Unveiling phenotypic differences among the genetically conserved *Mycobacterium tuberculosis* complex: Characterization of proteomes and post-translational modifications

Alemayehu Godana Birhanu

Thesis for the degree of Philosophiae Doctor (PhD)

UiO : Faculty of Medicine
University of Oslo

Institute of Clinical Medicine
Department of Microbiology
University of Oslo
January 2019

Acknowledgements

I want to express my gratitude to my supervisor Professor Tone Tønjum for letting me be part of her research group and for creating such a stimulating work environment. I am very grateful for her scientific inputs, encouragement and support throughout the whole study period.

I would also like to thank my co-supervisor Dr. Markos Abebe for his role in establishing the collaboration and his valuable inputs to my papers and thesis.

I wish to thank to all the co-authors for their invaluable contributions in my papers. My gratitude also goes to all the members of the Genome Dynamics group for being nice to me and for creating an excellent working environment.

My sincere thanks to Dr. Solomon Abebe, Dr. Getachew Tesfaye and Dr. Ephrem Debebe for their feedback on the dissertation.

It is a pleasure to thank my friends Mequannent, Meskerem, Addis and Hiwote for their nice friendship and making my stay in Oslo enjoyable.

I would like to thank the Norwegian State Educational Loan Fund (Lånekassen) for the financial support to pursue my study. I am grateful to the Pasteur Legacy and EMBO for supporting my study.

To my many friends and family, you should know that your support and encouragement worth more than I can express on paper.

Finally, my deep and sincere gratitude to my wife, Beselam Tadesse, for her continuous and unparalleled love, unconditional support, understanding and never ending patience. To my son Haniel and daughter Yohanna, you have made me stronger, better and more fulfilled than I could have ever imagined.

Alemayehu Godana Birhanu

Oslo, January 2019

TABLE OF CONTENTS

LIST OF PAPERS

This dissertation is based on the following papers, which will be referred to by their Roman numbers.

I. Yimer SA, **Birhanu AG**, Kalayou S, Riaz T, Zegeye ED, Beyene GT, Holm-Hansen C, Norheim G, Abebe M, Aseffa A, Tønjum T. Comparative Proteomic Analysis of *Mycobacterium tuberculosis* Lineage 7 and Lineage 4 Strains Reveals Differentially Abundant Proteins Linked to Slow Growth and Virulence. Frontiers in Microbiology. 2017;8.

II. **Birhanu AG**, Yimer SA, Holm-Hansen C, Norheim G, Aseffa A, Abebe M, Tønjum T. Nε- and O-Acetylation in *Mycobacterium tuberculosis* lineage 7 and lineage 4 Strains: Proteins Involved in Bioenergetics, Virulence, and Antimicrobial Resistance Are Acetylated. Journal of proteome research. 2017 Oct 4;16(11):4045-59.

III. **Birhanu AG**, Yimer SA, Kalayou S, Riaz T, Zegeye ED, Holm-Hansen C, Norheim G, Aseffa A, Abebe M, Tønjum T. Ample glycosylation in membrane and cell envelope proteins may explain the phenotypic diversity and virulence in the *Mycobacterium tuberculosis* complex. *Scientific Reports* 2019, 9(1):2927.

ABBREVIATIONS

AMR	antimicrobial resistance
BCG	bacilli Calmette-Guérin strain
CR	complement receptors
CTLs	cytotoxic $CD8^+$ T lymphocytes
DC	dendritic cell
DC-SIGN	DC-specific intercellular adhesion molecule (ICAM)-3-grabbing non-integrin
DR	direct repeat
GlcNAc	N-acetylglucosamine
IFN-γ	interferon-gamma
iNOS	inducible nitric oxide synthase
IS	insertion sequence
IS6110-RFLP	insertion sequence (IS) 6110 restriction fragment length polymorphism
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LAM	lipoarabinomannan
LM	lipomannan
mAGP	mycolyl arabinogalactan-peptidoglycan
ManLAM	mannose-capped lipoarabinomannan
Mce	mammalian cell entry protein
MDR	multidrug-resistance
MHC	major histocompatibility complex
MIRU-VNTR	mycobacterial interspersed repetitive-unit-variable-number tandem repeat
MR	mannose receptor
MØ	macrophage
Mtb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
MurNAc	N-acylated muramic acid
NK	natural killer
NLRs	nucleotide binding and oligomerization domain (NOD-) like receptors
NOD	nucleotide binding and oligomerization domain
NOX2	NADPH oxidase
PAMPs	pathogen-associated molecular patterns
PDIM	phthiocerol dimycocerosate

PG	peptidoglycan
PGL	phenolglycolipid
PIM	phosphatidyl-myo-inositol mannosides
PRRs	pattern recognition receptors
PTMs	post-translational modifications
RD	region of difference
ROI	reactive oxygen intermediates
RNI	reactive nitrogen intermediates
SNP	single nucleotide polymorphism
TB	tuberculosis
TDM	trehalose-6, 6-dimycolate (cord factor)
TLRs	toll-like receptors
TNF-α	tumour necrosis factor-alpha
WGS	whole genome sequencing
XDR	extensively drug-resistance

SUMMARY

The *Mycobacterium tuberculosis* complex (MTBC) is the causative agent of tuberculosis (TB), one of the world's deadliest communicable diseases with 10 million new cases and 1.6 million deaths in 2017. Despite the high degree of genomic conservation among members of the MTBC, TB infections exhibit vast discrepancies in clinical outcome and epidemiological behaviour. Bacterial factors that contribute to these phenotypic variability remain elusive and cannot be explained by MTBC genomics alone.

We hypothesize that qualitative and quantitative differences in the abundance of proteins and post-translational modifications (PTMs) may explain the phenotypic diversity across members of the MTBC. This study analyzed the proteomic and PTM (acetylation and glycosylation) profiles among members of the MTBC. These two PTMs are reported to be important determinants of MTBC virulence and pathogenicity.

Shotgun-based proteomics yielded the identification of 2867 MTBC proteins. Quantitative proteomic analysis of the MTBC reference strain H37Rv and slow-growing lineage 7 strains demonstrated a differential abundance of proteins involved in virulence growth and bioenergetics. The acetylome analysis identified 2490 class-I acetylation sites on 953 proteins. The MTBC proteins found to be acetylated are involved in core metabolic processes, bioenergetics, virulence, and drug resistance. Notably, O-acetylation of MTBC was described for the first time. Quantitative PTM analysis revealed reduced acetylation on 97.5% of the differentially acetylated virulence factors in MTBC lineage 7 strains as compared to H37Rv. The glycoproteomic analysis of clinical MTBC strains representing lineages 3, 4, 5 and 7 identified 2944 glycosylation events on 1325 proteins. These proteins are involved in MTBC cell envelope biogenesis, pathogen-host interaction, membrane transport, and pathogenesis. The study also provides the first report on N-linked glycosylation in MTBC and in Gram-positive bacteria. Quantitative analysis revealed differential glycosylation of 67 proteins involved in MTBC fitness and survival. Identification of glycoproteins and their function contributes to a better understanding of the pathogenesis and survival strategies adopted by MTBC, which is fundamental to diseases management.

The data represents the highest number of acetylated and glycosylated proteins in MTBC recorded to date. The presence of significant differences in the proteome and PTMs among MTBC lineages may directly influence the strain phenotype. This will enable improved understanding in the adaptive potential of the pathogen with new potential for identification of novel drug targets, vaccine candidates and efficient TB diagnosis.

1. INTRODUCTION

1.1. The *Mycobacterium tuberculosis* complex

The *Mycobacterium tuberculosis* complex (MTBC) is a closely related group of *Mycobacterium* species that can cause tuberculosis (TB) in both humans and animals [1]. Human TB is principally caused by infection via the lung with the acid-fast bacillus *Mycobacterium tuberculosis* (Mtb), a pathogen identified by Robert Koch in 1882 [2]. TB is one of the top 10 causes of mortality worldwide and the leading cause of deaths from infectious diseases [3]. In 2017, there were an estimated 10 million new cases of TB causing 1.6 million deaths [3]. *M. bovis* has a wider host range and is the main cause of TB in other animal species. Other members of the MTBC that occasionally cause human infections include *M. canettii* and *M. africanum*, and those species-specific to some animals such as *M. microti* (voles), *M. caprae* (goats), *M. pinnipedii* (seals), *Mycobacterium mungi* (mongooses), *Mycobacterium orygis* (oryx) and *Mycobacterium suricattae* (dassies) [4-6]. The term "MTBC" distinguishes the above listed mycobacteria from *M. leprae*, saprophytic mycobacteria (e.g., soil organisms) and atypical mycobacteria, such as *M. avium-intracellulare* [7], collectively termed non-tuberculous mycobacteria (NTM). The MTBC possess a complex and lipid-rich cell wall that is responsible for many of the characteristic properties of the bacterium. These include acid fastness, slow growth, resistance to anti-bacterial drugs, thee host immune response and antigenicity [8].

1.2. Epidemiology of tuberculosis

Despite the availability of the BCG vaccine and a number of drugs, TB continues to be the major cause of morbidity and mortality worldwide (Figure 1) [3, 9]. Furthermore, worldwide emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) MTBC strains, HIV-TB co-infection, latent and sub-clinical TB infection are becoming the hurdles against the efforts towards the control of TB [3, 9, 10].

Approximately one-third of the world's population is infected with Mtb; however, a large majority remain asymptomatic, a condition defined as latent TB [11]. TB is characterized by necrotizing granulomatous inflammation usually in the lungs (~80-85% of cases), although almost any extrapulmonary site can be involved [11-13]. Primary infection with Mtb will lead to clinical disease in only 5 to 15% of the cases, but the risk is markedly increased in patients with human immunodeficiency virus (HIV) and other medical conditions [3]. The outcome following

TB infection is determined by a complex and dynamic pathogen-host interaction involving the host genetic factors [10, 14], pathogen traits [15-19] and environmental conditions [20, 21].

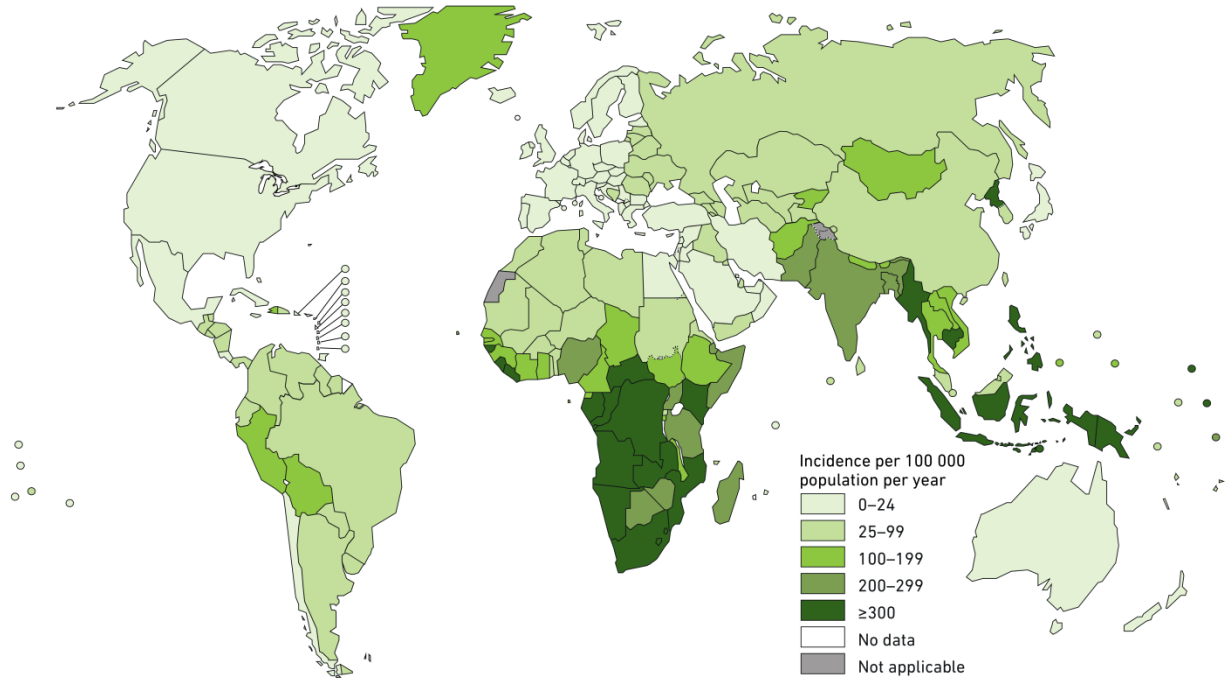


Figure 1. Estimated global tuberculosis incidence rates. Adapted with permission from WHO tuberculosis report, 2017 [3].

