv0431	--	-0.245	0.338	-1.397	0.338	-1.309	0.338	3	2	
Rv0440	*groEL1*	0.191	0.180	0.405	0.180	0.123	0.180	0	5	[275, 276, 395]
Rv0500	*proC*	-0.706	0.274	-0.063	0.274	0.025	0.274	7	2	

Rv0559c	--	-0.077	0.186	-0.757	0.186	-0.588	0.186	3	2	[275, 313, 395]
Rv0569	--	-0.652	0.191	-0.866	0.191	-0.543	0.191	10	3	[275, 395]
Rv0583c	*lpqN*	0.322	0.175	-0.388	0.175	-0.368	0.175	3	2	[313]
Rv0798c	*cfp29*	-0.177	0.171	0.577	0.171	0.490	0.171	0	1	[276]
Rv0884c	*serC*	-0.995	0.321	-0.987	0.321	-0.588	0.321	7	2	[313]
Rv0896	*gltA2*	0.687	0.391	1.440	0.391	1.097	0.391	7	9	[319]
Rv0934	*pstS1*	-0.210	0.247	-0.627	0.247	-0.398	0.247	3	8	[313]
Rv0951	*sucC*	-0.561	0.232	0.610	0.232	0.761	0.232	7	1	[319]
Rv0983	*pepD*	-0.059	0.272	-0.850	0.272	-0.822	0.272	7	3	
Rv1023	*eno*	0.046	0.203	0.569	0.203	0.402	0.203	7	1	
Rv1079	*metB*	0.604	0.220	0.144	0.220	0.109	0.220	7	3	
Rv1352	--	-0.807	0.253	-1.034	0.253	-0.732	0.253	10	1	[313]
Rv1392	*metK*	-0.535	0.231	0.850	0.231	-0.036	0.231	7	4	[313]
Rv1448c	*tal*	-2.636	0.951	0.448	0.951	0.976	0.951	7	1	[313]
Rv1844c	*gnd1*	1.040	0.391	0.743	0.391	1.064	0.391	7	2	
Rv1860	*modD*	0.544	0.193	0.277	0.193	-0.016	0.193	3	6	[313]
Rv1906c	--	-0.558	0.331	-1.389	0.331	-1.292	0.331	10	1	[313]
Rv1932	*tpx*	-0.453	0.193	0.137	0.193	0.110	0.193	0	1	[313]
Rv2031c	*hspX*	-0.117	0.689	1.284	0.689	1.954	0.689	0	4	[275]
Rv2110c	*prcB*	-0.565	0.161	-0.638	0.161	-0.597	0.161	7	2	[313]
Rv2204c	--	-0.709	0.178	0.301	0.178	0.150	0.178	10	4	
Rv2220	*glnA1*	-0.642	0.208	-0.604	0.208	-0.352	0.208	7	10	[313]
Rv2345		0.377	0.197	-0.627	0.197	-0.294	0.197	3	3	
Rv2376c	*cfp2*	1.109	0.360	0.344	0.360	0.169	0.360	3	4	[276, 313]
Rv2467	*pepN*	-0.029	0.130	0.618	0.130	0.218	0.130	7	4	

Rv2493	--	0.594	0.248	0.612	0.248	0.809	0.248	0	2	
Rv2721c	--	0.307	0.162	0.693	0.162	0.095	0.162	3	1	[313]
Rv2873	*mpt83*	1.383	0.355	0.588	0.355	0.134	0.355	3	1	[313]
Rv2895c	*viuB*	0.371	0.247	0.535	0.247	0.727	0.247	7	1	[319]
Rv3006	*lppZ*	-0.347	0.238	-0.895	0.238	-0.693	0.238	3	5	[313]
Rv3075c	--	0.059	0.126	0.761	0.126	0.761	0.126	10	1	[275]
Rv3267	--	-0.556	0.379	-0.869	0.379	-1.063	0.379	10	2	[313]
Rv3312A	--	-1.363	0.621	-2.035	0.621	-1.150	0.621	3	1	
Rv3354	--	-0.606	0.331	-1.029	0.331	-1.104	0.331	10	1	[313]
Rv3389c	--	0.390	0.210	0.708	0.210	0.013	0.210	7	2	[275]
Rv3503c	*fdxD*	-0.086	0.139	-0.526	0.139	-0.349	0.139	7	1	
Rv3587c	--	-0.193	0.338	-0.985	0.338	-1.434	0.338	3	4	[313]
Rv3682	*ponA2*	-0.286	0.305	-1.531	0.305	-0.930	0.305	3	2	[313]
Rv3874	*esxB*	0.294	0.210	0.795	0.210	-0.157	0.210	3	3	[275, 313]

Note: SE (Standard Error). Standard error are shown only for Log_2 ratio values of comparisons against BE. ND: Peak ion corresponding to the iTRAQ tag for the strain/protein was not detected by Mascot program. Proteins for which reference [319] is the only one available have not been described previously in the *Mtb* secreted fraction.

Table 4-4. Proteins identified in only one biological replicate and therefore not included in the quantitative analysis.

Rvnumber	Gene name	Fraction	Funtional category	Reference
Rv0203	--	CYT	3	[313]
Rv0269	--	CYT	10	
Rv0284	--	CYT	3	[319]
Rv0285	*pe5*	CFP	6	[313]
Rv0291	*mycP3*	CYT	7	[313]
Rv0352	*dnaJ1*	CYT	0	[319]
Rv0376c	--	CYT	10	
Rv0391	*metZ*	CFP	7	
Rv0455c	--	CFP	10	[313]
Rv0467	*icl*	CFP	7	
Rv0500A	--	CYT	10	
Rv0558	*ubiE*	CFP	7	
Rv0577	*tb27.3*	CYT	10	[276, 315]
Rv0629c	*recD*	CYT	2	[319]
Rv0634A	--	CYT	10	[321]
Rv0686	--	CYT	3	
Rv0700	*rpsJ*	CYT	2	[319]
Rv0705	*rpsS*	CYT	2	
Rv0715	*rplX*	CYT	2	[319]
Rv0721	*rpsE*	CYT	2	[276, 319]
Rv0730	--	CYT	10	
Rv0787A	--	CFP	10	[321]
Rv0815	*cysA2*	CYT	7	[275, 276, 319, 332]
Rv0863	--	CFP	10	
Rv0899	*ompA*	CFP	3	
Rv0921	--	CFP	5	
Rv0934	*pstS1*	CYT	3	[276, 313, 315, 319, 321]
Rv0968	--	CYT	10	[276]
Rv1038c	*esxJ*	CFP	3	[276, 313, 315]
Rv1077	*cbs*	CFP	7	[275, 276, 313]
Rv1097c	--	CFP	3	
Rv1109c	--	CYT	10	[276, 319]
Rv1174c	*TB8.4*	CFP	3	[313]
Rv1177	*fdxC*	CYT	7	
Rv1187	*rocA*	CFP	7	
Rv1195	*PE13*	CYT	6	
Rv1223	*htrA*	CFP	7	
Rv1262c	--	CFP	10	
Rv1270c	*lprA*	CFP	3	[313]
Rv1329c	*dinG*	CFP	2	

Rv1335	--	CFP	10	[401]
Rv1381	*pyrC*	CFP	7	
Rv1398c	--	CFP, CYT	0	
Rv1435c	--	CFP	3	[313]
Rv1488	--	CYT	3	[313, 319]
Rv1498A	--	CYT	10	[313, 321]
Rv1628c	--	CFP	10	
Rv1636	*TB15.3*	CFP	10	[275, 276, 315, 395]
Rv1655	*argD*	CFP	7	[276]
Rv1684	--	CFP	10	
Rv1810	--	CYT	10	[313]
Rv1826	*gcvH*	CYT	7	
Rv1844	*gnd1*	CYT	7	
Rv1876	*bfrA*	CYT	7	[313]
Rv1886c	*fbpB*	CFP	1	[275, 276, 313, 315, 395]
Rv1893	--	CYT	10	
Rv1911c	*lppC*	CFP	3	[313]
Rv1915	*aceAa*	CFP	7	
Rv1984c	*cfp21*	CFP, CYT	3	[276, 313, 315, 399]
Rv2050	--	CYT	3	
Rv2070c	*cobK*	CFP	7	
Rv2111c	*pup*	CFP	10	
Rv2187	*fadD15*	CYT	1	[276]
Rv2204	--	CYT	10	
Rv2210c	*ilvE*	CFP	7	
Rv2224c	--	CYT	3	[313, 319]
Rv2239	--	CYT	10	
Rv2241	*aceE*	CFP	7	[276, 313]
Rv2557	--	CYT	10	[401, 402]
Rv2588c	*yajC*	CYT	3	
Rv2672	--	CFP	7	[313]
Rv2694c	--	CYT	10	
Rv2721c	--	CYT	3	[313]
Rv2783c	*gpsI*	CYT	2	[276]
Rv2823c	--	CYT	10	[313, 319]
Rv2882c	*frr*	CFP	7	[313]
Rv2908c	--	CYT	10	
Rv2916c	*ffh*	CYT	3	[313, 319]
Rv2945c	*lppX*	CFP	3	[313]
Rv3001c	*ilvC*	CFP	7	[319]
Rv3033	--	CFP	10	[313]
Rv3046c	--	CFP	10	
Rv3196A	--	CYT	10	[321]

Rv3237c	--	CYT	10	
Rv3386	--	CYT	5	[319]
Rv3401	--	CFP	7	[276]
Rv3457c	*rpoA*	CYT	2	[275, 276, 315, 319]
Rv3459c	*rpsK*	CYT	2	
Rv3477	*PE31*	CFP	6	
Rv3592	*tb11.2*	CYT	10	[275, 276, 315, 395]
Rv3616	--	CYT	10	
Rv3628	*ppa*	CYT	7	[275, 319, 395]
Rv3705c	--	CFP	10	[313]
Rv3723	--	CYT	3	
Rv3774	*echA21*	CFP	1	
Rv3880c	--	CYT	10	
Rv3881c	--	CFP, CYT	10	[313, 401]
Rv3921c	--	CYT	3	

Note: proteins for which reference [319] is the only one available have not been described previously *Mtb* CFP.

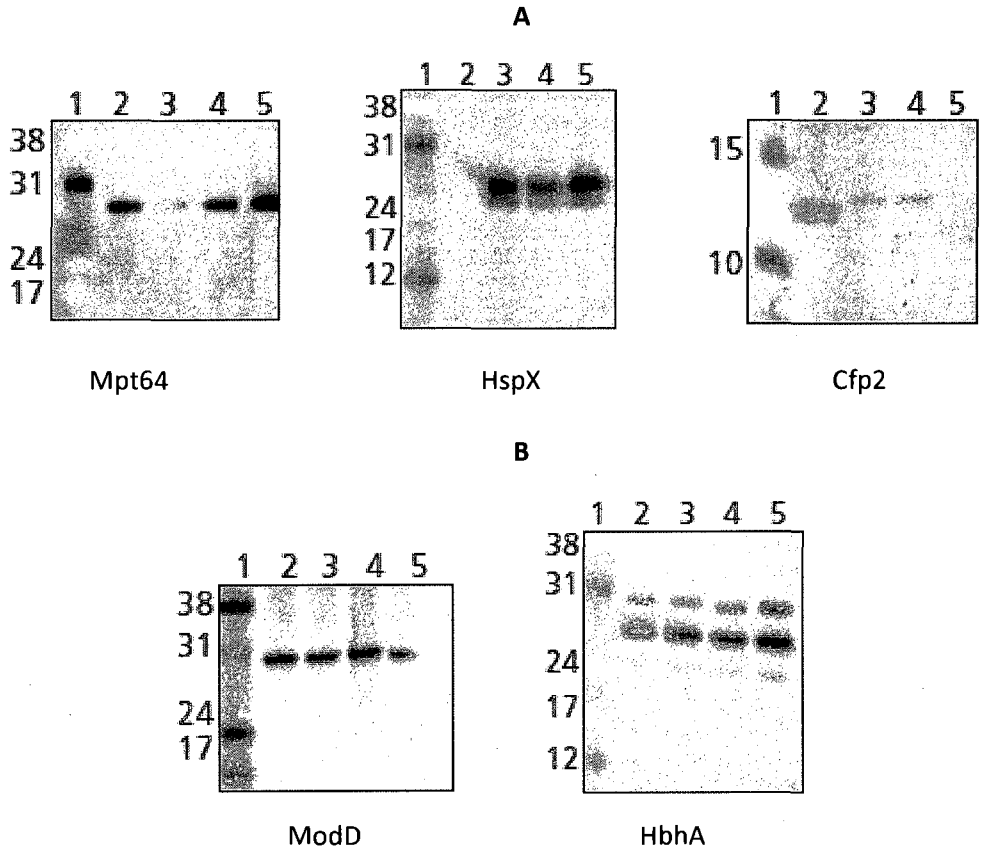


Figure 4-2. Representative western blot of selected proteins. Confirmation of the general trend of protein levels observed by iTRAQ.

A. Differential proteins by iTRAQ analysis (p-values <0.05). **B.** Proteins with no significant differences by iTRAQ analysis. Cfp2 was tested on the CFP, while the other four proteins were tested in the cytosol. **1:** Molecular ladder, **2:** CDC1551, **3:** BE, **4:** C28, **5:** H6. Molecular ladder all blots but Cfp2 corresponds to ECL Plex Rainbow marker (GE Healthcare). Molecular ladder of Cfp2 blot is Precision Plus Protein Standard Dual Color (Bio-Rad). Band sizes are shown in KDa. Note: All bands on HbhA blot were used for comparison.

4.4.6. 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 4-5). The majority of the variability was found between *Mtb* CDC1551 and the S75 group (BE, C28 and H6) in which 50 cytosolic proteins and 28 CFP proteins were found to be differentially expressed between *Mtb*

CDC1551 and at least one of the members of the S75 group. This represents 37.6% and 29.5% of the total number of significant comparisons in the cytosol and CFP, respectively.

Table 4-5. Summary of variability in protein expression between *Mtb* strains

	Number of significant comparisons	
Strain comparison	**CFP**	**CYT**
CDC1551 vs BE	15	35
CDC1551 vs C28	16	22
CDC1551 vs H6	12	24
BE vs C28	28	31
BE vs H6	21	11
C28 vs H6	3	10
Total number of significant comparisons (p-value <0.05)	**95**	**133**
Total number of differential proteins (p-value <0.05)	**44**	**63**

This table shows the number of significant comparisons (differential proteins) between strains for each of the fraction analyzed by iTRAQ – MS/MS

When variability between the closely related S75 group isolates was analyzed, few differences were found when comparing the *Mtb* H6 to *Mtb* C28 isolate. This analysis revealed only 13 differentially expressed proteins, 3 of which were secreted and 10 cytosolic proteins (Figure 4-3).

In contrast, *Mtb* BE presented the greatest variation in proteins when compared to C28 and H6 (Table 4-5). Specifically, 32 CFP (31.68%) and 35 cytosolic (25.18%) proteins were significantly different between *Mtb* 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 4-4 and Figure 4-5). Overall, differences between *Mtb* BE and at least one of the two other S75 strains were

responsible for 33.7% (32/95) and 26.3% (35/133) of the total number of significant comparisons in protein levels for secreted and cytosolic fractions, respectively.

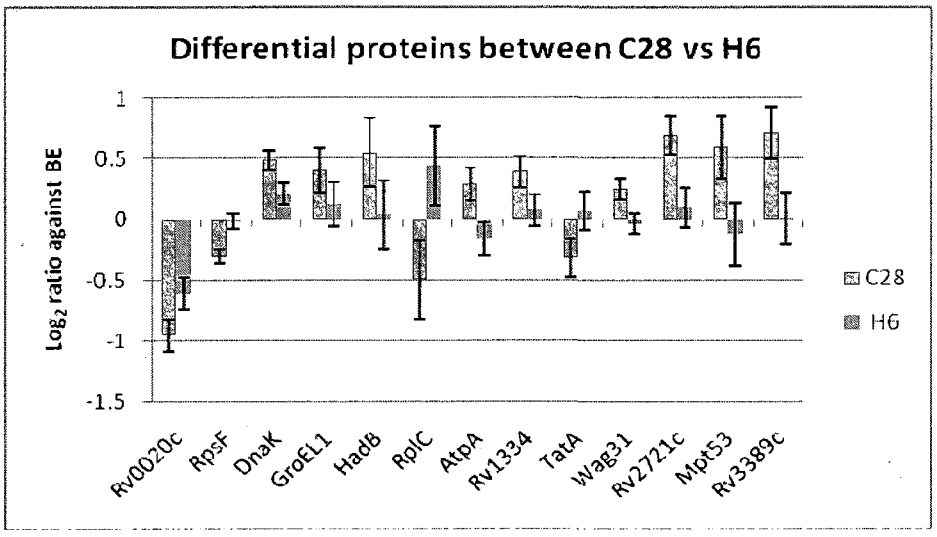


Figure 4-3. Differential proteins between *Mtb* strains C28 and H6.
Values are expressed as the Log_2 ratio against BE (BE=0). GroEL1, Rv2721c and Rv3389c were identified in the secreted fraction, while the remaining 10 proteins were identified in the cytosol. To facilitate graph interpretation, *Mtb* CDC1551 protein levels were excluded. Error bars represent standard error (SE) for each strain comparison against BE for each protein.

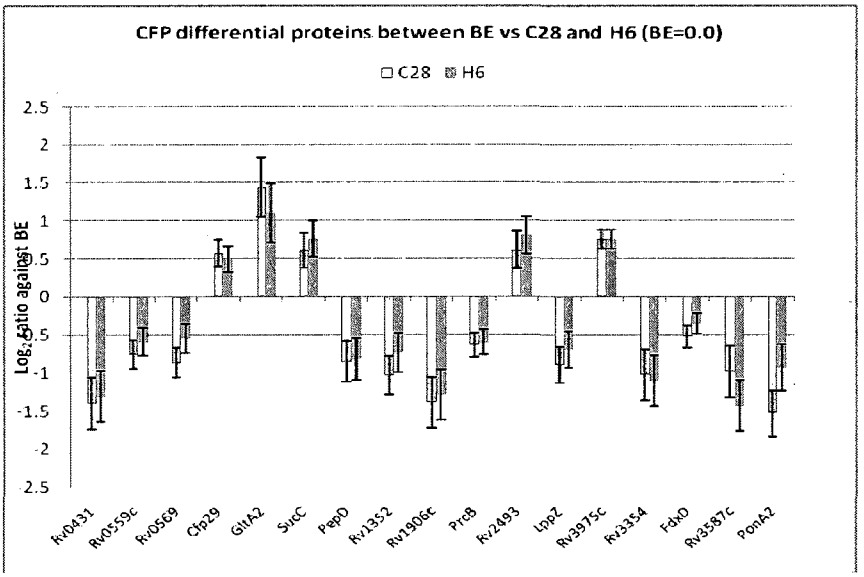


Figure 4-4. Differential CFP proteins between BE and both, C28 and H6.
Note: BE value = 0.0. PrcB, Rv0569 and Rv2493 expression values were also significantly different between BE and CDC1551. *Mtb* CDC1551 values are not including to facilitate the visualization of the data. Error bars represent standard error (SE) for each strain comparison against *Mtb* BE for each protein.

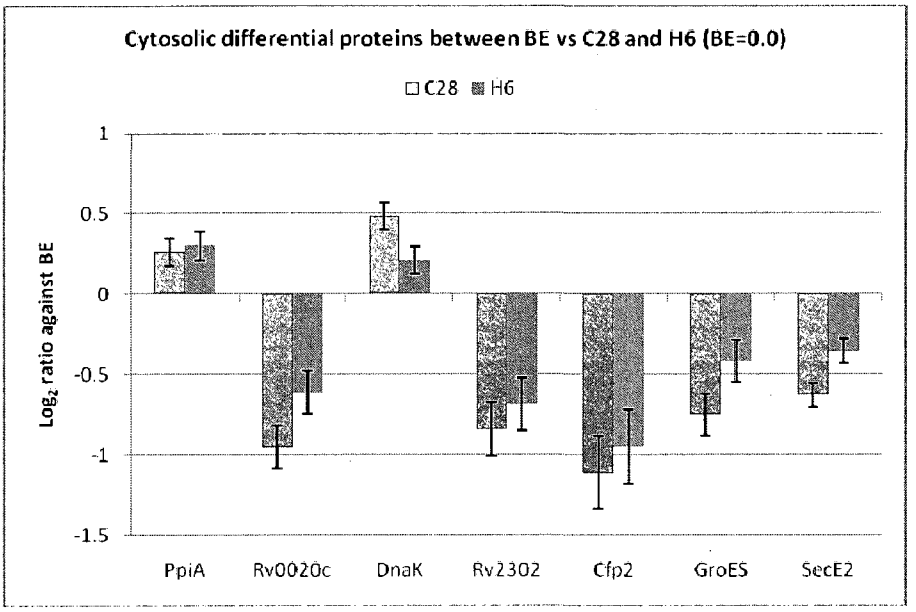


Figure 4-5. Differential cytosolic proteins between BE and both, C28 and H6.
Note: BE value = 0.0. All proteins, with exception of Rv2302 and Rv0020c also presented significantly different values between BE and CDC1551. *Mtb* CDC1551 values are not including to facilitate the visualization of the data. Error bars represent standard error (SE) for each strain comparison against *Mtb* BE for each protein.

Since *Mtb* CDC1551 was used as an outside member for comparison purposes, differences between *Mtb* 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 *Mtb* CDC1551 and at least one of the other three strains. For review purposes, Table 4-6 shows only proteins with p-value <0.01 for differential proteins between *Mtb* CDC1551 and at least one S75 strain or with p-values <0.05 for differential proteins between *Mtb* 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 *Mtb* 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 4-6).

Table 4-6. Differential proteins between CDC1551 and members of the S75 group.

Rv number	Product	Quantitative values expressed as ratio against BE (BE=1.0)		
		CDC1551	C28	H6
Higher levels in CDC1551				
Rv0398c	Unknown	1.48	0.85	0.76
Rv2873	Mpt83	2.61	1.50	1.09
Rv2889c	Probable elongation factor, Tsf	1.27	0.79	0.63
Rv0020c	Unknown	1.21	0.52	0.66
Rv0566c	Nucleotide binding protein	1.44	0.94	1.12
Rv0733	Adelynate kinase, Adk	1.36	1.27	1.17
Rv0854	Unknown	1.31	0.94	0.97
Rv2094c	Twin arginine translocase protein A, TatA	1.17	0.80	1.05
Rv2111c	Prokaryote ubiquitin-like protein, Pup	4.04	0.83	0.76
Rv2145c	Unknown, Wag31	1.63	1.18	0.98
Rv2244	Acyl carrier protein, AcpP	1.16	0.86	0.93
Rv2376c	Low molecular weight antigen, Cfp2	2.24	0.46	0.52
Rv2462c	Trigger factor, Tig	1.35	1.26	1.11
Rv3597c	Iron-regulated LSR2 protein precursor, Isr2	1.35	0.95	1.00
Lower levels in CDC1551				
Rv0500	Probable Pyrroline-5-5Carboxylate reductase, ProC	0.61	0.96	1.02
Rv0009	Probable iron regulated peptidyl-prolyl cis- trans isomerase A, PpiA	0.87	1.20	1.23
Rv0636	Single hotdog hydratase homodimer, HadB	0.74	1.46	1.03
Rv1448	Transaldolase, Tal	0.16	1.36	1.97
Rv3075c	Unknown	1.04	1.70	1.70
Rv1932	Thiol peroxidase, Tpx	0.73	1.09	1.08
Rv2204c	Unknown	0.61	1.23	1.11
Rv0350	Heat shock protein, chaperone DnaK	1.18	1.40	1.16
Rv0683	30S ribosomal protein S7, RpsG	0.55	0.82	0.97

Rv0363	Probable fructose-biphosphate aldolase, Fba	0.74	1.32	1.16
Rv0703	50S ribosomal protein L23, RplW	0.89	1.0	1.20
Rv0710	30S ribosomal protein S17, RpsQ	0.61	0.87	0.98
Rv0798c	29 Kda Antigen, Cfp29	0.85	1.49	1.40
Rv0951	Succinyl Co-A dehydrogenase, SucC	0.68	1.53	1.69
Rv1308	F0F1 ATP synthase subunit alpha, AtpA	0.86	1.22	0.89
Rv1630	30S ribosomal protein S1, RpsA	0.42	1.37	1.08
Rv2110c	Proteasome beta subunit, PrcB	0.68	0.64	0.66
Rv2412	30S ribosomal protein S20, RpsT	0.64	0.74	0.92
Rv3046c	Unknown	0.67	0.76	0.85
Rv0379	Protein transport protein, SecE2	0.79	0.65	0.78
Rv2744c	Immunogenic protein 30Kd_ag	0.68	1.23	1.54

Note: Only differential proteins with p-value <0.01 between CDC1551 and at least one of the other three strains or with p-values <0.05 between CDC1551 and all other three strains are shown. For Log2 values and standard error please refer to Table 4-2 and Table 4-3.

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

Identification of the most abundant proteins expressed by the mycobacteria 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 from composite MS/MS data [393]. From this 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 4-6).

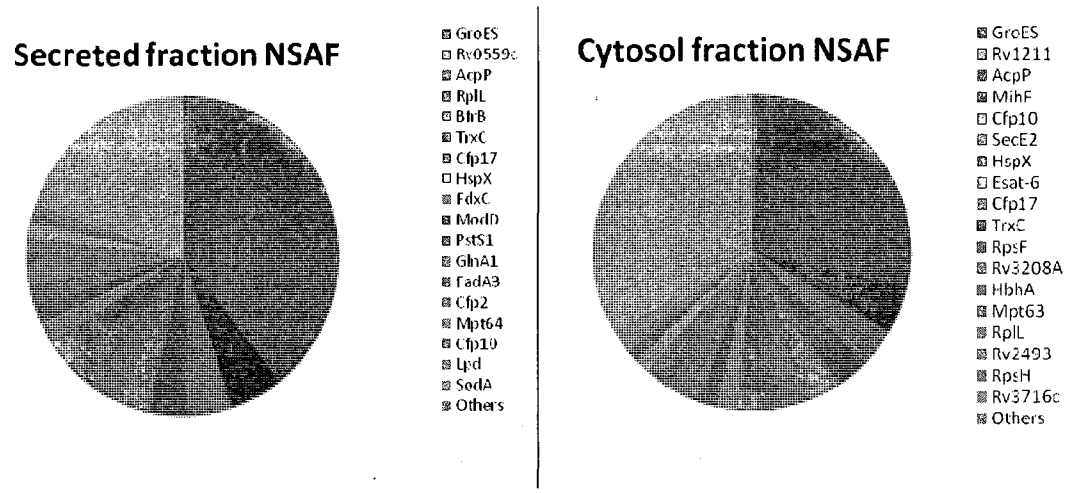


Figure 4-6. Relative abundance of proteins identified in the CFP (secreted) or cytosolic fraction. **NSAF**: Normalized Spectral-counts Abundance Factor.

4.4.8. 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).

4.4.9. 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 (Figure 4-7). 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 4-7).

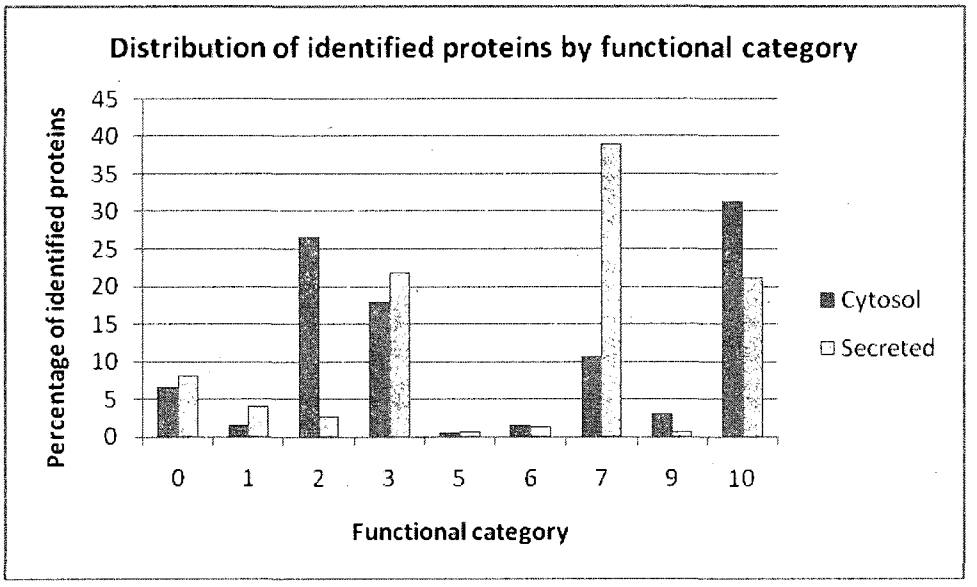


Figure 4-7. Distribution of identified proteins by functional category.
Functional categories are as follows: **0:** virulence, detoxification, adaptation; **1:** lipid metabolism; **2:** information pathways; **3:** cell wall and cell processes; **5**: insertion sequencess and phages; **6:** PE/PPE; **7:** intermediary metabolism and respiration; **9:** regulatory proteins; **10:** conserved hypotheticals.

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 4-8).

Cytosol CFP

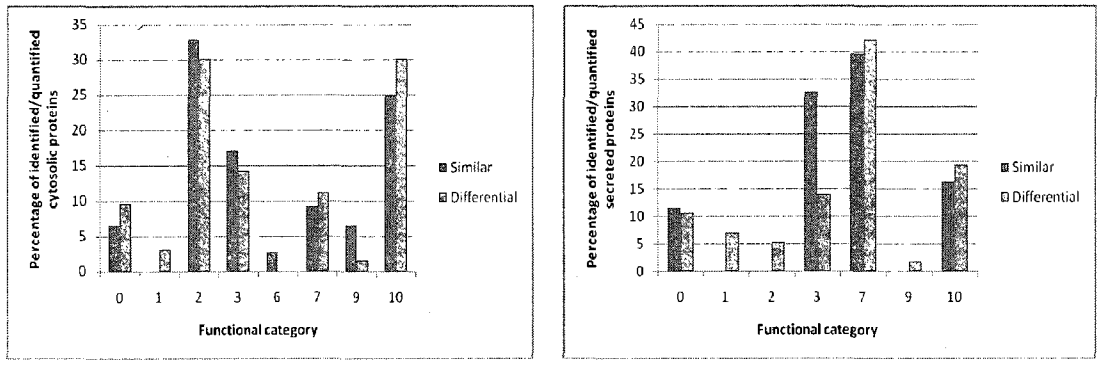


Figure 4-8. Distribution of cytosolic and CFP (secreted) proteins with similar or significantly differential levels between *Mtb* strains for each functional group.

Blue bars represent proteins without significant differences. Red bars represent proteins with significant differences between strains. Functional categories are as follows: **0:** virulence, detoxification, adaptation; **1:** lipid metabolism; **2:** information pathways; **3:** cell wall and cell processes; **5**: insertion sequencess and phages; **6:** PE/PPE; **7:** intermediary metabolism and respiration; **9:** regulatory proteins; **10:** conserved hypotheticals.

4.5. Discussion

In recent years, the global study of protein expression or proteomics has become a useful tool in the study of microbial physiology, in the context of virulence, pathogenesis, and environmental adaptation. Proteomic analysis via 2D-GE has been widely used to study many biological systems. In mycobacteriology, this technique has been used to assess differences in protein expression between different mycobacterial species [275, 314, 333], clinical isolates [249, 311, 341, 345, 347] and in response to certain stimuli [328, 330, 334, 336, 338].

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 [300, 313, 319].

Early studies employing the analysis of proteins by 2D-GE afforded global identification of mycobacterial proteins [249, 275, 276, 315, 395]. 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. However, 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 the present study, a straightforward 2D-GE 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 coomasie brilliant blue, which does not interfere with downstream mass spectrometry analysis [403].

In addition, 2D-GE was complemented with the iTRAQ method which offers a higher throughput analysis in a streamlined procedure.

4.5.1. Statistical analysis of iTRAQ data

Analysis methods for iTRAQ data vary considerably in the literature. One of the most common approaches is the arbitrary selection of upper and lower ratio thresholds [404, 405]. In other studies, thresholds are based on the global average ratio and standard deviation within samples [294]. More classical methods, such as those based on t-test for statistical significance, have also been described [406]. All of these studies perform the data analysis from single iTRAQ experiments on one protein at a time. In contrast, analysis of multiple iTRAQ experiments, in a single procedure can be performed using mixed models [302, 407]. In the present study, a rigorous statistical analysis using this

latter approach was developed and complemented the iTRAQ analysis of the CFP and cytosolic proteomes of *Mtb* clinical isolates. This type of statistical modeling provides a flexible framework to account for the effect of different peptides (representing the same protein) as well as biological and technical variation and the method by which samples representing technical replicates are grouped. As compared to the t-test, one of the major benefits offered by mixed models is the ability to account for variability between the peptides representing the same protein. Mixed models take into consideration this intra-protein variation (peptide-effects) when calculating the statistical significance to determine differences in protein levels between strains. While there are differences in the approaches used in previous studies using mixed models, [302, 407], both models include the effect of peptide variation (peptide effects).

The mixed model used in this study can be described as a linear regression model in which the data is fit using the least-squares method and then tested for statistical significance between protein levels in each strain by analysis of variance (ANOVA). The use of ANOVA allows the direct comparison of means between more than two groups at the same time. In contrast, a basic t-test allows the comparison of means between only two groups.

By using this mixed model, we can test if the relative abundance, calculated as a ratio (R), of a protein is equal to zero (R1=0; R2=0; R3=0), in which case we are directly comparing protein levels of each *Mtb* strain (CDC1551, C28, H6) against the reference strain (*Mtb* BE). Where R1, R2 and R3 will be the estimated ratio value of a particular protein for CDC1551/BE, C28/BE and H6/BE. In addition, the model also allows us to test if the relative abundance (R) of a protein from the non-reference strains is

significantly different against another non-reference strain. This is, in the same terms as above, if R1=R2, R1=R3 and R2=R3.

In other words, this method allows the interrogation of statistical significance not only between each of the samples (CDC1551, C28 and H6) in relation to the reference sample (i.e.BE) but also between each of the possible comparisons between the non-reference samples (CDC1551 vs C28, CDC1551 vs H6 and C28 vs H6).

As noted previously, one of the major benefits offered by mixed models is the ability to account for peptide differences. Song et al (2008) reported that there is a considerable variance in the iTRAQ ratios of individual peptide ions for a given protein [408]. This statement was evident in some of the proteins identified in our analysis when a pair of peptides representing the same protein showed opposite values, that is, one peptide presented evidence of up-regulation and the other indicated down-regulation as compared to BE. In such cases, performing a simple average of the ratio values while ignoring which peptide they relate to would result in the loss of information and possibly misleading results, especially considering that the observed peptides representing a given protein can vary from run to run. Using the mixed model analysis allows not only to account for potential peptide effects but also to identify this potential variation and flag proteins for further inspection.

4.5.2. iTRAQ analysis of *Mtb* proteome

In the present study, 146 (CFP) and 195 (Cytosol) proteins were identified by iTRAQ labeling followed by MS/MS analysis. Data analysis and interpretation by the rigorous statistical model described in the previous section allowed the identification,

quantification and statistical testing of 139 and 101 cytosolic and secreted proteins, respectively.

From the total number of identified proteins, only 31 proteins were identified in both fractions (Table 4-2, Table 4-3, Table 4-4). The majority of them have previously been shown to be present in the culture filtrates of *Mtb* [275, 276, 313, 395], 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 4-7). 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 [313]. 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 [276].

Variation in expression levels between proteins that were found in both fractions was observed. 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 *Mtb* CDC1551, *Mtb* C28 and *Mtb* H6 respectively. It is important to note that the protein expression values for each strain are not absolute but rather relative to strain *Mtb* BE which was taken as the

reference for analysis of iTRAQ data. Nonetheless, the variation between fractions within one single strain illustrates the utility of proteomics over mRNA arrays on the ability to localize and quantify protein expression in different cell compartments, and in this case, intracellular versus extracellular proteins.

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.

Finally, 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.

4.5.3. Proteomic analysis by 2D-GE versus iTRAQ

The value of using different proteomic approaches in a single study, as well as the advantages and disadvantages of 2D-GE and shotgun methods have been clearly shown elsewhere [296, 320, 409]. Therefore, the use of both 2D-GE 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, 2D-GE analysis

provided important information that otherwise would have been missing. For instance, both, 2D-GE and iTRAQ analyses clearly showed the increase expression and/or secretion of Cfp2 in *Mtb* CDC1551 (Figure 4-1, Figure 4-9, Table 4-1 and Table 4-3). However, only 2D-GE showed the presence of different isoforms of this protein, from which just one (Figure 4-1, spot #17) seems to account for the majority of variation in Cfp2 levels between *Mtb* isolates.