1.3. Genetic diversity and ancestral sequence inference of the MTBC species

The MTBC consists of bacteria that genetically share identical 16S rRNA sequences (excluding *M. canetti*), but differing in their epidemiological profiles, host ranges, pathogenicities, geographic distributions, and drug resistances [7, 22]. Members of the MTBC have a set of clustered regularly interspaced short palindromic repeat sequences, multiple copies of insertion sequences, which are used as target loci for strain typing [12]. Methods formerly used in MTBC strain typing include the insertion sequence (IS) 6110 restriction fragment length polymorphism (IS6110-RFLP), a method based on the analysis of mobile DNA elements [23] and the analysis of variations within the genomic direct repeat (DR) region by spoligotyping [24, 25].

Currently, mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) and whole genome sequencing (WGS) are used in strain typing [26]. MTBC strains show unidirectional loss of spacers, an indicator of the evolutionary trends between strains. MIRU-VNTR typing uses 40-100 bp tandem repeats located in microsatellite regions around the chromosome. The PCR-free IS6110-RFLP method has a better discriminatory power than the

MIRU-VNTR or spoligotyping, but combining MIRU-VNTR with spoligotyping can enhance the discriminatory power to even more than IS6110-RFLP [26-28].

WGS analysis of genomic deletions on region of differences (RDs) or large sequence polymorphisms (LSPs) and single nucleotide polymorphisms (SNPs) are useful for phylogenetic applications [29-31]. SNPs in *katG* (a gene encoding catalase-peroxidase) and a subunit of the DNA gyrase gene *gyrA* enable three principal MTBC genetic groups (PPGs) to be recognized [31]. Analysis of deletions using WGS can be used to discriminate members of the MTBC and to study their evolution [32]. Subsequently, a commercial multiplex PCR-based kit, the GenoType MTBC assay (Hain Lifescience GmbH, Nehren, Germany) was developed based on the detection of SNPs in *gyrB*, the 23S rRNA gene and the RD1 deletion [33-35]. Unlike WGS, the GenoType MTBC assay cannot differentiate *M. canetti* from Mtb and *M. africanum* type I from *M. pinnipedii* [36].

Comparative analysis of 20 variable genomic regions from 100 MTBC strains with different hosts and a broad range of geographical origins revealed that MTBC strains have undergone reductive evolution by the unidirectional and irreversible loss of chromosomal sequences [32]. Genome sequencing and deletion analysis indicated that *M. bovis* has undergone 11 deletions relative to Mtb, and, thus, Mtb is a more ancestral lineage than *M. bovis* [22, 37, 38]. This deletion is a biomarker of a clone derived from a single cell and all of its descendants [39].

The human-adapted MTBC comprises of seven phylogeographic lineages that diverged from a common ancestor and diversified in different regions of the world (Figure 2) [17]. Based on the presence or absence of a specific deletion (TbD1), MTBC strains can be referred to as evolutionarily “ancestral” or “ancient” and “modern” lineages, respectively [17, 32]. *M. africanum, M. bovis*, and other animal strains shared another deletion (RD9) not present in the modern strains or any other strains belonging to Mtb [32]. The geographical spread of these lineages differs markedly, with some lineages exhibiting a global distribution, while others are geographically restricted (Figure 2) [40]. Lineage 4 strains have been implicated in major outbreaks and are responsible for most TB cases in the world [17]. The more recently identified lineage 7 strains have been shown to be slow-growing, less virulent and drug sensitive [41].

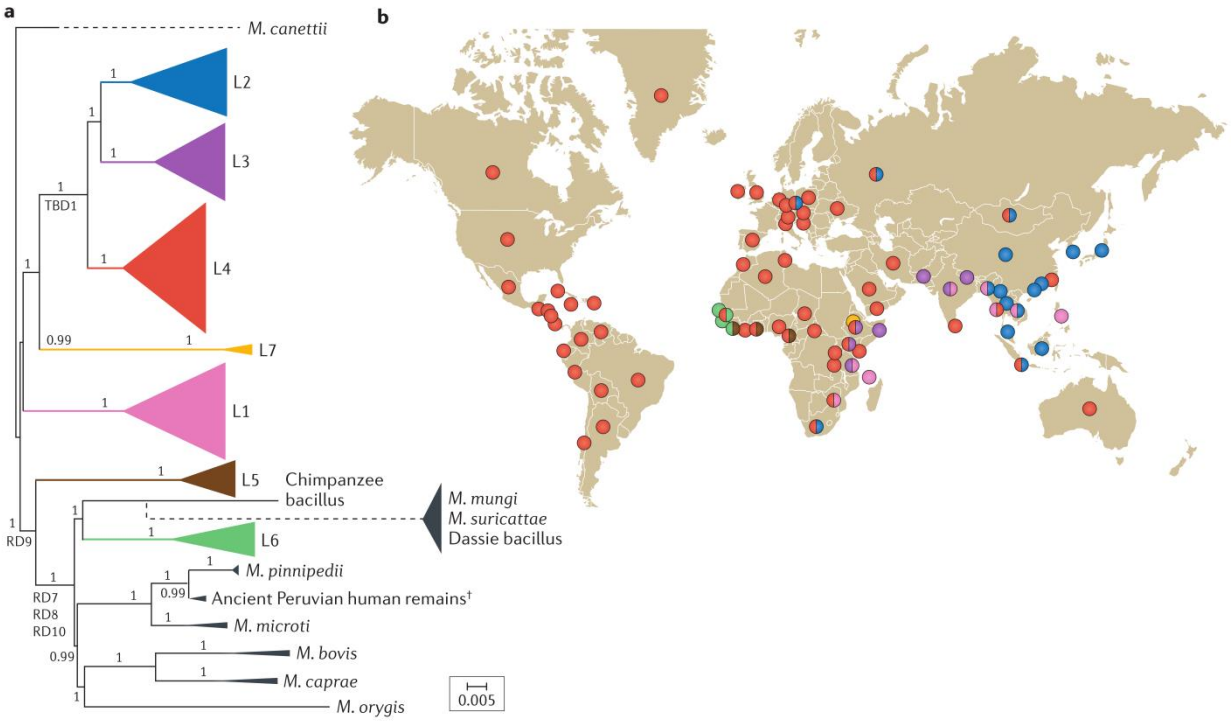


Figure 2. Global phylogeography of the human-adapted MTBC. (a) Phylogeny of 220 strains of *Mycobacterium tuberculosis* complex based on genome-wide studies [40, 42]. The MTBC comprises seven human-adapted lineages (in colour) and several lineages adapted to various wild and domestic animals (in grey). The tree is rooted using the ancestral *M. canettii*. The animal lineages represent a monophyletic branch in the complex. Based on *M. tuberculosis*-specific deletion 1 (TBD1) analysis, *M. tuberculosis* strains can be divided into ancestral and "modern" strains, the latter comprising lineage 2, lineage 3 and lineage 4 strains [32]. Similarly, the deletion of the region of difference 7 (RD7), RD8, RD9 and RD10 are indicated under the respective branches. The dagger indicates genomes generated from ~1,000-year-old MTBC DNA that was recovered from archaeological human remains in Peru [43]. Bootstrap confidence intervals are indicated. Scale bar represents number of nucleotide substitutions per site. **(b)** The global distribution of the seven main human-adapted MTBC lineages. Lineages were coloured as in a, and coloured dots represent the dominant MTBC lineage(s) in each country [16]. Adapted with permission from Ref. [44].

1.4. Clinical consequences of strain variation in the MTBC

The genetic homogeneity amongst members of the MTBC initially led to the assumption that the variations in the clinical or epidemiological outcomes of TB infection were largely related to host susceptibility and environmental factors, and not to the bacteria [45, 46]. However, subsequent studies have demonstrated that the clinical phenotypes manifested by members of the MTBC are influenced by the genetic and evolutionary background of the strains [16, 17]. Genotyping of MTBC clinical isolates has revealed a limited number of clades that are responsible for a large proportion of diseases outbreaks and multi-drug resistance, suggesting that such strains may be more virulent than others [47, 48].

Furthermore, studies have confirmed that different isolates of MTBC exhibit phenotypic variability in terms of clinical outcome and epidemiological behaviour [16, 18, 19, 49]. For

example, strains of the MTBC Beijing genotype are highly prevalent in the global TB community representing about 50% of strains in East Asia and at least 13% of strains worldwide, suggesting that the Beijing genotype may have a better adaptive potential over other MTBC genotypes and lineages [50-53]. Beijing genotype strains have been associated with extrapulmonary TB, human immunodeficiency virus (HIV) infection, outbreaks and AMR [54, 55]. Another outbreak strain, CDC1551 from lineage 4, was found to be resistant to RNI and ROI [48, 56]. The exact mechanism for these variations amongst different lineages is not fully understood and is likely to be a combination of factors [50, 57, 58]. Some of the proposed mechanisms include altered pathogen-host interactions, stress response and AMR [59, 60]. Despite these variability, former studies predominantly used the lineage 4 reference laboratory strain H37Rv from lineage 4 as a model in the proteomics and post-translational modification (PTM) studies [61]. It is therefore ideal to analyse the qualitative and quantitative differences in expressed proteome and PTM between members of the MTBC and investigate the contribution of these differences to the observed phenotypic variability.

1.5. Pathophysiology of tuberculosis

1.5.1 Natural course of tuberculosis infection

Most human TB infections arise through inhalation of Mtb bacilli-containing droplet nuclei of one to five micrometer (μm) in diameter expectorated from the lungs of an infected individual, typically through coughing and sneezing. Following entry, AMs ingest the bacilli, leading to one of four possible outcomes; immediate clearance of the organism, immediate onset of active disease (primary disease), latent infection or onset of active disease many years following a period of latent infection (reactivation disease) (Figure 3). Among the 5 to 15% of infected individuals who develop active TB disease, about half of them will have “early” progressive disease within five years after infection, while the rest have “late” diseases, which is caused by reactivation as long as several decades after the primary infection when the immune response is compromised [62-64]. In most of the TB cases, Mtb infection arises in the respiratory system; however, it can spread to other organs, such as the lymphatics, pleura, bones/joints, or meninges, and cause extrapulmonary TB in 15-20% of all cases [12, 13, 65].

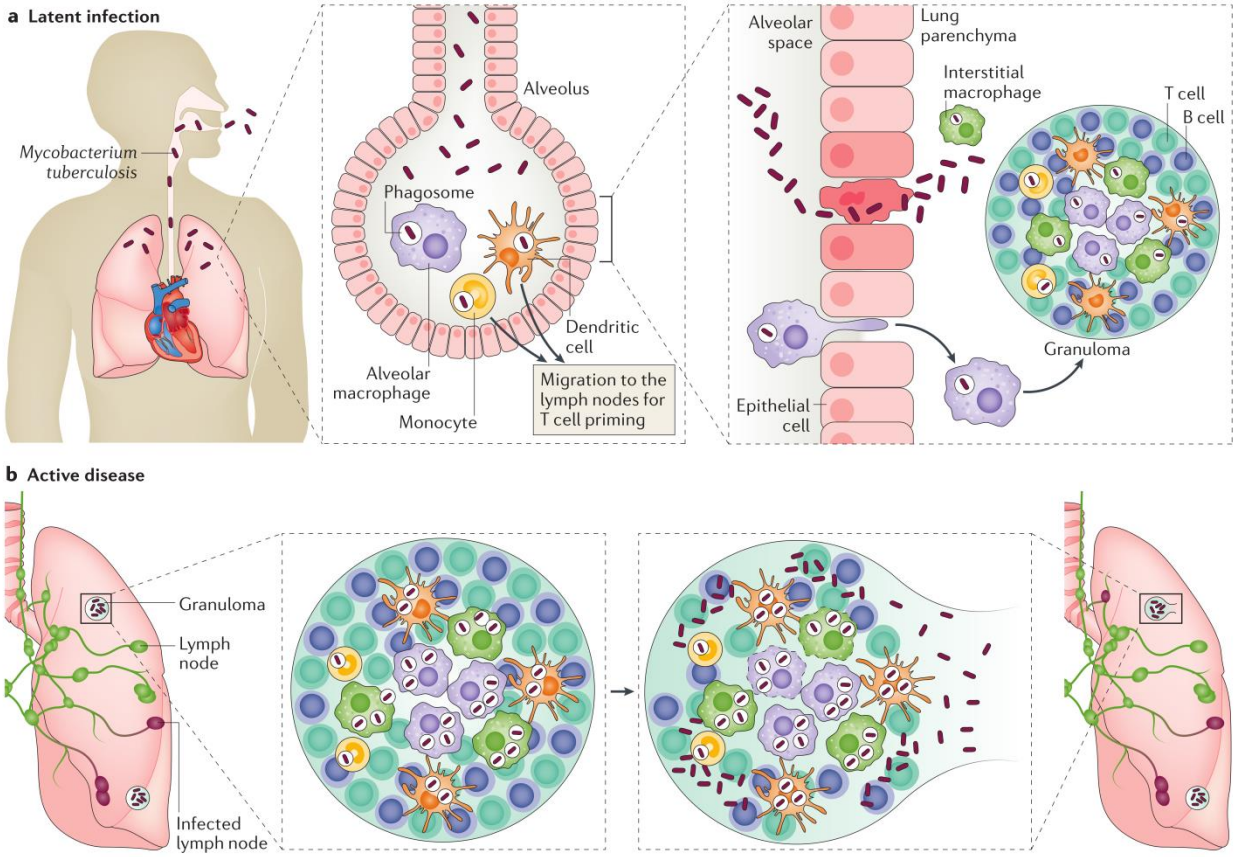


Figure 3. Pathophysiology of *Mycobacterium tuberculosis* infection. (**a**). Infection occurs through inhalation of Mtb bacilli containing droplet nuclei, reaching alveoli where it encounters the alveolar macrophages (AMs). If survived, Mtb invades the lung interstitial tissue via infected AMs or by directly infecting the alveolar epithelium. Subsequently, antigen presenting cells transport Mtb to lymph nodes for T cell priming, which leads to the recruitment of other immune cells to the lungs to form a granuloma, the hallmark of Mtb infection. (**b**). The bacilli replicate inside the granuloma followed by the burst and dissemination of Mtb to other organs. At this stage, the host is infectious, symptomatic and is said to have active TB diseases. Adapted with permission from Pai *et al.*, 2016 [66].

If not treated, TB causes death in up to 80% of cases while the remaining will develop chronic disease, characterized by repeated episodes of healing by fibrotic changes around the lesions and tissue breakdown, or recovery [67]. Reactivated TB tends to be localized at the lung apices with little regional lymph nodes involved, unless the host is severely immunosuppressed. Prior latent TB infection is known to confer protection against subsequent TB disease, while prior TB disease is associated with an increased risk of subsequent TB disease [68-70].

1.5.2. Innate immune responses to tuberculosis

Innate immunity has gained much more attention due to its profound role both in early responses and control of Mtb infection as well as in sustaining the T cell response [71]. Innate immunity contributes to the bactericidal activity through autophagy, apoptosis, inflammasome formation, and nitric oxide (NO) production thereby limiting the growth of Mtb [72]. For example, the

inflammasome, a multiprotein oligomer, is an important innate immune pathway involved in the maturation of two pro-inflammatory cytokines (IL-1β and IL-18) and induction of pyroptosis, which induces an inflammation-induced death of infected cells [73].

When the tubercle bacilli reach the lung alveoli, they are phagocytosed by resident AMs, the sentinels of Mtb infection [74]. Other cells involved in the uptake of Mtb bacilli include the alveolar endothelial cells, type I and II alveolar epithelial cells (pneumocytes), or DCs [75-78]. This will lead to a cascade of events that result in either successful containment of the infection or spread and progression to clinically active disease [75, 76, 78].

During the first contact, the pattern recognition receptors (PRRs) on innate immune cells (DCs or AMs) recognize the pathogen-associated molecular patterns (PAMPs) exposed on the Mtb cell wall, which is central to the initiation and coordination of the host innate immune response [79]. The host PRRs include the toll-like receptors (TLRs), nucleotide-binding oligomerization domain- (NOD-) like receptors (NLRs), and C-type lectins. The C-type lectins include the mannose receptor (CD207), DC-specific intercellular adhesion molecule (ICAM)-3-grabbing non-integrin (DC-SIGN) and dectin-1 [80, 81]. Other PRRs are the complement receptors, Fc receptors, scavenger receptors, surfactant protein A receptors and cholesterol receptors [82]. Mtb PAMPs include the lipoarabinomannan (LAM), lipomannan (LM), mannose-capped lipoarabinomannan (ManLAM) and phosphatidyl-myo-inositol mannosides (PIM), mammalian cell entry (Mce) proteins, LpqH and LprG [1, 81, 83, 84]. Specific cell types and receptors involved in phagocytic entry have a major impact on subsequent events [85, 86].

TLR signaling is the main arm of the innate immune response during Mtb infection [80, 81], which also induces cytokines and chemokines that are crucial for eliciting the adaptive immune response against the pathogen [87]. In addition to MØs and DCs, gamma delta T cells (γδ T cells), natural killer (NK) cells, and natural killer T (NKT) cells also participate in the innate immune response to tuberculosis. Engagement of PAMPs to the TLR2 and TLR4 (PRRs) on the surface of AMs induces the activation of nuclear transcription factor-kappa B (NF-κB) and production of pro-inflammatory cytokines, such as TNF-α, IFN-γ, IL-1β, IL-6, IL-12 and IL-18 [8, 80, 81, 84]. The molecular mechanisms behind the control of Mtb infection may involve production of RNI and ROI, chages in phagosomal pH inducing apoptosis and other factors [88].

IFN-γ and TNF-α are the main pro-inflammatory cytokines involved in the immune response against Mtb [89, 90]. IFN-γ is mainly secreted by Th1 and CD8+ cytotoxic lymphocytes

(CTLs), NK cells, NKT cells, γδ T cells, professional antigen-presenting cells (APCs), and, to a lesser extent, B cells [91]. IFN-γ stimulates MØs to produce TNF-α, an essential component of the innate defense mechanism. TNF-α, in combination with IFN-γ, upregulates nitric oxide synthase 2 (NOS2) expression and induces production of RNI within the phagolysosome, leading to increased phagosome-lysosome fusion and intracellular bacterial killing [8, 91, 92]. Too much induction TNF-α causes tissue damage and promotes Mtb growth and transmission [93]. In contrast, the anti-inflammatiory cytokine IL-10 downregulates activation of MØs and DCs, inhibiting IFN-γ inducible genes, RNI and ROI [94]. Thus, IL-10 is an important cytokine in modulating the balance between inflammatory and immunopathologic responses [94, 95].

IL-12 induce T helper type 1 (Th1) mediated IFN-γ production and stimulate the proliferation of antigen-specific cytolytic T cells and NK cells [96, 97]. The γδ T cells secrete TNF-α and IFN-γ, which induces the bactericidal activity of MØs and stimulates the migration of immune cells to the infection site [95, 98, 99]. These cascades of cytokine and chemokine signaling recruit other immune cells, including activated T cells and MØs, and facilitate the formation of granuloma, the pathological hallmark of TB [95]. Immune cells in the granuloma limit further replication and spread of Mtb [100].

1.5.3. Adaptive immune responses to tuberculosis

Furthermore, IFN-γ enhance the expression of MHC class II molecules on MØ, resulting in antigen presentation to T cells [81]. Most infected individuals develops cell-mediated immunity within approximately three weeks after infection [101]. APCs, such as MØs and DCs, carry phagocytized bacilli to draining lymph nodes where they present Mtb-derived antigens to T cells via the MHC class I and II molecules to CTLs and $CD4^+$ helper T cells, respectively. Lysis of infected cells via activated $CD8^+$ CTL involves a degranulation pathway that generates granulysin, perforin and granzymes, and a Fas-FasL-dependent pathway that induces apoptosis of the target cell [102, 103]. Granulysin, a protein found in the granules of $CD8^+$ CTLs T cells and NK cells, is involved in bacterial killing [104, 105].

Depending on the cytokine environment, the $CD4^+$ helper T cells differentiate either into Th1, Th2, Th17 and regulatory T cell (Treg). The Th1 cells secrete pro-inflammatory cytokines that promote stimulation of Th1 cells, $CD8^+$ CTLs, and maturation and activation of MØs as well as granulocytes, thereby mediating the killing of intracellular bacilli through RNI and ROI production. The $CD4^+$ Th1 cells produce a much stronger IFN-γ response than $CD8^+$ CTLs

following Mtb infection [106], and are therefore thought to play a prominent role in protection [107]. Although cellular-mediated immune response is far more critical for effective responses against Mtb, B cells and antibodies may play a significant role during Mtb infection [108, 109].

1.6. The unique mycobacterial cell envelope

The MTBC cell envelope is composed of three layers: an inner plasma membrane; a cell wall core with an outer mycomembrane; and an outermost layer (capsule), which is exceptionally rich in unusual lipids, glycolipids, polysaccharides and proteins (Figure 4) [37, 110]. The cell wall core is composed of peptidoglycan (PG) covalently linked via phosphoryl-N-acetylglucosaminosylrhamnosyl to arabinogalactan (AG), which, in turn, is esterified to long-chain mycolic acids (C_{60}-C_{90} fatty acids), forming the mycolyl arabinogalactan-peptidoglycan (mAGP) complex [110, 111].

This barrier is essential for bacterial viability and is responsible for many of the medically challenging physiological characteristics of TB, including resistance to antibiotics and host defense mechanisms [112, 113]. Mycolic acids constitute up to 60% of the cell dry weight [114]. PG is a polymer of alternating N-acylated muramic acid (MurNac) and N-acetylglucosamine (GlcNac) residues with cross-linked peptides of varying composition attached to the muramyl moieties [115, 116]. PG confers cell wall rigidity and is another contributor to the impermeability of mycobacterial wall [65]. In contrast to most other bacteria, muramic acid moieties are N-glycolylated (oxidized) in mycobacteria [111].