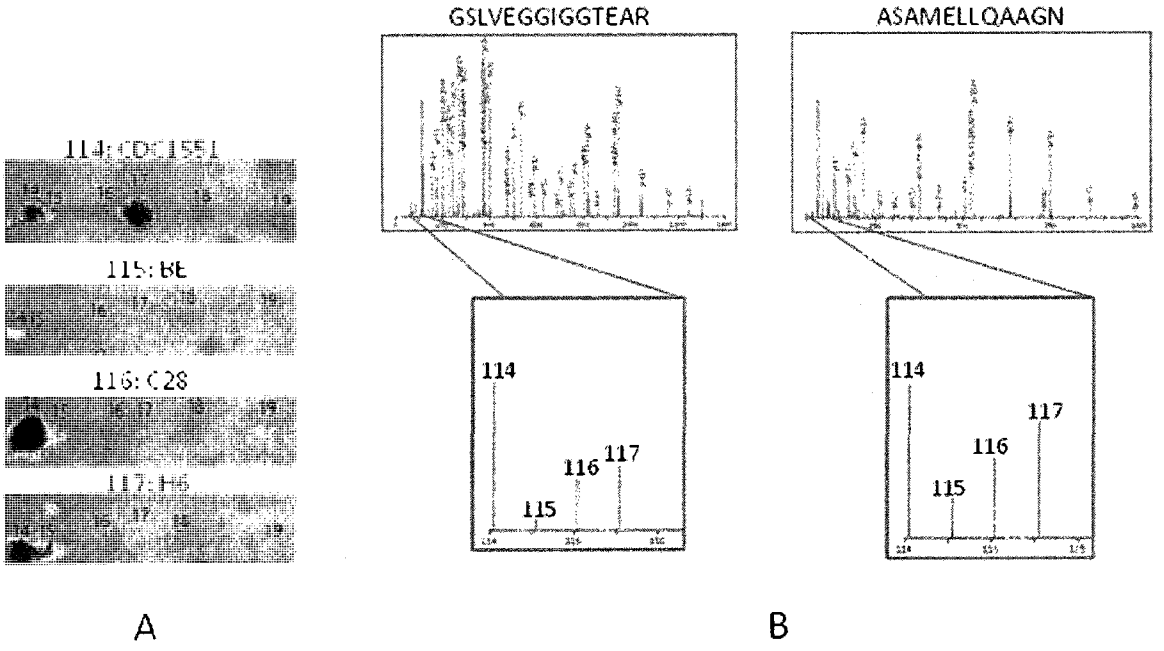


Figure 4-9. Comparison of Cfp2 levels between A: 2D-GE and B: iTRAQ.
Notes: Please refer to Figure 4-1and Table 4-1 for spot identification. Note that several peptides were used for identification and quantification of Cfp2 by iTRAQ. Only two are shown in this figure.

Interestingly, the differences between density levels of this Cfp2 isoform were statistically significant between *Mtb* CDC1551 and all three other *Mtb* isolates. However, comparison of Cfp2 quantity by iTRAQ labeling, which by default will show the total sum of different expression isoforms of this protein, was only statistically significant between *Mtb* CDC1551 and *Mtb* BE (p-value = 0.018) while the differences were not significant when *Mtb* CDC1551 was compared to either *Mtb* C28 (p-value = 0.17) or *Mtb*

H6 (p-value = 0.11). Sequence analysis of cfp2 identified a SNP in *Mtb* 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 but the pattern of Cfp2 expression levels between strains was similar between 2D-GE and iTRAQ.

Different isoforms of Rv1906c were identified by 2D-GE. General expression levels for this protein among strains were similar between 2D-GE and iTRAQ (Figure 4-1, Table 4-1 and Table 4-3).

A general agreement of protein level patterns between techniques was obtained for proteins identified in single spots such as CspA (Rv3648c) and GlnA1 (Rv2220c).

In addition to the identification of protein isoforms, 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. 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 [313, 410, 411], as well as in our 2D-GE, 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 reason for this finding is not clear. Nonetheless, 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 [278] 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.

Finally, 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.

The combination of iTRAQ and 2D-GE analyses allowed the identification of several differences in the protein expression of secreted and cytosolic proteins of four different *Mtb* clinical isolates. Three of these *Mtb* isolates (BE, C28 and H6) are genetically very closely related and have been previously grouped in an *Mtb* cluster, the S75 group [352]. Within this group, *Mtb* BE is the most common strain, identified in 73.2% of the patients in which the S75 group was initially described, followed by *Mtb* H6 and *Mtb* C28 (23.2% and 3.6%, respectively) [352].

This group of *Mtb* strains belongs to the Euro-American lineage (principal genetic group 2, Lineage IV). Strain *Mtb* CDC1551, which has been extensively characterized in previous studies and belongs to the same group but different lineage (V) was included in this study as a reference for S75 strains.

4.5.4. Proteomic variability between *Mtb* CDC1551 (Lineage V) and the S75 group (Lineage IV)

Strain *Mtb* CDC1551 was isolated from an outbreak on the Kentucky-Tennessee border between 1994 and 1996. It was initially described as hypervirulent in humans, because of its high rates of skin-test conversion among contacts of patients with active disease [193]. In addition, *Mtb* CDC1551 showed 100-fold higher numbers of bacilli compared to the numbers of bacilli isolated from the lungs of mice infected with *Mtb* Erdman strain [193]. However, additional studies in mouse and rabbit models suggested that *Mtb* CDC1551 is actually less virulent than other clinical laboratory strains, showing prolonged survival in infected mice [220] and slower growth in the rabbit lung [221, 222]. In addition, during mixed mouse infections with the hypervirulent strain *Mtb* HN878, *Mtb* CDC1551 grew more slowly and the recovery post-infection was poor [223]. Specifically, *Mtb* CDC1551 has been shown to induce higher levels of Th-1 cytokines (TNF-a, IL-6, IL-12) in both in vivo and vitro infection in mouse and human monocytes respectively, as well as after exposure of monocytes to *Mtb* CDC1551 extractable lipids [220, 224] . Combined, these characteristics have led others to classify this strain as both hyperimmunogenic and hypovirulent, especially when compared to the hypervirulent strain HN878 [223].

In agreement with this idea, *Mtb* CDC1551 presented higher levels of several immunogenic proteins (Figure 4-1, Table 4-1 and Table 4-6) that could be involved with its previously reported hyperimmunogenic phenotype. For instance, Mpt83 and Cfp2 levels were considerably higher in the strain *Mtb* CDC1551 versus the other *Mtb* strains. Cfp2 over-production or over-secretion by *Mtb* CDC1551 was evident by 2D-GE and corroborated by iTRAQ. 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* [412]. 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 [413, 414]. 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 *Mtb* 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 [415]. 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. In contrast to the general trend of other immunogenic proteins that were highly abundant in *Mtb* CDC1551 when compared to the other *Mtb* strains, the Ag85 complex presented lower levels in 2D-gels of this strain when compared to the S75 group. Similar to Mpt83, the variability in expression or secretion of these proteins in strains representing two different *Mtb* lineages could have a potential negative impact in the efficacy of subunit vaccines that include the members of the Ag85 complex. It would be interesting to evaluate the production of the Ag85 complex in other strains from the *Mtb* lineage V, as well as the expression of these proteins from different *Mtb* lineages during in vivo infections.

Another differential protein between *Mtb* CDC1551 and the S75 group was Rv2111c which showed 4 times higher levels in *Mtb* CDC1551 (Table 4-2). 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 [416]. Pup is specifically conjugated to proteasome substrates, targeting them for degradation. Thus it is possible for *Mtb* CDC1551 to present an increased level of protein degradation compared to *Mtb* BE, *Mtb* C28 and *Mtb* H6. Interestingly, protein degradation by the proteasome unit has been shown to be essential for *Mtb* virulence and nitric oxide resistance [417, 418]. 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 [416]. This contrasts our findings, and may indicate a loss of accessory molecules or products required for proper Pup interaction in *Mtb* CDC1551, resulting in the observed increase in Pup expression in *Mtb* CDC1551.

4.5.5. Proteomic variability between closely related *Mtb* clinical isolates (S75 group) and its potential association to virulence

Differences between *Mtb* CDC1551 and the S75 strains are notable because they might illustrate some of the protein variation between *Mtb* lineages. More interesting, and a major objective of the proteome analysis of these *Mtb* strains is the identification of proteins that are possibly involved in the differences in virulence that were observed between the closely related *Mtb* strains (BE, C28 and H6) in the guinea pig model (Figure 3-1). As discussed in Chapter III, *Mtb* strains C28 and H6 seem to be more virulent than *Mtb* BE as assessed by the survival of guinea pigs infected with these strains. Protein levels between clinical isolates *Mtb* C28 and *Mtb* H6 were very similar, with only a few proteins demonstrating significant variation between these two strains (Figure 4-3). These

differences are consistent with the genetic relatedness between these two *Mtb* isolates, which as mentioned in the introduction, appear to have evolved from *Mtb* BE. These minor differences are probably not biologically relevant since these two strains show the same level of virulence in the guinea pig model. In contrast, several secreted and cytosolic proteins presented differential expression in strain *Mtb* BE when compared to strains *Mtb* C28 and *Mtb* H6.

Specifically, 32 secreted and 35 cytosolic proteins showed differences between *Mtb* BE and at least one of the other two S75 strains, *Mtb* C28 and *Mtb* H6. Since differences between *Mtb* C28 and *Mtb* H6 in the guinea pig model were not statistically significant, we were interested in proteins that presented similar levels in these two *Mtb* strains but that differ from *Mtb* BE, thus having a potential role in the higher virulence of *Mtb* C28 and *Mtb* H6 in the guinea pig model (Figure 4-10). In order to facilitate visualization and interpretation of the differential proteins, they were grouped by functional categories according to the code used by Tuberculist (http://genolist.pasteur.fr/TubercuList/).

A pattern in the proteomes of strains *Mtb* H6 and *Mtb* C28 versus strain *Mtb* BE was identified when differentially expressed proteins were grouped by functional categories (Figure 4-10).

In general, *Mtb* C28 and *Mtb* 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) (Figure 4-10). Further, the differential expression of proteins belonging to group 7 (Intermediary metabolism and respiration) is found when the data is divided into subgroups.

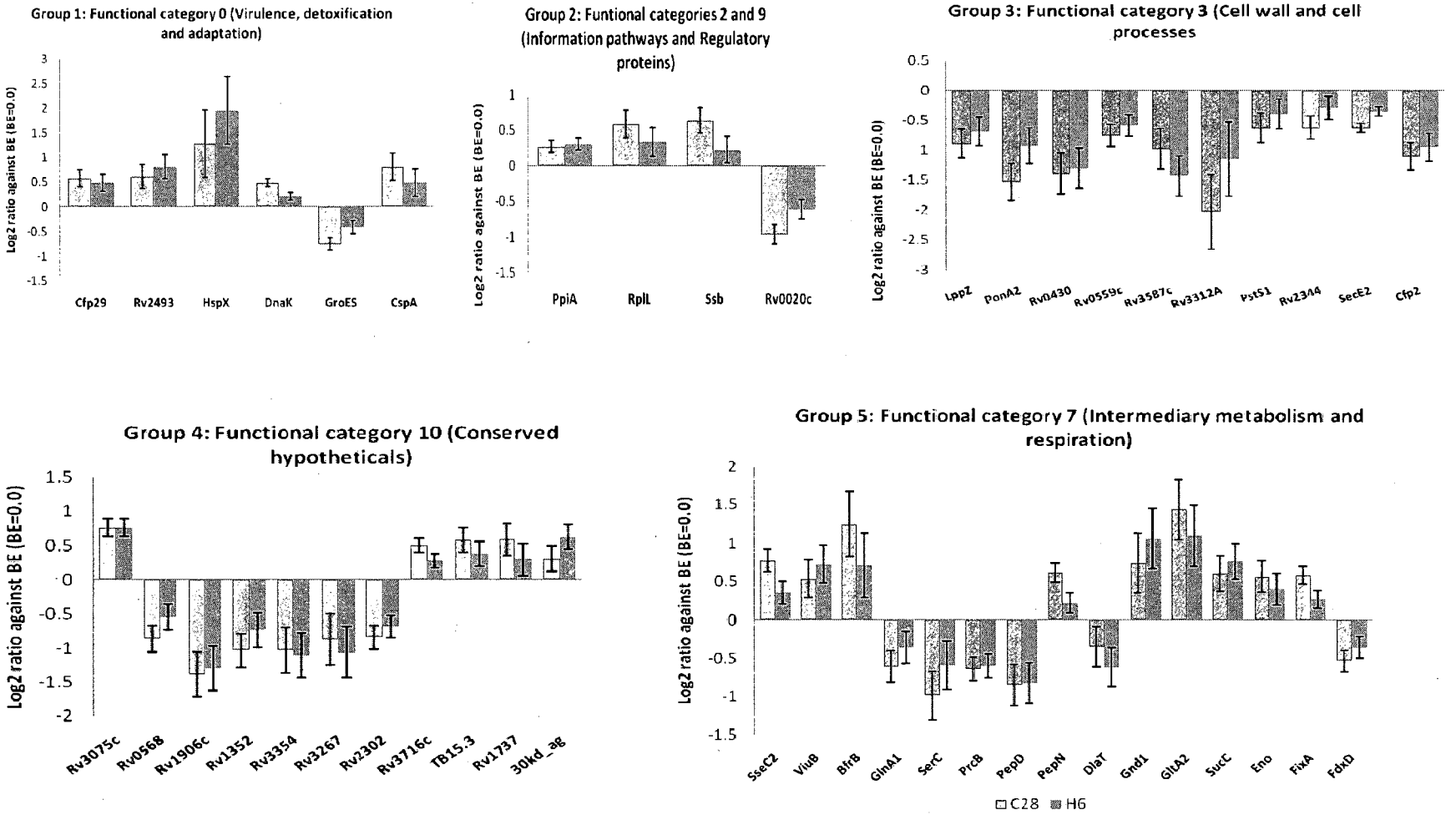


Figure 4-10. Differential proteins between BE and the other two members of the S75 group (C28 and H6) classified by functional category. Proteins Rv2492, DnaK, GroES, PpiA, SecE2, Cfp2, FixA, SseC2, GlnA1, SerC, PrcB, Gnd1, Rv0568, Rv1352, 30Kd_ag were also statistically significant between BE and CDC1551. Ratio values for *Mtb* CDC1551 were not included to facilitate visualization of the data. Error bars represent the standard error for each comparison against *Mtb* BE.

Proteins involved in metabolic pathways (i.e. Eno, SucC, GltA2, Gnd1), and nutrient acquisition (BfrB, ViuB, Sse2) are, in general, more abundant in strains *Mtb* C28 and *Mtb* 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. Lower abundance of several immunogenic proteins in strains *Mtb* C28 and *Mtb* H6 versus strain *Mtb* BE was also observed (Table 4-2 and Table 4-3). This trend was not universal; however, as BfrB and Cfp29, B- and T-cell antigens respectively [323, 419] were both more abundant in these two strains compared to the less virulent strain *Mtb* BE.

Proteins involved in metabolic pathways, such as glycolisis (i.e. Eno), the hexose monophosphate (HMP) shunt (i.e. Gnd1) and the TCA cycle (i.e. SucC, GltA2) were found in higher abundance in the more virulent strains *Mtb* C28 and *Mtb* H6. Glycolysis and to a less extent in *Mtb*, the HMP-shunt [420] are responsible for carbohydrate metabolism. In addition, the HMP-shunt provides ribose-5-phosphate (R5P) which is essential for nucleotides and nucleic acid metabolism. Enolase (Eno) is involved in the catalysis of 2-phospho-D-glycerate into phosphoenolpyruvate, the second to last step in glycolisis. Gnd1 is involved in the last step of the HMP-shunt pathway converting 6-phospho-D-gluconate into D-ribose 5-phosphate, generating CO_2 and NADPH. The tricarboxilic acid cycle (TCA) is not only involved in energy production, but it is important to generate precursors for amino acid biosynthesis. Succinyl CoA synthase, a heterodimer, is formed by SucC (beta chain) and SucD (alpha chain). Although only SucC expression demonstrates significant differences between *Mtb* C28 and *Mtb* H6 vs. *Mtb* BE, SucD expression shows a similar trend, with higher quantities in the two more virulent strains compared to *Mtb* BE (ratios against *Mtb* BE of 1.52 and 1.45 for *Mtb* C28

and *Mtb* H6, respectively). The other protein involved in the TCA cycle for which higher abundance was observed in C28 and H6 was GltA2. Interestingly, this protein is not only involved in the TCA cycle but is also required for the anaplerotic glyoxylate pathway. This pathway utilizes fatty acids as the carbon source and has been shown to be essential for growth of *Mtb* in the host [421]. Higher production of metabolic enzymes in *Mtb* strains *Mtb* C28 and *Mtb* H6 suggests that these isolates might be more metabolically active and might have an advantage for utilization of energy sources available within the host, which could be related to their higher virulence in the guinea pig model.

Proteins related to iron storage such as bacterioferritin (BfrB), and iron acquisition such as the possible mycobactin utilization protein (ViuB), and Tb15.3 which is an iron-regulated protein [346] are present in a higher abundance in strain *Mtb* C28 and *Mtb* H6. In addition, the sulfur uptake protein (SseC2) also demonstrates higher levels in *Mtb* C28 and H6. Iron and sulfur are essential for both *Mtb* survival and virulence in the host. Sulfur is required for the production of mycothiols and sulfolipids, which have been shown to be important in the anti-oxidant response and pathogenicity of *Mtb* respectively [422-424]. Likewise, iron is a required cofactor for the correct function of essential enzymes [425, 426]. Furthermore, iron acquisition has been shown to be indispensable for survival of *Mtb* within the host [427, 428]. These findings suggest that more virulent strains *Mtb* C28 and *Mtb* H6 might have an advantage in iron and sulfur acquisition.

In addition to these differences, Rv2493 was more abundant in secreted fraction of the more virulent strains C28 and H6 compared to BE. Rv2493 is the toxin component of a putative toxin-antitoxin (TA pair) [429]. Different functions for prokaryote TA pairs have been identified. In general, they function as growth regulator elements [430, 431], which

could provide an advantage to pathogens that inhabit variable and stressful environments to adapt and persist [432, 433]. Interestingly, some TA pairs have been shown to be expressed in broth culture, while only one of the TA components (either toxin or antitoxin) is expressed in human macrophages [434]. This suggests that TA pairs or their single components could be involved in TB adaptation to the host cells.

Although there is a clear trend in protein levels between the less virulent strain *Mtb* BE versus *Mtb* strains C28 and H6, some of the findings are puzzling. For instance, SodA, classified under group 1 for virulence, detoxification and adaptation and which has been shown to be an important virulence factor in *Mtb* [435] however it is less abundant in *Mtb* C28 and *Mtb* H6. DlaT, the enzyme responsible for converting pyruvate to acetyl-CoA in the first step of the TCA cycle, is also less abundant in *Mtb* C28 and *Mtb* H6. It may be that these *Mtb* strains rely on obtaining acetyl CoA through the oxidation of fatty acids, which has been shown to be important in *Mtb* [436], and may result in down regulation of DlaT expression.

The protein PonA2 is also present in lower levels in *Mtb* C28 and *Mtb* H6. This protein is involved in the final stages of peptidoglycan biosynthesis and has been recently shown to be important for complete virulence of *Mtb* in mice [437]. In addition, the same study showed that PonA2 mutants were hypersusceptible to antibiotics, heat, and H_2O_2 exposure. Some proteins involved in amino acid biosynthesis (i.e.GlnA1 and SerC) as well as some proteases (i.e. PrcB, PepD) demonstrated decreased abundance in *Mtb* C28 and *Mtb* H6. GlnA1 is involved in synthesis of L-glutamine and has been shown to be essential for *Mtb* survival in the macrophage [438-440]. At the same time, bacterial proteases have been suggested as important virulence factors [441-443].

While it is puzzling that some of these putative virulence factors are more abundant in the less virulent strain *Mtb* BE, their increased expression may be due to intrinsic physiological differences in strain *Mtb* BE versus *Mtb* strains C28 and H6. In support of this, our studies demonstrated that all of the differential proteins categorized under cell wall and cell processes were less abundant in strains *Mtb* C28 and *Mtb* H6 versus strain BE. Some of these proteins are recognized antigens (LppZ, tuberculin related peptide (Rv0431) and Rv3312A). LppZ has been described as one of the most immunogenic secreted proteins in TB [323, 444]. Rv3312A was recently described as part of a new *Mtb* structure with similarities to pili from other bacteria. In addition, this protein is recognized by sera from patients with active TB [84]. A decrease expression of immunogenic proteins and therefore potential decreased recognition by the immune system could favor virulent strains to establish and survive in the host.

4.6. Conclusion

Proteomics allows high throughput profiling of expressed proteins at cellular and subcellular levels. During the past decade, it has been dominated by two technologies: 2D-GE to separate proteins, and mass spectrometry to identify these proteins. Although mass spectrometry continues to be a central tool in proteomics, other technologies have been developed to either improve or replace 2D-GE.

In this study, a novel proteomic approach based on peptide labeling with isobaric tags (iTRAQ) in combination with liquid chromatography separation was used. This shotgun approach with iTRAQ labeling is more sensitive than traditional 2D gels and allows for the relative quantification and identification of hundreds of proteins in a single

experiment. Moreover, the use of three independent biological replicates per sample in addition to three technical replicates during mass spectrometry analysis increased the statistical confidence of identified proteins and their quantitative values. This data was coupled to a strict statistical analysis to test quantitative values for each protein and determine significant differences between samples. In this case, this analysis allowed the identification of several proteins with statistically significant quantitative differences between closely related clinical isolates of *Mtb*.

Similar to proteomic studies elsewhere [305, 445, 446], the use of two different techniques, in this case iTRAQ and 2D-GE, was complementary.

The majority of the differences observed in the protein levels were found between the relatively distant *Mtb* strains, *Mtb* CDC1551, which belong to the phylogenetic lineage V, and the S75 group which belongs to lineage IV. The most pronounced difference was the higher production or secretion by *Mtb* CDC1551 of the immunogenic protein Cfp2, which was especially evident by 2D-GE and validated by iTRAQ. The reason for this up-regulation/secretion of Cfp2 in *Mtb* CDC1551 is worthy of study.

Interesting observations were obtained when the proteomes of the closely related *Mtb* strains comprising the S75 group were compared. We not only identified several differential protein levels between the closely related *Mtb* strains, but given their differences in virulence in the guinea pig model, our findings suggest that the biological variability might be related to differences in their protein component. Based on the current hypothesis regarding the S75 group of *Mtb* strains [352], *Mtb* BE is the strain from which *Mtb* C28 and *Mtb* H6 originated. Interestingly, our proteomic findings seem to agree with this epidemiological data, showing significant differences in protein

expression levels between these isolates, especially between *Mtb* BE and the other two related strains *Mtb* C28 and *Mtb* H6. In addition, findings presented in the previous chapter suggest that there are also differences in the virulence of *Mtb* BE when compared to *Mtb* C28 and *Mtb* H6.

Higher levels of proteins involved in virulence, detoxification and adaptation, energy production, and nutrient acquisition were identified in *Mtb* C28 and *Mtb* H6 when compared to *Mtb* BE and may contribute to the advantage strains *Mtb* C28 and *Mtb* H6 demonstrated in establishing a lethal infection in the guinea pig model.

Some results obtained in this study relating to protein levels are puzzling. For example, higher levels of some proteins known to be involved in virulence, such as SodA, GlnA1 and some proteases, presented higher levels in the less virulent strain *Mtb* BE. It is important to note that the proteomic analysis conducted here was performed from in vitro growing mycobacteria. Therefore, the potential impact of these puzzling results on the biological variability of these *Mtb* strains, as well as the potential role for specific proteins in the higher virulence of strains *Mtb* C28 and *Mtb* H6 require further validation studies.

Several conserved hypotheticals and other proteins with unknown function, presented differential levels between strains, indicating that these could also have a potential role in the biological behavior of the analyzed strains. One of these hypothetical proteins, Rv1906c, was the only protein that showed significant differences between BE and the two *Mtb* related strains C28 and H6 by both 2D-GE and iTRAQ analysis. Functional studies on these hypothetical proteins, with special attention to virulence and pathology of *Mtb* represent an open field for research.

Finally, findings in this study comment to the multifactorial nature and complexity of virulence and pathogenicity mechanisms in *Mtb*, suggesting a required interconnection of multiple variables that will in the end, be responsible for sustaining a productive infection.

4.7. Supplementary information

Studies elsewhere have shown that *Mtb* clinical isolates may present different rates of 'in vitro' growth [249, 311, 349]. This in turn might affect the protein levels at different time points. Growth curve of *Mtb* CDC1551 and S75 strains shows that these strains present some differences in their growth during in vitro culture (Figure 4-11). Particularly, C28 appears to have a slower growth rate as compared to the other strains. This could explain preliminary data in which 2D-GE of *Mtb* C28 CFP harvested at two weeks of culture presented poor resolution and were not able to be analyzed (Data not shown).

At week 4, all strains were found to be in early stationary phase and no significant differences were detected. At this point CFPs and cells were finally harvested. In accordance, Pheiffer and colleagues also reported the use of four-week cultures for proteomic studies of *Mtb* clinical isolates in order to minimize growth-related differences between strains [311].

Western blot of selected proteins was performed on CFP harvested at day 10, 20 and 30 of culture. Some proteins were present at detectable levels only at day 30 (Figure 4-11). Protein expression variability for a few proteins (i.e. Mpt64, DnaK) was found (Data not shown), but no evidence of protein degradation along the time was observed. Finally,

protein levels at day 30 confirmed the differences identified by iTRAQ labeling (Figure 4-12, Table 4-3).

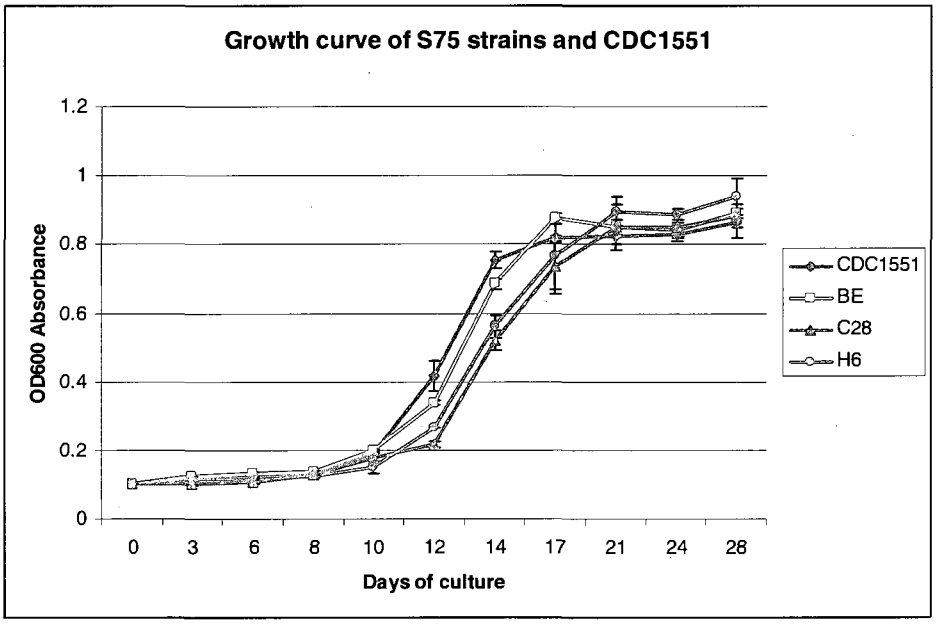


Figure 4-11. Growth curve of *Mtb* clinical isolates CDC1551, BE, C28 and H6.

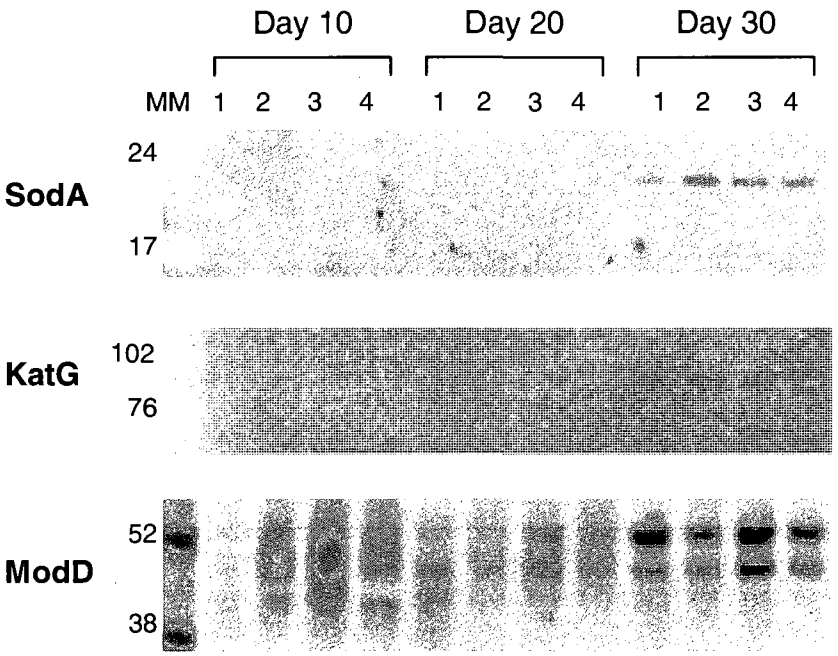


Figure 4-12. Western blot of selected proteins at days 10, 20 and 30 of in vitro culture. Graphic bars represent the quantitative values obtained by iTRAQ of SodA, KatG, and ModD. For the western blots, **MM:** Molecular ladder ECL Plex Rainbow marker (GE Healthcare), **1:** CDC1551, **2:** BE, **3:** C28, **4:** H6.

5. Chapter V: Preliminary study on Cfp2 secretion and function

5.1. Introduction

Secreted proteins of *Mtb* play important roles regarding nutrient uptake, adherence to host proteins, and modulation of the host immune response. Therefore, they are involved in the ultimate success of *Mtb* to colonize, survive and replicate in host cells [447, 448]. It has also been proposed that the differential expression of antigenic proteins might be related to differences observed in the immune response of TB patients and could in turn be responsible in part for the lack of a complete effective serodiagnostic method for TB [349, 449]. In addition, differential expression of immunogenic proteins by different *Mtb* strains might have a negative impact on the efficacy of subunit and rBCG vaccines.

Findings presented in the previous chapter indicate that several immunogenic proteins are expressed and/or secreted in higher quantity by *Mtb* CDC1551 (Lineage V) in relation to the S75 group of *Mtb* strains (Lineage IV) (Chapter IV, Figure 4-1, Table 4-1, Table 4-6). One of these proteins was Cfp2, an immunogenic protein with unknown function. The secreted and immunogenic protein Cfp2 (a.k.a. Mbt12) has been shown to induce lympho- proliferative responses and induction of IL-6 and TNF-α in cells obtained from TB patients [450, 451]. In addition, Cfp2 showed an increase induction of IFN-γ in TB patients after two months of treatment, which lead the authors to suggest that this antigen

could be useful to monitor host responses in TB patients undergoing drug treatment [451]. The low amount of Cfp2 in the S75 group of *Mtb* strains questions its possible efficacy as a target for immunodiagnostics.

Since the higher abundance of Cfp2 in the secreted fraction of *Mtb* CDC1551 was the most remarkable difference observed between *Mtb* strains from different phylogenetic groups, and given the importance that differential expression of immunogenic proteins can play on different aspects of TB control, we decided to preliminary investigate the secretion and possible function of this protein to understand the differences observed in the abundance of this protein between *Mtb* lineages.

Some intriguing aspects regarding the signal peptide of Cfp2 are interesting. The Cfp2 predicted signal sequence indicates a cleavage site at position 21 – 22 (AAA↓A) [452] which disagrees with the reported N-terminal sequence of the mature protein, which indicates a cleavage site at position 48-49 (LPL↓D) [450]. In addition, Cfp2 signal sequence is considerably longer than the ones frequently found in *Mtb* secreted proteins [452] and its cleavage site does not have the classical motif (i.e. AXA) that has been reported to be recognized by signal peptidase I (SPase I) in most microorganisms [453-455] and also correspond to the majority of cleavage sites in mycobacteria [450, 452].

These differences suggest that either SPase I in *Mtb* has different processing requirements or that Cfp2 might have a unique secretion mechanism. Supporting the former idea is the fact that mature Cfp2 shares the DP motif that has been reported in many *Mtb* secreted proteins and has also been suggested to play a role in protein localization [450].

In an effort to provide preliminary data for future research on this protein, a series of experiments were performed to generate tools and methods for the purification of native

and recombinant Cfp2. Site direct mutagenesis of key residues in the signal peptide of Cfp2 was performed as a preface to understand the processing of this unusual signal sequence. In addition, proteomic-based pull-down assays were performed to identify putative Cfp2 interacting proteins.

Results obtained from these pilot studies should guide future research regarding Cfp2 secretion and function, as the first step to understand the differences observed in the secretion of this protein between relatively distantly related *Mtb* strains (i.e. Lineages IV and V).

5.2. Hypotheses

Despite the apparent lack of homology between signal sequences in general, there is a well characterized motif at the cleavage site that seems to be essential for processing by the SPase I. This motif is found at positions -3 and -1 relative to the cleavage site and consist of small and neutral residues, with a predilection for alanine in both sites [456-458]. In addition, analysis of a subset of secreted proteins in *Mtb* also showed the predominance of alanine in both positions, especially in position -1, where this amino acid was present in 82% of the analyzed proteins [452], suggesting that residues at these positions are also important for cleavage and secretion in *Mtb*.

Consequently, we hypothesize that, if SPaseI is involved in cleavage of Cfp2, site directed mutagenesis involving positions -3 and -1 of Cfp2 signal peptide will reduced the secretion efficiency of this protein.

In addition, and even though no conserved motif has been described, amino acids upstream of the cleavage site have also been shown to play a role in protein secretion

[459, 460]. Therefore, and given the suggested role of the DP motif in mycobacterial secreted proteins [452], we hypothesize that mutations of position +1 (D) and +2 (P) in Cfp2 will also decrease the secretion of this protein in relation to the wild type.

Finally, identification of protein-protein interactions has been shown to be a useful strategy to elucidate potential roles of *Mtb* proteins for which functions are unknown [416, 461-464]. Thus, using a pull-down assay, we expect to identify potential proteins that interact with Cfp2 that could, in future studies, be validated by other more specific assays. In addition, we anticipate that identification of these interactions will generate a hypothesis regarding the possible role of Cfp2 in *Mtb* physiology.

5.3. Materials and methods

5.3.1. Extraction of mycobacterial DNA

Mtb strains BE, C28, H6 and CDC1551 were grown in 7H11/OADC plates for two weeks at 37°C. After this, colonies from agar were scraped and transferred into 500µl of sterile water. DNA extraction was performed as described elsewhere [465, 466] with some minor modifications. Briefly, cells were killed by heat shock at 80°C for 30min. Then, 70µl of 10% SDS and 50µl of proteinase K (10mg/ml) were added to the samples and incubated for 1 hour at 60°C. Samples were mixed by inversion every 10min. With samples still at 60°C, 100µl of pre-heated 5M NaCl and 100µl of pre-heated 10% CTAB (Sigma) were added and mixed by inversion. Samples were further incubated for 15 min at 60°C. Samples were then incubated at -70°C for 15 min, followed by thawing and reincubation at 60°C for additional 15 minutes.

At this point, samples were frozen overnight at -20°C, defrosted and mixed with 700ul of Chloroform:isoamyl alcohol (24:1 v/v). After centrifugation at 13000rpm for 10min, aqueous phase was recovered and transferred to a pre-chilled tube containing 700ul of isopropanol. After mixing several times by inversion, a DNA precipitate was obtained. At this point, samples were incubated at -20°C for at least 30 min and centrifuged at 13000 rpm for 10 min. DNA pellet was washed with 80% ethanol followed by centrifugation at 13000 rpm for 10min. Finally, the pellet was dried under vacuum and resuspended in ultrapure water. Quality of DNA was checked by agarose gel electrophoresis and ethidium bromide staining following standard procedures.

5.3.2. Production of *Mtb* electrocompetent cells

H37Rv cells were grown in 1L of 7H9/OADC/0.05% Tween until an O.D600 of 0.5-0.6. At this point, cells were pelleted for 10min at 3000rpm, 4°C; and then washed twice with 40ml of 10% glycerol. Finally cells were resuspended in 5ml of 10% glycerol. Stocks were made and stored at -80°C until further use. Cells were maintained on ice during the entire procedure.