Besides, the cell wall contains lipoglycans including LAM, its precursors LM, PIM and TDM which contribute to Mtb survival, trigger inflammatory host reactions and involved in pathogenesis [117, 118]. The mycolic acid-containing glycolipid trehalose-6,6-dimycolate (TDM), also known as cord factor, is an important immunomodulatory virulence factor in MTBC [119]. The soluble components of the cell envelope include free lipids, proteins, LAM, and PIMs, which are signaling effector molecules in bacterial pathogenesis and disease processes [111]. LAM, one of the key Mtb virulence factors, is involved in inhibition of phagosomal maturation and interferes with cell signaling leading to dampening of host pro-inflammatory response [65, 117]. The cell wall of MTBC also contains a 19-kDa lipoprotein (LpqH) which has been implicated in host cell death, and manipulation of bactericidal mechanisms [117, 120]. The nature and amounts of mycomembrane and capsular material vary among MTBC isolates, and is likely to alter the pathogen phenotype and the outcome of pathogen-host interaction [121, 122].

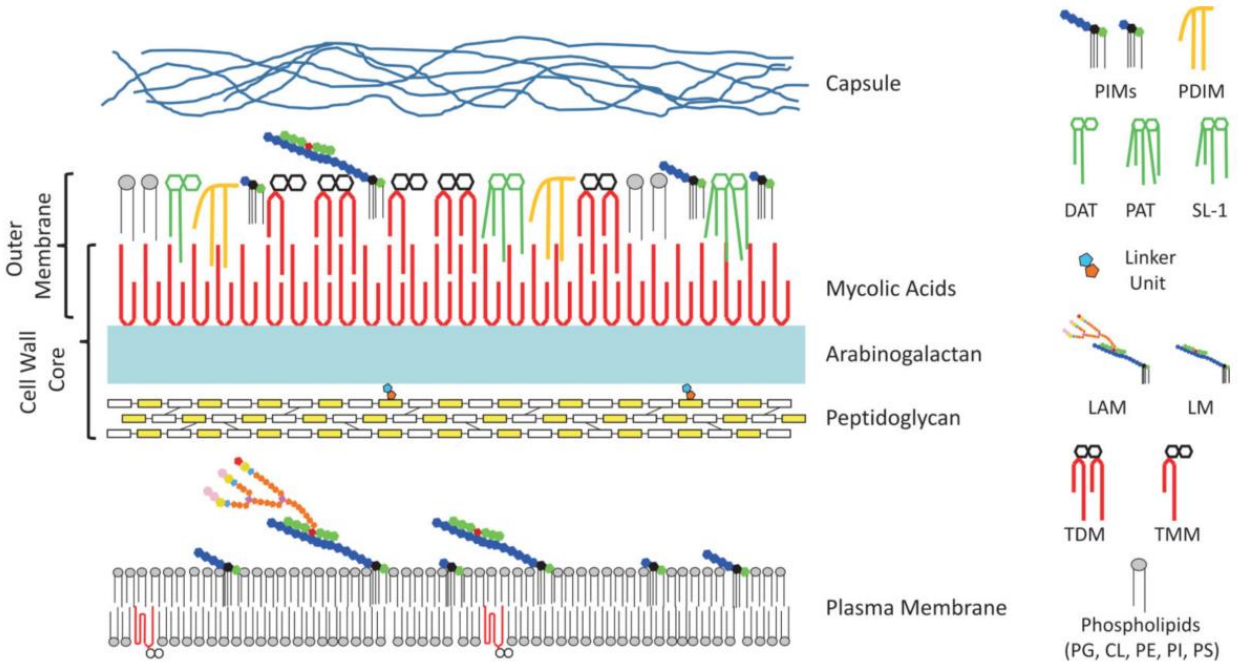


Figure 4. Schematic representation of the mycobacterial cell envelope. The peptidoglycan polymer is represented by yellow (GlcNAc) and white (MurNAc) rectangles, arabinogalactan by a blue region. Trehaloses (in green and red) are non-reducing disaccharides made of glucose. PDIM, phthiocerol dimycocerosate lipid; PE, phosphatidylethanolamine; PI, phosphatidyl-myo-inositol; CL, cardiolipin; PS, phosphatidylserine; PG, phosphatidylglycerol. Proteins and peptides are not shown for the sake of clarity. Adapted with permission from Angala *et al*., 2014 [110].

1.7. Virulence and pathogenicity in the MTBC

1.7.1. Defining virulence and pathogenicity in the MTBC

Pathogenicity can be defined as the capacity of a microbe to cause damage or disease in a host [123-125]. Virulence is the degree of pathogenicity or the relative capacity of a microbe to cause damage in the host as compared to similar pathogens [123]. The damage is caused by microbial factors and/or the host response to the infection [126]. The severity of disease resulting from an infection with the same pathogen depends on virulence factors, microbial components that are able to override the mechanisms of host defense and cause damage to the host [1, 123, 124]. Mtb cells employ a number of virulence factors to efficiently colonize, survive and propagate inside the host cell, which therefore are promising drug targets and vaccine candidates [76, 126-128].

1.7.2. Virulence factors in the MTBC

The main antimicrobial mechanisms of the mature macrophages (MØs) phagosome include acidification of the phagosome, production of antimicrobial peptides, activation of the NADPH oxidase (NOX2) and inducible nitric oxide synthase (iNOS) leading to the generation of reactive

oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI), respectively, and the expression of degradative enzymes such as cathepsins [129]. Mtb produces a variety of virulence factors to circumvent the host defense mechanism by facilitating bacterial adherence and invasion, and subvert host cell signaling that regulate intracellular microbial survival [130]. Major evasion mechanisms include resistance to ROIs, inhibition of phagosome-lysosome fusion, inhibition of phagosome acidification, and escaping the harmful phagosome to instead allow replication in the cytosol [129]. LAM is a major cell wall glycolipid involved in scavenging oxidative radicals produced by MØ [131]. Mycolic acids help the bacilli to resist ROI and inhibit the expression of IL-12, TNF-α and MCP-1 in TLR-2-dependent manner in MØs [1, 132, 133].

Mtb is internalized by alveolar macrophages (AMs) and dendritic cells (DCs) within 14 days after aerosol infection via a variety of cell surface receptors; the intracellular fate is thought to be influenced by the type of receptor employed [128, 134]. For example, when the bacilli enter the MØ via the complement receptors, it doesn't stimulate ROI production [135, 136]. Upon the fusion of the matured phagosome and lysosome, the lysosome releases microbicidal molecules that can kill the bacilli. Mtb can inhibit vacuolar acidification by blocking the phagosomal proton-ATPases [137, 138]. Phagocytosed Mtb secretes products that inhibit the conversion of Rab5 into GTP-bound Rab7 and the generation of phosphatidylinositol 3-phosphate (PI3P), which are required for phagosome maturation [139-141].

Cell envelope-associated virulence factors and immune evasion mechanisms. The mycobacterial cell wall components are the major virulence factor involved in MTBC pathogenesis (discussed in section 1.6) [1]. MTBC lipid products have been known for their role in TB pathogenesis since 1947 [142]. Later, the major toxic effect of the Mtb cord factor was recognized and tested in an animal model [143], and its chemical structure was identified to be mycolic acids covalently bound to trehalose, TDM [144]. Differences associated to structure, composition and metabolism of cell wall glycolipids, including TDM, contribute to the varation in bacterial virulence and clinical outcome in the host [119, 145, 146]. TDM in a micellar conformation is non-toxic and protects organisms from host defenses while it becomes highly toxic and immunogenic when released in a monolayer conformation [147]. Mtb with disrupted cyclopropane mycolic acid synthase 3 (*pcaA*), a gene involved in cording and the formation of cyclopropane rings in its α-mycolates, is attenuated and unable to block phagosome maturation [145, 148]. TDM induce the production of a number of chemokines (MCP-1, MIP-1alpha, IL-8) and cytokines (e.g., IL-12, IFN-gamma, TNF-alpha, IL-4, IL-6, IL-10) leading to stimulation of cellular and humoral immunity and granuloma formation [119].

Another group of lipids, lipoglycans, sulfolipids, phthiocerol dimycocerosates (PDIM) and ManLAM are major Mtb virulence factors involved in receptor-mediated phagocytosis of Mtb, prevention of phagosome acidification, modulation of innate and adaptive immune responses, and neutralization of the bactericidal activity of RNIs [1, 119, 149, 150]. Alkyl hydroperoxidase (AhpC) is a protein expressed by Mtb to protect bacterial and human cells against necrosis and apoptosis caused by RNIs [1, 151]. Mycolic acids were also shown to be part of the extracellular matrix of the Mtb biofilm *in vitro* [152]. The ManLAM on the Mtb cell envelope interacts with the MR on MØs, leading to phagocytosis of the bacilli and inhibition of MR-dependent IL-12 production [153, 154]. The ManLAM also inhibits phagolysosome maturation and phagosome-lysosome fusion, thereby promoting Mtb survival in MØs [154, 155].

Mtb has also evolved strategies that dampen the host immune responses to its advantage. Mtb blocks major histocompatibility complex (MHC) class II mediated antigen presentation that induce $CD4^+$ T cell responses, which is central to host resistance to Mtb [156-158]. This will lead to resistance to host acquired immunity and, hence, may contribute to bacterial persistence, a characteristic feature of Mtb. For example, prolonged TLR2 signaling by LAM, the 25 kDa glycoprotein and the 19 kDa lipoprotein LpqH inhibits MHC class II expression and antigen processing by MØ leading to immune evasion [83, 159].

Furthermore, the bacilli develop adaptive potential through differential expression of virulence factors. The *mce1* operon has been shown to be repressed during early stage infection in mice, and a strain of Mtb with a disrupted *mce1*operon over-expresses free mycolic acids, causing increased mortality in immunocompetent mice [160-162]. Mtb with a disrupted *mce1* operon also exhibited reduced expression of MmpL8, MmpL10, Stf0, Pks2 and PapA2 genes involved in transport and metabolism of pro-inflammatory lipids including DIM [161]. *Mce1A* encodes an integral cell wall protein and *mce1*disruption leads to enhanced virulence compared to the wild-type in mouse models [37, 160, 163]. An outbreak-associated Mtb lineage 2 strain, termed HN878, expresses a highly active lipid species called phenolic glycolipid (PGL), which inhibits pro-inflammatory cytokine release by MØs [164, 165].

Membrane vesicles. Membrane vesicles (MVs) are closed spherical nanostructures that are shed naturally by Mtb cells. MVs are vehicles for the selective packaging and release of Mtb virulence factors, such as toxins lipoproteins, enzymes and immunomodulatory molecules [166, 167]. Proteins within the MVs and the freely secreted proteins constitute the major set of

virulence factors produced by MTBC to counteract the host defense system. Phagocytosed dead bacilli, with no secreted proteins, lose their ability to hinder phagosome-lysosome fusion [168].

Metabolic adaptation. MØs limit the access of nutrient and carbon source (glucose and lipids) to Mtb cells engulfed in the phagosomal compartment. When carbon substrates for glycolysis become limited, Mtb exposes its vast and adaptive metabolic repertoire. Mtb preferentially metabolizes the host lipid droplets accumulated in the phagosome as a sole carbon source using the enzymes isocitrate lyases 1 and 2 (*ICL1/2*)[169-171]. The ESX-1 secretion system is another virulence factor involved in the secretion of immunodominant proteins ESAT-6 and CFP-10, and has been implicated in promoting escape of Mtb or its products from the phagosome into the cytosol [172, 173]. ESAT-6 has been shown to exert membrane-lysing activity *in vitro* [174, 175]. The ESX-1 secretion system is located in the Mtb chromosome locus called the region of difference (RD-1), but it is absent in all BCG vaccine strains, which is believed to be the basis for the attenuation of BCG [176, 177].

1.8. MTBC genome characteristics

About ~4,000 open reading frames were identified when the first WGS of Mtb was made available in 1998, out of which 61% were assigned with putative functions based on sequence homologies [37]. The Mtb genome has 4,411,529 base pairs with an average 65.6% G+C content [8, 37]. About 10% of the Mtb genes encode two families of glycine-rich proteins termed PE (proline-glutamine motifs) and PPE (proline-proline-glutamine motifs). These genes are composed of polymorphic G+C-rich repetitive sequences (PGRSs) and major polymorphic tandem repeats, which can also be used in Mtb strain-typing [178, 179]. PE/PGRS proteins are expressed inside the granuloma and serve as a virulence factor for *M. marinum* [180]. Notably, the 4,345,492 base pairs genome sequence of *M. bovis* was completed in 2003 and revealed deletion of genetic information leading to a reduced genome size [38]. After genome annotation, Mtb genes essential for *in vivo* survival have been identified using mutant analysis [181, 182]. However, only little of these proteins are well characterized for their role in Mtb infection and virulence [183-186].

About 59% of Mtb genes are transcribed in the same direction as the replication, compared to the 75% of *Bacillus subtilis*, which can probably be related to its peculiar slow growth and infrequent replication cycles [187]. The availability of high G+C in the Mtb genome has led to a significant statistical preference for amino acids encoded by G+C-rich codons, such

as alanine, glycine, proline, arginine and tryptophan and a comparative reduction in the use of amino acids encoded by A+T-rich codons, such as Asn, Ile, Lys, Phe and Tyr [187].

The plethora of data now available from WGS has created new research opportunities and improved our knowledge on the biology of microbial pathogens. The annotated genome sequences of Mtb and *M. bovis* have opened the door for dissection of both protein expression/regulation and function [37, 38]. Later, Mtb putative virulence factors have been identified using animal model [18, 188], comparative genomics [189, 190] and transposon site hybridization [181, 191]. Despite the opportunity gained from genomic data, there are still gaps in understanding the secret codes that makes only some of Mtb strains a successful human pathogen [37, 38, 189, 192, 193]. Genomic analysis alone can not provide a complete picture of the system-wide biological complexity of an organism, while the output from proteomics is variable among cell-to-cell and from time-to-time [194]. Furthermore, PTMs of proteins contribute to a much broader diversity of protein species than those that are predicted based on the the organism's genome alone [194]. Proteomics and PTMs have the potential to illustrate the dynamic regulation of virulence factors under different environmental conditions, which is descriptive of biological phenotype [126, 195, 196]. Proteomics is, therefore, a widely used method to characterize MTBC virulence factors through comparative expression proteomics, protein interactions, PTMs and cellular localization [9, 126, 197].

1.9. Mass spectrometry-based proteomics

Proteomics is the analysis of varieties, quantities, spatial distribution, protein response to environmental stimuli, roles and dynamics of all proteins in a cell, tissue or organism under a specific, defined set of conditions [198, 199]. Proteomic information has the potential to illustrate the dynamic regulation of virulence factors under different environmental conditions [195, 196]. It is the ultimate production of functional proteins that allows the pathogen to adapt and survive within the dynamics of pathogen-host interactions [126, 195]. Thus, proteomics is the most direct platform for measuring cellular activity [200]. Further layer of complexity is added to proteomics by PTMs of proteins and their organization in complexes, both highly relevant for functionality [200]. Furthermore, studies have shown that the cellular abundance of mRNA and protein encoded by the same locus does not always strictly correlate [201, 202].

Mass spectrometry (MS) is an analytical device for measuring the mass-to-charge ratio (m/z) of ionized molecules to extract qualitative and quantitative information [203]. An MS unit contains three main parts; an ionization source to produce gas phase ions from the sample, mass analyzer(s) to separate ions according to their m/z and a detector to 'count' the ions emerging from the last analyzer (Figure 5). The data processing system record the signal registered by the detector and produces the mass spectrum [203].

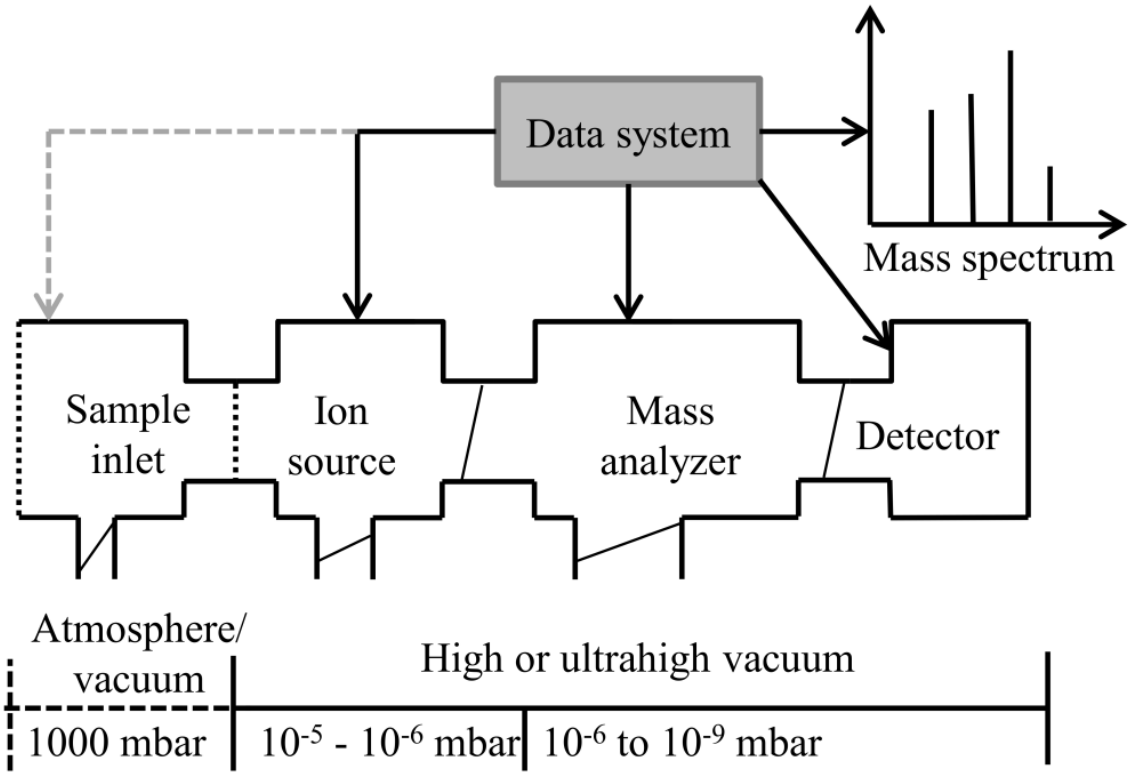


Figure 5. Outline of generic mass spectrometer components. The sample inlet systems introduce the sample into the ion source with minimal loss of vacuum. The molecules are converted to gas phase ions in the ion source. The mass analyzer separates the ions according to their mass-to-charge ratio. The detector records a signal that can be electronically amplified and stored [203, 204].

Additional separation techniques, such as liquid chromatography (LC), are often coupled to the MS prior to mass spectrometric analysis, and a whole series of MS spectra can be acquired for the duration of the LC separation of the peptide mixture. Typical LC-MS/MS analysis involves the acquisition of the intact (precursor) peptide mass spectrum (MS1), fragmentation of the isolated precursor ion of interest and sebsiquent mass analysis of the fragments (MS2) [203, 205]. Precursor ion fragmentation can be done by collision-induced dissociation (CID) or alternative techniques such as electron capture dissociation (ECD) and electron transfer dissociation (ETD). Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are the most commonly used soft ionization methods [206].

MS-based proteomics is emerging as a broadly effective means for identification, characterization, and quantification of proteins and PTMs in either a purified, enriched or complex protein mixture [203, 206, 207]. It is highly sensitive, can be completely automated and easy to hyphenate with different separation techniques that allow detection and quantification of the rich diversity of protein isoforms present in a biological sample, often spanning as much as 12 orders of magnitude in relative abundance [203, 206, 208].

Proteomic experiments can be performed through bottom-up (discovery proteomics) or top-down approaches. Bottom-up proteomics is the most commonly used method to identify components in a mixture of proteins and characterize their amino acid sequences and PTMs by proteolytic digestion of proteins followed by MS analysis (Figure 6) [209, 210]. Top-down proteomics to the analysis of intact proteins and protein complexes which can be subsequently fragmented inside the MS and the masses of the fragments recorded as well [209, 211]. It is applied for identification of protein isoforms arising from amino acid modifications, gene variants, transcript variation, and PTMs as well as proteolytic processing of proteins [206].

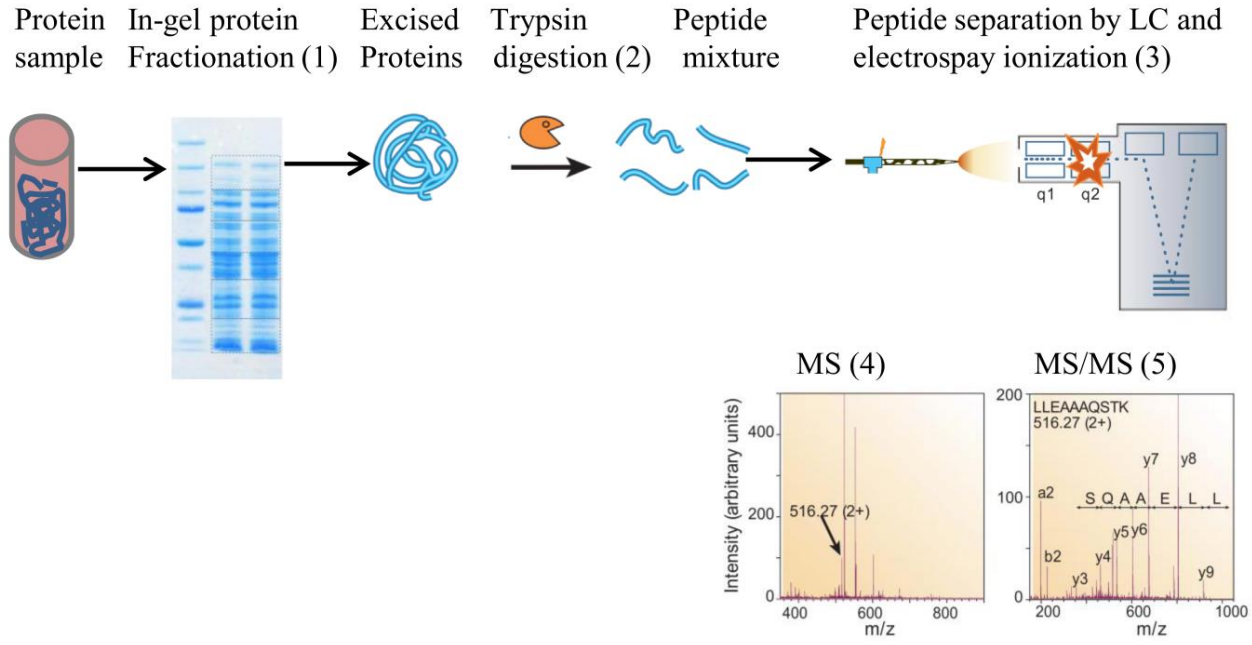


Figure 6. Generic mass spectrometry (LC-MS/MS)-based proteomics experiment. Proteins from cell lysate are typically fractionated by gel electrophoresis (1), and digested enzymatically to peptides (2), usually by trypsin. The peptides are further separated by one or more high pressure liquid chromatography and eluted into an electrospray ion source where they are nebulised in small, highly charged droplets containing peptides (3). After evaporation, multi-protonated peptides enter the mass spectrometer, generating a mass spectrum of the peptides eluting at this time point (MS1 spectrum or 'normal mass spectrum') (4). The computer generates a prioritized list of these peptides for fragmentation by energetic collision with gas, and recording of the tandem mass spectrometry (MS/MS) spectra. The resulting mass spectra are then stored for search against protein sequence databases and identification of the parent protein. Modified from Aebersold, R. and Mann, M., 2003 [205].