5.3.3. QRT-PCR of *cfp2*

Mtb strains BE, C28, H6 and CDC1551 were cultured in triplicate in 7H9/OADC/0.05% tween at 37°C in agitation until midlog phase (OD_{600}: 0.4-0.5). At this point cells were harvested and washed twice with sterile PBS (Invitrogen), followed by resuspension in 10ml of Trizol (Invitrogen). Cells were broken by sonication on ice (50% duty cycle, 1min on, 1 min off, 6 times). Subsequently, 2ml of chloroform were added to each

sample and then centrifuged for 30min at 27,0000 g, 4°C. The upper layer was transferred to a 50ml falcon tube and mixed with 5ml of isopropanol. RNA was precipitated overnight at -80°C. Then, samples were slightly thawed and centrifuged at 3500rpm for 50min at 4°C. The RNA pellet was resuspended in 700µl of 80% Ethanol. Any large insoluble debris was allowed to settle to the bottom of the tube before continuing. RNA final purification was performed using an RNeasy column (QIAgen) following the manufacturer's instructions. Finally, RNA was eluted in 30ul of RNAse/DNase free water (Invitrogen) and stored at -20°C. RNA quantification and agarose gel electrophoresis were performed following standard procedures.

250ng of each RNA sample was subjected to cDNA synthesis using the SuperScript III First strand synthesis system (Invitrogen) following manufacturer's instructions. cDNA was stored at -20°C until further use.

QRT-PCR was performed using the SYBR-green system (Bio-rad) as described by the manufacturer. Primer sequences for QRT-PCR of *cfp2* were cfp2-L (5'-GTT CAG CGT GAC GA-3') and cfp2-R (5'-ACG TGA CGT TCT GC-3'). In addition, housekeeping *sigA* gene was analyzed to normalize the data. Primer sequence for *sigA* were sigA-L (5'-CGA CGA CGA AGA CCA CGA AG-3') and sigA-R (5'-TGT CCT TTT CGG TGG GTT CA-3'). Every sample was analyzed in triplicate. Each amplification plate contained *Mtb* genomic DNA to perform a standard curve at the following concentrations: 100, 10, 1 and 0.1ng/µl. QRT-PCR was performed in an iCycler (Bio-Rad) using the following thermal profile: 55°C for 2min, 95°C for 2min, 45 cycles of: 95°C, 15sec; 60°C, 30sec; 72°C, 45sec, final extension at 72°C for 5min. Melting curve was obtained by performing 65 cycles of 30°C, 10 sec increasing the temperature by 1°C after cycle 2. Quality of the

reaction was performed by analysis of standard curve and primer efficiency. Quantity of each analyzed gene was obtained after normalization against *sigA* and statistical significance was determined by the t-student test.

5.3.4. DNA sequencing of *cfp2*

The entire coding region of *cfp2* (cfp2-cr) for each strain was amplified using the following primers (5'-ATGAAGATGGTGAAATCGATCGCCG-3' and 5'-TCAGTTCCCTGCGGCCTGCA-3'). In addition, 500 bp corresponding to the upstream region of cfp2 (cfp2-prom) were amplified using the following primers: (5'-CCTGAGCTGCCTCGGCTTGTT-3' and 5'-TCAGTTCCCTGCGGCCTGCA-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) following standard procedures. DNA sequencing was performed at the Proteomics and Metabolomics Core facility (Colorado State University) and sequences were analyzed using Vector NTI software (Invitrogen)

5.3.5. PCR of *cfp2* for recombinant expression

PCR amplification of *cfp2* gene corresponding to both, precursor (with signal sequence, Cfp2-wg) and mature protein (without signal sequence, Cfp2-nss) was performed. Primer

sequences to obtain cfp2-wg (i.e. Rv2376c-F; Rv2376c-R) and cfp2-nss (i.e. Rv2376c-F2; Rv2376c-R) are shown in Table 5-1. Recombination sites attB1 and attB2, necessary for upstream work with the Gateway system (Invitrogen) were included at 5' end of each forward and reverse primer respectively. In addition, NdeI and HindIII restriction sites were included in each forward and reverse primer respectively. Platinum Taq DNA polymerase High Fidelity (Invitrogen) was used for PCR amplification as suggested by the manufacturer and using the following protocols:

Thermal profile for Cfp2-wg: 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.

Thermal profile for Cfp2-nss: was performed by 35 cycles of a. 94°C for 30sec, and b.68°C for 1.5 min, Final elongation was done at 68°C for 3min.

Table 5-1. Primers used for PCR amplification of *cfp2* coding and upstream region.

Name	Sequence
2376-F	5'-GGGGCAACTTTGTACAAAAAAGTTGCCCATATGAAGATGGTGAAATCGATCG-3'
2376-F2	5'-GGGGCAACTTTGTACAAAAAAGTTGCCCATATGTGCCGTTGGACCCGG-3'
2376-R	5'-GGGGCAACTTTGTACAAGAAAGTTGCAACGTTGTTCCCTGCGGCCTG-3'

5.3.6. Cloning and expression of wild type Cfp2

PCR products from above (Cfp2-wg and cfp2-nss) were agarose-gel purified using the PureLink Quick extraction kit (Invitrogen) and then cloned using the Gateway system (Invitrogen). Briefly, each product was subjected to a first recombination reaction with the entry vector pDONR. Then, pDONR-Cfp2-wg and pDONR-Cfp2-nss were each used

to perform the second recombination reaction with the expression vector pET DEST 161 following manufacturer's instructions. Each product was transformed in *E.coli* DH5-α competent cells (Invitrogen) following standard procedures.

***Expression in E.coli*:** pETDEST 161-Cfp2-wg and pETDEST 161-Cfp2-nss were transformed into *E.coli* BL21 DE3 star (Invitrogen) following standard procedures. After transformation cells were incubated overnight and then upscaled to 1L of LB media containing 100µg/ml of ampiciline (Sigma). When this culture reached an OD600 of 0.5, rCfp2 expression was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5mM. Five hours after induction, cells and CFP were harvested. Pelleted cells were resuspended in 10ml of Ni-chromatography binding buffer (20mM tris-HCl, 500mM NaCl, 5mM imidazole, pH 7.9) containing EDTA-free protease inhibitor (Roche Diagnostics) and 200µg/ml lysosyme (Sigma). Cells were incubated at 37°C for 1h and then lysated by sonication on ice (50% duty cycle, 60 sec on, 90 sec off, 6 times). Whole cell lysates were centrifuged at 13.000 rpm for 90min. Supernatant containing soluble rCfp2 was stored at -20°C until further analysis. Pellet containing inclusion bodies and insoluble rCfp2 were incubated for 1h at 37°C in binding buffer complemented with EDTA-free protease inhibitor (Roche Diagnostics) 6M urea and lysosyme (100ug/ml). Insoluble rCfp2 was recovered by centrifugation at 13000 rpm for 30min. Presence of rCfp2 in whole cell lysates and inclusion bodies was determined by SDS-PAGE and western blot following standard procedures. Anti-penta His antibody was obtained from Qiagen. 2D-GE was performed as described in chapter III. N-terminal sequencing was performed as described earlier.

***Expression in Mtb H37Rv*:** Cfp2-wg was released from the corresponding pETDEST 161 after enzyme digestion with NdeI (Invitrogen) and HindIII (Invitrogen) following standard procedures. Products were agarose gel purified as described above and then ligated in plasmid pVV16 following standard procedures. Briefly, 90µl of H37Rv competent cells were gently mixed with 5ul of plasmid DNA (500ng) and incubated at RT for about 30min. The following parameters were used for electroporation: Voltage 1.25KV, Resistance 1000Ω and Capacitance 25µF. After electroporation, cells were immediately transferred to 5ml of 7H9/OADC/0.05% Tween and incubated at 37°C in agitation for 24h. After this time, each reaction was plated in 7H11 (Becton Dickinson) media containing 25µg/ml of kanamycin and 25µg/ml Hygromicin B. Approximately two weeks later, single colonies were picked and transfer to 10ml of GAS/0.05%Tween containing 50µg/ml kanamycin and incubated at 37°C in agitation for two weeks. Finally, cultures were upscaled two times to 100ml and then to1L of GAS/0.05%Tween/50µg/ml kanamycin. GAS media was prepared as described elsewhere [391]. Final upscaled cultures were incubated for two weeks in agitation at 37°C and then cells and CFP were harvested as described on Chapter III. Presence of rCfp2 was determined as before.

5.3.7. Generation of Cfp2-signal peptide mutants

Two hundred microliters of pVV16-Cfp2-wg were used as the DNA template for site-directed mutagenesis. Primers and expected mutations are shown in Table 5-2.

Table 5-2. Primers used for site directed mutagenesis

Mutation*	Name	Sequence
+2 (P>V)	Pos+2 P-V F	5'-/Phos/CCACTGCCGTTGGAGTGGCATCCGCCCCTGAC -3'
	Pos+2 P-V R	5'-/Phos/GTCAGGGGCGGATGCCACGTCCAACGGCAGTGG-3'
+1 (D>E)	cleavE-F	5'-/Phos/GCGCCACTGCCGTTGGAGCCGGCATCCGCC -3'
	cleavE-R	5'-/Phos/GGCGGATGCCGGCTCCAACGGCAGTGGCGC-3'
+1 (D>L)	cleavL-F	5'-/Phos/GCGCCACTGCCGTTGCTGCCGGCATCCGCC-3'
	cleavL-R	5'-/Phos/GGCGGATGCCGGCAGCAACGGCAGTGGCGC-3'
-1 (L>V)	Pos -1V-F	5'-/Phos/CGGCGCGCCACTGCCGGTGGACCCGGCATCC-3'
	Pos -1V-R	5'-/Phos/GGATGCCGGGTCCACCGGCAGTGGCGCGCCG-3'
-1 (L>S)	Pos -1S-F	5'-/Phos/CGGCGCGCCACTGCCGTCGGACCCGGCATCC-3'
	Pos -1S-R	5'-/Phos/GGATGCCGGGTCCGACGGCAGTGGCGCGCCG-3'
-3 (L>V)	Pos -3V-F	5'-/Phos/CGGCGCGCCAGTGCCGTTGGACCCGGCATCC-3'
	Pos -3V-R	5'-/Phos/GGATGCCGGGTCCAACGGCACTGGCGCGCCG-3'
-3 (L>V)	Pos -3S-F	5'-/Phos/CGGCGCGCCATCGCCGTTGGACCCGGCATCC-3'
	Pos -3S-R	5'-/Phos/GGATGCCGGGTCCAACGGCGATGGCGCGCCG-3'
DelAAA -27 to -29	DelAAA F	5'-/Phos/GCTGCAATCGGCGCCGGTGTGACTTGGATC-3'
	DelAAA R	5'-/Phos/GATCCAAGTCACAGGGGCGCCGATTGCAGC-3'

*All mutation positions are relative to Cfp2 experimental cleavage site (pos 48-49).

Site directed mutagenesis reactions and *E.coli* transformations were performed using the Quick exchange –XL Site Directed mutagenesis kit and XL10-gold competent cells respectively (Stratagene), following manufacturer's instructions. Before electrotransformation of *Mtb* was performed, all mutations were confirmed by DNA sequencing.

5.3.8. Purification of native Cfp2

Briefly, CFP from H37Rv was subjected to 40% ammonium sulfate and proteins precipitated by centrifugation at 27,000 x 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 antiserum. Fractions containing Cfp2 were pooled together and protein polishing accomplished by size exclusion chromatography (Superdex 75, 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.

25µg of Cfp2 were analyzed by 2D-GE as described before (Chapter III). After electrophoresis, gel was blotted into a PVDF membrane following standard procedures and then spots were visualized with Safe Blue (Bio-rad). Spots from membrane were

sequenced for at least 6 cycles by Erdman degradation at the Proteomics and Metabolomics Core facility at CSU.

5.3.9. Nickel chromatography

Purification of both wild type and mutated rCfp2 was performed as follows. Appropriate volume of resin was packed in a poly-prep column and washed with endotoxin-free water. Resin was charged by running 2.5 column volumes (CV) of charge buffer (50mM $NiSO_4$) through the column and then equilibrated with 1.5 CV of binding buffer (20mM Tri-HCl, 500mM NaCl, 5mM Imidazole, pH 7.9). After this, sample (in binding buffer) was applied to the column and then washed with 10 CV of binding buffer followed by 10 CV of wash buffer (20 mM Tris-HCl, 500 mM NaCl, 60 mM Imidazole, pH 7.9), 10 CV of 10mM Tris-HCl, pH 8.0 and 10 CV of 0.05% ASB-14 in 10mM Tris-HCl. Final wash was 10 CV of 10mM Tris-HCl, followed by elution with 1M imidazole in 10mM Tris-HCl, pH 8.0. Purified recombinant protein was dialyzed against 10mM Ammonium bicarbonate for 48h. Protein concentration was determined by the BCA method (Pearce).

5.3.10. Protein-protein interaction Pull down assays

Twenty five micrograms of the rCfp2-wg produced in *E.coli* was mixed with 250µg of different subcellular fractions obtained from *M.tb* H37Rv (i.e. cytosol, membrane, cell wall and CFP). The mix was incubated overnight at 4°C with gentle agitation. A control assay was performed by incubating each of the subcellular fractions without the recombinant protein. Each mixture was applied to a 0.2ml Ni- column and processed as described above. Following elution, samples were dried down and resuspended in 6M

guanidine hydrochloride, 0.6M Tris-HCl and then reduced by adding 4M DTT (Sigma) and incubating at room temperature for 3hr. Samples were then alkylated by adding 160mM iodoacetamide (Sigma) and incubated in the dark at 37°C for 30min. Samples were desalted by dialyzing against 10mM Ammonium bicarbonate for 48hr. Finally, samples were dried under vacuum and resuspended in 100µl of 10% Acetonitrile, 0.2M ammonium bicarbonate and digested overnight with 4µg of trypsin (Roche).

Each sample was analyzed by Tandem LTQ/MS/MS as described before (Chapter III). Spectra were analyzed by interrogation of the *Mtb* genome using Sequest and data collated using Scaffold.

5.4. Results

5.4.1. Bioinformatic analysis of *Rv2376c* and Cfp2

Rv2376c (*cfp2*), the gene encoding the immunogenic protein Cfp2 is apparently located as a single gene operon [467] between Rv2375, which encodes a hypothetical conserved protein, and *mbtG*, which is the last gene in the mycobactin operon (Figure 5-1). Even though *cfp2* is adjacent to the mycobactin operon, there is no evidence that *cfp2* gene is co-transcribed with this operon [467].

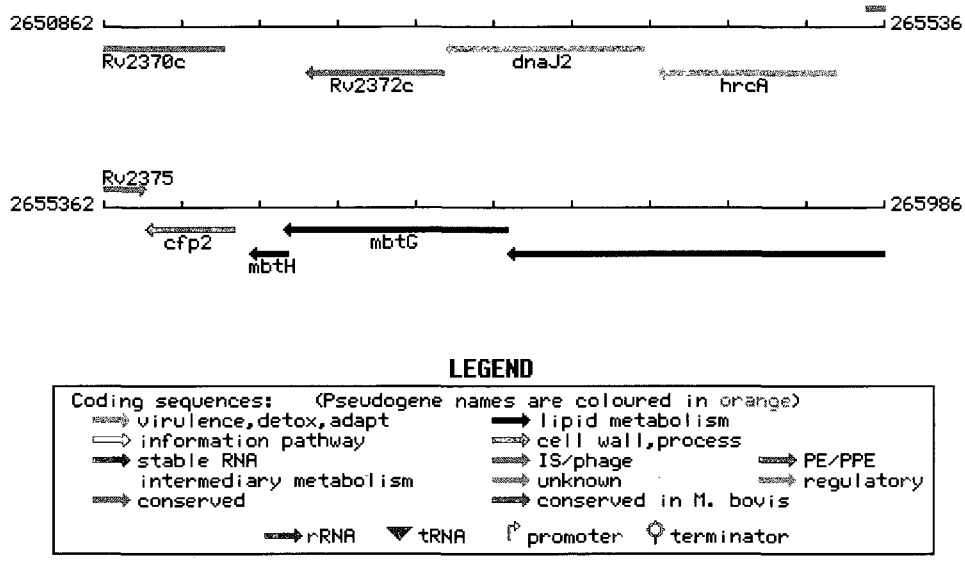


Figure 5-1. Localization of *cfp2* in the *Mtb* genome.
Image obtained from Tuberculist (http://genolist.pasteur.fr/TubercuList/)

Sequence of the *cfp2* gene, including its 500bp upstream region, was obtained after PCR amplification of genomic DNA from each of the *Mtb* strains (CDC1551, BE, C28 and H6). These sequences were also compared against *Mtb* H37Rv genome using the Blastn program [468]. The 500bp- upstream region of *cfp2* was analyzed to identify possible mutations in the regulatory region of this gene in *Mtb* CDC1551 that could be responsible for the higher abundance of Cfp2 in this strain. The 500bp-upstream region was identical in all of the strains, including *Mtb* H37Rv. In contrast, a single point mutation previously reported for the coding region of *cfp2* in *Mtb* CDC1551 was confirmed at position 426 (A>G), resulting in an amino acid change (Q131R). Sequences of *Mtb* clinical isolates BE, C28 and H6 did not present this SNP, and were identical to *cfp2* from *Mtb* H37Rv. This single mutation does not have a major effect on the predicted molecular weight of the protein (16.635 vs 16.663 KDa), however, the predicted isoelectric point is higher for *Mtb* CDC1551 Cfp2 (6.72) compared to Cfp2 from the S75 strains (6.04).

Cfp2 amino acid sequence was analyzed by different bioinformatic programs to obtain more information regarding this protein.

A prediction of transmembrane helices was performed using the program TMHMM ver2.0 (http://www.cbs.dtu.dk/services/TMHMM/) [469, 470]. This prediction revealed a possible transmembrane region from residues number 7 to 29 of the Cfp2 sequence. In accordance with the secretion nature of this protein, residues 30 to 168 are predicted to be extracellular (Figure 5-2). Predicted transmembrane regions at the N-terminus of secreted proteins are generally considered to represent the signal peptide due to the hydrophobic, h-region, of the signal peptide. Predictions using other programs, such as DAS for prokaryote transmembrane helices produced similar results (http://www.sbc.su.se/~miklos/DAS) [471] (Data not shown).

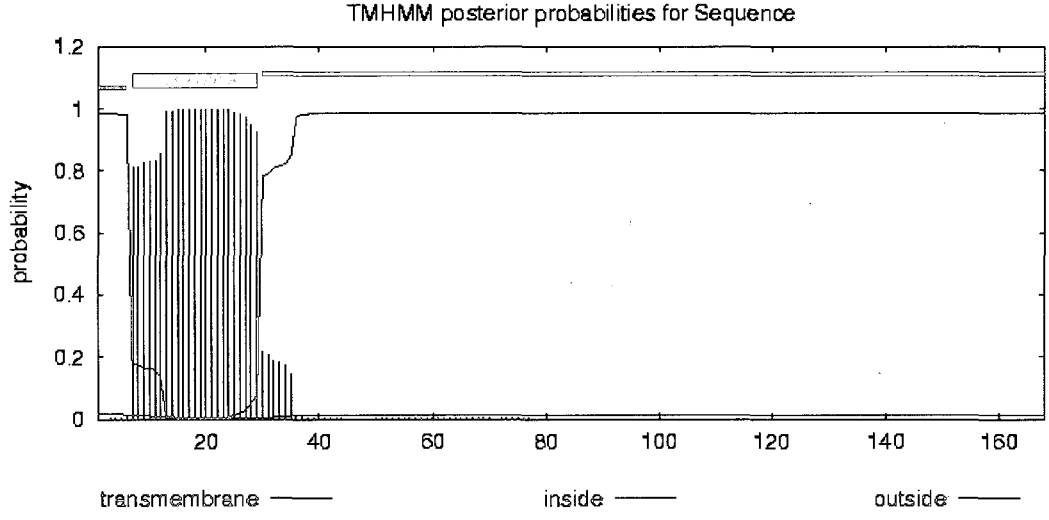


Figure 5-2. Transmembrane helices prediction of Cfp2.
Prediction was performed using TMHMM ver2.0 (http://www.cbs.dtu.dk/services/TMHMM/)

Transmembrane helices predictions are based on the hydrophobicity of the sequence and in other parameters such as charge bias and helix length, but hydrophobicity analysis using the Kyte & Doolittle scale [472] and the ProtScale program

(http://www.expasy.ch/tools/protscale.html) retrieve similar results (Figure 5-3). In the Kyte & Doolittle scale, the closer a group of residues is to 2.0, the higher the probability that this region is hydrophobic. In contrast, if the score is close to -2.0, the region is considered to be hydrophilic. Similar to the transmembrane prediction, Figure 5-3 shows a hydrophobic region at the N-terminus of Cfp2, ending around residue number 30. However, the hydrophobic analysis shows another short, but highly hydrophobic region from residue 40-45.

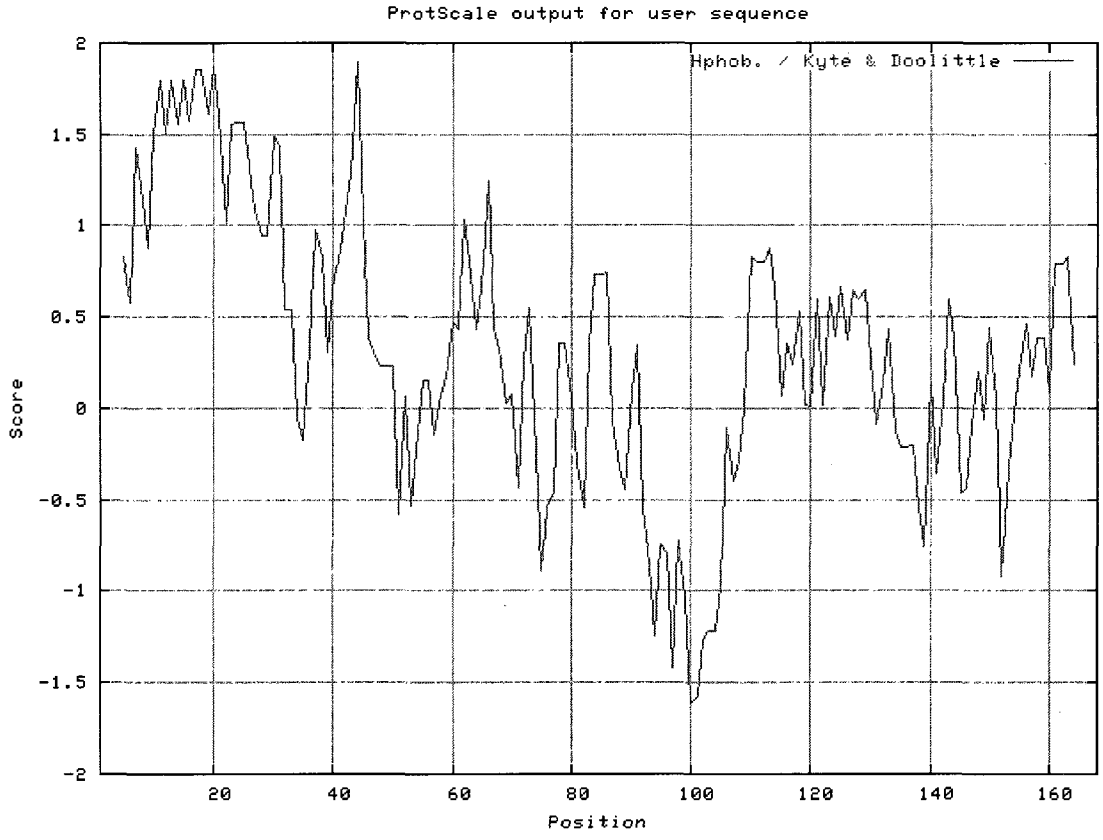


Figure 5-3. Hydrophobicity prediction of Cfp2
Prediction was performed using ProtScale (http://www.expasy.ch/tools/protscale.html) and the Kyte and Doolittle scale.

Cfp2 sequence was analyzed by SignalP ver3.0 (http://www.cbs.dtu.dk/services/SignalP/) using both neural networks and hidden Markov model algorithms [473-475]. This program predicts not only if a protein is secreted or not, but also the most likely cleavage site in signal peptides of predicted secreted proteins. Since the algorithms used for prediction have not been trained on mycobacterial proteins, both gram negative (Figure 5-4) and gram positive (Figure 5-5) datasets were used.

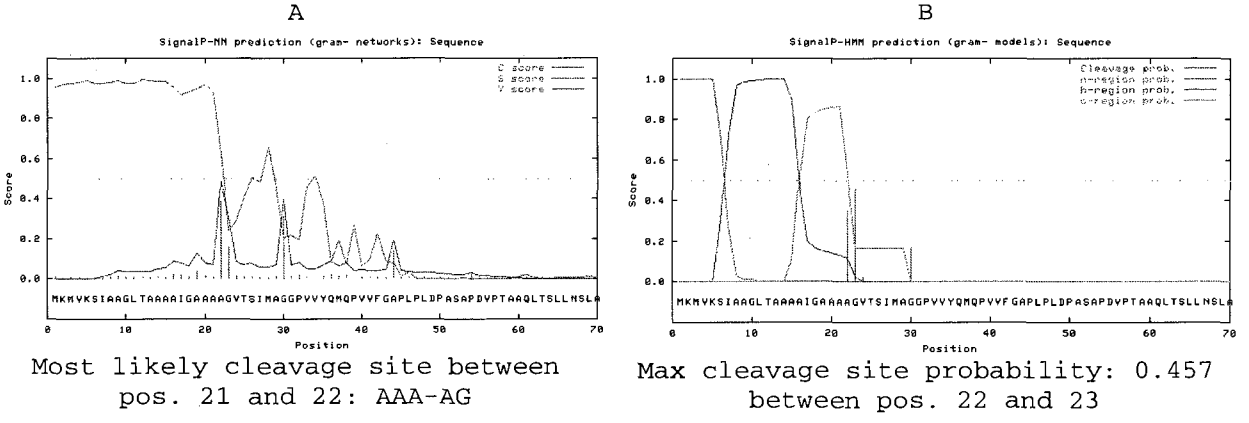


Figure 5-4. Prediction of signal peptide and cleavage site of Cfp2 by SignalP using the gram negative database.

A. Prediction by Neural Networks algorithm, **B.** Prediction by Hidden Markov Model algorithm.

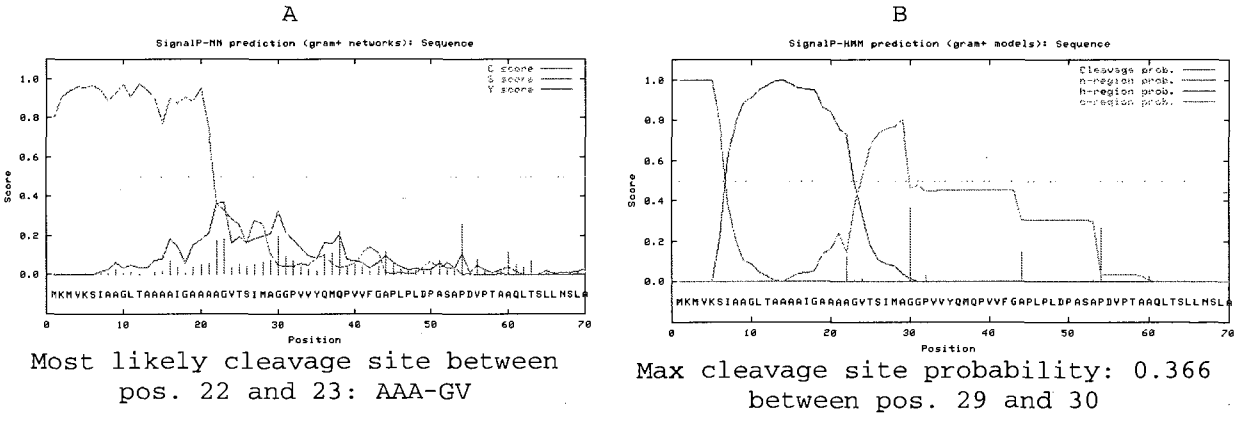


Figure 5-5. Prediction of signal peptide and cleavage site of Cfp2 by SignalP using the gram positive database.

A. Prediction by Neural Networks algorithm, **B.** Prediction by Hidden Markov Model algorithm.

Most of the predictions calculate the cleavage site close to residue 22, with the exception of the hidden markov model using the gram positive database, which predicts the cleavage site between positions 29 and 30. These results seem to agree with the transmembrane helices prediction, which indicated that the first 30 residues may correspond to the signal peptide. However, none of the algorithms predicted the correct cleavage site of Cfp2 signal peptide, which has been experimentally determined to be between positions 48 and 49. As mentioned previously, hydrophobicity analysis shows a hydrophobic region between residues 40 to 45, indicating that the h-region of the signal peptide of Cfp2 might be longer than predicted by SignalP ver3.0, and may explain in part the discrepancy between predicted and observed signal peptide of Cfp2.

5.4.2. Analysis of *cfp2* gene expression

In order to evaluate if the higher production and/or secretion of Cfp2 by *Mtb* CDC1551 was due to a higher expression of the *cfp2* gene, QRT-PCR was performed on cDNA obtained from each *Mtb* strain. Three biological replicates, in addition to three technical replicates were analyzed. Normalization of *cfp2* transcription levels was performed using the housekeeping gene *sigA*. Even though there was high variation of *cfp2* levels between biological replicates of the S75 group of *Mtb* strains, it was determined that expression of *cfp2* by *Mtb* CDC1551 was significantly lower than the expression of this gene by *Mtb* BE and *Mtb* H6 (p-values of 0.031 and 0.038, respectively) (Figure 5-6). Expression of *cfp2* was not statistically different for *Mtb* strain CDC1551 versus C28 or H37Rv. However, *cfp2* levels in *Mtb* C28 and *Mtb* H37Rv were higher than in *Mtb* CDC1551. These results were unexpected because Cfp2 protein levels are higher in *Mtb* CDC1551

in both cytosol and CFP in comparison to the S75 group (Chapter IV, Figure 4-9, Table 4-2, Table 4-3) and CFP of *Mtb* H37Rv (data not shown).

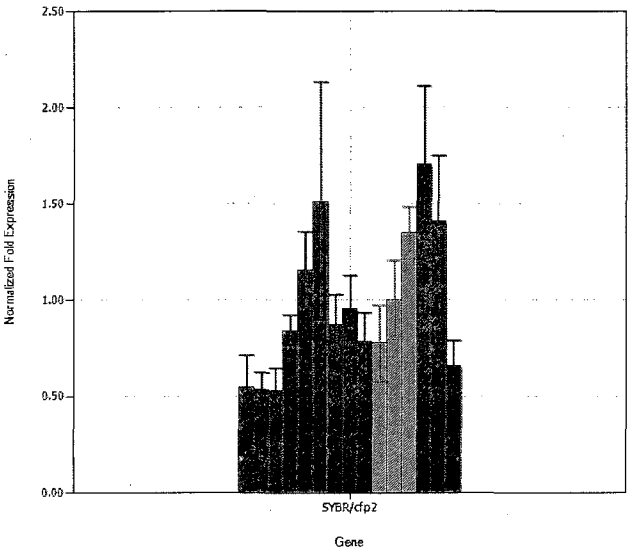


Figure 5-6. Normalized fold expression (against *sigA*) of *cfp2* transcripts in *Mtb* strains.
Blue: Mtb CDC1551, Green: Mtb BE, Red: Mtb C28, Orange: Mtb H6, Purple: Mtb H37Rv. Error bars indicate the standard deviation of technical replicates for each biological replicate.

The lack of correlation between *cfp2* transcript and Cfp2 protein levels suggest that the higher abundance of Cfp2 in *Mtb* CDC1551 could be associated with post-translational regulatory mechanisms, including a possible higher rate of secretion of this protein. Since the processing of this protein and its signal peptide during secretion is unknown, we decided to perform preliminary experiments to understand more about the Cfp2 secretion mechanism.

5.4.3. Production and analysis of native Cfp2

Native Cfp2 was purified from the CFP of *Mtb* H37Rv following a three step separation process. Precipitation of *Mtb* CFP with 40% ammonium sulfate is regularly used in the native purification of the Ag85 complex [476]. Unpublished observations in our

laboratory have identified Cfp2 in the supernatant of the 40% ammonium sulfate cut during standard preparation of Ag85 complex. Sequential rounds of ammonium sulfate precipitation have been demonstrated to be a useful strategy to decrease the complexity of protein fractions from *Mtb* [323]. Therefore, our first step in the purification procedure of native Cfp2 was the precipitation of the CFP with 40% ammonium sulfate, followed by concentration of the obtained supernatant adding ammonium sulfate to 70% saturation point. The second step for native Cfp2 purification was a cation exchange chromatography. SP-sepharose fast flow (Amersham) resin was used to bind Cfp2 while allowing passage of contaminating proteins. Several loading and elution gradients at different pH levels were attempted and their performance evaluated by analyzing the eluting fractions by SDS-PAGE followed by silver staining and western blot using a polyclonal Anti-Cfp2 mouse antibody (Data not shown). The recommended loading buffer pH for this type of strong cation exchange resin is at least one unit below the isoelectric point of the protein of interest. Mature Cfp2 has a predicted isoelectric point of 5.1. However, loading buffers with a pH below 4.5 did not show a good separation of Cfp2 from other contaminating proteins (data not shown). The loading buffer with the best performance was 50mM Sodium acetate, pH 4.5. Bound proteins were eluted using a linear gradient from 10% - 50% elution buffer (Sodium acetate, pH 5.5, 1M NaCl). At this point, most of the contaminating proteins, as determined by visual analysis of protein fractions by SDS-PAGE, were at least 6kDa bigger than Cfp2 (12kda). Therefore, the last step in the purification process was polishing by size exclusion chromatography using Superdex 75 resin (GE-Healthcare), which can be used to separate proteins in a molecular mass ranging from 3 to 70kDa.

Fractions from this fractionation were resolved by SDS-PAGE and stained with coomassie brilliant blue (Figure 5-7). In addition to the band corresponding to Cfp2, two additional bands with a lower intensity were also observed in the Cfp2 enriched fractions. All of them were subjected to 'in gel' trypsin digestion followed by identification by Tandem MS/MS.

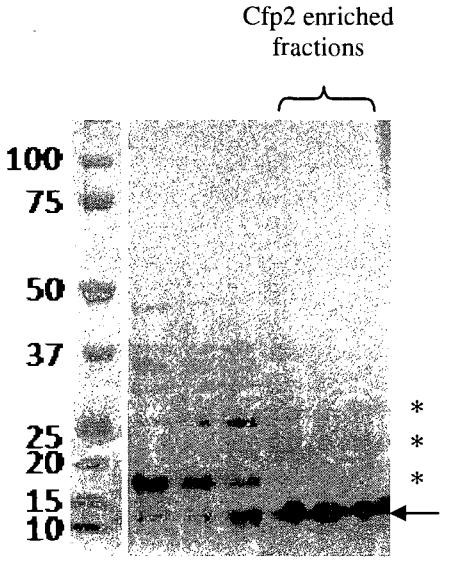
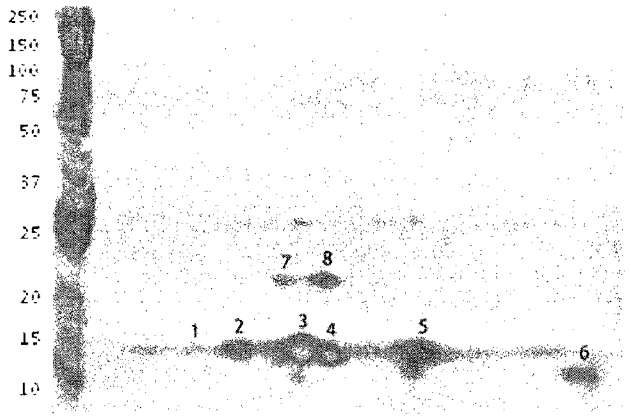


Figure 5-7. Size Exclusion Chromatography, final step in native Cfp2 purification.
Arrow shows band corresponding to Cfp2. Asterisks (*) show bands with a Mr higher than expected but with the presence of Cfp2 confirmed by tandem MS/MS

Tandem MS/MS confirmed the presence of Cfp2 in all analyzed bands. Interestingly, this protein was also present in bands appearing at a molecular weight higher than expected (Figure 5-7) suggesting the presence of Cfp2 multimers. In some cases, tandem MS/MS analysis indicated that thioredoxin (TrxC) co-purified with Cfp2. Two contaminating spots, most likely corresponding to a mixture of proteins due their ambiguous N-terminal sequence, were observed when fraction 3 (Figure 5-7, last lane) was resolved by 2D-GE (Figure 5-8). However, N-terminal sequencing of major spots from fraction 3 (Figure 5-7, last lane), showed the unequivocal presence of Cfp2 (Figure 5-8).