1.9.1. Quantitative proteomics

Quantitative proteomic measurements by MS can be performed using either label-free or label-based (stable isotope labelling) methods [212, 213]. In label-free methods, two or more experiments to be compared are individually injected into the MS. The relative difference in protein abundance is calculated by comparing the intensity of the mass spectrometric ion signal for any given peptide or by mass spectra matching, considering the number of acquired spectra matching to a peptide/protein as an indicator for their respective amounts in a given sample [206, 208, 213]. In the stable isotope labelling methods, the samples are differentially labelled, combined and subjected to MS analysis. The ratio of peak intensity between the light and heavy peptide ions gives the relative abundance of the protein which the peptide is derived from. These isotope mass tags can be introduced into proteins/peptides (i) metabolically, (ii) chemically, (iii) enzymatically or (iv) provided by spiked synthetic peptide standards [213]. MS based quantification methods have their particular strengths and weaknesses. Label-free strategies are cost-efficient, much more simple and provide higher dynamic range of quantification than label-based MS, but more error-prone due to systematic variations among the individual runs [213]. Absolute protein quantification can be performed by adding a known amount of a stable isotope-labeled standard peptide to a protein digest and subsequent comparison of the mass spectrometric signal to the endogenous peptide in the sample [213].

1.9.2. Proteomic data analysis

Peptide identification and quantification. After the LC-MS/MS analysis, the raw mass spectral data is analyzed via different bioinformatics tools. Peptide identification utilizes genomics data by matching mass measurements for intact peptides and MS/MS fragment ions to theoretical sequences derived from genomic data [206]. Therefore, in each search, the algorithm matches the mass of the peptide and resulting MS/MS spectrum against a proteomic sequence database to find a peptide that maximizes the peptide-spectrum matching score [214, 215]. A peptide-spectrum match (PSM) is the match between the spectrum and the highest-scoring peptide. Database search algorithms differ in the scoring function of PSMs (the similarities between calculated and experimental spectra) and in the statistical validation of results [216-218]. The scoring function evaluates the matching quality between a peptide and an MS/MS spectrum. Commonly used search engines include Mascot [219], SEQUEST [220], OMSSA [221], X!tandem [222] and Andromeda of MaxQuant [223, 224]. Scoring of PSMs is the most defining aspect of a search algorithm [225]. The search algorithm filter peptides at MS1 level based on the precursor mass

tolerance specified in the m/z, mass, or parts-per-million (ppm) and calculates a score from the number of matched peaks compared to peaks present in MS/MS (MS2) spectrum, which is linked to a probability [225]. The selection of precursor ions for fragmentation is performed with low resolution to ensure high sensitivity. In this case, co-eluting peptides with similar masses are often co-fragmented resulting in "chimeric" MS/MS spectra, which usually distort the detection and quantitation of peptides. The search engine Andromeda employs an algorithm that detects the "second" peptide and uses this information to increase the peptide identification rate [226]. The high-scoring PSM for a given set of spectra differs markedly among various search engine results. Combining multiple search engines in parallel increases the coverage and confidence of the PSMs, distinct peptide sequences and proteins identified [227]. Proteins are identified based on the fragment/peptide mass and fragmentation patterns from different cleaved peptides.

The database search algorithms produce a match for almost every input MS/MS spectrum [215]. The matches assigned by the search algorithms can have false positives and statistical validation of these false matches is required [228]. In certain cases, especially when the datasets are generated using low mass accuracy instrumentation, incorrect PSMs are the majority [215]. Other reasons for identification of false PSMs include low-quality spectrum, peptides not in the database, and imperfect scoring function. Thus, there is a need for the development of methods for assessing the confidence of PSMs and to estimate for the number of false positives present in the data processed. The most commonly used and accepted statistical confidence measures are the false discovery rate (FDR) and the posterior error probability (PEP) or local FDR for individual PSMs [215]. FDR, the expected proportion of incorrect PSMs and the total number of PSMs above a given score threshold, is the metric for global confidence assessment of a large-scale proteomics dataset [215, 228]. The PEP is the probability that the observed PSM is incorrect. Unlike FDR, PEP measures the error rate associated with a single PSM [228]. The target-decoy database search strategy has been widely used to estimate the FDR, where the database search is carried out on the true (target) as well as null (decoy) database [215, 218]. The basic assumption is that matches to decoy peptide sequences and false matches to sequences from the target database follow the same distribution [215, 218]. A decoy database can be created by shuffling, randomizing or by simple reversal of the target database.

Post-translational modifications (PTMs). Likewise, PTMs are identified either through the differences in observed and expected m/z (mass shift) or by their unique fragment ion spectra (reporter ion) [229]. Global identification of PTMs remains to be challenging for many reasons. This is primarily due to the diversity and complexity of PTMs, its complex fragmentation pattern

in tandem MS (poses a significant challenge for subsequent data analysis) and its low-abundance [230]. Furthermore, global proteomic studies are often limited to a specific PTM due to the prerequisite of effective enrichment strategies that target specific PTMs [229]. PTM identification by these algorithms is based on the alignment of MS/MS spectra to the protein sequences bearing a few specified PTMs attached to specific amino acids [230]. But when it comes to the identification of too many PTM sites, all methods face the same issue of the combinatorial explosion of theoretical peptides (increased number of potential molecular states), which increases the search time and the false positive rate [231]. Such complexity may provide the foundation for sophisticated forms of cellular information processing that are essential for the emergence of organismal complexity [230, 232]. It is therefore suggested to include a limited number of variable modifications during database search using conventional search engines [230]. The spectrum match algorithms are based on the similarity of mass shifts and retention times between the unmodified form and its modified counterpart [230, 231]. This may sometimes result in incorrect localization of PTM sites within the peptide and, thus, it is highly recommended to look for corroborating evidence via manual inspection of spectra, MS/MS analysis of the synthetic peptide with a specific modification or site-directed mutagenesis [230, 233]. The search algorithms assess the probability of correct site localization based on the presence and intensity of site-determining ions, as expressed by the localization probability score in Amdromeda [231].

High-resolution mass analyzers are essential to guarantee a mass accuracy sufficient to differentiate two nearly isobaric PTMs, such as acetylation and tri-methylation, 42.0106 Da and 42.0470 Da, respectively. Other considerations in any proteomic study include assignment of common chemical modifications like Gln→PyroGlu, Glu→PyroGlu, Oxidation (M), Carbamidomethylation (C), Acetyl (N-terminus), the use of unique and/or razor peptides for protein identification and assessment of data quality.

1.10. MTBC proteomics

MTBC proteomic studies are still in the infancy. The primary studies on Mtb proteomics were performed by two-dimensional gel electrophoresis (2D-GE)- based approach, which resulted in identification of limited amounts of Mtb proteins [234-237]. Further developments in the field of proteomic research have led to the establishment advanced technologies, such as liquid chromatography-tandem mass spectrometry (LC-MS/MS), which resulted in a better resolution and much higher protein identification than the 2D-GE based approach. Recent shotgun

proteomic studies on in-gel fractionated Mtb lysates resulted in identification of up to 3290 proteins, which corresponds to 82% of the 4023 theoretical Mtb proteome [9, 192, 238].

Expression proteomics also exploits the differential abundance of Mtb proteins as biomarkers for early and efficient diagnosis. Recent technological developments in clinical proteomics have improved our understanding of virulence, pathogenicity, pathogen-host interactions and drug resistance, thereby defining novel diagnostic markers, vaccine candidates and potential therapeutic targets [239]. The consequent comprehensive view of the changes in the proteomics and post-translational modifications that occur across MTBC together with genomic information and database comparisons, can reveal evidence of potential virulence factors and other clinically important proteins that can be further investigated through functional studies [240]. This, in turn, will have important implications not only in new drug and diagnostic target discovery and for the development of new drugs and vaccines, but also shed some light on pathways used by Mtb to successfully infect its host [241].

1.11. Post-translational modifications (PTMs) in MTBC pathogenesis

MTBC employs several mechanisms to evade the hostile host environment. PTMs confer diversity and complexity to the simpler genomes and proteomes of MTBC and play a significant role by regulating a myriad of cellular processes (Figure 7) [242]. PTMs influence the protein's active state, half-lives, compartmentalization, turnover, and/or interactions with other proteins [242, 243]. These modifications allow bacteria to rapidly alter protein activity in response to its environmental host factors and have been implicated in Mtb virulence and pathogenicity [126, 244, 245]. It is, therefore, a potentially central field of research to study PTMs in Mtb to characterize the complex layers of regulation contributing to its success as a pathogen [246]. Furthermore, it is becoming clear that prokaryotic modifications occur in parallel to create even more heterogeneity in the protein arsenal [229, 247] and are capable of cross-talk [248, 249] which also can interfere with host PTMs [250, 251], adding further complexity to the scope of regulatory modifications. Widely studied MTBC PTMs include phosphorylation, acetylation, glycosylation, methylation, deamidation, formylation, pupylation and lipidation [126, 243].

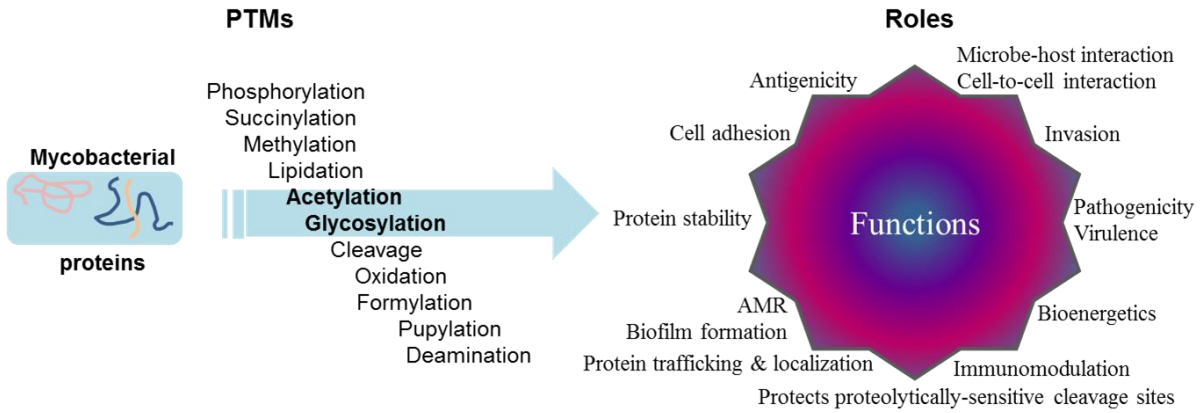


Figure 7. Diversity of functions regulated by mycobacterial post-translational modifications (PTMs) [252].