1. DPASAPD
2. DPASAPDVPTA
3. DPASAPDVPTA
4. DPASAPDVPTA
5. ASAPDVP
6. SLVNSLA
7. (A/S)VESXM
8. (A/S/V)(V/L)(E/I)(S/L)(Q/N)M

Figure 5-8. 2D-GE of native Cfp2 and N-terminal sequencing of Cfp2 isoforms.
For fraction information, refer to last lane, Figure 5-1. Based on N-terminal sequence, spots 1-6 correspond to Cfp2. Spots 7-8 correspond most likely to a mixture of proteins. Fifth residue of spot 7 was not identified and is represented with the letter X.

All selected spots, except number 5 and 6 correspond to the expected mature protein after cleavage of the signal sequence, DPASAPD [450]. Surprisingly, spot 5 resulted in an N-terminal sequence in which the first two residues (DP) are removed from the mature protein, suggesting an additional post-translational modification. Spot 6 did not present a clean sequence, however the main sequence was called SLVNSLA which does not match Cfp2 completely, but is similar to residues 64 – 70 of preprotein Cfp2 (i.e.SLLNSLA). Interestingly, analysis of the third position of this sequence presents peaks corresponding to both valine and leucine, this last one having a height of almost twice the value of the valine peak. This could be due to retention of the leucine present at position 2, however, it is most likely due to the actual presence of leucine at position 3 (data not shown). Therefore, this spot probably belongs to a shorter Cfp2 in which the first 63 amino acids are missing. It is unknown if this isoform represents a real post translational modification or if it is due to protein fragmentation caused by other reasons.

5.4.4. Production and analysis of recombinant Cfp2

High yield production of *Mtb* recombinant proteins in *E.coli* has been previously demonstrated elsewhere [346, 442, 477-479]. This and the extensive experience that our laboratory has on the production of *Mtb* recombinant proteins, led us to chose *E.coli* for the production of rCfp2. Both forms of Cfp2, whole protein (Cfp2-wg) and protein lacking the signal sequence (Cfp2-nss), were successfully produced in *E.coli.* However, the construct corresponding to the mature protein (Cfp2-nss) showed some level of toxicity to *E.coli* cells, and therefore, the production levels for this construct were lower than for the clone expressing the precursor form of Cfp2 (Cfp2-wg) (data not shown).

As expected, both constructs of recombinant Cfp2 were detected only in the cell lysate and missing from the culture filtrate, indicating that *E.coli* is unable to export rCfp2. Interestingly, expression of Cfp2-wg (precursor) in *E.coli* resulted in the production of two different forms of Cfp2, with a mass difference of approximately 4KDa (Figure 5-9). rCfp2 was also found in the inclusion bodies containing insoluble proteins (data not shown).

Even though *E.coli* was chosen with the sole purpose of obtaining rCfp2, finding two different forms of this protein in *E.coli* cell lysates led us to explore the possibility that these forms correspond to both precursor and mature Cfp2.

To confirm if *E.coli* is able to cleave rCfp2-wg, the purified protein was separated by 2D-GE and blotted into PVDF membrane. Spots were selected based on molecular weight and subjected to N-terminal sequencing (Figure 5-10)

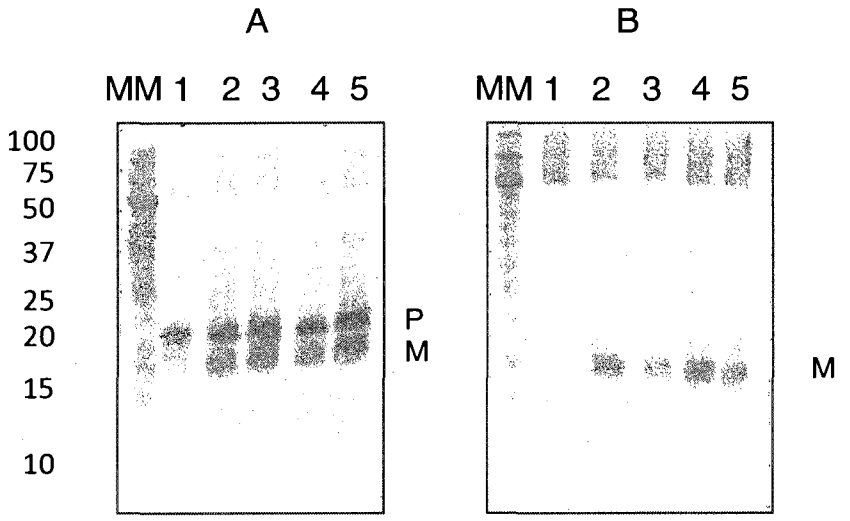
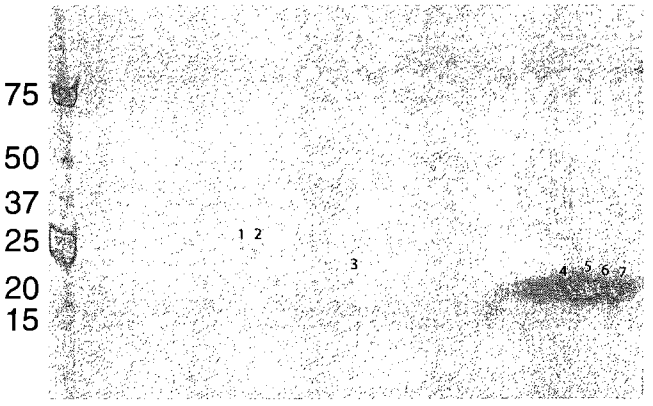


Figure 5-9. Western blot of *E. coli* BL21 DE3 star whole cell lysates (5ug per lane).
A. Lysate from pET DEST 161-Cfp2-wg. **B.** Lysate from pET DEST 161-Cfp2-nss. **Lanes 1-5** correspond to time points 0 – 4 hours after induction. Primary antibody: Anti-Penta His (Qiagen). **MM:** Precision Plus Protein Standard Dual Color (Bio-Rad), P= precursor, M= mature protein



Spot #	N-terminus sequence
1	MKVAKD
2	MKVAKD
3	(M/S/L)KIFDF
4	MQPVVF
5	MQPVVF
6	MQPVVF
7	MQPVVF

Figure 5-10. N-terminal sequence of rCfp2-wg produced in *E . coli*.
2D-GE of rCfp2-wg produced in *E. coli* (25µg). Spots 4 – 7 correspond to Cfp2, starting at residue 36 of the preprotein. Spots 1-3 correspond to *E.coli* contaminating proteins. Blastp analysis indicates that these proteins probably correspond to peptidylprolyl isomerase FKBP-type for spots 1 and 2 and fructose-bisphosphate aldolase, class II for spot 3.

Based on the N-terminal sequence, spots 1 to 3 correspond to contaminating proteins. N-terminal sequence of spots spots 1-2 present 100% homology with the N-terminus of a peptidylprolyl isomerase FKBP-type from *E. coli* BL21(DE3). This protein has high histidine content at the C-terminus, explaining its presence in the eluate from the Ni-

chromatography. Spot 3 is also a contaminating protein and probably belongs to a fructose-bisphosphate aldolase, class II from *E. coli* BL21(DE3). The reason for the presence of this protein in the eluate is possibly due to unspecific interaction with the Ni-column.

Spots 4 – 7 correspond to rCfp2, lacking the first 36 residues. This is in contrast to *Mtb* mature Cfp2 which lacks the first 48 amino acids. This finding suggests that the cleavage site in *E.coli* might be different. Another possible explanation for this finding could be an alternate translational start site, in which case, no cleavage of rCfp2 is performed by *E.coli*. Shine-Dalgarno (SD) sequences [480], AGGAGGU, upstream of the initiation codon are required for ribosome binding and translation. Therefore, if an alternative translational start site is the reason for the apparent truncation of rCfp2 in *E.coli*, a SD sequence should be located intragenically in the 5' region of Cfp2. However, no SD sequence was identified in the intragenic sequence of Cfp2.

5.4.5. Mutagenesis of Cfp2 signal sequence

Cfp2 signal sequence was subjected to site directed mutagenesis to obtain two different mutations (i.e. one neutral and one radical) at each of the following targeted positions: -3, -1 and +1. Mutations at these positions were primarily selected on the basis of amino acid physicochemical characteristics (polarity, charge and size), by choosing an amino acid with either similar properties to the wild type residue (i.e neutral change) or with different characteristics (i.e. radical). In addition to these parameters, mutations for aminoacids at position -3 and -1 were also selected based on previous studies where the most frequent aminoacids at these positions have been identified [455, 459, 474]. Valine

was chosen as a neutral substitution for positions -3 and -1 based on its similar characteristics to leucine and because it is frequently present at these positions in signal peptides from other bacteria. Serine, in contrast to leucine, is a polar amino acid, and although it has been reported at position -3 in other prokaryotes, it is not frequently found at either -3 or -1 positions in *Mtb* signal peptides [452]. Serine was chosen as a more radical substitution to leucine at both -3 and -1 positions. Position +1 (D) was substituted by either a glutamic acid (neutral substitution) or a leucine (radical substitution). One mutation at position +2 (P>V) was performed. Selection of a valine for this substitution was based on previous reports that indicate valine is not frequently found at position +2 in *Mtb* proteins [313, 452]. Finally, deletion of three alanines at position -27 to -29 with relation to cleavage site was also evaluated. Position -27 to -29 correspond to one of the predicted cleavage sites by SignalP ver3.0. Each of the pvv16-Cfp2 mutants were transformed and expressed in *Mtb* H37Rv competent cells.

The presence of rCfp2 in the culture filtrate of each *Mtb* H37Rv transformed with rCfp2 wild type, and each of the mutagenesis constructs, was determined by Anti-His western blot (Figure 5-11). When compared to the relative amount of rCfp2-wild type, none of the mutations completely abolished secretion of rCfp2, however, when a serine was introduced at position -3, instead of leucine, a marked decreased in secretion of rCfp2 was observed (Figure 5-11, lane 6). In contrast, introduction of a valine instead of leucine at position -1 seems to have a positive impact on the secretion of Cfp2 as compared to the rCfp2 wild type (Figure 5-11, lane 5).

Signal cleavage site was confirmed by N-terminal sequencing of all rCfp2 mutants, in addition to the wild type. Cfp2 bearing mutations at pos-3, pos -1 and deletion of three

alanines at pos-17-19, as well as the wild type, presented the expected DPASAP, sequence. Cfp2 mutants bearing changes at pos +1 and pos+2 also presented the expected sequence EPASAP, LPASAP and DVASAP for pos+1 (D>E), pos+1 (D>L) and pos+2 (P>V) respectively.

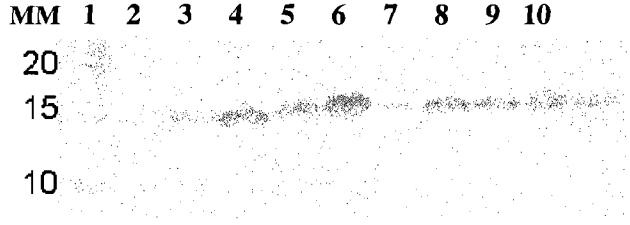


Figure 5-11. Western blot of CFP obtained from *Mtb* H37Rv expressing Cfp2 signal peptide mutants. **1.** Control pvv16, **2.** pvv16-Cfp2-wg (pos +1, D>E), **3.** pvv16-Cfp2-wg (pos +1, D>L), **4.** pvv16-Cfp2-wg (pos -1, L>S), **5.** pvv16-Cfp2-wg (pos -1 , L>V), **6.** pvv16-Cfp2-wg (pos -3, L>S), **7.** pvv16-Cfp2-wg (pos -3, L>V), **8.** pvv16-Cfp2-wg (Wild type), **9.** pvv16-Cfp2-wg, (deletion AAA, pos-27-29), **10.** pvv16-Cfp2-wg (pos+2, P>V). CFP was obtained after two weeks of culture.

5.4.6. Cfp2 interacting proteins

A systematic pull-down assay was performed to detect Cfp2 interacting proteins. Cytosol, membrane, cell wall and CFP were each incubated with rCfp2-wg (obtained from *E.coli*) and purified by Ni-chromatography. In addition, blank controls without the bait protein were performed. Several proteins were identified by tandem MS/MS in the eluate of each assay. Cell wall and cell membrane controls also resulted in the identification of several proteins. This could be due to unspecific interaction with the resin or to retention of some proteins due to some insolubility of these samples. Nonetheless, only proteins that elute with Cfp2 and were absent from these negative controls were considered as possible interacting proteins (Table 5-3).

Interestingly, even though, Cfp2 is a secreted protein, most of the interacting proteins were identified when Cfp2 was incubated with cell wall and membrane fractions.

Table 5-3. Putative Cfp2 interacting proteins.

Protein ID	Accession Number	Molecular Weight	Fraction	NSAF*
Hypothetical protein Rv0580c	gi|15607720|	18	CW Mem	83.32 27.78
Hypothetical protein Rv1738	gi|15608876|	11	CW Cyt Mem	63.64 27.27 18.18
Heat shock protein, HspX	gi|15609168|	16	Cyt	62.50
6 KDA early secretory antigen, Esat-6	gi|57117165|	10	CW	40.00
50S ribosomal protein L15, RlpO	gi|15607863|	16	Mem Cyt	37.50 6.25
50S ribosomal protein L14, RplN	gi|15607854|	13	Mem	30.77
30S ribosomal protein S15, RpsO	gi|15609922|	10	Cyt CW	30.00 20.00
Hypothetical protein Rv2632c	gi|15609769|	10	Cyt	30.00
Preprotein translocase subunit, SecE	gi|57116765|	17	CW	23.53
Hypothetical protein Rv2298	gi|15609435|	35	CW	22.86
50S ribosomal protein L28, RpmB	gi|15609195|	9	Mem	22.22
Hypothetical protein Rv0088	gi|15607230|	25	CW	20.00
Possible Fatty acid synthase, Rv3720	gi|15610856|	47	CW	19.15
Probable-DNA binding protein, HupB	gi|15610123|	22	CW	18.18
Hypothetical protein Rv1919c	gi|15609056|	17	CW	17.65
Probable conserved membrane protein, Rv0479c	gi|15607620|	37	CW	16.22
Possible ribonucleotide transport ATP-binding protein, MKL	gi|15607795|	39	CW	15.39
30S ribosomal protein S2, RpsB	gi|15610027|	31	CW	12.90

Hypothetical protein Rv2216	gi|15609353|	32	CW	12.50
Possible conserved membrane protein, Rv2520c	gi|15609657|	8	Mem	12.50
Possible protein transport protein, SecE2	gi|57116727|	8	Mem	12.50
3-ketoacyl-(acyl-carrier-protein) reductase, FabG4	gi|15607383|	47	CW	10.64
Hypothetical protein Rv3269	gi|15610405|	10	Mem	10
Possible conserved membrane protein, Rv3869	gi|15611005|	51	CW	9.80
elongation factor Tu, EfTu	gi|15607825|	44	Mem	9.090
Iron-Regulated heparin binding hemagglutinin, HbhA	gi|15607616|	22	Mem	9.090
Possible two component transcriptional regulatory protein LuxR-family, Rv0195	gi|15607336|	23	CFP Cyt	8.69 8.69
acyl carrier protein, AcpP	gi|15609381|	13	Cyt	7.69

Note: Spectral counts were obtained from the Scaffold program and then normalized into NSAF (Normalized Spectral-count Abundance factor) by taking into consideration the protein length of each protein as suggested elsewhere [393]. Only proteins with NSAF higher than 7 in at least one of the four fractions are shown.

5.5. Discussion

5.5.1. Production of rCfp2 in *E.coli*

Production of *Mtb* proteins in *E.coli* is a common strategy to obtain high yield recombinant proteins [346, 442, 477-479]. In most cases, secreted proteins are expressed without the signal peptide in order to obtain the mature protein that is similar to that present in *Mtb* culture filtrates.

In this study, two different forms of Cfp2, one containing the signal peptide (Cfp2-wg) and one without it (Cfp2-nss), were successfully produced in *E.coli.* However, the construct corresponding to the mature protein (nss) showed some level of toxicity to *E.coli* cells, and therefore, the production levels were higher for the Cfp2-wg construct. Recombinant Cfp2 was not found in the culture filtrate of either construct, suggesting that *E.coli* does not have the secretion machinery to export this protein to the extracellular milleu. However, in the lysate obtained from the Cfp2–wg construct, two different bands, matching the molecular weights of mature and precursor Cfp2 were observed (Figure 5-9). These findings are similar to those obtained elsewhere in which the signal peptide of several *Mtb* secreted proteins was cleaved from a β-lactamase fusion protein in *E.coli.* In these cases, the mature protein (based on Mr) was found in the periplasm and lysates but absent from the culture filtrate [481]. Even though the authors did not confirm the N-terminal sequence of the expressed proteins, this suggests that SPase-I of *E.coli* recognizes and cleaves *Mtb* signal peptides but the secreted protein is not released to the extracellular milleu.

E.coli was initially chosen with the sole purpose of producing recombinant Cfp2. However, the observation that two different forms of rCfp2 were present in the lysates led us to further study the N-terminal sequence of rCfp2 produced by *E.coli*. When the N-terminus sequence of the purified rCfp2-wg was determined, it was found to be MQPVVF which corresponds to residues 37 to 42 of the preprotein, which is, 12 aa upstream of the cleavage site observed in *Mtb* (Figure 5-12). This indicates that the cleavage site recognized by *E.coli* secretion system, and probably the SPase I, differs significantly from the cleavage site in *Mtb*. This suggests that SPase I may present a universal promiscuity in which not only additional residues are allowed at positions -1 and -3, but also, inter-specie cleavage can be performed. If this hypothesis is correct it would also indicate that *E.coli*, although able to cleave rCfp2, is unable to export the protein to the extracellular milleu. This in turn would suggest that signal peptide requirements for translocation and/or export differ between *E.coli* and *Mtb* and might be due to inherent differences in the cell membranes of these organisms.

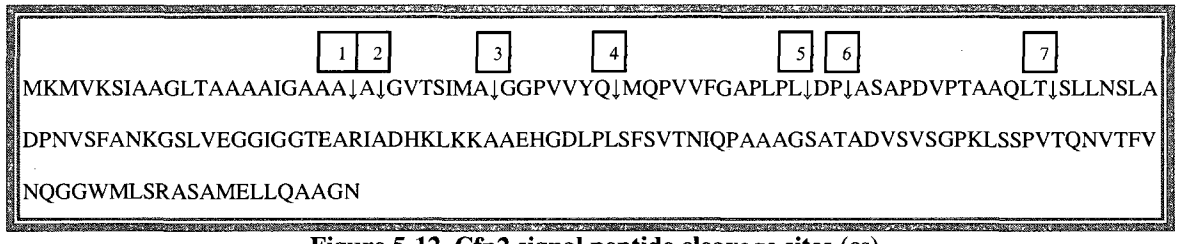


Figure 5-12. Cfp2 signal peptide cleavage sites (cs).
Boxes 1, 2 and 3: predicted cs by SignalP v3.0 determined by neural network and hidden Markov Model algorithms. **Box 4:** Cleavage site or alternative translational start site in *E.coli*. **Box 5:** *Mtb* cleavage site. **Box 6:** Alternative cleavage site in *Mtb* or post-translational modification by other peptidases. **Box 7:** Alternative N-terminal sequence presence in one Cfp2 isoform.

Our findings indicate that when *Mtb* secreted proteins are cloned and expressed in *E.coli*, it is ideal to experimentally confirm the N-terminal sequence of the resultant recombinant

protein. Our results indicate that *E.coli* is not a good model to study secretion of recombinant *Mtb* proteins.

5.5.2. Impact of signal peptide mutations on Cfp2 secretion

Signal peptides have at least three functions in protein transport. First, they are recognized by receptors of the secretion machinery, allowing the protein to be directed to the translocon [455, 482, 483]. Second, signal peptides serve as a determinant of protein orientation in the membrane, resulting in the initiation of protein movement from the C-terminal region while the N-terminal region, containing the signal peptide, remains located in the cytoplasmic side of the membrane [455, 484, 485]. Finally, signal peptides can participate in the inhibition of folding of nascent polypeptide chains in the cytoplasm, therefore, retaining the capability of translocation [455, 486, 487].

Consequently, it is not surprising to find a conserved structure that applies to most signal peptides. In general, signal peptides consist of three distinct regions: N-, H- and C-regions. The N- region corresponds to a positively charged stretch of polar residues at the N-terminus. The hydrophobic core, or H-region consists of approximately 10-15 amino acids and the C-region contains the recognition site for the signal peptidase I [455, 459].

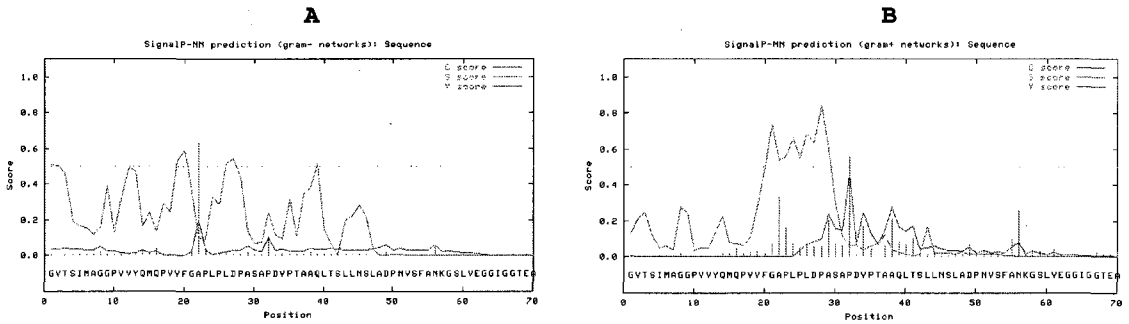
In both G+ and G- organisms the C-region favors small, aliphatic residues; particularly alanine and glycine, at positions -3 and -1 relative to the cleavage site [455]. In *Mtb*, this general rule has also been observed [452], predominantly for position -1, where alanine is the most frequent residue, followed by glycine.

Various bacterial species have more than one SPase-I. For instance, cyanobacteria, possesses two signal peptidases, each of which has been linked to the subcellular

localization of different subsets of proteins (i.e. secreted and tylakoid- associated) [488]. Similarly, *Bacillus sp.* have been shown to contain up to seven SPase-I proteins and their differential use is linked to sporulation and vegetative states in which different subset of proteins are secreted [455, 489, 490]. *Mtb* has only one identified SPase-I (i.e. LepB, Rv2903), suggesting that in *Mtb,* most secreted proteins with identified signal sequences are processed by this peptidase.

Bioinformatic prediction of Cfp2 signal peptide by SignalP v3.0, results in the identification of the first 21-23 amino acids as the putative signal peptide, with a cleavage site of AAA↓AG or AAA↓GV depending on the algorithm used (Figure 5-4 and Figure 5-5). Hidden markov model trained in protein sequences from gram positive bacteria predicts a cleavage site between pos 29 and 30 (IMA↓G) (Figure 5-5). Nevertheless, previous studies [450], as well as findings presented here, showed that Cfp2 mature protein has a N-terminal sequence corresponding to DPASAP which suggests a signal peptide of 48 residues, with a unusual cleavage site harboring Leucine residues at both -3 and -1 positions.

SignalP v3.0 has not been trained on mycobacterial proteins, however, it has been shown to accurately predict signal peptides in *Mtb* proteins [452]. Interestingly, when the Cfp2 sequence lacking the first 22 amino acids, is used again for signal prediction, additional cleavage sites are identified when the neural networks algorithm is used (Figure 5-13).



Cleavage site at pos 21-22 (This corresponds to residues 43-44 of the entire Cfp2 sequence)

Cleavage site at pos 31-32 (This corresponds to residues 53-54 of the entire Cfp2 sequence)

Figure 5-13. SignalP v3.0 prediction of signal peptide in Cfp2 sequence lacking the first 22 amino acids.

A: Prediction performed by neural networks and database trained on G- organisms. **B:** Prediction performed by neural networks and database trained on G+ organisms.

These findings led us to study the C-region of Cfp2 signal peptide by site directed mutagenesis. Two different mutations at positions -3 (L>V and L>S), -1 (L>V and L>S) and +1 (D>E and D>L) were analyzed. Residues at positions -3 and -1 have been shown to be essential for cleavage by Spase-I in other bacterial species [455, 459, 474]. If Spase-I is involved in the processing of Cfp2, it is expected that substitutions made at these positions would decrease the cleavage and secretion of Cfp2 in *Mtb*. If so, it is also expected that more radical substitutions (L>S) in comparison to neutral substitutions (L>V) at positions -3 and -1 will have a greater negative impact on Cfp2 processing, as measured by a reduction of Cfp2 secretion. Our observations indicate that one of the mutations (pos-3 L>S) resulted in decreased secretion of rCfp2 when compared to the wild type. Serine has been reported to be relatively common at this position in other bacteria [455]. This amino acid was chosen to substitute leucine at pos -3 because it presents different physicochemical characteristics and is not frequently found at this position in *Mtb* [452]. When leucine at pos -1 is changed by a valine, the secretion of

rCfp2 seems to increase. Interestingly, after alanine, valine is one of the most common residues found at position -1 in other bacteria.

The negative impact on Cfp2 secretion shown by the substitution at pos-3 (L>S) and the positive effect that mutation at pos-1 (L>V) had on secretion of Cfp2 suggest that SPase-I is involved in the cleavage of Cfp2 signal peptide. However, additional mutations as well as determination of secretion rate over time should be performed in order to confirm this suggestion.

As mentioned above, serine is frequently found at position -3 in other bacteria [455]. This, in addition to the unusual Cfp2 cleavage site (LPL), indicates that if *Mtb* SPase-I is truly involved in Cfp2 cleavage, it might show different requirements for the recognition of cleavage sites in comparison to its homologue in other bacteria.

In addition, our findings suggest that site directed mutations in *Mtb* signal peptides could increase the production of recombinant proteins which could have a positive impact on design and production of recombinant BCG and subunit vaccines.

Mutations at the DP motif (pos+1 and +2), which has been proposed as a possible sorting signal [452], do not seem to have an impact on secretion of rCfp2 suggesting that this motif might not be directly related to processing and secretion of extracellular proteins. However, further studies on other secreted proteins bearing the DP motif, in conjunction to additional mutations, including deletion of the entire motif are required in order to confirm this assumption.

At the beginning of this study, one of the working hypotheses, based on the predictions made by SignalP v.3.0 regarding the cleavage site, was the possibility that Cfp2 was cleaved twice during the secretion process. However, the deletion of three alanines at

positions -27 to – 29, which corresponds to one of the cleavage site predicted by SignalP v3.0, did not affect the secretion of rCfp2, indicating that this site is probably not subjected to cleavage by SPase-I. Unfortunately, several attempts to delete residues -20 to -22 (IMA) which correspond to the cleavage site predicted by the hidden markov model in the gram positive trained dataset (Figure 5-5) were unsuccessful. Therefore, the possibility of an additional cleavage at this site cannot be discarded.

In addition to the role of SPase-I in the cleavage of the Cfp2 signal peptide, other possibilities should be contemplated. For example, processing of pre-pro-proteins involves the sequential cleavage of two different signal peptides required for maturation and secretion of the protein. In general, this type of protein is more common in eukaryotes. In bacteria pre-pro-proteins are mainly restricted to proteases [491], however, a few others have been identified [491]. One of these is a nuclease from *Staphylococcus aureus* called NucA. Interestingly, the protease involved in the processing of pro-Nuc (after the first signal peptide has been cleaved) has been shown to be HtrA [492], a homologue of the periplasmic protease of *E.coli*, DegP. In the same study, HtrA was also shown to be involved in the processing of a native autolysin from *Lactococcus lactis* after exported via a classical signal peptide. The *Mtb* genome contains several genes that encode different types of proteases. Several of these proteases, including HtrA2 (PepD), HtrA1, HtrA4 (PepA) among others are membrane bound and bear homology to DegP from *E.coli.* The precise function of these proteases in the context of cell physiology of the bacilli is unknown. However, since HtrA homologues seem to play a role in the processing of secreted proteins in other bacteria, it may be possible that one of these

proteases is involved in the cleavage of Cfp2 during or after secretion. Future studies to evaluate this possibility are required.

5.5.3. Identification of Cfp2 interacting proteins

Cfp2 is one of the many *Mtb* proteins for which function has not been assigned. In addition to its role as an immunogenic protein [450, 451], only few other characteristics of this protein are known. For instance, Gao and colleagues (2004) [273] determined that cfp2 is overexpressed in *Mtb* H37Rv as compared to H37Ra, suggesting some role in virulence. In addition, they catalogued Cfp2 under functional category 3: Cell wall and cell wall processes.

The gene encoding Cfp2 has also been shown to be one of the genes regulated by PhoP [215, 493], which also regulates several genes involved in fatty acid metabolism, suggesting that Cfp2 could play a role, directly or indirectly, in the biosynthesis of fatty acids [214, 215]. Finally, in addition to being secreted, Cfp2 has been found in the cell membrane [494] and recently, was predicted to be an outer membrane protein (OMP) [83].

Interestingly, identification of Cfp2 interacting proteins by pull-down assays coupled to tandem MS/MS seems to support the possible role of Cfp2 in the biosynthesis of fatty acids, as well as a possible function associated with the cell membrane.

Rv0580, which co-eluted with Cfp2 and was present in the highest levels in the cell wall and also in higher levels in the membrane, is annotated as a hypothetical protein. However, when a homology search is performed using the blastp algorithm [468] against single genus genomes (i.e. *Corynebacterium, Nocardia, Rhodococcus, Streptomyces*),

this protein was found to be similar to GrhN from *Streptomyces sp.* JP95 (E value= 8 e-14). While GrhN does not have a function assigned perse, it is part of a cluster of genes responsible for the production of the aromatic spiroketal polyketide griseorhodinA [495]. With this information, another blast search was performed using Cfp2 sequence against the database corresponding to *Streptomyces sp.* JP95. In this case, none of the proteins reached the standard cutoff value, however, Cfp2 seems to have some homology with the protein GrhP which is annotated as a putative class II glutamine amidotransferase, specifically as a asparagine synthetase, and is also part of the gene cluster responsible of the biosynthesis of polyketide griseorhodinA [495]. Interestingly, GrhN and GrhP are not only part of this cluster, but are found adjacent to each other in the genome of *Streptomyces sp.* JP95, which remarks the relevance of identifying Rv0580c as a putative interacting protein of Cfp2.

Other hypothetical proteins that were found to co-elute with Cfp2 also present similarities to proteins involved in fatty acid biosynthesis. Following the same blast search strategy mentioned above, Rv2298, which co-eluted with Cfp2 in the cell wall fraction, was found not only to have similarity to an aldo-keto reductase of *Corynebacterium* (E value= 1e-21), but also bears the conserved domains associated with aldo-keto reductases. This type of enzymes reduces aldehydes and ketones to primary and secondary alcohols [496] and have been shown to regulate human lipid synthesis [497, 498]. Similarly, Rv1919c was found to present conserved domains associated with the polyketide cyclase 2 superfamily, specifically with cyclase/dehydrase lipid transport proteins.

In addition to the findings associated with these hypothetical proteins, three other proteins known to be involved in fatty acid biosynthesis were also identified as putative

Cfp2 interacting proteins. These correspond to possible fatty acid synthase (Rv3720), 3-ketoacyl-(acyl-carrier-protein) reductase FabG4 (Rv0242) and acyl carrier protein AcpP (Rv2244). All of this supports the hypothesis that Cfp2 could be potentially involved, directly or indirectly in the synthesis of fatty acids.

Despite these consistent findings, other observations are puzzling. For instance, class II glutamine amidotransferase proteins present a cysteine in the N-terminus of the enzyme, which is essential for both, its enzymatic function and acyl binding [499]. However, Cfp2, as well as its possible homologue GrhP lack this residue, which led to Li and colleagues to hypothesize that GrhP might be a non-functional enzyme [495]. Cfp2 is also much shorter than GrhP, which also adds to the idea of Cfp2 being non-functional as well. However, given the number of proteins related to fatty acid synthesis, it is also possible that Cfp2 functions as a heterodimer with another *Mtb* protein(s).

Surprisingly, several ribosomal proteins were also found as putative Cfp2 interacting proteins from the cell membrane and cell wall fractions. At the same time, protein transport proteins SecE and SecE2 were identified as possible Cfp2 interacting proteins in the cell wall and cell membrane respectively. At this point, the reason why these proteins might interact with Cfp2 is not clear, and may be due to an unspecific affinity to Cfp2 by SecE and SecE2 given the secretory nature of this protein. Interestingly, ribosomal proteins associated with the cell membrane [500-502], and more specifically with the SecYEG translocon have been described before [503, 504]. This association has been linked to protein synthesis and also to biogenesis of integral membrane proteins [501, 505]. Therefore, it is possible that ribosomal proteins identified in this study are interacting with SecE and/or SecE2 rather than presenting a direct association with Cfp2.

Finally, other hypothetical proteins identified in the pull-down assays were subjected to blast search analysis. Unfortunately, the hypothetical protein Rv1738, which co-eluted with Cfp2 from cell wall, membrane, and cytosol and presented the second highest levels after Rv0580c, did not present significant similarity to any protein with a known function. It is interesting that this hypothetical protein is similar to Rv2632c, which was also identified as putative Cfp2 interacting protein. Remarkably, Rv2632c has also been shown to be regulated by PhoP [215], suggesting a possible relation to Cfp2.

5.6. Conclusions

Secreted mycobacterial proteins represent one of the most important aspects of the tubercle bacilli physiology due to their potential impact as antigens, virulence determinants, and role in host-cell interactions.

However, the role of many of these proteins in the context of the cell biology of the mycobacteria is still unknown. In addition, the exact mechanisms of protein secretion, particularly the processing of signal peptides, have not been well studied in mycobacteria. In the previous chapter, a higher abundance of the secreted protein Cfp2 was observed in *Mtb* CDC1551 when compared to the S75 group of *Mtb* strains and it was the most pronounced difference between these relatively distant *Mtb* strains. This resulted in an increased interest in understanding the higher expression or secretion of this protein in *Mtb* CDC1551. Surprisingly, *cfp2* gene expression seems to be lower in *Mtb* CDC1551 when compared to the S75 group. Consequently, we decided to focus our attention on the study of the secretion mechanism of Cfp2.

In the present study, alternative methods for the production and purification of recombinant and native Cfp2 were developed. Additional experiments to determine the subcellular localization of rCfp2 in *E.coli* (cytoplasm vs. periplasm) as well as the use of different expression vectors are necessary to further understand the apparent alternative cleavage site of this protein in *E.coli*.

Site directed mutagenesis of the Cfp2 signal peptide cleavage site offered initial insights on the processing of the unusual signal peptide of Cfp2 in conjunction with novel information regarding the sequence requirements of SPase-I in *Mtb*. Decrease secretion of the rCfp2 bearing a serine at pos -3 and increase secretion of the rCfp2 mutant bearing a valine at pos -1 of the signal peptide suggest that SPase-I is involved in the cleavage of Cfp2 during export. Additional mutations, including deletions of the -3 and -1 residues, as well as changes in other residues at both ends of the cleavage site are necessary to fully understand the cleavage process by SPase-I in *Mtb*. Other mechanisms for the processing of Cfp2, such as the potential involvement of some proteases should also be studied.

Finally, protein-protein interaction assays to determine the interacting proteins of Cfp2 and gain insight into its possible role within the cell were performed. As mentioned in the discussion, findings presented here, as well as gene expression studies presented elsewhere [214, 215, 493], suggest the possible function of Cfp2 in the biosynthesis of fatty acids. Although this hypothesis is not conclusive, these findings set the precedent for future studies and remark the importance of continuing research on functional validation of Cfp2 and other secreted *Mtb* proteins.

6. Chapter VI: Final discussion

Mycobacterium tuberculosis is an ancient pathogen that has had a major impact on human population from ancestral times to modern society. As discussed in chapter I, the understanding of the pathology and physiology of this bacillus, as well as TB itself has come a long way since the first descriptions of the illness. However, even though this disease was thought to be close to eradication some decades ago; has reemerged and presents new challenges to both scientists and physicians.

Today, the development of more effective drugs, better vaccines and faster diagnostic methods represents a key step towards the control of this disease.

In this regard, drug, vaccine and molecular diagnostic targets can be affected by strain to strain variation, which is the reason why some authors consider the survey of genetic diversity among clinical isolates very useful [273] in order to obtain information of the extent and types of variability that can occur and to apply this knowledge on the designing of the control tools described above.

As discussed in chapter II, strain to strain variation can be assessed at different levels. For instance, genomic studies can provide insights on the deletion/insertion of target genes, while gene expression studies provide information about genes that are upregulated or down regulated under certain conditions. Gene expression studies at the mRNA level have the advantage that virtually all the genes in the genome can be tested given a specific set of experimental conditions.

However, due to potential differences in mRNA processing, post-translational modifications and protein stability, it is difficult to conclude if the genes that demonstrate differential expression maintain the same pattern at the protein level.