To date, MS remains to be the only technique used in large scale PTM studies and its potential have been explored in several reviews [253, 254]. Despite the vast multitude of proteomic studies performed over the last two decades, our insight into MTBC PTMs is still far from complete [243]. Furthermore, previous PTM studies have involved laborious isolation of a single modified protein, which sometimes have been discovered by chance [126]. Recent advances in the field of MS-based proteomics have enabled us to identify several thousands of MTBC proteins and unraveled the rich nature of PTMs in creating diversity and complexity of prokaryote gene products [246]. This in turn will lead to the identification of hundreds of putative virulence determinants, as well as an unsurpassed coverage of PTMs and, thereby, a better understanding of the pathogen and management of the disease [126].

This study analyzed two abundant PTMs, acetylation [126, 255, 256] and glycosylation in MTBC [113, 121, 126]. These modifications were selected as they are expected to play key roles in MTBC physiology. Acetylation is an important determinant of bacterial pathogenicity [248] that regulates the function of metabolic enzymes and carbon metabolism [257, 258]. On the other hand, carbohydrates are structurally diverse molecules commonly presented on the cell surface where they directly mediate specific cell-to-cell interactions [259].

1.11.1. Protein acetylation in the MTBC

Nε-acetylation, which is acetylation at the ε-amine of lysine (K) residues, is an abundant and evolutionarily conserved PTM that regulates a broad range of biological functions in bacteria, including motility and chemotaxis, transcription, metabolism, DNA metabolism, siderophore biosynthesis, and stress responses [246, 260, 261]. Increasing evidence supports the presence and role of lysine acetylation in Mtb [257, 262, 263]. In Mtb, lysine acetylation presumably confers

protein stability and compartmentalization, thereby modulating diverse cellular processes [243, 256, 264]. Nε-acetylation of Mtb histone-like nucleoid protein MtHU modulates DNA binding and genome organization [265]. It has also been shown that reversible lysine acetylation regulates fatty acid, acetate and propionate metabolism in mycobacteria [257, 258]. Liu *et al*. have demonstrated the regulatory role of lysine acetylation in the immunogenicity of the secreted protein HspX in Mtb [255]. Nε-lysine acetylation is modulated via both enzymatic and non-enzymatic mechanisms [261]. The enzymatic mechanism is regulated by the opposing actions of acetyltransferases and deacetylases. Nε-lysine acetylation can also occur non-enzymatically via direct interaction of the target protein and AcCoA [266].

O-acetylation is the acetylation at the -OH group of serine (S), threonine (T) and tyrosine (Y) by a bacterial acetyltransferase. In 2006, Mukherjee *at al.* discovered that YopJ, a secreted virulence factor from *Yersinia pestis,* acetylates critical serine and threonine residues and thereby inhibits kinase activation in the host by blocking phosphorylation [267]. That study was the first report regarding the presence and function of O-acetylation. Current evidence suggests that O-acetylation becomes even more important when acetylation takes place on kinase substrates, which is known to regulate myriads of signaling pathways in Mtb [244, 268]. The transfer of an acetyl group to the -OH group of serine [269] and glycoconjugates (peptidoglycans) [270, 271] has also been reported in bacteria. Nevertheless, O-acetylation has not been formerly described in MTBC.

1.11.2. Protein glycosylation in the MTBC

Protein glycosylation predominantly occurs in surface-exposed cell envelope and membrane proteins, which holds the key roles to MTBC microbe-host interactions, pathogenesis and antimicrobial resistance (AMR) [110, 113, 272]. The unique structure, antigenicity and essentiality of Mtb cell envelope glycoconjugates in pathogen-host interactions and mycobacterial growth provide opportunities for the development of novel diagnostics, biomarkers, drugs and vaccines [110, 273]. MTBC glycoproteins are involved in cell adhesion and invasion, protein stability, localization, function and life time, maintenance of protein conformation, cellular signaling and adhesion, AMR, immunomodulation, intracellular bacterial survival, biofilm and protein complex formation, protein antigenicity, pathogenicity and virulence [113, 121, 126, 274, 275]. Post-translational glycosylation has been associated with low cell envelope permeability and AMR in *M. abscessus* [276]. Furthermore, glycosylation protects proteolytically

sensitive cleavage site, thereby maintaining the membrane association of the protein by its lipid anchor. It may also be linked to protein export and other functions [274, 277-279].

Glycoproteomics is likely to identify Mtb virulence factors because glycoproteins on the bacterial cell envelope are the ones used by mycobacteria to enter its main host cell, the MØs [259]. It has been proposed that Mtb interacts with mannose receptors on host cells via its mannosylated proteins to enter the MØs [280]. Despite their vital importance in Mtb pathogenesis, our knowledge of Mtb glycoproteins is limited and only a few secreted and cell wall-associated MTBC glycoproteins have so far been described [126, 259, 273]. It is demanding to fully characterize protein glycosylation due to the size, heterogeneity and complexity of the modifications [203].

1.12. Comparative proteomics to characterize MTBC virulence factors

Model approaches used for characterization of MTBC virulence factors include animal models, macrophages, molecular profiling (genomics, transcriptomics, some proteomics, lipidomics, etc), gene disruption and overexpression techniques, and other methods [76]. The genetic variation across members of the MTBC has been studied in depth, but the clinical and epidemiological consequences of the relatively sparse genetic differences detected is poorly understood [281]. Compared to genomics and transcriptomics, proteomics has the advantage of defining proteins that are differentially expressed, localized or secreted [127]. Proteomics is a powerful tool to identify potential biomarkers and virulence factors involved in bacterial adaptation to antibiotics, host factors or other environmental stimuli [282], as well as host responses against the bacilli [283]. The post-translationally modified proteins cannot be assessed by using genomics or transcriptomics. The expression of virulence factors is often highly variable within the dynamic pathogen-host interaction, and their mere presence is, therefore, an insufficient indicator of virulence [284]. Furthermore, strain virulence is determined by the interaction between a multitude of virulence factors that are better assessed globally through large-scale comparison of their relative abundance [285]. The proteomic approach is particularly important when comparing the molecular mechanisms of virulence in closely related strains which differ in virulence or phenotype [9, 197, 286, 287]. Comparative shotgun proteomic analysis with a gel-based or gel-free fractionation approach have been widely used in identification of virulence factors and potential therapeutic targets in a number of pathogenic bacteria, including MTBC [127, 207, 288].

Current knowledge on comparative proteomic studies in MTBC has been limited to the reference strains including H37Rv, H37Ra and BCG [9, 285-287]. These studies have identified several differentially abundant virulence factors across different strains (Table 1). Thus, including representative clinical strains exhibiting different clinical phenotype would provide a better understanding of MTBC virulence factors. Likewise, only limited knowledge is available on the global analysis of the two physiologically important MTBC PTMs, acetylation [255, 256] and glycosylation [273, 289]. The former studies have focused on the global profiling of protein Nε-acetylation and O-glycosylation using reference strains (Table 1) [255, 256, 273, 289]. Most glycoproteomics studies are based on the analysis of hexose (mannose) residues attached to proteins. We suggest that global analysis of both N- and O-forms of acetylation and an extended armory of glycosylation variants will broaden our understanding of the role of these PTMs in MTBC fitness and survival. Until now, comparative PTM studies have not been conducted in the various members of the MTBC. Proteomic and PTM studies using valid clinical representatives of the MTBC lineages are therefore highly warranted. Current knowledge on MS-based proteomic and PTM in MTBC is presented under Table 1 below.

Table 1. Summary of mass spectrometry (MS)-based proteomic and post-translational modification (PTM) studies on the *Mycobacterium tuberculosis* complex (MTBC) and their respective number of identified proteins, PTMs and differentially abundant proteins.

Studies	**Strains used in the study**	**Proteins**	**References**
Proteomics	H37Rv (M)	32	[290]
	H37Rv	3176 and 1051	[192, 291]
	M. bovis BCG (M)	351	[292]
	H37Rv (M)	1417	[293]
	H37Rv	2884	[238]
	H37Rv and *M. bovis* BCG	2458 and 2518	[294]
Studies	**Strains used in the study**	**DA proteins**	**References**
Comparative proteomics	Hypo- and Hypervirulent Beijing Isolates	101/1668	[295]
	M. bovis, *M. bovis* BCG, CAS, H37Rv, Beijing, LAM	23/3788	[9]
	H37Rv, H37Ra, CAS (BND-433 and JAL-2287)	257/2161	[286]
	H37Rv and H37Ra	627/2709	[296]
	H37Rv and H37Ra	Not specified	[287]
	Beijing BO/W148 and H37Rv	266/1868	[297]
	H37Rv and H37Ra (M)	29/1771	[298]
	H37Rv and *M. bovis* BCG (M)	294/2203	[299]
PTMs	**Strains used in the study**	**PTM sites**	**References**
Nε-acetylation	H37Ra	226	[255]
	H37Rv	1128	[256]
O-acetylation	Not reported	Not reported	
O-glycosylation	H37Rv (CFP)	NA*	[289]
	Mtb (CFP)	34	[273]
N-glycosylation	Not reported	Not reported	

DA: differentially abundant, CFP: culture filtrate protein, M: membrane fraction, NA*: 41 glycoproteins were identified and the number of PTM sites were not stated, Central Asian spoligotype family (CAS): lineage 3, Latin American-Mediterranean (LAM): lineage 4, Beijing: lineage 2.