Therefore, measuring variability and diversity of *Mtb* strains at the protein level provides an advantage over genomic and mRNA studies. As described previously, the development of proteomic tools has had a major impact on the study of infectious and non-infectious diseases. In *Mtb*, several proteomic studies have described, to a great extent, the secreted proteins produced by this pathogen [249, 275, 276, 313-318]. However, only few studies have analyzed protein variability between clinical isolates.

In this study, the protein variability and virulence of very closely related clinical strains were evaluated using novel proteomic methods and the guinea pig model of TB respectively. In addition, preliminary characterization studies of a secreted and antigenic protein, Cfp2, which also showed differential levels between strains, were performed.

6.1. Relevance of biological findings and future studies of S75 group clinical isolates

Clinical manifestation of tuberculosis might be influenced by numerous factors unrelated to the pathogen, including the host's genetic background, immune and nutritional status, and social and economic environment [237, 506, 507]. Therefore, it is difficult to directly link differences in the infecting *Mtb* strain to variations in the course and outcome of human tuberculosis.

However, clinical isolates of *Mtb* have been shown to behave differently in various models of virulence, leading to the conclusion that there are convincing reasons to believe that strain-specific attributes contribute directly to virulence and disease outcome of TB [223]. In addition, isolates of *Mtb* cultured from individual patients vary in their ability to survive and grow under different conditions in the laboratory and to cause progressive infection in animals [89, 238].

Supporting this idea, our findings showed the biological diversity that exists between clinical isolates, three of which are very closely related in terms of both genetics and evolution. Even though all four analyzed strains seem very successful in causing lethal disease in the guinea pig model, some of them (*Mtb* C28, and to a less extent *Mtb* H6) produced more rapid death in infected animals. Meanwhile, *Mtb* BE was linked to longer survival times.

Interestingly and puzzling, in humans, *Mtb* BE appears to be the most successful strain of the S75 group to infect and/or transmit and subsequently cause disease [352].

This led us to hypothesize that *Mtb* BE might be associated with a less severe disease in humans, which in turn could be translated in a delayed time to seek medical help and therefore an increased chance to spread the disease.

Delayed medical consultation has a negative impact on TB control because it increases the possibility of transmission to others and therefore the reason of this delay has been a focus of study. Some studies indicate that delay in seeking medical help is a multifactorial issue; however, one of these factors was found to be the type of pulmonary TB [508, 509]. Patients presenting as smear negative TB, which may be linked to milder symptoms, seem to wait longer to visit a health provider. However, it has also been

shown that smear negative patients present a relative transmission rate of only 22% when compared to smear positive patients [510].

In this regard, the hypothesis could be generated from the other end of the spectrum. In this case, *Mtb* C28, which in the initial study describing the S75 group was related to only two TB cases [352], could be associated with a much more aggressive disease which could lead to either rapid death or shorter time to seek medical help and therefore treatment, both resulting in decreased transmission rates. This is supported by the host-pathogen co-evolutionary paradigm in which a successful pathogen should not only be able to colonize and cause disease in one host but must also be capable to infect new hosts [511]. If this is the case the higher infectious rates of *Mtb* BE in humans are only apparent due to the increased pathogencity of *Mtb* C28 leading to a defect of this strain to infect new hosts.

It would be interesting to perform a retrospective epidemiological study to evaluate the clinical characteristics of the patients from where S75 strains were first identified and isolated. Analysis of the different symptoms, as well as time to seek medical help and to receive treatment for each of these patients could provide relevant information to either validate or reject these hypotheses.

On the other hand it can also be argued that the differences in prevalence between these strains in humans are due to intrinsic factors related to the bacilli. As a result, *Mtb* BE would present an inherent advantage, when compared to *Mtb* C28 and *Mtb* H6, to transmit to and colonize new hosts.

Interestingly, in a recent study in which the famous Riley experiment on TB transmission [512, 513] was recreated, guinea pigs were exposed to air from a tuberculosis ward. In

this study, less than 9% of the TB isolates were shown to be responsible for more than 98% of all TB cases in the guinea pigs [514], suggesting a relation between *Mtb* strain and transmission. Unfortunately, direct assessment of natural transmission in a controlled environment represents a challenge for all current animal models of TB.

Fennelly and colleagues presented a novel method to study infectiousness of different *Mtb* strains. In this study, isolation and culture of *Mtb* directly from concentrated cough-generated air droplet nuclei from TB patients was possible [515]. Even though this model does not measure transmission, it could be modified in a way such that guinea pigs could be exposed to these concentrated air droplets from different TB patients. However, this method would not be applicable to current identified strains.

In the meantime, if an ideal animal model to measure differences in transmission between *Mtb* strains is developed, the S75 group would be a good representation, given its epidemiology in humans and biological characteristics in the guinea pigs.

These strains also present a good source to investigate differences in immune responses elicited by different degrees of virulence. As mentioned in chapter III, survival rates could not be correlated to differences in growth between the strains. It would be interesting to determine if despite the similar growth of these strains in the guinea pigs, they induce differential expression of relevant cytokines such as IL-12, IL-10, IFN-γ and TNF-α, which could have an impact not only in the ability of the immune system to contain the disease, but also in the minor differences observed in the lung pathology induced by each strain. In this regard, lung and spleen tissue obtained from guinea pigs infected with each of the strains was collected, and future studies will include the analysis of gene expression of selected cytokines. In addition, proteome analysis of host proteins

is also an interesting possibility, which would not only address the possible differences in the expression of cytokines and chemokines, but would also provide information regarding other proteins that might be directly and/or indirectly related to the immune response and that could play a role in the final outcome of the disease.

6.2. Relevance of proteomic findings on the biology of the S75 group

Several differences in the protein expression profiles of *Mtb* strains belonging to the S75 group, as well as *Mtb* CDC1551 were identified by a combination of classical 2D-GE analysis and the more novel ITRAQ method. Interestingly, this study not only demonstrated that differences in the protein levels between closely related *Mtb* strains occur, but also showed a clear trend on the expression of several protein groups between more virulent (*Mtb* C28 and *Mtb* H6) versus less virulent (*Mtb* BE) isolates.

For instance, *Mtb* C28 and *Mtb* H6 presented higher levels of proteins involved in metabolic pathways as well as proteins involved in iron uptake and storage and sulfur acquisition, suggesting that these *Mtb* strains might be more metabolically active and have an advantage in nutrient acquisition. Similarly, several proteins classified under virulence, detoxification and adaptation showed elevated levels in *Mtb* C28 and *Mtb* H6 as compared to *Mtb* BE.

Taken together, biological and proteomic findings suggest that the proteins expressed differentially in *Mtb* C28 and *Mtb* H6 could be associated with their higher virulent phenotype and are good candidates for functional validation as determinants of virulence.

In addition, several hypothetical proteins were also found to present a clear pattern of expression, being upregulated (30Kd-ag (Rv2744c), Rv1738, TB15.3 (Rv1636), Rv3716c, Rv3075c) or down regulated (Rv2302, Rv3267, Rv3354, Rv1352, Rv1906c, Rv0569) in both *Mtb* C28 and *Mtb* H6 when compared to *Mtb* BE. These proteins could be playing a role in the biological differences observed between these strains and therefore should also be included for functional validation, not only in the context of virulence and pathogenicity, but also in the context of the physiology of *Mtb*.

As discussed in chapter IV, some of the findings regarding differentially expressed proteins between the more virulent strains *Mtb* C28 and *Mtb* H6 and the less virulent strain *Mtb* BE are puzzling and emphasize the complexity of the biology and physiology of *Mtb*. These findings included the lower amounts in strains *Mtb* C28 and *Mtb* H6 of some proteins that have been previously characterized as virulence factors in *Mtb* and other bacteria such as SodA [435], PonA2 [437], GlnA1 [438-440], and several proteases [441-443]. It is possible that the differential expression of some of these proteins could result from intrinsic phenotypic differences between *Mtb* strains during in vitro culture.

Proteomic analysis from *Mtb* cultured 'in vitro' is facilitated by the easy acquisition of high yield protein levels with no contamination of host-proteins. Moreover, 'in vitro' analyses require a less complex infrastructure which is translated in lower costs and faster results. Therefore, they offer a good starting point for the identification of virulence determinants, as well as drug, vaccine and diagnostic targets that can then be validated by different analyses focused on single proteins.

In addition to the intrinsic characteristic of in vitro studies, it is important to highlight that some differences in protein levels could be due to normal phenotypic variation without any particular association with virulence and disease outcome.

Consequently, future studies are essential, not only to corroborate the expression 'in vivo' of the differentially expressed proteins, but also to validate their potential implication in virulence and disease outcome. These studies should be focused to ultimately answer the question: Does the over expression and/or down regulation of these proteins have a significant effect on the biological properties of the different isolates?

A scheme outlining a proposed plan to validate the differential proteins identified in this study and with a clear pattern of expression between more virulent *Mtb* C28 and *Mtb* H6 and less virulent *Mtb* BE is presented in Figure 6-1 and discussed in the following lines.

As mentioned above, one of the most remarkable findings was the identification of higher levels in *Mtb* C28 and *Mtb* H6 of enzymes involved in metabolic pathways and in iron and sulfur uptake. Further validation of this finding should include not only the evaluation of 'in vivo' protein levels but also the determination of enzymatic or protein activity, which in *Mtb* C28 and *Mtb* H6 should be higher than in *Mtb* BE.

For instance, cell-free extracts of *Mtb* C28 and *Mtb* H6, when compared to *Mtb* BE, should present a higher enolase activity as measured by the conversion of phosphoglycerate to phosphoenolpyruvate and NADH to NAD. Similarly, these extracts should present higher activity of 6-Phosphogluconate dehydrogenase (Gnd1), succinyl-CoA Synthetase (SucC) and citrate synthase I (GltA2). In vitro, fluorometric assays to measure these activities have been described previously elsewhere [516-520].

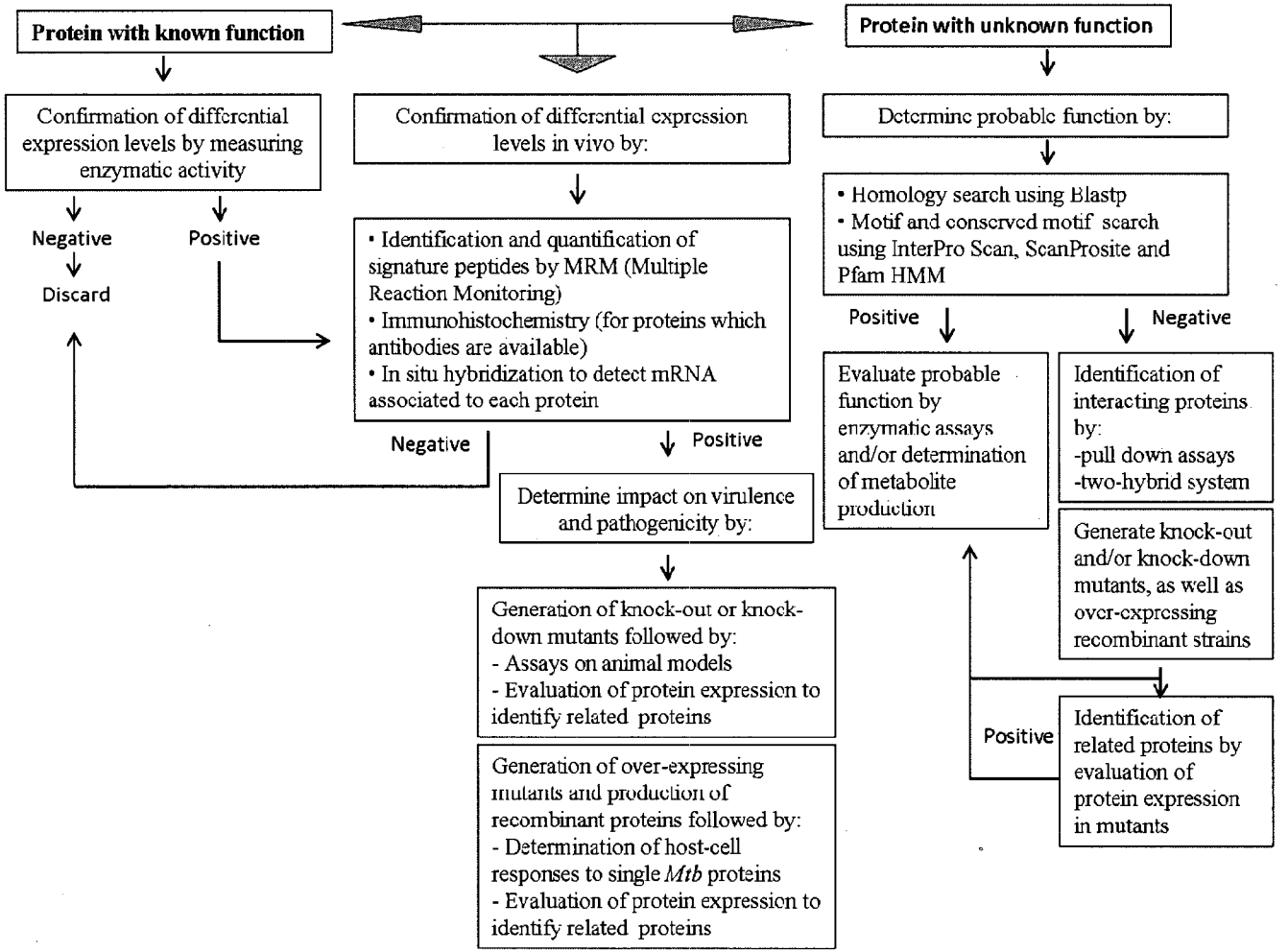


Figure 6-1. Scheme for further validation of differential proteins.

In addition to the inherent enzymatic activity, some of these proteins could be playing an additional role in *Mtb* physiology and therefore it would be interesting to further investigate this possibility. This is based on observations by other authors in which metabolic enzymes, such as malate synthase [521] from *Mtb* and enolase from *Streptococcus pneumonie* [522], function as laminin and plaminogen binding proteins respectively, in addition to the activity related to their glycolytic pathways. Interestingly, these enzymes, which were believed to be strictly intracellular, are also present in the secreted fraction [317, 523], or exposed on the cell surface [516]. Similarly, in this study we reported the presence of Eno, SucC and GltA2 in the secreted fraction of *Mtb*, supporting the idea that these proteins might present additional roles, unrelated to metabolic pathways.

Higher expression of BfrB, ViuB and SseC2 by *Mtb* C28 and *Mtb* H6 can also be confirmed by directly measuring iron and sulfur uptake rates for each strain. Methods to measure iron uptake in *Mycobacterium* have been described previously and involve the use of ^{55}Fe and detection by scintillation counting [524]. In a similar way, sulfur uptake could be assessed by the decrease of free $^{32}SO42-$ in the medium.

Other functions besides iron uptake and storage have been associated with bacteroferritin [525]. This protein might be involved in protection of cells from radical oxygen species (ROS). If this is the case, a higher level of this protein in strains *Mtb* C28 and *Mtb* H6 might be related to a higher resistance to ROS as compared to *Mtb* BE and therefore, determination of resistance to ROS could provide and indirect confirmation of the importance of the difference of BfrB levels between strains.

As shown in Figure 6-1, validation of proteomic findings in an 'in vivo' model is an essential step towards the functional characterization of differential proteins as molecular determinants of virulence. Different approaches can be taken to directly identify and quantify the presence of these proteins during in vivo infection. For instance, identification and absolute quantification of selected *Mtb* proteins isolated directly from infected lung tissue can be facilitated by mass spectrometry methods, such as Multiple Recognition Monitoring (MRM) which is focused on the detection of the most common and yet unique peptides associated with each of the proteins to be tested. The specificity of this method overcomes the difficulty of detecting *Mtb* specific proteins in a complex mixture with high amounts of contaminating host- proteins. In addition, this technique could be applied to protein pools isolated directly from individual granulomas which would facilitate validation and quantification of many *Mtb* proteins.

MRM utilizes triple quadrupole instruments. When functioning in MRM mode, the mass of each specific target is used to select the precursor ion in the first quadrupole (Q1), then, the ion is transmitted and fragmented in the second quadrupole (Q2) by CID. Finally, the resulting fragmented ions are analyzed in the third quadrupole (Q3). The mass of these ions are used for identification, while the precursor fragment transitions are used for quantification purposes [526].

MRM is a rapid and cost-effective method for absolute quantification of proteins and peptides in complex samples, and therefore this technology has been successfully used for the validation of cancer [527] and plasma biomarkers [528-530] and virulence factors of pathogenic bacteria [531].

As mentioned above, validation and quantification of proteins that might be involved in *Mtb* virulence is possible by the MRM approach due to the identification of proteotypic peptides for each of the selected proteins.

Proteotypic peptides are those that after trypsin digestion and mass spectrometry analysis are preferentially observed in a single protein [527]. Prediction software has been developed in order to identify these proteotypic peptides for validation of selected biomarkers [532, 533]. Table 6-1 provides a list of the proteotypic peptides identified from in silico analysis of proteins highlighted in our iTRAQ analysis. In addition to in silico predictions, empirical detection of these peptides is one of the advantages of pre-validation stages by proteomic methods such as the one applied in the present study for the analysis of *Mtb* clinical isolates proteomes.

One of the advantages of MRM over immune-based techniques for protein identification is that up to 50 peptide candidates can be simultaneously detected and quantified in a single experiment [534], which makes the validation step faster and simpler.

However, some of the limitations of this technique are the need for careful standardization of the mass spectrometry methods, including upstream procedures such as liquid chromatography and other purification steps. Therefore, other approaches should also be included for the validation of selected proteins and their relative expression during in vivo infection.

One of these techniques, immunohistochemistry, uses antibodies coupled to fluorescence or chromogen detection systems to directly detect specific proteins in histology slides. This approach has been successfully used to detect certain *Mtb* proteins in tissue from infected guinea pigs [535].

Table 6-1. Signature peptides of proteins with differential expression between C28 and H6 versus BE.

Gene	Number of Peptides	Functional category	Proteotypic peptide	Mascot ion score	Mascot identity score	X! tandem	Observed m/z	Actual peptide mass (AMU)
DnaK	7	0	LLGSFELTGIPPAPR	50.12	28.52	2.60	571.34	1710.99
Cfp29	1	0	EIPDVISQALSELR	39.28	28.74	3.68	857.48	1712.94
HspX	4	0	AELPGVDPDkDVDIMVR	49.55	29.06	5.70	1079.08	2156.15
Rv2493	2	0	TTLDLDDDVIAAAR	58.62	28.52	5.52	816.93	1631.85
GroES	8	0	ILVQANEAETTTASGLVIPDTAK	59.77	29.83	5.80	877.83	2630.46
CspA	2	0	WFNAEK	31.91	27.35	0.38	541.80	1081.59
PpiA	4	2	HTIFGEVIDAESQR	53.55	28.75	6.24	873.45	1744.89
Ssb	2	2	EAAENVAESLTR	69.63	28.39	4.19	717.38	1432.74
RplL	6	2	LEAAGATVTVK	71.30	28.48	5.00	674.41	1346.80
SecE2	4	3	VIDIIGTSPTSWEQAAAEAVQR	81.75	29.97	8.57	1243.66	2485.31
Rv0431	2	3	TLGAAVELR	61.74	28.77	3.00	537.33	1072.64
Rv0559c	2	3	GTTQGQAFQFLGAAIDHYcPEHVGVLQR	56.15	30.37	4.52	1078.54	3232.59
PstS1	8	3	GLGEAQLGNSSGNFLLPDAQSIQAAAAGFASK	43.95	30.77	4.51	1127.26	3378.74
Rv2345	3	3	TSELGNYDALLAVATTGR	108.88	29.10	8.42	998.53	1995.05
Cfp2	1	3	GSLVEGGIGGTEAR	79.85	28.47	2.48	723.89	1445.76
LppZ	5	3	TVIPVDPAGDGGLMDIVLSPTYSQDR	38.30	30.37	2.19	954.16	2859.45
Rv3312A	1	3	DYPGPILEGPVLDDPGAAPPPPAAGGGA	49.50	30.16	9.77	1356.69	2711.37
Rv3587c	4	3	LWSNLDCAPSNETLVK	50.28	29.25	4.54	1062.54	2123.07
PonA2	2	3	VPSVAGLDVDAAR	97.43	28.61	5.51	707.40	1412.78
SseC2	4	7	LLDSSDEFTAEWASATGDFR	46.53	29.68	3.19	792.07	2373.17
SerC	2	7	SQVVGTIDFVDDVDAGTVAK	40.01	29.38	1.43	775.41	2323.22
GltA2	9	7	ELEEAALTDDYFIER	65.53	29.02	3.89	979.49	1956.96

SucC	1	7	ITLDGNADFR	51.24	27.97	3.96	633.34	1264.66
PepD	3	7	VDDRPINSADALVAAVR	74.22	28.89	4.09	642.70	1925.07
Eno	1	7	AVQAVLDEIGPAVIGLNADDQR	39.47	29.41	3.34	803.43	2407.28
Gnd1	2	7	SALDLGVPVTGIAEAVFAR	54.75	29.31	1.92	677.39	2029.14
PrcB	2	7	VAVEALYDAADDDSATGGPDLVR	79.94	29.68	6.54	1232.60	2463.19
DlaT	1	7	AFSVQMPALGESVTEGTVTR	70.84	29.48	ND	747.39	2239.14
GlnA1	10	7	SVFDDGLAFDGSSIR	92.79	29.10	9.70	864.44	1726.86
PepN	4	7	FVDPVDGETYLYSQFETADAK	38.96	30.09	2.92	895.10	2682.29
ViuB	1	7	VLGGSGFDTFVPSDFTDSYIK	48.68	29.81	1.64	880.46	2638.37
FixA	5	7	AVEEALQIR	40.53	27.92	1.42	586.85	1171.68
FdxD	1	7	TDPIPVDQEDLAEQAIAECPR	74.82	29.89	3.64	834.07	2499.20
BfrB	3	7	EALALALDQER	63.07	28.60	3.92	686.89	1371.75
Rv0020c	2	9	VPGYAPQGGGYAEPAGR	86.36	28.71	5.21	895.95	1789.89
Rv0569	3	10	GLIIEVR	42.06	27.87	2.20	472.31	942.60
Rv1352	1	10	TEGPSNPLILVFGR	46.60	28.84	4.26	822.46	1642.90
TB15.3	3	10	LIIASAYLPQHEDAR	56.84	29.00	2.39	614.34	1840.01
Rv1738	2	10	NVPEIGDELSVAR	81.88	28.47	4.82	771.92	1541.82
Rv1906c	1	10	MGNPDGALIDNALSK	42.81	28.74	4.85	902.48	1802.94
Rv2302	3	10	VGDYLVVK	33.52	27.36	1.43	590.87	1179.72
30kd_ag	2	10	QALTLADQATAAGDAAK	37.18	28.87	ND	635.35	1903.03
Rv3075c	1	10	RLPNVPIVALVETAR	36.76	28.74	ND	598.04	1791.11
Rv3267	2	10	ELGGLPVVADASIAPGSVR	64.83	29.28	4.28	976.55	1951.08
Rv3354	1	10	ALGQSVcPILAEPGGSFNTAVASVVAR	64.17	30.25	10.06	1402.73	2803.45
Rv3716c	3	10	VVDPDDIETLQDLIVGAMR	80.32	29.54	4.19	748.40	2242.17

Note: In general, proteins identified with more than 4 peptides have more than one proteotypic peptide. In these cases, the peptide with the higher scores is shown.

In addition, an automated quantitative analysis (AQUA) system has been developed to facilitate the relative quantification of biomarker proteins in tissue sections [536]. The evident limitations of this method are the requirement of monoclonal antibodies, which are costly and time consuming to obtain, and the low sensitivity to detect scarce proteins. However, this technique is a good option to validate in vivo protein levels of proteins with high expression and with availability of monoclonal antibodies for their detection.

Finally, another technique that can be used to indirectly measure protein expression during in vivo infection is in situ hybridization of mRNA. In this approach, histological slides are subjected to hybridization procedures using probes specific to each targeted gene. However, even though this technique has been used successfully to detect mRNA of *Mtb* genes in human tuberculous granulomas [537], the correlation between mRNA and final protein levels is always questionable.

After validation of in vivo expression of selected proteins with a putative association to virulence, proteins that are confirmed as presenting differential levels between strains can now be subjected to functional validation to assess their direct impact on virulence and pathogenicity. In addition, proteins with unknown function should also be a focus of study to determine their potential role in the mycobacterial physiology and possible linked to virulence.

In this regard, the most informative approach would be the generation of single gene knockouts to assess the biological impact of each of the proteins encoded by these genes. Most of the differential proteins identified in the present study are not essential for survival [98, 271] and therefore the generation and testing of knockout mutants in animal models of infection is be possible.

Nevertheless, it is important to note that this approach, although very valuable, is minimalistic, and our findings discussed in chapter IV suggest that the overall phenotypic behavior of each strain might be due to a combination of differential expression levels of certain proteins, rather than a complete absence of specific proteins.

Therefore, the evaluation of other approaches, such as, antisense RNA and conditional knock-down mutants using inducible repressor promoters may be worthwhile, as it allows the partial inhibition of gene expression and, in theory, could be used to test the effect that differences in expression levels of proteins have in terms of virulence in the animal model. It is important to note, however, that in mycobacteria, these technologies have not been used extensively, and have only been used in vitro [538-541]. The development of this approach for in vivo experiments might be very complex, requiring molecular biology procedures that are not currently available.

Both knock-out and knock down mutants can also provide important information for proteins with unknown function. In this case, the effect of conditional inhibition of protein expression, or the absolute absence of a specific protein could be assessed by proteomic and metabolomic approaches to identify potential related proteins, as well as production of specific metabolites which could offer insights into the potential role of each studied protein.

Several hypothetical proteins presented differential levels between *Mtb* C28 and *Mtb* H6 as compared to *Mtb* BE. As shown on Figure 6-1, one of the strategies to determine the possible function of these proteins is the bioinformatic search for protein homologues and conserved domains and motifs. A preliminary exploration using a systematic approach, in which specific proteins are subjected to blastp search against individual databases from

genus related to *Mycobacterium*, was done (Table 6-2). The exclusion of *Mycobacterium* databases from the search decreases the likelihood of finding significant results to only other hypothetical proteins present in other *Mycobacterium* species. Similarly, the exclusion of databases from other distantly related genus is performed to reduce the complexity of the search and increase the probability of finding significant matches to other proteins that have been already characterized.

It is important to state that even though, this strategy might increase the chance of finding positive results leading to hypothesis generation, it is also possible that the exclusion from the search of other more well characterized microorganism such as *E. coli, Staphylococcus sp, Streptococcus sp.* and *Bacillus sp.* between others, might result in negative results. In this case, additional searches, including these and other important species should be done.

Table 6-2. Homology search using the blastp program of differential proteins classified under conserved hypotheticals.

Protein	Identified conserved domain or family	E-value	Protein homologues	Organism	E-value
TB15.3 (Rv1636)	Universal stress protein, Usp-like (UspA)		Possible universal stress protein, UspA	*Corynebacterium glucoranolyticum*	8e-33
	CD00293	7e-13	Stress-inducible protein	*Streptomyces avermitilis* MA-4680	5e-07
	PFAM00582	4e-18	UspA domain protein	*Rhodococcus erythropolis* SK121	2e-46
30Kd-ag (Rv2744)	Phage shock protein A, PspA (IM30)		Possible phage shock protein	*Rhodococcus jostii* RHA1	1e-108
	PFAM04012	6e-27	PspA/IM30 family protein	*Corynebacterium glutamicum*	3e-105
	COG1842	2e-21	Phage shock protein A (IM30)	*Streptomyces pristinaespiralis*	0.002
Rv2302	None	--	DNA-binding protein	*Streptomyces coelicolor* A3(2)	1e-08
Rv0569	None	--	DNA-binding protein	*Streptomyxes coelicolor* A3(2)	5e-07
Rv3075	Citrate lyase beta subunit, COG2301	2e-55	Putative citrate	*Nocardia farcinica* IFM 10152	3e-15

	HpcH/Hpal aldolase/citrate lyase family, PFAM03328	1e-23	Lyase beta subunit Citrate lyase	*Rhodococcus jostii* RHA1	6e-70
			Citrate lyase beta chain	*Streptomyces ghanaensis*	2e-20
			Citrate	*Corynebacterium amycolatum* SK46	2e-14
Rv3267	LytR, Transcriptional regulator		Transcriptional regulator, LytR family	*Corynebacterium amycolatum* SK46	5e-120
	COG1316		Putative	*Nocardia farcinica*	1e-138
	PFAM03816	7e-51	transcriptional	IFM 10152	
	PRK09379 TIGR00350	2e-44	regulator	*Rhodococcus erythropolis*	4e-129
		3e-22	Transcriptional		
		3e-28	regulator, LytR family	*Streptomyces sp.*	3e-65
			CpsA	Mg1	
				Streptomyces	3e-38
			Transcriptional regulator	*sviceus*	

Note: For a list of selected hypothetical proteins, refer to Figure 4-10

Six of the eleven hypothetical proteins with differential expression between *Mtb* BE and both *Mtb* C28 and *Mtb* H6 (Figure 4-10) presented some homology to other proteins with assigned function. In addition, four of them presented significant homology to conserved domains or motifs and therefore are targets for further functional validation.

Interestingly, TB15.3 and 30Kd-ag presented significant homology to stress inducible proteins. TB15.3, which in *Mtb* is catalogued as an iron regulated protein, bears homology to conserved domains from the Universal stress protein A. UspA proteins have been characterized in *E. coli* [542-544] and *Salmonella sp* [545] and are shown to respond to several types of stress conditions. In addition, UspA of *Salmonella typhimuriun* has been associated with virulence in mice [545]. Similarly, 30Kda-ag presents homology to conserved domains from PspA which is part of the Phage shock protein (Psp) system, associated to responses against extracytoplasmic stress [546, 547]. In addition, this system and PspA itself have been associated to virulence in *Yersinia enterocolitica* [548] and *Salmonella typhimuriun* [549] respectively, and therefore it has

been proposed to play a role in other pathogens [546, 548]. Moreover, it was demonstrated recently that PspA has a direct role on maintaining the membrane potential through phospholipid binding [550]. Psp system functions as a carefully regulated operon [548], and therefore, it is possible that the 30Kd-antigen is a remnant from a ancient complex of genes and as a result it might not be functional in *Mtb*. However, given the direct role of PspA in *Salmonella* virulence and membrane function, it would be interesting to determine if this protein is playing a similar role in *Mtb*, including a potential function in response to other stress conditions.

Due to the putative role of TB15.3 and the 30Kd-ag as stress response proteins, we proposed their categorization under Functional category 0 (Virulence, detoxification and adaptation). Interestingly, these two proteins add to the general pattern of over expression of this type of proteins by the more virulent strains *Mtb* C28 and *Mtb* H6, as compared to BE (Figure 4-10).

Similarly, our findings regarding the hypothetical protein Rv3075 as a possible citrate lyase, support our observation that enzymes involved in metabolic pathways seem to present higher levels in C28 and H6, as compared to BE. Citrate lyases are involved in the conversion of citrate into Acetyl-CoA and oxoloacetate. In general they are believed to play a role in lipid biosynthesis (i.e. generating Acetyl-CoA) [551, 552]. In *Mtb* there is another citrate lyase (CitE), which also shows homology to Rv3075. Based on structural data, it has been proposed that this enzyme produces free acetyl-CoA instead of protein-bound acetyl-CoA [551]. The authors also proposed that this resulting product could be directly used in the first step of fatty acid biosynthesis.

Remarkably, if Rv3075 functions similar to its homologue CitE, it could explain in part why, DlaT, the enzyme responsible for converting pyruvate to acetyl-CoA in the first step of the TCA cycle, is less abundant in *Mtb* C28 and *Mtb* H6.

As mentioned in the discussion on chapter IV, we hypothesized that strains H6 and C28 may rely on obtaining acetyl CoA through the oxidation of fatty acids; however, it could be possible that the higher levels of Rv3075 in these strains could be the source of Acetyl-CoA in the first step of the TCA cycle, compensating then, for the lower levels of DlaT. Therefore, this protein is also a good target for functional validation, and further studies are needed to confirm not only its role within the *Mtb* physiology, but also the hypotheses proposed here as a result of our observations on global protein expression.

Finally, Rv3267 showed significant homology to conserved domains from the LytR/CpsA/Psr family. Proteins from this family seem to play a role in maintaining the proper cell wall structure [553] and are required for the normal septum formation in *Streptococcus pneumoniae* [554]. It is not clear why this protein presents differential levels in the S75 strains; however, it fits the pattern of expression of proteins classified under cell wall and cell processes which are in general down regulated in *Mtb* C28 and *Mtb* H6.

Continuing with the proposed design to understand the role of the identified differential proteins, the next step is the functional characterization of proteins for which the homology search retrieved useful information. In this case, purification of native or recombinant proteins is necessary to test for specific enzymatic activities or other protein functions. An example of the literature search that is needed in order to identify possible

functions in the context of the cell biology of the organisms was presented above for the some of the hypothetical proteins that presented homology to other known proteins.

If the bioinformatic search does not produce results that generate hypotheses regarding possible function for these proteins, such as in the case of the remaining hypothetical proteins listed on Figure 4-10, it is necessary to consider other experimental procedures, including the generation of knock-out or knock-down mutants and their downstream analysis and characterization in relation to the wild type strain. Other strategies, such as protein-protein interaction assays, should also be considered.

Finally, functional validation of both set of proteins (i.e. with and without known function) should also include the identification of possible host-pathogen interactions, in which the production of *Mtb* native or recombinant proteins and their subsequent exposure to human or murine macrophages would provide information regarding the possible role of these proteins in the context of the infection and connection to the immune response.

6.3. Relevance of Cfp2 preliminary studies on the functional characterization of this protein

The proteomic findings presented in chapter IV regarding the substantial presence of Cfp2 in *Mtb* CDC1551 identified by 2D-GE analysis and corroborated by the iTRAQ approach represented the most remarkable difference between strains from different phylogenetic lineages resulting in an interest for this immunogenic protein. Differences in Cfp2 between *Mtb* strains were specially highlighted in the secreted fraction. Therefore, as a first step to understand the reason for this difference, preliminary studies to

characterize the secretion mechanisms and possible function of this protein were initiated. Analysis of the impact of single mutations in the Cfp2 signal peptide cleavage site on protein secretion, as well as protein-protein interaction assays using a proteomic platform were performed.

Interestingly, despite the general rule of the conservation of -3 and -1 sites (i.e. AXA) in the cleavage site of SPase-I processed signal peptides [456-458], most of the mutations did not have any effect on the secretion of recombinant Cfp2. Nevertheless, one of the mutations (L>S at pos-3) resulted in a marked decrease of rCfp2 secretion. This suggests that *Mtb* SPase-I, requires some conservation at the cleavage site, but is more flexible in recognizing this region than its homologues in other bacteria. However, a more comprehensive analysis, including a much greater array of mutations should be done to confirm or reject this hypothesis, as well as to increase our understanding of signal peptide processing in mycobacteria.

One of the issues when studying signal peptidases is that, with some exceptions, they are usually present as a single copy in the genome and consequently they play an essential role in the cellular biology of the bacteria. This is the case of SPase I and II in *Mtb* and therefore, generation of knockout strains for these proteins is not a possibility. However, it would be interesting to study both of these SPases in the context of conditional knock-downs, using some of the approaches discussed in the previous section. This would permit the proteome analysis of the secreted proteins in *Mycobacteria spp* bearing reduced expression of Spases, and therefore, the identification of the subset of proteins processed by these peptidases.

Secretion of Cfp2 was affected by the change of Leucine by Serine at pos-3, suggesting that SPase-I is involved in the cleavage of this protein. However, deletion of the three alanines predicted by SignalP v3.0 as being the cleavage site of Cfp2 did not have an impact on rCfp2 secretion, indicating that this predicted site is probably not recognized by SPase-I, thus ruling out the possibility of two cleavage events by this peptidase.

Given the Cfp2 unique signal peptide, it could be possible that other peptidases, with different requirements, are involved in the processing of Cfp2, prior to the cleavage by SPase-I.

The possible role of membrane bound proteases, such as HtrA, in the processing of Cfp2 was discussed in Chapter V. In addition, it would be interesting to study the possible role of other *Mtb* proteins in the processing of Cfp2 and other secreted proteins, including lipoproteins. For instance, Rv2869c which is annotated as a membrane bound metalloprotease, presents significant homology (E-value: 5e-10) to Eep (enhanced expression of pheromone) from *Streptococcus uberis*, which has been implicated in alternative lipoprotein processing in this bacterium [555].

Similarly, Rv3886c which is annotated as a probable alanine and proline rich membrane-anchored mycosin, MycP2 (Serine protease) presents significant homology towards the conserved domain of the protein NisP from *Lactoccocus lactis* which is involved in the cleavage of antimicrobial peptides [556, 557]. Even though there are no reports regarding a possible function of this protein in the processing of non-bacteriocin related proteins, it would be interesting to determine its function within the physiology of *Mtb*.

Finally, a pull-down approach was performed to identify possible Cfp2 interacting proteins. Several Cfp2-coeluting proteins were identified by mass spectrometry with at

least 95% confidence at both peptide and protein level. Interestingly, several of these proteins are related to fatty acid biosynthesis and therefore we hypothesize that Cfp2 might play a role, directly or indirectly, in the biosynthesis of fatty acids.

One of the limitations of this preliminary study is the previously reported identification of false positives inherent to the pull-down approach [558] which could lead to misleading results. In this regard, it is expected that some of the identified proteins are not authentic Cfp2 interacting proteins, even though they are absent from the negative controls. However, relative quantification of co-eluting proteins by spectral counting provides important information that can be used to select proteins that, based on their high amount in the eluates, represent the most likely genuine Cfp2 interacting proteins. In addition, it is important to note that other proteins, such as HspX, due to its inherent nature as a chaperone [331] might be interacting with Cfp2 but this interaction may not be directly associated to Cfp2 function. Similarly, the large number of identified ribosomal proteins, as discussed in chapter V, might be due to interactions with other proteins such as those involved in protein secretion. Regardless, it is important to maintain open the possibility that these proteins could be directly interacting with Cfp2 for reasons that remain to be understood.

All of this highlights the importance of validating these preliminary results by other techniques with higher specificity, such as the two-hybrid system. In addition, pull-down assays using native Cfp2, which might present important PTM involved in protein-protein interactions, could provide additional and valuable information that might have been missed from the studies presented here due to the use of recombinant Cfp2.

In the meantime, functional analysis of the possible role of Cfp2 in the biosynthesis of fatty acids could be facilitated by the generation of a Cfp2 knock-out mutant and comparison of lipid profiles with a wild type strain as well as microscopy studies to evaluate if the lack of Cfp2 has any impact on the morphology of the bacilli.

6.4. Final conclusions

Current clinical isolates of *Mycobacterium tuberculosis* associated with clinical and epidemiological differences are an important source of study to identify possible molecular determinants that could add to our understanding of the involvement of the bacillus in disease presentation, disease outcome and transmission. This in turn could translate to interventions with a positive impact on the control of the disease.

In this study we presented a comprehensive proteomic analysis of closely related clinical isolates BE, C28 and H6 and demonstrated a significant variation in protein levels between this isolates. In addition, the guinea pig model of TB was used to study the biological behavior of these isolates, showing that these strains also present different degrees of virulence. Based on these results we hypothesize that these biological differences might be related to the variability observed at the protein level. Moreover, we presented several proteins that might represent good targets for further validation as virulence determinants.

In addition, preliminary studies on the low molecular weight antigen, Cfp2, provided evidence regarding the importance of the cleavage site in the processing of secreted proteins in *Mtb* and the possible function of this protein in the biosynthetic pathway of fatty acids.

In summary, studies presented in this document demonstrate that, although unexpected, protein variability between closely related strains of *Mtb*, not only exists, but might also be associated to differences in their biology and disease presentation. These studies also generated several hypotheses that can form the basis for further research aimed at increasing our understanding on the phenotypic variability observed between TB patients. Finally, standardization of the shotgun proteomic technique, iTRAQ, which incorporated systematic data and statistical analyses, provides a useful tool for the study of *Mtb* proteomes, which may be applied for the identification of discreet differences between samples (i.e. clinical isolates, mutants, environmental conditions).

7. Literature cited

1. WHO (2005) *Global Tuberculosis Control:Surveillance, Planning,Financing.* **Report WHO/HTM/TB/2005.**

2. WHO (2009) *Global tuberculosis control - epidemiology, strategy, financing WHO Report.* **WHO/HTM/TB/2009.411**.

3. Ahmad S, A.H., Thole JER, Young DB, Mustafa AS, *Identification of a novel protein antigen encoded by a Mycobacterium tuberculosis-Specific RD1 Region Gene.* Scand J Immunol, 1999. **49**: p. 515-522.

4. Dye, C., *Global epidemiology of tuberculosis.* Lancet, 2006. **367**: p. 938-940.

5. Arachi, A., *The global tuberculosis situation and the new control strategy of the World Health Organization.* Tubercle, 1991. **72**: p. 1-6.

6. Dye, C., et al., *Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project.* JAMA, 1999. **282**(7): p. 677-86.

7. Kim, J.S., A; Castro, A; Vande, C; Farmer, P (2009) *Global Public Goods for Health - A reading companion, Session 3. Tuberculosis control.*

8. Corbett, E.L., et al., *The growing burden of tuberculosis: global trends and interactions with the HIV epidemic.* Arch Intern Med, 2003. **163**(9): p. 1009-21.

9. Murray, J.F., *Cursed duet: HIV infection and tuberculosis.* Respiration, 1990. **57**(3): p. 210-20.

10. Daniel, T.M., *The history of tuberculosis.* Respir Med, 2006. **100**(11): p. 1862-70.

11. Hayman, J., *Mycobacterium ulcerans: an infection from Jurassic time?* Lancet, 1984. **2**(8410): p. 1015-6.

12. Gutierrez, M.C., et al., *Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis.* PLoS Pathog, 2005. **1**(1): p. e5.

13. Sreevatsan, S., et al., *Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination.* Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9869-74.

14. Morse, D., D.R. Brothwell, and P.J. Ucko, *Tuberculosis in Ancient Egypt.* Am Rev Respir Dis, 1964. **90**: p. 524-41.

15. Nerlich, A.G., et al., *Molecular evidence for tuberculosis in an ancient Egyptian mummy.* Lancet, 1997. **350**(9088): p. 1404.

16. Salo, W.L., et al., *Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy.* Proc Natl Acad Sci U S A, 1994. **91**(6): p. 2091-4.

17. Filliol, I., et al., *Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set.* J Bacteriol, 2006. **188**(2): p. 759-72.

18. Meachen, G., *A short history of tuberculosis,* 1936, London: Staples Press Limited.

19. Coar, T.B., AB, *The aphorisms of Hippocrates with a Translation into Latin, and English*, ed. G. Editions. 1982.

20. Herzog, H., *History of tuberculosis.* Respiration, 1998. **65**(1): p. 5-15.

21. Daniel, T.M., *The impact of tuberculosis on civilization.* Infect Dis Clin North Am, 2004. **18**(1): p. 157-65.

22. Grigg, E.R., *The arcana of tuberculosis with a brief epidemiologic history of the disease in the U.S.A.* Am Rev Tuberc, 1958. **78**(2): p. 151-72 contd.

23. von Pirquet, C., *Frequency of tuberculosis in childhood.* J Am Med Assoc, 1909. **52**: p. 675-678.

24. Grigg, E.R., *The arcana of tuberculosis with a brief epidemiologic history of the disease in the U.S.A. IV.* Am Rev Tuberc, 1958. **78**(4): p. 583-603.

25. Davies, R.P., et al., *Historical declines in tuberculosis in England and Wales: improving social conditions or natural selection?* Int J Tuberc Lung Dis, 1999. **3**(12): p. 1051-4.

26. Daniel, T.M., *Robert Koch and the pathogenesis of tuberculosis.* Int J Tuberc Lung Dis, 2005. **9**(11): p. 1181-2.

27. Brewer, T.F., *Preventing tuberculosis with bacillus Calmette-Guerin vaccine: a meta-analysis of the literature.* Clin Infect Dis, 2000. **31 Suppl 3**: p. S64-7.

28. Colditz, G.A., et al., *Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature.* JAMA, 1994. **271**(9): p. 698-702.

29. Ritz, N., et al., *Influence of BCG vaccine strain on the immune response and protection against tuberculosis.* FEMS Microbiol Rev, 2008. **32**(5): p. 821-41.

30. Zignol, M., et al., *Global incidence of multidrug-resistant tuberculosis.* J Infect Dis, 2006. **194**(4): p. 479-85.

31. Blower, S.M. and T. Chou, *Modeling the emergence of the 'hot zones': tuberculosis and the amplification dynamics of drug resistance.* Nat Med, 2004. **10**(10): p. 1111-6.

32. Dukes Hamilton, C., et al., *Extensively Drug-Resistant Tuberculosis: Are We Learning from History or Repeating It?* Clinical Infectious Diseases, 2007. **45**(3): p. 338-342.

33. Lehmann, K.N., R. 1896. , *Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik.*, ed. K. Lehmann. 1896, Munich.

34. Shinnick, T.M. and R.C. Good, *Mycobacterial taxonomy.* Eur J Clin Microbiol Infect Dis, 1994. **13**(11): p. 884-901.

35. Griffith, D.E., et al., *An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases.* Am J Respir Crit Care Med, 2007. **175**(4): p. 367-416.

36. Levy-Frebault, V.V. and F. Portaels, *Proposed minimal standards for the genus Mycobacterium and for description of new slowly growing Mycobacterium species.* Int J Syst Bacteriol, 1992. **42**(2): p. 315-23.

37. Jarzembowski, J.A. and M.B. Young, *Nontuberculous mycobacterial infections.* Arch Pathol Lab Med, 2008. **132**(8): p. 1333-41.

38. Cole, S.T., N. Honore, and K. Eiglmeier, *Preliminary analysis of the genome sequence of Mycobacterium leprae.* Lepr Rev, 2000. **71 Suppl**: p. S162-4; discussion S164-7.

39. Marri, P.R., J.P. Bannantine, and G.B. Golding, *Comparative genomics of metabolic pathways in Mycobacterium species: gene duplication, gene decay and lateral gene transfer.* FEMS Microbiol Rev, 2006. **30**(6): p. 906-25.

40. Havlir, D.V., *Mycobacterium avium complex: advances in therapy.* Eur J Clin Microbiol Infect Dis, 1994. **13**(11): p. 915-24.

41. Swords, W.E., et al., *Mycobacterium xenopi multiplies within human macrophages and enhances HIV replication in vitro.* Microb Pathog, 2006. **40**(2): p. 41-7.

42. Griffith, D.E., *Therapy of nontuberculous mycobacterial disease.* Curr Opin Infect Dis, 2007. **20**(2): p. 198-203.

43. Woods, G.L. and J.A. Washington, 2nd, *Mycobacteria other than Mycobacterium tuberculosis: review of microbiologic and clinical aspects.* Rev Infect Dis, 1987. **9**(2): p. 275-94.

44. Huard, R.C., et al., *The Mycobacterium tuberculosis complex-restricted gene cfp32 encodes an expressed protein that is detectable in tuberculosis patients and is positively correlated with pulmonary interleukin-10.* Infect Immun, 2003. **71**(12): p. 6871-83.

45. van Soolingen, D., et al., *A novel pathogenic taxon of the Mycobacterium tuberculosis complex, Canetti: characterization of an exceptional isolate from Africa.* Int J Syst Bacteriol, 1997. **47**(4): p. 1236-45.

46. Cousins, D.V., et al., *DNA fingerprinting of Australian isolates of Mycobacterium avium subsp paratuberculosis using IS900 RFLP.* Aust Vet J, 2000. **78**(3): p. 184-90.

47. Aranaz, A., et al., *Elevation of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to species rank as Mycobacterium caprae comb. nov., sp. nov.* Int J Syst Evol Microbiol, 2003. **53**(Pt 6): p. 1785-9.

48. van Soolingen, D., et al., *Diagnosis of Mycobacterium microti infections among humans by using novel genetic markers.* J Clin Microbiol, 1998. **36**(7): p. 1840-5.

49. Fabre, M., et al., *High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a large collection of "Mycobacterium canettii" strains indicates that the M. tuberculosis complex is a recently emerged clone of "M. canettii".* J Clin Microbiol, 2004. **42**(7): p. 3248-55.

50. Goh, K.S., et al., *Study of the gyrB gene polymorphism as a tool to differentiate among Mycobacterium tuberculosis complex subspecies further underlines the older evolutionary age of 'Mycobacterium canettii'.* Mol Cell Probes, 2006. **20**(3-4): p. 182-90.

51. Montali, R.J., S.K. Mikota, and L.I. Cheng, *Mycobacterium tuberculosis in zoo and wildlife species.* Rev Sci Tech, 2001. **20**(1): p. 291-303.

52. Rahim, Z., et al., *Assessment of population structure and major circulating phylogeographical clades of Mycobacterium tuberculosis complex in Bangladesh suggests a high prevalence of a specific subclade of ancient M. tuberculosis genotypes.* J Clin Microbiol, 2007. **45**(11): p. 3791-4.

53. Kubica, T., S. Rusch-Gerdes, and S. Niemann, *Mycobacterium bovis subsp. caprae caused one-third of human M. bovis-associated tuberculosis cases reported in Germany between 1999 and 2001.* J Clin Microbiol, 2003. **41**(7): p. 3070-7.

54. Prodinger, W.M., et al., *Infection of red deer, cattle, and humans with Mycobacterium bovis subsp. caprae in western Austria.* J Clin Microbiol, 2002. **40**(6): p. 2270-2.

55. de Kantor, I.N., et al., *Human Mycobacterium bovis infection in ten Latin American countries.* Tuberculosis, 2008. **88**(4): p. 358-365.

56. Hlavsa, M.C., et al., *Human tuberculosis due to Mycobacterium bovis in the United States, 1995-2005.* Clin Infect Dis, 2008. **47**(2): p. 168-75.

57. Brennan, P.J. and H. Nikaido, *The envelope of mycobacteria.* Annu Rev Biochem, 1995. **64**: p. 29-63.

58. Daffe, M. and P. Draper, *The envelope layers of mycobacteria with reference to their pathogenicity.* Adv Microb Physiol, 1998. **39**: p. 131-203.

59. Alderwick, L.J., et al., *Structure, function and biosynthesis of the Mycobacterium tuberculosis cell wall: arabinogalactan and lipoarabinomannan assembly with a view to discovering new drug targets.* Biochem Soc Trans, 2007. **35**(Pt 5): p. 1325-8.

60. Draper, P.D., M. 261-273. , *Cell envelope of Mycobacterium tuberculosis*, in *Tuberculosis and the tubercle bacillus*, S.E. Cole, KD; McMurray, DN; Jacobs, WR Jr., Editor. 2005, ASM Press p. 261-273.

61. Dover, L.G., et al., *Comparative cell wall core biosynthesis in the mycolated pathogens, Mycobacterium tuberculosis and Corynebacterium diphtheriae.* FEMS Microbiol Rev, 2004. **28**(2): p. 225-50.

62. Draper, P., *The outer parts of the mycobacterial envelope as permeability barriers.* Front Biosci, 1998. **3**: p. D1253-61.

63. Barry, C.E., 3rd and K. Mdluli, *Drug sensitivity and environmental adaptation of mycobacterial cell wall components.* Trends Microbiol, 1996. **4**(7): p. 275-81.

64. Clemens, D.L., *Mycobacterium tuberculosis: bringing down the wall.* Trends Microbiol, 1997. **5**(10): p. 383-5.

65. Barry, C.E., 3rd, et al., *Mycolic acids: structure, biosynthesis and physiological functions.* Prog Lipid Res, 1998. **37**(2-3): p. 143-79.

66. Watanabe, M., et al., *Location of functional groups in mycobacterial meromycolate chains; the recognition of new structural principles in mycolic acids.* Microbiology, 2002. **148**(Pt 6): p. 1881-902.

67. Watanabe, M., et al., *Separation and characterization of individual mycolic acids in representative mycobacteria.* Microbiology, 2001. **147**(Pt 7): p. 1825-37.

68. Barkan, D., et al., *Mycolic acid cyclopropanation is essential for viability, drug resistance, and cell wall integrity of Mycobacterium tuberculosis.* Chem Biol, 2009. **16**(5): p. 499-509.

69. Guerardel, Y., et al., *Structural study of lipomannan and lipoarabinomannan from Mycobacterium chelonae. Presence of unusual components with alpha 1,3-mannopyranose side chains.* J Biol Chem, 2002. **277**(34): p. 30635-48.

70. Khoo, K.H., et al., *Inositol phosphate capping of the nonreducing termini of lipoarabinomannan from rapidly growing strains of Mycobacterium.* J Biol Chem, 1995. **270**(21): p. 12380-9.

71. Bernardo, J., et al., *Differential responses of human mononuclear phagocytes to mycobacterial lipoarabinomannans: role of CD14 and the mannose receptor.* Infect Immun, 1998. **66**(1): p. 28-35.

72. Schlesinger, L.S., et al., *Differences in mannose receptor-mediated uptake of lipoarabinomannan from virulent and attenuated strains of Mycobacterium tuberculosis by human macrophages.* J Immunol, 1996. **157**(10): p. 4568-75.

73. Sidobre, S., et al., *Lipoglycans are putative ligands for the human pulmonary surfactant protein A attachment to mycobacteria. Critical role of the lipids for lectin-carbohydrate recognition.* J Biol Chem, 2000. **275**(4): p. 2415-22.

74. Shabaana, A.K., et al., *Mycobacterial lipoarabinomannans modulate cytokine production in human T helper cells by interfering with raft/microdomain signalling.* Cell Mol Life Sci, 2005. **62**(2): p. 179-87.

75. Welin, A., et al., *Incorporation of Mycobacterium tuberculosis lipoarabinomannan into macrophage membrane rafts is a prerequisite for the phagosomal maturation block.* Infect Immun, 2008. **76**(7): p. 2882-7.

76. Daffe, M. and G. Etienne, *The capsule of Mycobacterium tuberculosis and its implications for pathogenicity.* Tuber Lung Dis, 1999. **79**(3): p. 153-69.

77. Ortalo-Magne, A., et al., *Molecular composition of the outermost capsular material of the tubercle bacillus.* Microbiology, 1995. **141 (Pt 7)**: p. 1609-20.

78. Rezwan, M., et al., *Breaking down the wall: fractionation of mycobacteria.* J Microbiol Methods, 2007. **68**(1): p. 32-9.

79. Stahl, C., et al., *MspA provides the main hydrophilic pathway through the cell wall of Mycobacterium smegmatis.* Mol Microbiol, 2001. **40**(2): p. 451-64.

80. Senaratne, R.H., et al., *Expression of a gene for a porin-like protein of the OmpA family from Mycobacterium tuberculosis H37Rv.* J Bacteriol, 1998. **180**(14): p. 3541-7.

81. Singh, J.P. and R. Verma, *Cloning and molecular characterization of outer membrane protein A (ompA) gene from Mycobacterium bovis.* Microbiol Immunol, 2008. **52**(8): p. 410-7.

82. Siroy, A., et al., *Rv1698 of Mycobacterium tuberculosis represents a new class of channel-forming outer membrane proteins.* J Biol Chem, 2008. **283**(26): p. 17827-37.

83. Song, H., et al., *Identification of outer membrane proteins of Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2008. **88**(6): p. 526-44.

84. Alteri, C.J., et al., *Mycobacterium tuberculosis produces pili during human infection.* Proc Natl Acad Sci U S A, 2007. **104**(12): p. 5145-50.

85. Wallgren, A., *The time-table of tuberculosis.* Tubercle, 1948. **29**(11): p. 245-51.

86. Banchereau, J., et al., *Immunobiology of dendritic cells.* Annu Rev Immunol, 2000. **18**: p. 767-811.

87. North, R.J. and Y.J. Jung, *Immunity to tuberculosis.* Annu Rev Immunol, 2004. **22**: p. 599-623.

88. Brill, K.J., et al., *Human natural killer cells mediate killing of intracellular Mycobacterium tuberculosis H37Rv via granule-independent mechanisms.* Infect Immun, 2001. **69**(3): p. 1755-65.

89. Janulionis, E., et al., *Survival and replication of clinical Mycobacterium tuberculosis isolates in the context of human innate immunity.* Infect Immun, 2005. **73**(5): p. 2595-601.

90. Roach, D.R., et al., *TNF regulates chemokine induction essential for cell recruitment, granuloma formation, and clearance of mycobacterial infection.* J Immunol, 2002. **168**(9): p. 4620-7.

91. Dannenberg, A.M., Jr., *Roles of cytotoxic delayed-type hypersensitivity and macrophage-activating cell-mediated immunity in the pathogenesis of tuberculosis.* Immunobiology, 1994. **191**(4-5): p. 461-73.

92. Lazarevic, V., D. Nolt, and J.L. Flynn, *Long-term control of Mycobacterium tuberculosis infection is mediated by dynamic immune responses.* J Immunol, 2005. **175**(2): p. 1107-17.

93. Gomez, J.E. and J.D. McKinney, *M. tuberculosis persistence, latency, and drug tolerance.* Tuberculosis (Edinb), 2004. **84**(1-2): p. 29-44.

94. Rustad, T.R., et al., *Hypoxia: a window into Mycobacterium tuberculosis latency.* Cell Microbiol, 2009. **11**(8): p. 1151-9.

95. Via, L.E., et al., *Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates.* Infect Immun, 2008. **76**(6): p. 2333-40.

96. Medlar, E.M., *The behavior of pulmonary tuberculous lesions; a pathological study.* Am Rev Tuberc, 1955. **71**(3, Part 2): p. 1-244.

97. Beste, D.J., et al., *The genetic requirements for fast and slow growth in mycobacteria.* PLoS One, 2009. **4**(4): p. e5349.

98. Sassetti, C.M., D.H. Boyd, and E.J. Rubin, *Genes required for mycobacterial growth defined by high density mutagenesis.* Mol Microbiol, 2003. **48**(1): p. 77-84.

99. Smith, I., *Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence.* Clin Microbiol Rev, 2003. **16**(3): p. 463-96.

100. Brennan, P.J., *Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2003. **83**(1-3): p. 91-7.

101. Chan, J., et al., *Lipoarabinomannan, a possible virulence factor involved in persistence of Mycobacterium tuberculosis within macrophages.* Infect Immun, 1991. **59**(5): p. 1755-61.

102. Kang, P.B., et al., *The human macrophage mannose receptor directs Mycobacterium tuberculosis lipoarabinomannan-mediated phagosome biogenesis.* J. Exp. Med., 2005. **202**(7): p. 987-999.

103. Okamoto, Y., et al., *Mycobacterial sulfolipid shows a virulence by inhibiting cord factor induced granuloma formation and TNF-alpha release.* Microb Pathog, 2006. **40**(6): p. 245-53.

104. Reed, M.B., et al., *A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response.* Nature, 2004. **431**(7004): p. 84-7.

105. Indrigo, J., R.L. Hunter, Jr., and J.K. Actor, *Cord factor trehalose 6,6'-dimycolate (TDM) mediates trafficking events during mycobacterial infection of murine macrophages.* Microbiology, 2003. **149**(Pt 8): p. 2049-59.

106. Henao-Tamayo, M., et al., *A mutant of Mycobacterium tuberculosis lacking the 19-kDa lipoprotein Rv3763 is highly attenuated in vivo but retains potent vaccinogenic properties.* Vaccine, 2007. **25**(41): p. 7153-9.

107. Noss, E.H., et al., *Toll-like receptor 2-dependent inhibition of macrophage class II MHC expression and antigen processing by 19-kDa lipoprotein of Mycobacterium tuberculosis.* J Immunol, 2001. **167**(2): p. 910-8.

108. Vergne, I., et al., *Mechanism of phagolysosome biogenesis block by viable Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2005. **102**(11): p. 4033-8.

109. Walburger, A., et al., *Protein kinase G from pathogenic mycobacteria promotes survival within macrophages.* Science, 2004. **304**(5678): p. 1800-4.

110. Kumar, A., et al., *Mycobacterium tuberculosis DosS is a redox sensor and DosT is a hypoxia sensor.* Proc Natl Acad Sci U S A, 2007. **104**(28): p. 11568-73.

111. Roberts, D.M., et al., *Two sensor kinases contribute to the hypoxic response of Mycobacterium tuberculosis.* J Biol Chem, 2004. **279**(22): p. 23082-7.

112. Hu, Y., et al., *Deletion of the Mycobacterium tuberculosis alpha-crystallin-like hspX gene causes increased bacterial growth in vivo.* Infect Immun, 2006. **74**(2): p. 861-8.

113. Edwards, K.M., et al., *Iron-cofactored superoxide dismutase inhibits host responses to Mycobacterium tuberculosis.* Am J Respir Crit Care Med, 2001. **164**(12): p. 2213-9.

114. Munoz-Elias, E.J., et al., *Role of the methylcitrate cycle in Mycobacterium tuberculosis metabolism, intracellular growth, and virulence.* Mol Microbiol, 2006. **60**(5): p. 1109-22.

115. Ratledge, C. and M. Ewing, *The occurrence of carboxymycobactin, the siderophore of pathogenic mycobacteria, as a second extracellular siderophore in Mycobacterium smegmatis.* Microbiology, 1996. **142 (Pt 8)**: p. 2207-12.

116. Gobin, J. and M.A. Horwitz, *Exochelins of Mycobacterium tuberculosis remove iron from human iron-binding proteins and donate iron to mycobactins in the M. tuberculosis cell wall.* J Exp Med, 1996. **183**(4): p. 1527-32.

117. Guinn, K.M., et al., *Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis.* Mol Microbiol, 2004. **51**(2): p. 359-70.

118. Lewis, K.N., et al., *Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guerin attenuation.* J Infect Dis, 2003. **187**(1): p. 117-23.

119. Pathak, S.K., et al., *Direct extracellular interaction between the early secreted antigen ESAT-6 of Mycobacterium tuberculosis and TLR2 inhibits TLR signaling in macrophages.* Nat Immunol, 2007. **8**(6): p. 610-8.

120. Wards, B.J., G.W. de Lisle, and D.M. Collins, *An esat6 knockout mutant of Mycobacterium bovis produced by homologous recombination will contribute to the development of a live tuberculosis vaccine.* Tuber Lung Dis, 2000. **80**(4-5): p. 185-9.

121. Chitale, S., et al., *Recombinant Mycobacterium tuberculosis protein associated with mammalian cell entry.* Cell Microbiol, 2001. **3**(4): p. 247-54.

122. Joshi, S.M., et al., *Characterization of mycobacterial virulence genes through genetic interaction mapping.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11760-5.

123. Mueller-Ortiz, S.L., A.R. Wanger, and S.J. Norris, *Mycobacterial protein HbhA binds human complement component C3.* Infect Immun, 2001. **69**(12): p. 7501-11.

124. Pethe, K., et al., *The heparin-binding haemagglutinin of M. tuberculosis is required for extrapulmonary dissemination.* Nature, 2001. **412**(6843): p. 190-4.

125. Kehinde, A.O., et al., *Diagnosis of tuberculosis: urgent need to strengthen laboratory services.* J Natl Med Assoc, 2005. **97**(3): p. 394-6.

126. Hanna, B.A., et al., *Multicenter evaluation of the BACTEC MGIT 960 system for recovery of mycobacteria.* J Clin Microbiol, 1999. **37**(3): p. 748-52.

127. Teixeira, H.C., C. Abramo, and M.E. Munk, *Immunological diagnosis of tuberculosis: problems and strategies for success.* J Bras Pneumol, 2007. **33**(3): p. 323-34.

128. Gilpin, C., et al., *Critical appraisal of current recommendations and practices for tuberculosis sputum smear microscopy.* Int J Tuberc Lung Dis, 2007. **11**(9): p. 946-52.

129. Perkins, M.D., G. Roscigno, and A. Zumla, *Progress towards improved tuberculosis diagnostics for developing countries.* Lancet, 2006. **367**(9514): p. 942-3.

130. Gopinath, K. and S. Singh, *Urine as an adjunct specimen for the diagnosis of active pulmonary tuberculosis.* Int J Infect Dis, 2009. **13**(3): p. 374-9.

131. Eisenach, K.D., et al., *Polymerase chain reaction amplification of a repetitive DNA sequence specific for Mycobacterium tuberculosis.* J Infect Dis, 1990. **161**(5): p. 977-81.

132. Hermans, P.W., et al., *Specific detection of Mycobacterium tuberculosis complex strains by polymerase chain reaction.* J Clin Microbiol, 1990. **28**(6): p. 1204-13.

133. Richardson, E.T., D. Samson, and N. Banaei, *Rapid Identification of Mycobacterium tuberculosis and nontuberculous mycobacteria by multiplex, real-time PCR.* J Clin Microbiol, 2009. **47**(5): p. 1497-502.

134. Saves, I., et al., *Specificities and functions of the recA and pps1 intein genes of Mycobacterium tuberculosis and application for diagnosis of tuberculosis.* J Clin Microbiol, 2002. **40**(3): p. 943-50.

135. Stauffer, F., et al., *Genus level identification of mycobacteria from clinical specimens by using an easy-to-handle Mycobacterium-specific PCR assay.* J Clin Microbiol, 1998. **36**(3): p. 614-7.

136. Pietzka, A.T., et al., *Rapid identification of multidrug-resistant Mycobacterium tuberculosis isolates by rpoB gene scanning using high-resolution melting curve PCR analysis.* J Antimicrob Chemother, 2009. **63**(6): p. 1121-7.

137. Hazbon, M.H. and D. Alland, *Hairpin primers for simplified single-nucleotide polymorphism analysis of Mycobacterium tuberculosis and other organisms.* J Clin Microbiol, 2004. **42**(3): p. 1236-42.

138. Gopinath, K. and S. Singh, *Multiplex PCR assay for simultaneous detection and differentiation of Mycobacterium tuberculosis, Mycobacterium avium complexes and other Mycobacterial species directly from clinical specimens.* J Appl Microbiol, 2009.

139. Tsara, V., E. Serasli, and P. Christaki, *Problems in diagnosis and treatment of tuberculosis infection.* Hippokratia, 2009. **13**(1): p. 20-2.

140. Piersimoni, C. and C. Scarparo, *Relevance of commercial amplification methods for direct detection of Mycobacterium tuberculosis complex in clinical samples.* J Clin Microbiol, 2003. **41**(12): p. 5355-65.

141. Woldehanna, S. and J. Volmink, *Treatment of latent tuberculosis infection in HIV infected persons.* Cochrane Database Syst Rev, 2004(1): p. CD000171.

142. Fine, P.E., et al., *Delayed-type hypersensitivity, mycobacterial vaccines and protective immunity.* Lancet, 1994. **344**(8932): p. 1245-9.

143. Andersen, P., et al., *Specific immune-based diagnosis of tuberculosis.* Lancet, 2000. **356**(9235): p. 1099-104.

144. Markowitz, N., et al., *Tuberculin and anergy testing in HIV-seropositive and HIV-seronegative persons. Pulmonary Complications of HIV Infection Study Group.* Ann Intern Med, 1993. **119**(3): p. 185-93.

145. Huebner, R.E., M.F. Schein, and J.B. Bass, Jr., *The tuberculin skin test.* Clin Infect Dis, 1993. **17**(6): p. 968-75.

146. Schluger, N.W., *Assessing tuberculosis transmission and virulence: the vanishing tuberculin skin test.* Am J Respir Crit Care Med, 2006. **173**(9): p. 942-3.

147. Pai, M., A. Zwerling, and D. Menzies, *Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update.* Ann Intern Med, 2008. **149**(3): p. 177-84.

148. Pai, M. and R. O'Brien, *New diagnostics for latent and active tuberculosis: state of the art and future prospects.* Semin Respir Crit Care Med, 2008. **29**(5): p. 560-8.

149. Harboe, M., et al., *Evidence for occurrence of the ESAT-6 protein in Mycobacterium tuberculosis and virulent Mycobacterium bovis and for its absence in Mycobacterium bovis BCG.* Infect Immun, 1996. **64**(1): p. 16-22.

150. Kallenius, G., et al., *Evolution and clonal traits of Mycobacterium tuberculosis complex in Guinea-Bissau.* J Clin Microbiol, 1999. **37**(12): p. 3872-8.

151. de Jong, B.C., et al., *Mycobacterium africanum: a new opportunistic pathogen in HIV infection?* AIDS, 2005. **19**(15): p. 1714-5.

152. Vincenti, D., et al., *Response to region of difference 1 (RD1) epitopes in human immunodeficiency virus (HIV)-infected individuals enrolled with suspected active tuberculosis: a pilot study.* Clin Exp Immunol, 2007. **150**(1): p. 91-8.

153. Mazurek, G.H. and M.E. Villarino, *Guidelines for using the QuantiFERON-TB test for diagnosing latent Mycobacterium tuberculosis infection. Centers for Disease Control and Prevention.* MMWR Recomm Rep, 2003. **52**(RR-2): p. 15-8.

154. Mazurek, G.H., et al., *Guidelines for using the QuantiFERON-TB Gold test for detecting Mycobacterium tuberculosis infection, United States.* MMWR Recomm Rep, 2005. **54**(RR-15): p. 49-55.

155. Ulrichs, T., et al., *Numbers of IFN-gamma-producing cells against ESAT-6 increase in tuberculosis patients during chemotherapy.* Int J Tuberc Lung Dis, 2000. **4**(12): p. 1181-3.

156. Chegou, N.N., et al., *Host markers in QuantiFERON supernatants differentiate active TB from latent TB infection: preliminary report.* BMC Pulm Med, 2009. **9**: p. 21.

157. Jones, D., et al., *Control of Gram-Negative Bacteria in Experimental Animals by Streptomycin.* Science, 1944. **100**(2588): p. 103-105.

158. Middlebrook, G. and S.H. Dressler, *Clinical evaluation of isoniazid.* Am Rev Tuberc, 1954. **70**(6): p. 1102-3.

159. Maggi, N., et al., *Rifampicin: a new orally active rifamycin.* Chemotherapy, 1966. **11**(5): p. 285-92.

160. Delaude, A., M.F. Jorda, and G. Moreau, *[Action of ethambutol on experimental tuberculosis in the guinea pig.].* Rev Tuberc Pneumol (Paris), 1962. **26**: p. 519-24.

161. Guy, E.S. and A. Mallampalli, *Managing TB in the 21st century: existing and novel drug therapies.* Ther Adv Respir Dis, 2008. **2**(6): p. 401-8.

162. Blumberg, H.M., et al., *American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis.* Am J Respir Crit Care Med, 2003. **167**(4): p. 603-62.

163. WHO (2003) *Treatment of Tuberculosis: Guidelines for National Programmes*

164. WHO (1998) *Policy Statement on Preventive Therapy Against Tuberculosis in People Living With HIV.* **WHO/TB/98.255.**

165. Bucher, H.C., et al., *Isoniazid prophylaxis for tuberculosis in HIV infection: a meta-analysis of randomized controlled trials.* AIDS, 1999. **13**(4): p. 501-7.

166. Grant, A.D., et al., *Effect of routine isoniazid preventive therapy on tuberculosis incidence among HIV-infected men in South Africa: a novel randomized incremental recruitment study.* JAMA, 2005. **293**(22): p. 2719-25.

167. *Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs--worldwide, 2000-2004.* MMWR Morb Mortal Wkly Rep, 2006. **55**(11): p. 301-5.

168. Crofton, J., P. Chaulet, D. Maher, J. Grosset, W. Harris, H. Norman, M. D. Iseman, and B. Watt. (1997) *Guidelines for the management of multidrugresistant tuberculosis.*

169. Bifani, P.J., et al., *Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family.* JAMA, 1996. **275**(6): p. 452-7.

170. *Transmission of Mycobacterium tuberculosis associated with failed completion of treatment for latent tuberculosis infection--Chickasaw County, Mississippi, June 1999-March 2002.* MMWR Morb Mortal Wkly Rep, 2003. **52**(11): p. 222-4.

171. WHO (2006) *Instructions for applying to the Green Light Committee for access to secondline anti-tuberculosis drugs.* **WHO/HTM/TB/2006.369.**

172. Laurenzi, M., A. Ginsberg, and M. Spigelman, *Challenges associated with current and future TB treatment.* Infect Disord Drug Targets, 2007. 7(2): p. 105-19.

173. Raviglione, M.C. and A. Pio, *Evolution of WHO policies for tuberculosis control, 1948-2001.* Lancet, 2002. **359**(9308): p. 775-80.

174. WHO (2008) *Tuberculosis facts.*

175. WHO (2009) *Global tuberculosis control - epidemiology, strategy, financing WHO Report 2009.* **WHO/HTM/TB/2009.411**.

176. Alix, E., S. Godreuil, and A.B. Blanc-Potard, *Identification of a Haarlem genotype-specific single nucleotide polymorphism in the mgtC virulence gene of Mycobacterium tuberculosis.* J Clin Microbiol, 2006. **44**(6): p. 2093-8.

177. Filliol, I., et al., *Snapshot of moving and expanding clones of Mycobacterium tuberculosis and their global distribution assessed by spoligotyping in an international study.* J Clin Microbiol, 2003. **41**(5): p. 1963-70.

178. Kremer, K., et al., *Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility.* J Clin Microbiol, 1999. **37**(8): p. 2607-18.

179. Gagneux, S., et al., *Variable host-pathogen compatibility in Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2869-73.

180. Cave, M.D., et al., *IS6110: conservation of sequence in the Mycobacterium tuberculosis complex and its utilization in DNA fingerprinting.* Mol Cell Probes, 1991. **5**(1): p. 73-80.

181. Kamerbeek, J., et al., *Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology.* J Clin Microbiol, 1997. **35**(4): p. 907-14.

182. Mazars, E., et al., *High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology.* Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1901-6.

183. van Embden, J.D., et al., *Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology.* J Clin Microbiol, 1993. **31**(2): p. 406-9.

184. van Soolingen, D., et al., *Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of Mycobacterium tuberculosis.* J Clin Microbiol, 1993. **31**(8): p. 1987-95.

185. Rhee, J.T., et al., *Use of multiple markers in population-based molecular epidemiologic studies of tuberculosis.* Int J Tuberc Lung Dis, 2000. **4**(12): p. 1111-9.

186. Rhee, J.T., et al., *Molecular epidemiologic evaluation of transmissibility and virulence of Mycobacterium tuberculosis.* J Clin Microbiol, 1999. **37**(6): p. 1764-70.

187. Stermann, M., et al., *Polymorphic nucleotide within the promoter of nitrate reductase (NarGHJI) is specific for Mycobacterium tuberculosis.* J Clin Microbiol, 2003. **41**(7): p. 3252-9.

188. Sreevatsan, S., et al., *Identification of a polymorphic nucleotide in oxyR specific for Mycobacterium bovis.* J Clin Microbiol, 1996. **34**(8): p. 2007-10.

189. Lee, J.S., et al., *Mutation in the transcriptional regulator PhoP contributes to avirulence of Mycobacterium tuberculosis H37Ra strain.* Cell Host Microbe, 2008. **3**(2): p. 97-103.

190. Sola, C., et al., *Genotyping of the Mycobacterium tuberculosis complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics.* Infect Genet Evol, 2003. **3**(2): p. 125-33.

191. Fang, Z., et al., *IS6110 transposition and evolutionary scenario of the direct repeat locus in a group of closely related Mycobacterium tuberculosis strains.* J Bacteriol, 1998. **180**(8): p. 2102-9.

192. Gagneux, S. and P.M. Small, *Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development.* Lancet Infect Dis, 2007. **7**(5): p. 328-37.

193. Valway, S.E., et al., *An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis.* N Engl J Med, 1998. **338**(10): p. 633-9.

194. Das, S., et al., *IS6110 restriction fragment length polymorphism typing of clinical isolates of Mycobacterium tuberculosis from patients with pulmonary tuberculosis in Madras, south India.* Tuber Lung Dis, 1995. **76**(6): p. 550-4.

195. Small, P.M., et al., *Exogenous reinfection with multidrug-resistant Mycobacterium tuberculosis in patients with advanced HIV infection.* N Engl J Med, 1993. **328**(16): p. 1137-44.

196. Sonnenberg, P., et al., *HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: a cohort study in South African mineworkers.* Lancet, 2001. **358**(9294): p. 1687-93.

197. Barnes, P.F., et al., *Patterns of tuberculosis transmission in Central Los Angeles.* JAMA, 1997. **278**(14): p. 1159-63.

198. Small, P.M., et al., *The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods.* N Engl J Med, 1994. **330**(24): p. 1703-9.

199. Vynnycky, E. and P.E. Fine, *The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection.* Epidemiol Infect, 1997. **119**(2): p. 183-201.

200. Li, Q., et al., *Differences in rate and variability of intracellular growth of a panel of Mycobacterium tuberculosis clinical isolates within a human monocyte model.* Infect Immun, 2002. **70**(11): p. 6489-93.

201. Dormans, J., et al., *Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different Mycobacterium tuberculosis genotypes in a BALB/c mouse model.* Clin Exp Immunol, 2004. **137**(3): p. 460-8.

202. Anderson, S.T., et al., *Transmission of Mycobacterium tuberculosis undetected by tuberculin skin testing.* Am J Respir Crit Care Med, 2006. **173**(9): p. 1038-42.

203. Manca, C., et al., *Virulence of a Mycobacterium tuberculosis clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha /beta.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5752-7.

204. Rajavelu, P. and S.D. Das, *Th2-type immune response observed in healthy individuals to sonicate antigen prepared from the most prevalent Mycobacterium tuberculosis strain with single copy of IS6110.* FEMS Immunol Med Microbiol, 2005. **45**(1): p. 95-102.

205. Domenech, P., et al., *The role of MmpL8 in sulfatide biogenesis and virulence of Mycobacterium tuberculosis.* J Biol Chem, 2004. **279**(20): p. 21257-65.

206. Roach, T.I., et al., *Macrophage activation: lipoarabinomannan from avirulent and virulent strains of Mycobacterium tuberculosis differentially induces the early genes c-fos, KC, JE, and tumor necrosis factor-alpha.* J Immunol, 1993. **150**(5): p. 1886-96.

207. Steenken, W.G., LU., *History of H37 strain of tubercle bacillus.* American Review Tuberculosis, 1946. **54**: p. 62-66.

208. Hoal-van Helden, E.G., et al., *Mycobacterial growth in human macrophages: variation according to donor, inoculum and bacterial strain.* Cell Biol Int, 2001. **25**(1): p. 71-81.

209. Jacobs, W., et al., *Comparative sequencing.* Science, 1996. **274**(5284): p. 17-8.

210. Andreu, N. and I. Gibert, *Cell population heterogeneity in Mycobacterium tuberculosis H37Rv.* Tuberculosis (Edinb), 2008. **88**(6): p. 553-9.

211. North, R.J. and A.A. Izzo, *Mycobacterial virulence. Virulent strains of Mycobacteria tuberculosis have faster in vivo doubling times and are better equipped to resist growth-inhibiting functions of macrophages in the presence and absence of specific immunity.* J Exp Med, 1993. **177**(6): p. 1723-33.

212. Freeman, S., et al., *Mycobacterium tuberculosis H37Ra and H37Rv differential growth and cytokine/chemokine induction in murine macrophages in vitro.* J Interferon Cytokine Res, 2006. **26**(1): p. 27-33.

213. Silver, R.F., et al., *Human alveolar macrophage gene responses to Mycobacterium tuberculosis strains H37Ra and H37Rv.* Am J Respir Cell Mol Biol, 2009. **40**(4): p. 491-504.

214. Gonzalo Asensio, J., et al., *The virulence-associated two-component PhoP-PhoR system controls the biosynthesis of polyketide-derived lipids in Mycobacterium tuberculosis.* J Biol Chem, 2006. **281**(3): p. 1313-6.

215. Walters, S.B., et al., *The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis.* Mol Microbiol, 2006. **60**(2): p. 312-30.

216. Chesne-Seck, M.L., et al., *A point mutation in the two-component regulator PhoP-PhoR accounts for the absence of polyketide-derived acyltrehaloses but not that of phthiocerol dimycocerosates in Mycobacterium tuberculosis H37Ra.* J Bacteriol, 2008. **190**(4): p. 1329-34.

217. Gonzalo-Asensio, J., et al., *PhoP: a missing piece in the intricate puzzle of Mycobacterium tuberculosis virulence.* PLoS One, 2008. **3**(10): p. e3496.

218. Sohn, H., et al., *Induction of cell death in human macrophages by a highly virulent Korean Isolate of Mycobacterium tuberculosis and the virulent strain H37Rv.* Scand J Immunol, 2009. **69**(1): p. 43-50.

219. Tanveer, M., et al., *Reduced TNF-alpha and IFN-gamma responses to Central Asian strain 1 and Beijing isolates of Mycobacterium tuberculosis in comparison with H37Rv strain.* Trans R Soc Trop Med Hyg, 2009. **103**(6): p. 581-7.

220. Manca, C., 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**(11): p. 6740-6.

221. Bishai, W.R., et al., *Virulence of Mycobacterium tuberculosis CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method.* Infect Immun, 1999. **67**(9): p. 4931-4.

222. Tsenova, L., 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**(1): p. 98-106.

223. Barczak, A.K., et al., *In vivo phenotypic dominance in mouse mixed infections with Mycobacterium tuberculosis clinical isolates.* J Infect Dis, 2005. **192**(4): p. 600-6.

224. Manca, C., et al., *Differential monocyte activation underlies strain-specific Mycobacterium tuberculosis pathogenesis.* Infect Immun, 2004. **72**(9): p. 5511-4.

225. Glynn, J.R., et al., *Worldwide occurrence of Beijing/W strains of Mycobacterium tuberculosis: a systematic review.* Emerg Infect Dis, 2002. **8**(8): p. 843-9.

226. Caws, M., et al., *Beijing genotype of Mycobacterium tuberculosis is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis.* J Clin Microbiol, 2006. **44**(11): p. 3934-9.

227. Friedman, C.R., et al., *Widespread dissemination of a drug-susceptible strain of Mycobacterium tuberculosis.* J Infect Dis, 1997. **176**(2): p. 478-84.

228. Chan, J., et al., *Killing of virulent Mycobacterium tuberculosis by reactive nitrogen intermediates produced by activated murine macrophages.* J Exp Med, 1992. **175**(4): p. 1111-22.

229. O'Brien, L., et al., *Strains of Mycobacterium tuberculosis differ in susceptibility to reactive nitrogen intermediates in vitro.* Infect Immun, 1994. **62**(11): p. 5187-90.

230. MacMicking, J.D., et al., *Identification of nitric oxide synthase as a protective locus against tuberculosis.* Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5243-8.

231. Rhoades, E.R. and I.M. Orme, *Susceptibility of a panel of virulent strains of Mycobacterium tuberculosis to reactive nitrogen intermediates.* Infect Immun, 1997. **65**(4): p. 1189-95.

232. Kim, S.J., et al., *Transmission of Mycobacterium tuberculosis among high school students in Korea.* Int J Tuberc Lung Dis, 2001. **5**(9): p. 824-30.

233. Bloom, B.R. and C.J. Murray, *Tuberculosis: commentary on a reemergent killer.* Science, 1992. **257**(5073): p. 1055-64.

234. Hernandez-Pando, R., et al., *Pathogenesis of tuberculosis in mice exposed to low and high doses of an environmental mycobacterial saprophyte before infection.* Infect Immun, 1997. **65**(8): p. 3317-27.

235. Nardell, E.A., *Tuberculosis in homeless, residential care facilities, prisons, nursing homes, and other close communities.* Semin Respir Infect, 1989. **4**(3): p. 206-15.

236. Collins, F.M., *Mycobacterial disease, immunosuppression, and acquired immunodeficiency syndrome.* Clin Microbiol Rev, 1989. **2**(4): p. 360-77.

237. Bellamy, R., et al., *Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans.* N Engl J Med, 1998. **338**(10): p. 640-4.

238. Ordway, D.J., et al., *Drug-resistant strains of Mycobacterium tuberculosis exhibit a range of virulence for mice.* Infect Immun, 1995. **63**(2): p. 741-3.

239. Medini, D., et al., *The microbial pan-genome.* Curr Opin Genet Dev, 2005. **15**(6): p. 589-94.

240. Cole, S.T., et al., *Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.* Nature, 1998. **393**(6685): p. 537-44.

241. Mahairas, G.G., et al., *Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis.* J Bacteriol, 1996. **178**(5): p. 1274-82.

242. Behr, M.A., et al., *Comparative genomics of BCG vaccines by whole-genome DNA microarray.* Science, 1999. **284**(5419): p. 1520-3.

243. Gordon, S.V., et al., *Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays.* Mol Microbiol, 1999. **32**(3): p. 643-55.

244. Newton, S.M., et al., *A deletion defining a common Asian lineage of Mycobacterium tuberculosis associates with immune subversion.* Proc Natl Acad Sci U S A, 2006. **103**(42): p. 15594-8.

245. Brosch, R., et al., *Genomic analysis reveals variation between Mycobacterium tuberculosis H37Rv and the attenuated M. tuberculosis H37Ra strain.* Infect Immun, 1999. **67**(11): p. 5768-74.

246. Brosch, R., et al., *A new evolutionary scenario for the Mycobacterium tuberculosis complex.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3684-9.

247. Sorensen, A.L., et al., *Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis.* Infect Immun, 1995. **63**(5): p. 1710-7.

248. Lazzarini, L.C., et al., *Discovery of a novel Mycobacterium tuberculosis lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil.* J Clin Microbiol, 2007. **45**(12): p. 3891-902.

249. Betts, J.C., et al., *Comparison of the proteome of Mycobacterium tuberculosis strain H37Rv with clinical isolate CDC 1551.* Microbiology, 2000. **146 Pt 12**: p. 3205-16.

250. Fleischmann, R.D., et al., *Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.* J Bacteriol, 2002. **184**(19): p. 5479-90.

251. Hirsh, A.E., et al., *Stable association between strains of Mycobacterium tuberculosis and their human host populations.* Proc Natl Acad Sci U S A, 2004. **101**(14): p. 4871-6.

252. Ho, T.B., et al., *Comparison of Mycobacterium tuberculosis genomes reveals frequent deletions in a 20 kb variable region in clinical isolates.* Yeast, 2000. **17**(4): p. 272-82.

253. Kato-Maeda, M., et al., *Comparing genomes within the species Mycobacterium tuberculosis.* Genome Res, 2001. **11**(4): p. 547-54.

254. Smith, G.A., et al., *The two distinct phospholipases C of Listeria monocytogenes have overlapping roles in escape from a vacuole and cell-to-cell spread.* Infect Immun, 1995. **63**(11): p. 4231-7.

255. George, K.M., et al., *Mycolactone: a polyketide toxin from Mycobacterium ulcerans required for virulence.* Science, 1999. **283**(5403): p. 854-7.

256. Betts, J.C., et al., *Signature gene expression profiles discriminate between isoniazid-, thiolactomycin-, and triclosan-treated Mycobacterium tuberculosis.* Antimicrob Agents Chemother, 2003. **47**(9): p. 2903-13.

257. Colangeli, R., et al., *Transcriptional regulation of multi-drug tolerance and antibiotic-induced responses by the histone-like protein Lsr2 in M. tuberculosis.* PLoS Pathog, 2007. **3**(6): p. e87.

258. Wilson, M., et al., *Exploring drug-induced alterations in gene expression in Mycobacterium tuberculosis by microarray hybridization.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12833-8.

259. Dasgupta, N., et al., *Characterization of a two-component system, devR-devS, of Mycobacterium tuberculosis.* Tuber Lung Dis, 2000. **80**(3): p. 141-59.

260. Kinger, A.K. and J.S. Tyagi, *Identification and cloning of genes differentially expressed in the virulent strain of Mycobacterium tuberculosis.* Gene, 1993. **131**(1): p. 113-7.

261. Rindi, L., N. Lari, and C. Garzelli, *Search for genes potentially involved in Mycobacterium tuberculosis virulence by mRNA differential display.* Biochem Biophys Res Commun, 1999. **258**(1): p. 94-101.

262. Gao, Q., et al., *Comparative expression studies of a complex phenotype: cord formation in Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2004. **84**(3-4): p. 188-96.

263. Voskuil, M.I., et al., *Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program.* J Exp Med, 2003. **198**(5): p. 705-13.

264. Voskuil, M.I., K.C. Visconti, and G.K. Schoolnik, *Mycobacterium tuberculosis gene expression during adaptation to stationary phase and low-oxygen dormancy.* Tuberculosis (Edinb), 2004. **84**(3-4): p. 218-27.

265. Reed, M.B., et al., *The W-Beijing lineage of Mycobacterium tuberculosis overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated.* J Bacteriol, 2007. **189**(7): p. 2583-9.

266. Sherman, D.R., et al., *Regulation of the Mycobacterium tuberculosis hypoxic response gene encoding alpha -crystallin.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7534-9.

267. Li, A.H., W.L. Lam, and R.W. Stokes, *Characterization of genes differentially expressed within macrophages by virulent and attenuated Mycobacterium tuberculosis identifies candidate genes involved in intracellular growth.* Microbiology, 2008. **154**(Pt 8): p. 2291-303.

268. Rehren, G., et al., *Differential gene expression between Mycobacterium bovis and Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2007. **87**(4): p. 347-59.

269. Lam, T.H., et al., *Differential fadE28 expression associated with phenotypic virulence of Mycobacterium tuberculosis.* Microb Pathog, 2008. **45**(1): p. 12-7.

270. Munoz-Elias, E.J. and J.D. McKinney, *Carbon metabolism of intracellular bacteria.* Cell Microbiol, 2006. **8**(1): p. 10-22.

271. Sassetti, C.M. and E.J. Rubin, *Genetic requirements for mycobacterial survival during infection.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12989-94.

272. Flores, J. and C. Espitia, *Differential expression of PE and PE_PGRS genes in Mycobacterium tuberculosis strains.* Gene, 2003. **318**: p. 75-81.

273. Gao, Q., et al., *Gene expression diversity among Mycobacterium tuberculosis clinical isolates.* Microbiology, 2005. **151**(Pt 1): p. 5-14.

274. Rindi, L., et al., *Variation of the expression of Mycobacterium tuberculosis ppe44 gene among clinical isolates.* FEMS Immunol Med Microbiol, 2007. **51**(2): p. 381-7.

275. Jungblut, P.R., et al., *Comparative proteome analysis of Mycobacterium tuberculosis and Mycobacterium bovis BCG strains: towards functional genomics of microbial pathogens.* Mol Microbiol, 1999. **33**(6): p. 1103-17.

276. Rosenkrands, I., et al., *Towards the proteome of Mycobacterium tuberculosis.* Electrophoresis, 2000. **21**(17): p. 3740-56.

277. Cavalieri, D. and C. De Filippo, *Bioinformatic methods for integrating whole-genome expression results into cellular networks.* Drug Discov Today, 2005. **10**(10): p. 727-34.

278. Boehm, A.M., et al., *Precise protein quantification based on peptide quantification using iTRAQ.* BMC Bioinformatics, 2007. **8**: p. 214.

279. Boehm, A.M., F. Grosse-Coosmann, and A. Sickmann, *Command line tool for calculating theoretical MS spectra for given sequences.* Bioinformatics, 2004. **20**(16): p. 2889-91.

280. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

281. Issaq, H.J., et al., *Separation, detection and quantitation of peptides by liquid chromatography and capillary electrochromatography.* J Chromatogr A, 2009. **1216**(10): p. 1825-37.

282. Koichi Tanaka, et al., *Protein and polymer analyses up to <I>m/z</I> 100 000 by laser ionization time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 1988. **2**(8): p. 151-153.

283. Bienvenut, W.M., M; Palagi, PM; Gasteiger, E; Heller, M; Jung, E; Giron, M; Gras, R; Gay, S; Binz, P-A; Hughes, GJ; Sanchez, J-C; Appel, RD; Hochstrasser, DF., *Proteomics and mass spectrometry.* Acceleration and improvement of protein identification by mass spectrometry, ed. W. Bienvenut. 2005, Springer, Netherlands.

284. Arnott, D., *Basics of Triple-Stage Quadrupole/Ion-Trap Mass Spectrometry: Precursor, Product and neutral loss scanning. Electrospray Ionisation and Nanospray Ionisation.* Proteome Research: Mass Spectrometry., ed. P. James. 2001, Germany: Springer.

285. Schwartz, J.C., M.W. Senko, and J.E. Syka, *A two-dimensional quadrupole ion trap mass spectrometer.* J Am Soc Mass Spectrom, 2002. **13**(6): p. 659-69.

286. Want, E.J., B.F. Cravatt, and G. Siuzdak, *The expanding role of mass spectrometry in metabolite profiling and characterization.* Chembiochem, 2005. **6**(11): p. 1941-51.

287. Griffin, T.J., et al., *iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer.* J Proteome Res, 2007. **6**(11): p. 4200-9.

288. Schwartz, J.S., JP.; Quarmby, ST. *In Improving the Fundamentals of Msn on 2D Ion Traps: New Ion Activation and Isolation Techniques.* in *53rd ASMS Conference on Mass Spectrometry.* 2005. San Antonio, Texas.

289. Bantscheff, M., et al., *Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer.* Mol Cell Proteomics, 2008. **7**(9): p. 1702-13.

290. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole-time-of-flight mass spectrometry.* J Mass Spectrom, 2001. **36**(8): p. 849-65.

291. Morris, H.R., et al., *High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer.* Rapid Commun Mass Spectrom, 1996. **10**(8): p. 889-96.

292. Steen, H., B. Kuster, and M. Mann, *Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning.* J Mass Spectrom, 2001. **36**(7): p. 782-90.

293. Lund, T.C., et al., *iTRAQ is a useful method to screen for membrane-bound proteins differentially expressed in human natural killer cell types.* J Proteome Res, 2007. **6**(2): p. 644-53.

294. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.

295. Molloy, M.P., et al., *Overcoming technical variation and biological variation in quantitative proteomics.* Proteomics, 2003. **3**(10): p. 1912-9.

296. Choe, L.H., et al., *A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in Escherichia coli cells.* Electrophoresis, 2005. **26**(12): p. 2437-49.

297. Shadforth, I.P., et al., *i-Tracker: for quantitative proteomics using iTRAQ.* BMC Genomics, 2005. **6**: p. 145.

298. Wu, K.L., WT; Hung, WN; Yian, YH; Chen, YR; Chen, YJ; Sung, TY; Hsu, WL., *MassTRAQ: A Fully Automated Tool for iTRAQlabeled Protein Quantification.* IEEE Computer Society, ed. D. Martin. 2005, Stanford, USA.

299. D'Ascenzo, M., L. Choe, and K.H. Lee, *iTRAQPak: an R based analysis and visualization package for 8-plex isobaric protein expression data.* Brief Funct Genomic Proteomic, 2008. **7**(2): p. 127-35.

300. Radosevich, T.J., et al., *Proteome and differential expression analysis of membrane and cytosolic proteins from Mycobacterium avium subsp. paratuberculosis strains K-10 and 187.* J Bacteriol, 2007. **189**(3): p. 1109-17.

301. Seth, M., et al., *Biomarker discovery in subclinical mycobacterial infections of cattle.* PLoS One, 2009. **4**(5): p. e5478.

302. Hill, E.G., et al., *A statistical model for iTRAQ data analysis.* J Proteome Res, 2008. **7**(8): p. 3091-101.

303. Oberg, A.L., et al., *Statistical analysis of relative labeled mass spectrometry data from complex samples using ANOVA.* J Proteome Res, 2008. **7**(1): p. 225-33.

304. Taylor, R.D., et al., *Proteomic analyses of Fusarium graminearum grown under mycotoxin-inducing conditions.* Proteomics, 2008. **8**(11): p. 2256-65.

305. Tafelmeyer, P., et al., *Comprehensive proteome analysis of Mycobacterium ulcerans and quantitative comparison of mycolactone biosynthesis.* Proteomics, 2008. **8**(15): p. 3124-38.

306. Shui, W., et al., *Quantitative proteomic profiling of host-pathogen interactions: the macrophage response to Mycobacterium tuberculosis lipids.* J Proteome Res, 2009. **8**(1): p. 282-9.

307. Yi, E.C., et al., *Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme.* Proteomics, 2005. **5**(2): p. 380-7.

308. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.

309. Hughes, M.A., et al., *Quantitative proteomic analysis of drug-induced changes in mycobacteria.* J Proteome Res, 2006. **5**(1): p. 54-63.

310. de Souza, G.A., et al., *High accuracy mass spectrometry analysis as a tool to verify and improve gene annotation using Mycobacterium tuberculosis as an example.* BMC Genomics, 2008. **9**: p. 316.

311. Pheiffer, C., et al., *Protein expression by a Beijing strain differs from that of another clinical isolate and Mycobacterium tuberculosis H37Rv.* Microbiology, 2005. **151**(Pt 4): p. 1139-50.

312. Roberts, A.D., et al., *Characteristics of protective immunity engendered by vaccination of mice with purified culture filtrate protein antigens of Mycobacterium tuberculosis.* Immunology, 1995. **85**(3): p. 502-8.

313. Malen, H., et al., *Comprehensive analysis of exported proteins from Mycobacterium tuberculosis H37Rv.* Proteomics, 2007. **7**(10): p. 1702-18.

314. Mattow, J., et al., *Comparative proteome analysis of culture supernatant proteins from virulent Mycobacterium tuberculosis H37Rv and attenuated M. bovis BCG Copenhagen.* Electrophoresis, 2003. **24**(19-20): p. 3405-20.

315. Rosenkrands, I., et al., *Mapping and identification of Mycobacterium tuberculosis proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection.* Electrophoresis, 2000. **21**(5): p. 935-48.

316. Samanich, K.M., et al., *Delineation of human antibody responses to culture filtrate antigens of Mycobacterium tuberculosis.* J Infect Dis, 1998. **178**(5): p. 1534-8.

317. Sonnenberg, M.G. and J.T. Belisle, *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**(11): p. 4515-24.

318. Urquhart, B.L., S.J. Cordwell, and I. Humphery-Smith, *Comparison of predicted and observed properties of proteins encoded in the genome of Mycobacterium tuberculosis H37Rv.* Biochem Biophys Res Commun, 1998. **253**(1): p. 70-9.

319. Mawuenyega, K.G., et al., *Mycobacterium tuberculosis functional network analysis by global subcellular protein profiling.* Mol Biol Cell, 2005. **16**(1): p. 396-404.

320. Schmidt, F., et al., *Complementary analysis of the Mycobacterium tuberculosis proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology.* Mol Cell Proteomics, 2004. **3**(1): p. 24-42.

321. Jungblut, P.R., et al., *Proteomics reveals open reading frames in Mycobacterium tuberculosis H37Rv not predicted by genomics.* Infect Immun, 2001. **69**(9): p. 5905-7.

322. Malen, H., et al., *Membrane and membrane-associated proteins in Triton X-114 extracts of Mycobacterium bovis BCG identified using a combination of gel-based and gel-free fractionation strategies.* Proteomics, 2008. **8**(9): p. 1859-70.

323. Sartain, M.J., et al., *Disease state differentiation and identification of tuberculosis biomarkers via native antigen array profiling.* Mol Cell Proteomics, 2006. **5**(11): p. 2102-13.

324. Gonzalez-Zamorano, M., et al., *Mycobacterium tuberculosis glycoproteomics based on ConA-lectin affinity capture of mannosylated proteins.* J Proteome Res, 2009. **8**(2): p. 721-33.

325. Sieling, P.A., et al., *Conserved mycobacterial lipoglycoproteins activate TLR2 but also require glycosylation for MHC class II-restricted T cell activation.* J Immunol, 2008. **180**(9): p. 5833-42.

326. Dobos, K.M., et al., *Evidence for glycosylation sites on the 45-kilodalton glycoprotein of Mycobacterium tuberculosis.* Infect Immun, 1995. **63**(8): p. 2846-53.

327. Espitia, C. and R. Mancilla, *Identification, isolation and partial characterization of Mycobacterium tuberculosis glycoprotein antigens.* Clin Exp Immunol, 1989. **77**(3): p. 378-83.

328. Betts, J.C., et al., *Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling.* Mol Microbiol, 2002. **43**(3): p. 717-31.

329. Boon, C., et al., *Proteins of Mycobacterium bovis BCG induced in the Wayne dormancy model.* J Bacteriol, 2001. **183**(8): p. 2672-6.

330. Starck, J., et al., *Comparative proteome analysis of Mycobacterium tuberculosis grown under aerobic and anaerobic conditions.* Microbiology, 2004. **150**(Pt 11): p. 3821-9.

331. Yuan, Y., D.D. Crane, and C.E. Barry, 3rd, *Stationary phase-associated protein expression in Mycobacterium tuberculosis: function of the mycobacterial alpha-crystallin homolog.* J Bacteriol, 1996. **178**(15): p. 4484-92.

332. Florczyk, M.A., et al., *Identification and characterization of mycobacterial proteins differentially expressed under standing and shaking culture conditions, including Rv2623 from a novel class of putative ATP-binding proteins.* Infect Immun, 2001. **69**(9): p. 5777-85.

333. Monahan, I.M., et al., *Differential expression of mycobacterial proteins following phagocytosis by macrophages.* Microbiology, 2001. **147**(Pt 2): p. 459-71.

334. Rosenkrands, I., et al., *Hypoxic response of Mycobacterium tuberculosis studied by metabolic labeling and proteome analysis of cellular and extracellular proteins.* J Bacteriol, 2002. **184**(13): p. 3485-91.

335. Cho, S.H., D. Goodlett, and S. Franzblau, *ICAT-based comparative proteomic analysis of non-replicating persistent Mycobacterium tuberculosis.* Tuberculosis (Edinb), 2006. **86**(6): p. 445-60.

336. Garbe, T.R., N.S. Hibler, and V. Deretic, *Response of Mycobacterium tuberculosis to reactive oxygen and nitrogen intermediates.* Mol Med, 1996. **2**(1): p. 134-42.

337. Stewart, G.R., et al., *Overexpression of heat-shock proteins reduces survival of Mycobacterium tuberculosis in the chronic phase of infection.* Nat Med, 2001. **7**(6): p. 732-7.

338. Wong, D.K., et al., *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**(1): p. 327-36.

339. Lee, B.Y. and M.A. Horwitz, *Identification of macrophage and stress-induced proteins of Mycobacterium tuberculosis.* J Clin Invest, 1995. **96**(1): p. 245-9.

340. Mattow, J., et al., *Proteins unique to intraphagosomally grown Mycobacterium tuberculosis.* Proteomics, 2006. **6**(8): p. 2485-94.

341. Jiang, X., et al., *Comparison of the proteome of isoniazid-resistant and -susceptible strains of Mycobacterium tuberculosis.* Microb Drug Resist, 2006. **12**(4): p. 231-8.

342. Mdluli, K., et al., *Inhibition of a Mycobacterium tuberculosis beta-ketoacyl ACP synthase by isoniazid.* Science, 1998. **280**(5369): p. 1607-10.

343. Wang, R. and E.M. Marcotte, *The proteomic response of Mycobacterium smegmatis to anti-tuberculosis drugs suggests targeted pathways.* J Proteome Res, 2008. **7**(3): p. 855-65.

344. He, X.Y., et al., *Comparative proteome analysis of culture supernatant proteins of Mycobacterium tuberculosis H37Rv and H37Ra.* Microbes Infect, 2003. **5**(10): p. 851-6.

345. Jiang, X., et al., *Identification of unique genetic markers in Rv0927c among Mycobacterium tuberculosis W-Beijing strains.* Microbes Infect, 2007. **9**(2): p. 241-6.

346. Bahk, Y.Y., et al., *Antigens secreted from Mycobacterium tuberculosis: identification by proteomics approach and test for diagnostic marker.* Proteomics, 2004. **4**(11): p. 3299-307.

347. Ryoo, S.W., 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**(3): p. 268-71.

348. Cherepkova, O.A., et al., *Chaperone-like activity of macrophage migration inhibitory factor.* Int J Biochem Cell Biol, 2006. **38**(1): p. 43-55.

349. Pheiffer, C., et al., *Protein expression in Mycobacterium tuberculosis differs with growth stage and strain type.* Clin Chem Lab Med, 2002. **40**(9): p. 869-75.

350. Shimono, N., et al., *Hypervirulent mutant of Mycobacterium tuberculosis resulting from disruption of the mce1 operon.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15918-23.

351. Tsolaki, A.G., et al., *Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis.* J Clin Microbiol, 2005. **43**(7): p. 3185-91.

352. Mathema, B., et al., *Identification and evolution of an IS6110 low-copy-number Mycobacterium tuberculosis cluster.* J Infect Dis, 2002. **185**(5): p. 641-9.

353. Manca, C., et al., *Hypervirulent M. tuberculosis W/Beijing strains upregulate type I IFNs and increase expression of negative regulators of the Jak-Stat pathway.* J Interferon Cytokine Res, 2005. **25**(11): p. 694-701.

354. Izzo, A., et al., *NIH pre-clinical screening program: overview and current status.* Tuberculosis (Edinb), 2005. **85**(1-2): p. 25-8.

355. Kaplan, E.L. and P. Meier, *Nonparametric Estimation from Incomplete Observations.* Journal of the American Statistical Association, 1958. **53**(282): p. 457-481.

356. Mantel, N., *Evaluation of survival data and two new rank order statistics arising in its consideration.* Cancer Chemother Rep, 1966. **50**(3): p. 163-70.

357. Slayden, R.A.B.r., C.E., *Analysis of the Lipids of Mycobacterium tuberculosis*, in *Mycobacterium tuberculosis Protocols* T.S. Parish, N.G. , Editor. 2001, Humana Press Inc: Towata NJ. p. 229-246.

358. Besra, G.S., *Preparation of Cell-Wall Fractions from Mycobacteria*, in *Methods in Molecular Biology, Vol 101: Mycobacterial Protocols*, T.S. Parish, N.G., Editor. 1998, Humana Press Inc: Towata NJ. p. 91-107.

359. Hunter, R.L., N. Venkataprasad, and M.R. Olsen, *The role of trehalose dimycolate (cord factor) on morphology of virulent M. tuberculosis in vitro.* Tuberculosis (Edinb), 2006. **86**(5): p. 349-56.

360. Boshoff, H.I. and C.E. Barry, 3rd, *Tuberculosis - metabolism and respiration in the absence of growth.* Nat Rev Microbiol, 2005. **3**(1): p. 70-80.

361. Gil, O., et al., *Intragranulomatous necrosis in lungs of mice infected by aerosol with Mycobacterium tuberculosis is related to bacterial load rather than to any one cytokine or T cell type.* Microbes Infect, 2006. **8**(3): p. 628-36.

362. McMurray, D.N., et al., *Pathogenesis of experimental tuberculosis in animal models.* Curr Top Microbiol Immunol, 1996. **215**: p. 157-79.

363. Baldwin, S.L., et al., *Evaluation of new vaccines in the mouse and guinea pig model of tuberculosis.* Infect Immun, 1998. **66**(6): p. 2951-9.

364. Cardona, P.J., et al., *Evolution of granulomas in lungs of mice infected aerogenically with Mycobacterium tuberculosis.* Scand J Immunol, 2000. **52**(2): p. 156-63.

365. Lefford, M., *Diseases in mice and rats*, in *The Mycobacteria: A Source Book*, G.P.W. Kubica, L.G. , Editor. 1984, Marcel Dekker Inc: New York. p. 947-977.

366. Turner, O.C., R.J. Basaraba, and I.M. Orme, *Immunopathogenesis of pulmonary granulomas in the guinea pig after infection with Mycobacterium tuberculosis.* Infect Immun, 2003. **71**(2): p. 864-71.

367. Grange, J.M., *Immunophysiology and immunopathology of tuberculosis*, in *Clinical Tuberculosis*, P.D.O. Davies, Editor. 1998, Chapman & Hall,: London. p. 113-127.

368. Larsen, M.H., et al., *Efficacy and safety of live attenuated persistent and rapidly cleared Mycobacterium tuberculosis vaccine candidates in non-human primates.* Vaccine, 2009. **27**(34): p. 4709-17.

369. Pal, P.G. and M.A. Horwitz, *Immunization with extracellular proteins of Mycobacterium tuberculosis induces cell-mediated immune responses and substantial protective immunity in a guinea pig model of pulmonary tuberculosis.* Infect Immun, 1992. **60**(11): p. 4781-92.

370. Hamasur, B., et al., *Mycobacterium tuberculosis arabinomannan-protein conjugates protect against tuberculosis.* Vaccine, 2003. **21**(25-26): p. 4081-93.

371. Basaraba, R.J., et al., *Decreased survival of guinea pigs infected with Mycobacterium tuberculosis after multiple BCG vaccinations.* Vaccine, 2006. **24**(3): p. 280-6.

372. Olsen, A.W., et al., *Protective effect of a tuberculosis subunit vaccine based on a fusion of antigen 85B and ESAT-6 in the aerosol guinea pig model.* Infect Immun, 2004. **72**(10): p. 6148-50.

373. Palanisamy, G.S., et al., *Disseminated disease severity as a measure of virulence of Mycobacterium tuberculosis in the guinea pig model.* Tuberculosis (Edinb), 2008. **88**(4): p. 295-306.

374. Palanisamy, G.S., et al., *Clinical strains of Mycobacterium tuberculosis display a wide range of virulence in guinea pigs.* Tuberculosis (Edinb), 2009. **89**(3): p. 203-9.

375. Martin, C., et al., *The live Mycobacterium tuberculosis phoP mutant strain is more attenuated than BCG and confers protective immunity against tuberculosis in mice and guinea pigs.* Vaccine, 2006. **24**(17): p. 3408-19.

376. North, R.J., et al., *Growth rate of mycobacteria in mice as an unreliable indicator of mycobacterial virulence.* Infect Immun, 1999. **67**(10): p. 5483-5.

377. Kaushal, D., et al., *Reduced immunopathology and mortality despite tissue persistence in a Mycobacterium tuberculosis mutant lacking alternative sigma factor, SigH.* Proc Natl Acad Sci U S A, 2002. **99**(12): p. 8330-5.

378. Medina, E., et al., *Superior virulence of Mycobacterium bovis over Mycobacterium tuberculosis (Mtb) for Mtb-resistant and Mtb-susceptible mice is manifest as an ability to cause extrapulmonary disease.* Tuberculosis (Edinb), 2006. **86**(1): p. 20-7.

379. Sun, R., et al., *Mycobacterium tuberculosis ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set.* Mol Microbiol, 2004. **52**(1): p. 25-38.

380. Michiels, S., et al., *Meta-analysis when only the median survival times are known: a comparison with individual patient data results.* Int J Technol Assess Health Care, 2005. **21**(1): p. 119-25.

381. Zhang, M., et al., *Enhanced capacity of a widespread strain of Mycobacterium tuberculosis to grow in human macrophages.* J Infect Dis, 1999. **179**(5): p. 1213-7.

382. Hiromatsu, K., et al., *Induction of CD1-restricted immune responses in guinea pigs by immunization with mycobacterial lipid antigens.* J Immunol, 2002. **169**(1): p. 330-9.

383. Barry, C.E., 3rd, *Interpreting cell wall 'virulence factors' of Mycobacterium tuberculosis.* Trends Microbiol, 2001. **9**(5): p. 237-41.

384. D'Ascenzo, M., L. Choe, and K.H. Lee, *iTRAQPak: an R based analysis and visualization package for 8-plex isobaric protein expression data.* Brief Funct Genomic Proteomic, 2008. **7**(2): p. 127-135.

385. Chen, Y., et al., *Differential expression of novel tyrosine kinase substrates during breast cancer development.* Mol Cell Proteomics, 2007. **6**(12): p. 2072-87.

386. Gagne, J.P., et al., *Comparative proteome analysis of human epithelial ovarian cancer.* Proteome Sci, 2007. **5**: p. 16.

387. Drummelsmith, J., et al., *Comparative proteomics analyses reveal a potential biomarker for the detection of vancomycin-intermediate Staphylococcus aureus strains.* J Proteome Res, 2007. **6**(12): p. 4690-702.

388. Lenco, 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**(10): p. 2875-82.

389. Zeaiter, Z., et al., *CagA of Helicobacter pylori alters the expression and cellular distribution of host proteins involved in cell signaling.* FEMS Microbiol Lett, 2008. **288**(2): p. 227-34.

390. Gu, S., et al., *Comprehensive Proteomic Profiling of the Membrane Constituents of a Mycobacterium tuberculosis Strain.* Mol Cell Proteomics, 2003. **2**(12): p. 1284-1296.

391. Takayama, K., et al., *Site of inhibitory action of isoniazid in the synthesis of mycolic acids in Mycobacterium tuberculosis.* J Lipid Res, 1975. **16**(4): p. 308-17.

392. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.

393. Zybailov, B., et al., *Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae.* J Proteome Res, 2006. **5**(9): p. 2339-47.

394. Neter, J., Wasserman, W., Kutner, M.H., Nachstheim, C., *Applied Linear Statistical Models.* 1996, McGraw-Hill Higher Education.

395. Mollenkopf, H.J., et al., *A dynamic two-dimensional polyacrylamide gel electrophoresis database: the mycobacterial proteome via Internet.* Electrophoresis, 1999. **20**(11): p. 2172-80.

396. Skjot, R.L., 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**(1): p. 214-20.

397. Berthet, F.X., et al., *A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10).* Microbiology, 1998. **144 (Pt 11)**: p. 3195-203.

398. Gan, C.S., et al., *Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ).* J Proteome Res, 2007. **6**(2): p. 821-7.

399. Weldingh, K., et al., *Two-dimensional electrophoresis for analysis of Mycobacterium tuberculosis culture filtrate and purification and characterization of six novel proteins.* Infect Immun, 1998. **66**(8): p. 3492-500.

400. Mattow, J., et al., *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**(4): p. 494-507.

401. Mattow, J., et al., *Identification of proteins from Mycobacterium tuberculosis missing in attenuated Mycobacterium bovis BCG strains.* Electrophoresis, 2001. **22**(14): p. 2936-46.

402. Betts, J.C., *Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis.* IUBMB Life, 2002. **53**(4-5): p. 239-42.

403. Scheler, C., et al., *Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS).* Electrophoresis, 1998. **19**(6): p. 918-27.

404. Salim, K., et al., *Identification of differentiating neural progenitor cell markers using shotgun isobaric tagging mass spectrometry.* Stem Cells Dev, 2006. **15**(3): p. 461-70.

405. Seshi, B., *An integrated approach to mapping the proteome of the human bone marrow stromal cell.* Proteomics, 2006. **6**(19): p. 5169-82.

406. Unwin, R.D., et al., *Quantitative proteomic analysis using isobaric protein tags enables rapid comparison of changes in transcript and protein levels in transformed cells.* Mol Cell Proteomics, 2005. **4**(7): p. 924-35.

407. Keshamouni, V.G., et al., *Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype.* J Proteome Res, 2006. **5**(5): p. 1143-54.

408. Song, X., et al., *iTRAQ experimental design for plasma biomarker discovery.* J Proteome Res, 2008. **7**(7): p. 2952-8.

409. Sophie Alvarez, et al., *Comprehensive analysis of the Brassica juncea root proteome in response to cadmium exposure by complementary proteomic approaches.* Proteomics, 2009. **9**(9): p. 2419-2431.

410. Nagai, S., et al., *Isolation and partial characterization of major protein antigens in the culture fluid of Mycobacterium tuberculosis.* Infect Immun, 1991. **59**(1): p. 372-82.

411. Wiker, H.G., M. Harboe, and S. Nagai, *A localization index for distinction between extracellular and intracellular antigens of Mycobacterium tuberculosis.* J Gen Microbiol, 1991. **137**(4): p. 875-84.

412. Hewinson, R.G., et al., *Molecular characterization of MPT83: a seroreactive antigen of Mycobacterium tuberculosis with homology to MPT70.* Scand J Immunol, 1996. **43**(5): p. 490-9.

413. Charlet, D., 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**(5): p. 1302-13.

414. Said-Salim, B., et al., *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**(5): p. 1251-63.

415. Cai, H., et al., *A combined DNA vaccine-prime, BCG-boost strategy results in better protection against Mycobacterium bovis challenge.* DNA Cell Biol, 2006. **25**(8): p. 438-47.

416. Pearce, M.J., et al., *Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis.* Science, 2008. **322**(5904): p. 1104-7.

417. Darwin, K.H., et al., *The proteasome of Mycobacterium tuberculosis is required for resistance to nitric oxide.* Science, 2003. **302**(5652): p. 1963-6.

418. Pearce, M.J., et al., *Identification of substrates of the Mycobacterium tuberculosis proteasome.* EMBO J, 2006. **25**(22): p. 5423-32.

419. Rosenkrands, I., 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**(6): p. 2728-35.

420. Jayanthi Bai, N., et al., *Pathways of carbohydrate metabolism in mycobacterium tuberculosis H37Rv1.* Can J Microbiol, 1975. **21**(11): p. 1688-91.

421. Munoz-Elias, E.J. and J.D. McKinney, *Mycobacterium tuberculosis isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence.* Nat Med, 2005. **11**(6): p. 638-44.

422. Fan, F., et al., *Steady-state and pre-steady-state kinetic analysis of Mycobacterium smegmatis cysteine ligase (MshC).* Biochemistry, 2007. **46**(40): p. 11421-9.

423. Schelle, M.W. and C.R. Bertozzi, *Sulfate metabolism in mycobacteria.* Chembiochem, 2006. **7**(10): p. 1516-24.

424. Zolotarev, A.S., et al., *Increased sulfate uptake by E. coli overexpressing the SLC26-related SulP protein Rv1739c from Mycobacterium tuberculosis.* Comp Biochem Physiol A Mol Integr Physiol, 2008. **149**(3): p. 255-66.

425. Braun, V., *Iron uptake mechanisms and their regulation in pathogenic bacteria.* Int J Med Microbiol, 2001. **291**(2): p. 67-79.

426. Monfeli, R.R. and C. Beeson, *Targeting iron acquisition by Mycobacterium tuberculosis.* Infect Disord Drug Targets, 2007. **7**(3): p. 213-20.

427. De Voss, J.J., et al., *The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages.* Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1252-7.

428. Gold, B., et al., *The Mycobacterium tuberculosis IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages.* Mol Microbiol, 2001. **42**(3): p. 851-65.

429. Arcus, V.L., P.B. Rainey, and S.J. Turner, *The PIN-domain toxin-antitoxin array in mycobacteria.* Trends Microbiol, 2005. **13**(8): p. 360-5.

430. Gupta, A., *Killing activity and rescue function of genome-wide toxin-antitoxin loci of Mycobacterium tuberculosis.* FEMS Microbiol Lett, 2009. **290**(1): p. 45-53.

431. Robson, J., et al., *The vapBC Operon from Mycobacterium smegmatis Is An Autoregulated Toxin-Antitoxin Module That Controls Growth via Inhibition of Translation.* J Mol Biol, 2009.

432. Gerdes, K., S.K. Christensen, and A. Lobner-Olesen, *Prokaryotic toxin-antitoxin stress response loci.* Nat Rev Microbiol, 2005. **3**(5): p. 371-82.

433. Pederson, K.J., et al., *Intracellular localization modulates targeting of ExoS, a type III cytotoxin, to eukaryotic signalling proteins.* Mol Microbiol, 2002. **46**(5): p. 1381-90.

434. Korch, S.B., H. Contreras, and J.E. Clark-Curtiss, *Three Mycobacterium tuberculosis Rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages.* J Bacteriol, 2009. **191**(5): p. 1618-30.

435. Braunstein, M., et al., *SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium tuberculosis.* Mol Microbiol, 2003. **48**(2): p. 453-64.

436. Wheeler, P.R., Blanchard, J. S., *General Metabolism and Biochemical Pathways of Tubercle Bacilli*, in *Tuberculosis and the Tubercle Bacillus*, S.T. Cole, Eisenach, K. D., McMurray, D. N., Jacobs, W. R., Editor. 2005, American Society for Microbiology: Washington, DC. p. 309-340.

437. Vandal, O.H., et al., *Acid-susceptible mutants of Mycobacterium tuberculosis share hypersusceptibility to cell wall and oxidative stress and to the host environment.* J Bacteriol, 2009. **191**(2): p. 625-31.

438. Harth, G. and M.A. Horwitz, *An inhibitor of exported Mycobacterium tuberculosis glutamine synthetase selectively blocks the growth of pathogenic mycobacteria in axenic culture and in human monocytes: extracellular proteins as potential novel drug targets.* J Exp Med, 1999. **189**(9): p. 1425-36.

439. Harth, G., et al., *All four Mycobacterium tuberculosis glnA genes encode glutamine synthetase activities but only GlnA1 is abundantly expressed and essential for bacterial homeostasis.* Mol Microbiol, 2005. **58**(4): p. 1157-72.

440. Tullius, M.V., G. Harth, and M.A. Horwitz, *Glutamine synthetase GlnA1 is essential for growth of Mycobacterium tuberculosis in human THP-1 macrophages and guinea pigs.* Infect Immun, 2003. **71**(7): p. 3927-36.

441. Ribeiro-Guimaraes, M.L. and M.C. Pessolani, *Comparative genomics of mycobacterial proteases.* Microb Pathog, 2007. **43**(5-6): p. 173-8.

442. Skeiky, Y.A., et al., *Cloning, expression, and immunological evaluation of two putative secreted serine protease antigens of Mycobacterium tuberculosis.* Infect Immun, 1999. **67**(8): p. 3998-4007.

443. Travis, J., J. Potempa, and H. Maeda, *Are bacterial proteinases pathogenic factors?* Trends Microbiol, 1995. **3**(10): p. 405-7.

444. Malen, H., T. Softeland, and H.G. Wiker, *Antigen analysis of Mycobacterium tuberculosis H37Rv culture filtrate proteins.* Scand J Immunol, 2008. **67**(3): p. 245-52.

445. Alvarez, S., et al., *Comprehensive analysis of the Brassica juncea root proteome in response to cadmium exposure by complementary proteomic approaches.* Proteomics, 2009. **9**(9): p. 2419-2431.

446. Thon, J.N., et al., *Comprehensive proteomic analysis of protein changes during platelet storage requires complementary proteomic approaches.* Transfusion, 2008. **48**(3): p. 425-35.

447. Leversen, N.A., et al., *Evaluation of signal peptide prediction algorithms for identification of mycobacterial signal peptides using sequence data from proteomic methods.* Microbiology, 2009. **155**(Pt 7): p. 2375-83.

448. Raghavan, S., et al., *Secreted transcription factor controls Mycobacterium tuberculosis virulence.* Nature, 2008. **454**(7205): p. 717-21.

449. Lyashchenko, K., et al., *Heterogeneous antibody responses in tuberculosis.* Infect Immun, 1998. **66**(8): p. 3936-40.

450. Webb, J.R., et al., *Molecular cloning, expression, and immunogenicity of MTB12, a novel low-molecular-weight antigen secreted by Mycobacterium tuberculosis.* Infect Immun, 1998. **66**(9): p. 4208-14.

451. Lee, J.S., et al., *Ex vivo responses for interferon-gamma and proinflammatory cytokine secretion to low-molecular-weight antigen MTB12 of Mycobacterium tuberculosis during human tuberculosis.* Scand J Immunol, 2006. **64**(2): p. 145-54.

452. Wiker, H.G., M.A. Wilson, and G.K. Schoolnik, *Extracytoplasmic proteins of Mycobacterium tuberculosis - mature secreted proteins often start with aspartic acid and proline.* Microbiology, 2000. **146 (Pt 7)**: p. 1525-33.

453. Tjalsma, H., et al., *Proteomics of protein secretion by Bacillus subtilis: separating the "secrets" of the secretome.* Microbiol Mol Biol Rev, 2004. **68**(2): p. 207-33.

454. Tjalsma, H., et al., *Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome.* Microbiol Mol Biol Rev, 2000. **64**(3): p. 515-47.

455. van Roosmalen, M.L., et al., *Type I signal peptidases of Gram-positive bacteria.* Biochim Biophys Acta, 2004. **1694**(1-3): p. 279-97.

456. Perlman, D. and H.O. Halvorson, *A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides.* J Mol Biol, 1983. **167**(2): p. 391-409.

457. Tuteja, R., *Type I signal peptidase: an overview.* Arch Biochem Biophys, 2005. **441**(2): p. 107-11.

458. von Heijne, G., *Signal sequences. The limits of variation.* J Mol Biol, 1985. **184**(1): p. 99-105.

459. Choo, K.H. and S. Ranganathan, *Flanking signal and mature peptide residues influence signal peptide cleavage.* BMC Bioinformatics, 2008. **9 Suppl 12**: p. S15.

460. Kajava, A.V., et al., *Processing of Escherichia coli alkaline phosphatase. Sequence requirements and possible conformations of the -6 to -4 region of the signal peptide.* J Biol Chem, 2002. **277**(52): p. 50396-402.

461. Klepp, L.I., et al., *Identification of two proteins that interact with the Erp virulence factor from Mycobacterium tuberculosis by using the bacterial two-hybrid system.* BMC Mol Biol, 2009. **10**: p. 3.

462. O'Hare, H., et al., *A split-protein sensor for studying protein-protein interaction in mycobacteria.* J Microbiol Methods, 2008. **73**(2): p. 79-84.

463. Raman, K. and N. Chandra, *Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance.* BMC Microbiol, 2008. **8**: p. 234.

464. Villarino, A., et al., *Proteomic identification of M. tuberculosis protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions.* J Mol Biol, 2005. **350**(5): p. 953-63.

465. Ausubel, F.M., Brent, R., Kingston R.E., Moore, D.D., Smith J.A., Seidman J.G., Struhl, K., *Current protocols in molecular biology.* 1988, Brooklyn, N.Y.: Greene Publishing Associates.

466. van Soolingen, D., et al., *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**(11): p. 2578-86.

467. Roback, P., et al., *A predicted operon map for Mycobacterium tuberculosis.* Nucleic Acids Res, 2007. **35**(15): p. 5085-95.

468. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

469. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden markov model: application to complete genomes.* Journal of Molecular Biology, 2001. **305**(3): p. 567-580.

470. Sonnhammer, E.L., G. von Heijne, and A. Krogh, *A hidden Markov model for predicting transmembrane helices in protein sequences.* Proc Int Conf Intell Syst Mol Biol, 1998. **6**: p. 175-82.

471. Cserzo, M., et al., *Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method.* Protein Eng, 1997. **10**(6): p. 673-6.

472. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* Journal of Molecular Biology, 1982. **157**(1): p. 105-132.

473. Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0.* J Mol Biol, 2004. **340**(4): p. 783-95.

474. Nielsen, H., et al., *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.* Protein Eng, 1997. **10**(1): p. 1-6.

475. Nielsen, H. and A. Krogh, *Prediction of signal peptides and signal anchors by a hidden Markov model.* Proc Int Conf Intell Syst Mol Biol, 1998. **6**: p. 122-30.

476. Belisle, J.T., et al., *Role of the major antigen of Mycobacterium tuberculosis in cell wall biogenesis.* Science, 1997. **276**(5317): p. 1420-2.

477. Koul, A., et al., *Cloning and characterization of secretory tyrosine phosphatases of Mycobacterium tuberculosis.* J Bacteriol, 2000. **182**(19): p. 5425-32.

478. Singh, J., M.C. Joshi, and R. Bhatnagar, *Cloning and expression of mycobacterial glutamine synthetase gene in Escherichia coli.* Biochem Biophys Res Commun, 2004. **317**(2): p. 634-8.

479. Jiang, X.Y., et al., *Cloning and expression of Mycobacterium bovis secreted protein MPB83 in Escherichia coli.* J Biochem Mol Biol, 2006. **39**(1): p. 22-5.

480. Shine, J. and L. Dalgarno, *Determinant of cistron specificity in bacterial ribosomes.* Nature, 1975. **254**(5495): p. 34-8.

481. Chubb, A.J., et al., *Identification of Mycobacterium tuberculosis signal sequences that direct the export of a leaderless beta-lactamase gene product in Escherichia coli.* Microbiology, 1998. **144 (Pt 6)**: p. 1619-29.

482. Dalbey, R.E., et al., *Understanding the insertion of transporters and other membrane proteins.* Curr Opin Cell Biol, 2000. **12**(4): p. 435-42.

483. Dalbey, R.E. and A. Kuhn, *Evolutionarily related insertion pathways of bacterial, mitochondrial, and thylakoid membrane proteins.* Annu Rev Cell Dev Biol, 2000. **16**: p. 51-87.

484. Andersson, H., E. Bakker, and G. von Heijne, *Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in Escherichia coli.* J Biol Chem, 1992. **267**(3): p. 1491-5.

485. Dalbey, R.E., *Positively charged residues are important determinants of membrane protein topology.* Trends Biochem Sci, 1990. **15**(7): p. 253-7.

486. Liu, G., T.B. Topping, and L.L. Randall, *Physiological role during export for the retardation of folding by the leader peptide of maltose-binding protein.* Proc Natl Acad Sci U S A, 1989. **86**(23): p. 9213-7.

487. Park, S., et al., *Modulation of folding pathways of exported proteins by the leader sequence.* Science, 1988. **239**(4843): p. 1033-5.

488. Rajalahti, T., et al., *Proteins in different Synechocystis compartments have distinguishing N-terminal features: a combined proteomics and multivariate sequence analysis.* J Proteome Res, 2007. **6**(7): p. 2420-34.

489. Bolhuis, A., et al., *Bacillus subtilis can modulate its capacity and specificity for protein secretion through temporally controlled expression of the sipS gene for signal peptidase I.* Mol Microbiol, 1996. **22**(4): p. 605-18.

490. Tjalsma, H., et al., *Bacillus subtilis contains four closely related type I signal peptidases with overlapping substrate specificities. Constitutive and temporally controlled expression of different sip genes.* J Biol Chem, 1997. **272**(41): p. 25983-92.

491. Shinde, U. and M. Inouye, *Intramolecular chaperones: polypeptide extensions that modulate protein folding.* Semin Cell Dev Biol, 2000. **11**(1): p. 35-44.

492. Poquet, I., et al., *HtrA is the unique surface housekeeping protease in Lactococcus lactis and is required for natural protein processing.* Mol Microbiol, 2000. **35**(5): p. 1042-51.

493. Gonzalo-Asensio, J., et al., *The Mycobacterium tuberculosis phoPR operon is positively autoregulated in the virulent strain H37Rv.* J Bacteriol, 2008. **190**(21): p. 7068-78.

494. Gu, S., et al., *Comprehensive proteomic profiling of the membrane constituents of a Mycobacterium tuberculosis strain.* Mol Cell Proteomics, 2003. **2**(12): p. 1284-96.

495. Li, A. and J. Piel, *A gene cluster from a marine Streptomyces encoding the biosynthesis of the aromatic spiroketal polyketide griseorhodin A.* Chem Biol, 2002. **9**(9): p. 1017-26.

496. Hyndman, D., et al., *The aldo-keto reductase superfamily homepage.* Chem Biol Interact, 2003. **143-144**: p. 621-31.

497. Ma, J., et al., *Aldo-keto Reductase Family 1 B10 Affects Fatty Acid Synthesis by Regulating the Stability of Acetyl-CoA Carboxylase-{alpha} in Breast Cancer Cells.* J. Biol. Chem., 2008. **283**(6): p. 3418-3423.

498. Wang, C., et al., *Aldo-keto reductase family 1 member B10 promotes cell survival by regulating lipid synthesis and eliminating carbonyls.* J. Biol. Chem., 2009: p. M109.022897.

499. Van Heeke, G. and S.M. Schuster, *Expression of human asparagine synthetase in Escherichia coli.* J Biol Chem, 1989. **264**(10): p. 5503-9.

500. Green, P. and M. Inouye, *In vivo and in vitro systems for studying bacterial membrane biogenesis.* Methods Enzymol, 1983. **96**: p. 74-84.

501. Herskovits, A.A., et al., *Accumulation of endoplasmic membranes and novel membrane-bound ribosome-signal recognition particle receptor complexes in Escherichia coli.* J Cell Biol, 2002. **159**(3): p. 403-10.

502. Randall, L.L. and A.J. Hardy, *Analysis of the ribosomes engaged in the synthesis of the outer membrane proteins of Escherichia coli.* Mol Gen Genet, 1975. **137**(2): p. 151-60.

503. Prinz, A., et al., *Evolutionarily conserved binding of ribosomes to the translocation channel via the large ribosomal RNA.* EMBO J, 2000. **19**(8): p. 1900-6.

504. Prinz, A., E. Hartmann, and K.U. Kalies, *Sec61p is the main ribosome receptor in the endoplasmic reticulum of Saccharomyces cerevisiae.* Biol Chem, 2000. **381**(9-10): p. 1025-9.

505. Walter, P. and A.E. Johnson, *Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane.* Annu Rev Cell Biol, 1994. **10**: p. 87-119.

506. Malik, S., et al., *Alleles of the NRAMP1 gene are risk factors for pediatric tuberculosis disease.* Proc Natl Acad Sci U S A, 2005. **102**(34): p. 12183-8.

507. Weiss, R.A. and A.J. McMichael, *Social and environmental risk factors in the emergence of infectious diseases.* Nat Med, 2004. **10**(12 Suppl): p. S70-6.

508. Mesfin, M.M., et al., *Delayed consultation among pulmonary tuberculosis patients: a cross sectional study of 10 DOTS districts of Ethiopia.* BMC Public Health, 2009. **9**: p. 53.

509. Madebo, T. and B. Lindtjorn, *Delay in Treatment of Pulmonary Tuberculosis: An Analysis of Symptom Duration Among Ethiopian Patients.* MedGenMed, 1999: p. E6.

510. Behr, M.A., et al., *Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli.* Lancet, 1999. **353**(9151): p. 444-9.

511. Brunham, R.C., F.A. Plummer, and R.S. Stephens, *Bacterial antigenic variation, host immune response, and pathogen-host coevolution.* Infect Immun, 1993. **61**(6): p. 2273-6.

512. Riley, R.L., et al., *Aerial dissemination of pulmonary tuberculosis. A two-year study of contagion in a tuberculosis ward. 1959.* Am J Epidemiol, 1995. **142**(1): p. 3-14.

513. Riley, R.L., et al., *Infectiousness of air from a tuberculosis ward. Ultraviolet irradiation of infected air: comparative infectiousness of different patients.* Am Rev Respir Dis, 1962. **85**: p. 511-25.

514. Escombe, A.R., et al., *The infectiousness of tuberculosis patients coinfected with HIV.* PLoS Med, 2008. **5**(9): p. e188.

515. Fennelly, K.P., et al., *Cough-generated aerosols of Mycobacterium tuberculosis: a new method to study infectiousness.* Am J Respir Crit Care Med, 2004. **169**(5): p. 604-9.

516. Pancholi, V. and V.A. Fischetti, *alpha-enolase, a novel strong plasmin(ogen) binding protein on the surface of pathogenic streptococci.* J Biol Chem, 1998. **273**(23): p. 14503-15.

517. Graybill, E.R., et al., *Functional comparison of citrate synthase isoforms from S. cerevisiae.* Arch Biochem Biophys, 2007. **465**(1): p. 26-37.

518. Srere, P.A. and M.L. John, *[1] Citrate synthase : [EC 4.1.3.7. Citrate oxaloacetate-lyase (CoA-acetylating)]*, in *Methods in Enzymology.* 1969, Academic Press. p. 3-11.

519. Bridger, W.A., et al., *[14] Succinyl coenzyme a synthetase from Escherichia coli : [EC 6.2.1.5 Succinate: CoA ligase (SDP)]*, in *Methods in Enzymology.* 1969, Academic Press. p. 70-75.

520. Sly, L.I. and H.W. Doelle, *6-Phosphogluconate dehydrogenase in cell free extracts of Escherichia coli K-12.* Archives of Microbiology, 1968. **63**(3): p. 214-223.

521. Kinhikar, A.G., et al., *Mycobacterium tuberculosis malate synthase is a laminin-binding adhesin.* Mol Microbiol, 2006. **60**(4): p. 999-1013.

522. Whiting, G.C., et al., *Purification of native alpha-enolase from Streptococcus pneumoniae that binds plasminogen and is immunogenic.* J Med Microbiol, 2002. **51**(10): p. 837-43.

523. Samanich, K., J.T. Belisle, and S. Laal, *Homogeneity of antibody responses in tuberculosis patients.* Infect Immun, 2001. **69**(7): p. 4600-9.

524. Stephenson, M.C. and C. Ratledge, *Iron transport in mycobacterium smegmatis: uptake of iron from Ferriexochelin.* J Gen Microbiol, 1979. **110**(1): p. 193-202.

525. Carrondo, M.A., *Ferritins, iron uptake and storage from the bacterioferritin viewpoint.* EMBO J, 2003. **22**(9): p. 1959-68.

526. Kitteringham, N.R., et al., *Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics.* Journal of Chromatography B, 2009. **877**(13): p. 1229-1239.

527. Yang, X. and I.M. Lazar, *MRM screening/biomarker discovery with linear ion trap MS: a library of human cancer-specific peptides.* BMC Cancer, 2009. **9**: p. 96.

528. Griffiths, J.R., et al., *The Application of a Hypothesis-driven Strategy to the Sensitive Detection and Location of Acetylated Lysine Residues.* Journal of the American Society for Mass Spectrometry, 2007. **18**(8): p. 1423-1428.

529. Latterich, M., M. Abramovitz, and B. Leyland-Jones, *Proteomics: New technologies and clinical applications.* European Journal of Cancer, 2008. **44**(18): p. 2737-2741.

530. Sahana Mollah, et al., *Targeted mass spectrometric strategy for global mapping of ubiquitination on proteins.* Rapid Communications in Mass Spectrometry, 2007. **21**(20): p. 3357-3364.

531. Lange, V., et al., *Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring.* Mol Cell Proteomics, 2008. **7**(8): p. 1489-500.

532. Blonder, J. and T.D. Veenstra, *Computational prediction of proteotypic peptides.* Expert Rev Proteomics, 2007. **4**(3): p. 351-4.

533. Mallick, P., et al., *Computational prediction of proteotypic peptides for quantitative proteomics.* Nat Biotechnol, 2007. **25**(1): p. 125-31.

534. Anderson, L. and C.L. Hunter, *Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins.* Mol Cell Proteomics, 2006. **5**(4): p. 573-88.

535. Sharma, D., et al., *Expression of mycobacterial cell division protein, FtsZ, and dormancy proteins, DevR and Acr, within lung granulomas throughout guinea pig infection.* FEMS Immunol Med Microbiol, 2006. **48**(3): p. 329-36.

536. McCabe, A., et al., *Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis.* J Natl Cancer Inst, 2005. **97**(24): p. 1808-15.

537. Fenhalls, G., et al., *In situ detection of Mycobacterium tuberculosis transcripts in human lung granulomas reveals differential gene expression in necrotic lesions.* Infect Immun, 2002. **70**(11): p. 6330-8.

538. Harth, G., et al., *Targeting the Mycobacterium tuberculosis 30/32-kDa mycolyl transferase complex as a therapeutic strategy against tuberculosis: Proof of principle by using antisense technology.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15614-9.

539. Harth, G. and M.A. Horwitz, *Inhibition of Mycobacterium tuberculosis glutamine synthetase as a novel antibiotic strategy against tuberculosis: demonstration of efficacy in vivo.* Infect Immun, 2003. **71**(1): p. 456-64.

540. Klotzsche, M., S. Ehrt, and D. Schnappinger, *Improved tetracycline repressors for gene silencing in mycobacteria.* Nucleic Acids Res, 2009. **37**(6): p. 1778-88.

541. Gandotra, S., et al., *In vivo gene silencing identifies the Mycobacterium tuberculosis proteasome as essential for the bacteria to persist in mice.* Nat Med, 2007. **13**(12): p. 1515-20.

542. Gustavsson, N., A. Diez, and T. Nystrom, *The universal stress protein paralogues of Escherichia coli are co-ordinately regulated and co-operate in the defence against DNA damage.* Mol Microbiol, 2002. **43**(1): p. 107-17.

543. Nystrom, T. and F.C. Neidhardt, *Expression and role of the universal stress protein, UspA, of Escherichia coli during growth arrest.* Mol Microbiol, 1994. **11**(3): p. 537-44.

544. Yohannes, E., D.M. Barnhart, and J.L. Slonczewski, *pH-dependent catabolic protein expression during anaerobic growth of Escherichia coli K-12.* J Bacteriol, 2004. **186**(1): p. 192-9.

545. Liu, W.T., et al., *Role of the universal stress protein UspA of Salmonella in growth arrest, stress and virulence.* Microb Pathog, 2007. **42**(1): p. 2-10.

546. Darwin, A.J., *The phage-shock-protein response.* Mol Microbiol, 2005. **57**(3): p. 621-8.

547. Kleerebezem, M., W. Crielaard, and J. Tommassen, *Involvement of stress protein PspA (phage shock protein A) of Escherichia coli in maintenance of the protonmotive force under stress conditions.* EMBO J, 1996. **15**(1): p. 162-71.

548. Darwin, A.J. and V.L. Miller, *The psp locus of Yersinia enterocolitica is required for virulence and for growth in vitro when the Ysc type III secretion system is produced.* Mol Microbiol, 2001. **39**(2): p. 429-44.

549. Hassani, A.S., N. Amirmozafari, and A. Ghaemi, *Virulence Increasing of Salmonella typhimurium in Balb/c Mice After Heat-Stress Induction of Phage Shock Protein A.* Curr Microbiol, 2009.

550. Kobayashi, R., T. Suzuki, and M. Yoshida, *Escherichia coli phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes.* Mol Microbiol, 2007. **66**(1): p. 100-9.

551. Goulding, C.W., et al., *The structure and computational analysis of Mycobacterium tuberculosis protein CitE suggest a novel enzymatic function.* J Mol Biol, 2007. **365**(2): p. 275-83.

552. Kanao, T., et al., *Kinetic and biochemical analyses on the reaction mechanism of a bacterial ATP-citrate lyase.* Eur J Biochem, 2002. **269**(14): p. 3409-16.

553. Chatfield, C.H., H. Koo, and R.G. Quivey, Jr., *The putative autolysin regulator LytR in Streptococcus mutans plays a role in cell division and is growth-phase regulated.* Microbiology, 2005. **151**(Pt 2): p. 625-31.

554. Johnsborg, O. and L.S. Havarstein, *Pneumococcal LytR, a protein from the LytR-CpsA-Psr family, is essential for normal septum formation in Streptococcus pneumoniae.* J Bacteriol, 2009.

555. Denham, E.L., P.N. Ward, and J.A. Leigh, *Lipoprotein signal peptides are processed by Lsp and Eep of Streptococcus uberis.* J Bacteriol, 2008. **190**(13): p. 4641-7.

556. Ye, S.Y., et al., *NisP is related to nisin precursor processing and possibly to immunity in Lactococcus lactis.* J Tongji Med Univ, 1995. **15**(4): p. 193-7.

557. Kuipers, A., et al., *Sec-mediated transport of posttranslationally dehydrated peptides in Lactococcus lactis.* Appl Environ Microbiol, 2006. **72**(12): p. 7626-33.

558. Zhang, B., et al., *From pull-down data to protein interaction networks and complexes with biological relevance.* Bioinformatics, 2008. **24**(7): p. 979-86.

List of Abbreviations

2D-GE:	Two dimensional gel electrophoresis
AFB:	Acid fast bacilli
AG:	Arabinogalactan
AIDS:	Acquired Immuno-Deficiency Syndrome
AM:	Arabinomannan
BCG:	Bacille Calmette-Guérin
CE:	Collision energy
CFU:	Colony forming units
CMI:	Cell-mediated immunity
CID:	Collision-induced dissociation
CFP:	Culture Filtrate Proteins
DC:	Dendritic cells
DIGE:	Differential gel electrophoresis
DOTS:	Direct observed treatment-short course
DTH:	Delayed type hypersensitivity
EDTA:	Ethylen-ediamine-tetra acetic acid
ELISA:	Enzyme-linked immunosorbent assay
ELISPOT:	Enzyme-linked immunospot
EMB:	Ethambutol
ESI:	Electrospray ionization
FDA:	Federal Drug Association
GAS:	Glycerol Alanine Salts
HIV:	Human Immunodeficiency Virus
HMP:	Hexose monophosphate
iCAT:	Isotope-Coded- Affinity Tags
IEF:	Isoelectric focusing
IFN-γ:	Interferon gamma
IGRA:	Interferon gamma release assay
INH:	Isoniazid
IL:	Interleukin
IP:	Isoelectric point
iTRAQ:	Isobaric tag for relative and absolute quantification
LAM:	Lipoarabinomannan
LTQ:	Linear trap-quadrupole
MAC:	*Mycobacterium avium* complex
mAGP:	Mycolyl-arabinogalactan-peptidoglycan
man-LAM:	Mannan capped-Lipoarabinomannan

MALDI: Matrix-assisted laser desorption/ionization
MDR-TB: Multi-drug resistant tuberculosis
MIRU: Mycobacterial interspersed repetitive units
MΦ: Macrophage
Mr: Molecular weight
MRM: Multiple reaction monitoring
MS: Mass spectrometry
MS/MS: Tandem mass spectrometry
***Mtb*:** *Mycobacterium tuberculosis*
MTC: *Mycobacterium tuberculosis* complex
MWCO: Molecular weight cut-off
NAAT: Nucleic acid amplification test
NK: Natural Killer cells
NO: Nitric oxide
NTM: Non-tuberculous Mycobacteria
OADC: Oleic acid Albumin Dextrose Catalase
PAMPs: Pathogen-associated molecular patterns
PBMC: Peripheral blood mononuclear cell
PBS: Phosphate buffered saline
PCR: Polymerase chain reaction
PDIM: Phthiocerol dimycocerosate
PG: Peptidoglycan
PGL-Tb: Phenolic glycolipid of *Mycobacterium tuberculosis*
PIM: Phosphotidyl-myo-inositol mannan
PPD: Purified protein derivative
PQD: Pulsed- Q-dissociation
PRRs: Pattern recognition receptors
PTM: Post-translational modifications
PZA: Pyrazinamide
Q-TOF: Quadrupole Time-of-Flight
RD: Region of difference
RF: Radio frequency
RFLP: Restriction fragment length polymorphism
RIF: Rifampin
RNI: Reactive nitrogen intermediates
ROS: Reactive oxygen species
SDS-PAGE: Sodium dodecyl sulfate- polyacrilamide gel electrophoresis
SILAC: Stable isotope labeling with amino acids in culture
SL: Sulfolipid
SNP: Single nucleotide polymorphism
TB: Tuberculosis
TCA: Tricarboxylic acid cycle
TDM: Trehalose dimicolate mannoside
TFA: Trifluoroacetic acid
TLC: Thin layer chromatography
TLR: Toll-like receptors

TNF-α:	Tumor necrosis factor - alpha
TST:	Tuberculin skin test
VNTR:	Variable number tandem repeats
WHO:	World Health Organization
XDR-TB:	Extensively drug resistant tuberculosis