1.13. Conceptual framework of the study

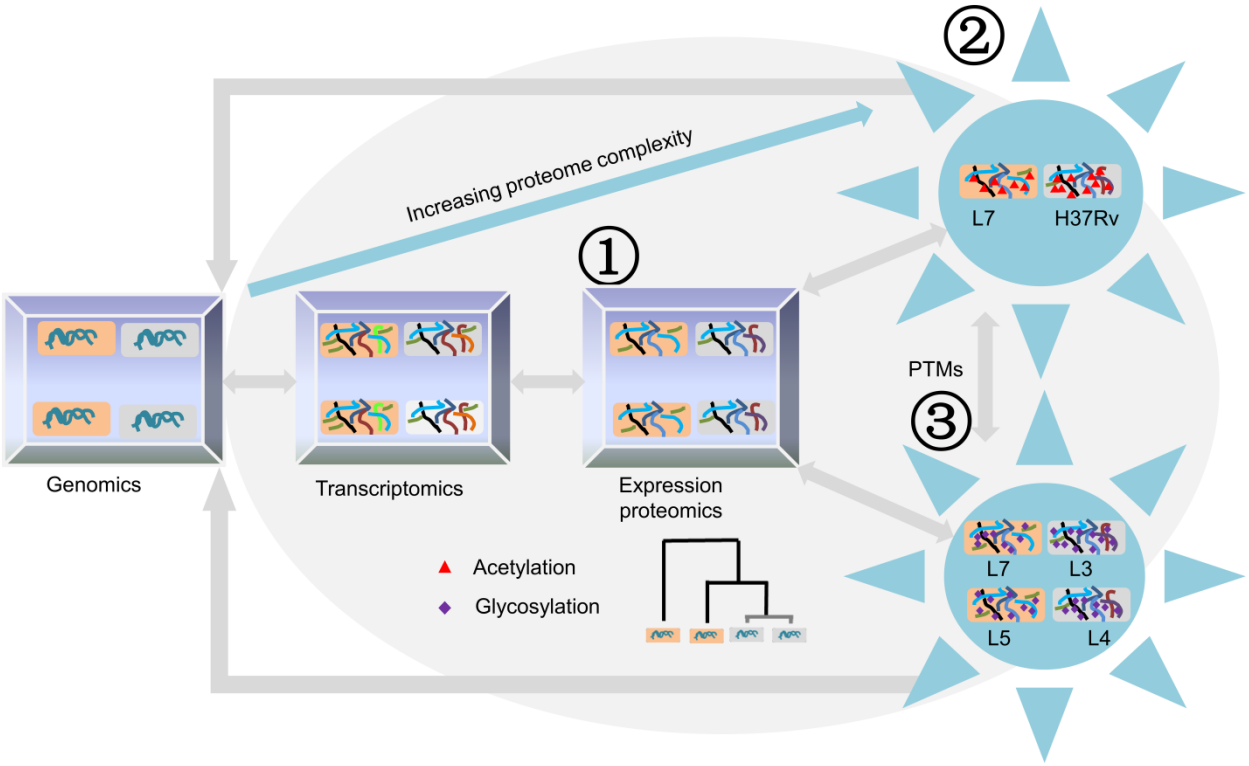


Figure 8. A schematic model of the global levels of regulatory mechanisms in MTBC.This work was performed pursuing the hypothesis that MTBC strain phenotypic diversity across lineages is vast and is greater than what can be predicted based on genomics and transcriptomics alone. Changes in the expression of genes and production of functional proteins allow the pathogen to develop an adaptive potential, enabling it to survive within the dynamics of pathogen-host interactions. The presence or absence as well as differential abundance of proteins among MTBC may contribute to the fitness and survival of the bacilli (**objective 1**). Furthermore, protein function is regulated via PTMs, including acetylation (**objective 2**) and glycosylation (**objective 3**), which can result in multiplying the different variants of the encoding gene. (L3, lineage 3; L4, lineage 4; L5, lineage 5 and L7, lineage 7).

2. AIM OF THE STUDY

The main aim of this study was to analyse the qualitative and quantitative differences in the proteome and PTM profiles of representative MTBC members so as to elucidate the phenotypic heterogeneities among these genetically conserved lineages. This core knowledge will expand the mycobacterial proteome and PTM catalog, which in turn will lead to a better understanding of the pathogen and potentially to the identification of novel drug targets and vaccine candidates, as well as efficient diagnostics for TB disease.

Specific objectives

1) To compare the whole cell proteome profiles of the reference laboratory strain H37Rv (lineage 4) and slow-growing clinical lineage 7 strains of MTBC.

2) To characterize the Nε-acetylation signatures and search for the presence of O-acetylation in reference laboratory strain H37Rv (lineage 4) and clinical lineage 7 strains of MTBC.

3) To explore the O-linked glycoproteomic patterns and search for the presence of N-linked glycosylation across clinical strains of the MTBC lineages 3, 4, 5 and 7.

3. SUMMARY OF PAPERS

Paper I: Comparative expression proteomics

In order to decipher the nature of the slowly growing MTBC lineage 7, the differentially abundant proteins of MTBC lineage 7 and lineage 4H37Rv strains were defined. Comparative proteomic analysis by mass spectrometry (MS) was employed to identify, quantitate and compare the protein profiles of strains from the two MTBC lineages.

Label-free peptide quantification of whole cells from MTBC lineage 7 and H37Rv yielded the identification of 2825 and 2541 proteins, respectively. A combined total of 2867 protein groups covering 71% of the predicted Mtb proteome were identified. The abundance of 125 proteins in MTBC lineage 7 and H37Rv strains was significantly altered. Notably, the analysis showed that a number of proteins involved in growth and virulence were less abundant in lineage 7 strains compared to H37Rv strains. These include the ABC transporter proteins, phosphate binding proteins and a number of ESX-3 secretion system proteins.

This proteomic analysis provided an insight into the lineage 7-specific protein profile which may provide clues to understanding the differential properties of lineage 7 strains in terms of slow growth, fitness for survival and pathogenesis.

Paper II: Nε- and O-acetylation of MTBC proteins

Protein Nε-acetylation is an important determinant of bacterial virulence and pathogenicity. However, the role of O-acetylation in Mtb is poorly understood. This study presents the global Nε- and O-acetylome profile, compare the variation in the acetylation profile and elucidate its potential biological significance in MTBC lineage 7 and H37Rv.

MS-based global Nε- and O-acetylome analysis resulted in identification of 2490 class-I acetylation sites derived from 953 proteins. The acetylated proteins identified are involved in carbon metabolism, translation, stress responses and drug resistance. Out of 165 differentially acetylated proteins, 161 proteins were found to be hypo-acetylated in lineage 7 strains than in H37Rv. These hypo-acetylated proteins are involved in growth, virulence, energy metabolism, pathogen-host interaction and stress responses. Furthermore, Gene Ontology (GO) analysis of exclusively identified acetylated proteins revealed strain specific enrichment of specific

biological processes. For example, carbon and fatty acid metabolism were enriched in lineage 7 strains, whereas translation and amino acid metabolism were enriched in H37Rv.

The study provides the first report on the presence and global analysis of O-acetylated proteins in Mtb. This acetylome data presents the abundance and diversity of acetylated proteins and opens new avenue of research in exploring the role of protein acetylation in MTBC physiology.

Paper III: N- and O-glycosylation of MTBC proteins

Protein glycosylation is one of the PTMs involved in regulating a myriad of cellular processes in MTBC, including pathogen-host interaction. Despite the vital importance in Mtb virulence and pathogenesis, our current knowledge of these Mtb glycoproteins is still in its infancy, and only a few secreted and cell wall-associated glycoproteins have been described to date. The study used qualitative and quantitative mass spectrometry to define the global N- and O-linked glycosylation profiles, compare the variation in glycosylation profile and elucidate its biological significance in MTBC lineages 3, 4, 5 and 7 clinical strains.

Glycoproteomic analysis revealed 2944 high confidence glycosylation events derived from 1325 unique proteins. This data set represents the highest number of glycosylated proteins identified in Mtb to date and provides the first report on N-glycosylation of MTBC proteins. Bioinformatics analysis revealed that a number of surface-exposed proteins involved MTBC fitness and survival are glycosylated. Quantitative glycoproteomics revealed that 101 events on 67 proteins were differentially glycosylated between the four MTBC lineages. Notably, more than 64% of the differentially glycosylated proteins were found to be cell envelope-associated proteins involved in cell wall synthesis, protein secretion, virulence, detoxification and adaptation. The differential glycosylation pattern may contribute to the phenotypic variability found among the genetically conserved MTBC. This study thus opens a new avenue of research by elucidating the role of protein glycosylation in the physiology of MTBC clinical strains.

4. GENERAL DISCUSSION

Members of the MTBC have relatively static genomes exhibiting low rates of mutation and recombination, with 99.9% nucleotide sequence identity [9]. Despite the high degree of genetic conservation, members of the MTBC display variable geographical distribution and clinical presentations [9, 18, 19, 47, 49]. Studies have also shown that only certain strains of Mtb are clinically predominant, with only a few representatives causing most of the disease outbreaks and multi-drug resistance [47, 48]. The contribution of genetic/genomic studies in explaining clinical and epidemiological consequences between members of the MTBC is limited [281]. However, compared to genomics, MTBC transcriptomics reveal a vivid variability in their RNA levels. Ultimately, the vast molecular changes variability at the level of proteomics and PTMs increase the MTBC cellular complexity and diversity even further, thereby providing the most direct and most precise insight into the finite nature of MTBC cells. Proteomics is a powerful tool to identify bacterial virulence factors, biomarkers, differentially expressed proteins, their localization and analyse PTMs, regulators of major cellular processes [127, 282]. Deciphering the contribution of quantitative differences in proteome and PTM profiles to the variability in phenotype and virulence across members of the MTBC is still at a pre-mature stage [9]. The present study analysed the proteomics of the slowly growing MTBC lineage 7 as well as the profiles of two abundant PTMs, protein acetylation and glycosylation, which are regulators of metabolic enzymes and carbon metabolism [257, 258] and mediators of cell-to-cell interactions [259], respectively. The MTBC strains included in this study are all pathogenic, however, the extent to which they cause disease in humans greatly varies.

The study identified qualitative and quantitative differences in protein and PTM (acetylation and glycosylation) abundances, and a number of uniquely identified proteins and PTMs among different lineages of MTBC. Quantitative proteomic analysis showed differential abundance of MTBC proteins involved in phosphate transport, type 7 secretion system, cell wall biosynthesis and energy metabolism in the slowly growing lineage 7 versus lineage 4. These proteins have been shown to be implicated in MTBC virulence and growth [1, 300-307] (**Paper I**). Most of these proteins were downregulated in the slow growing and less virulent lineage 7 strains than in H37Rv [41]. The acetylome analysis revealed that a number of cytosolic proteins involved in protein synthesis, energy metabolism and fatty acid metabolism were acetylated. Furthermore, the study showed that proteins involved in MTBC virulence, AMR, stress response and pathogen-host interaction were differentially acetylated between lineage 7 and lineage 4 H37Rv strains (**Paper II**). Protein acetylation has been shown to be a determinant of bacterial

pathogenicity [245]. Notably, **Paper II** presented the first evidence for the presence of protein O-acetylation in bacteria/MTBC. In contrast to the findings on acetylation, the combined N- and O-glycoproteomic study revealed that a number of cell envelope- and membrane-associated proteins with a prime role in virulence and pathogenesis (pathogen-host interaction) were glycosylated (**Paper III**). Quantitative glycoproteomic analysis demonstrated differential glycosylation of cell surface-associated virulence factors, with the pattern similar to their respective phylogeny. **Paper III** presented the first evidence for the presence of N-linked protein glycosylation in MTBC and in Gram-positive bacteria. The presence qualitative and quantitative differences in physiologically important proteins and PTMs, with high level PTM variability between MTBC lineages indicate the potential role of proteomic and PTMs studies in explaining phenotypic variability between MTBC strains. The findings presented here shed light on the possible role of selected proteins and PTMs in MTBC fitness and survival as well as in new drug target discovery and for the development of new drugs and vaccines. The MS proteomics data for the three manuscripts were deposited in PRIDE/ProteomeXchange with the accession numbers PXD006117 (Paper I), PXD006630 (Paper II) and PXD009676 (Paper III).

4.1. Comparative expression proteomics

4.1.1. Mapping the MTBC proteome by discovery mass spectrometry

In-gel based shotgun proteomics yielded the identification of 2867 Mtb proteins, which accounts for 71.8% of the 3993 annotated Mtb proteins at a dynamic range of six orders of magnitude. Recent studies have reported even higher proteome coverages between 77-82% of the predicted Mtb proteome [9, 192], while others have reported relatively lower proteome coverages of 46% [297], 54% [286] and 62% [294]. Quantitative analysis showed differential abundance of proteins involved in virulence and growth, such as membrane transport proteins (phosphate transporters, type 7 secretion system proteins), cell wall biosynthesis and bioenergetics.

4.1.2. Differential abundance of transport proteins

Proteins involved in membrane transport including phosphate-specific transport (Pst) system proteins, type 7 secretion system proteins and transmembrane transporter MmpL5 were differentially regulated between lineage 7 and H37Rv strains. The Pst system proteins PstS2, PstB, and PstS1 were downregulated in lineage 7 strains. ABC transporter proteins PstA, PstC, PstS, and PstB are essential for inorganic phosphate (Pi) uptake in Mtb [308]. Pi is an essential component of DNA, RNA, ATP, phospholipids, and proteins, and is crucial for energy transfer,

protein activation, and carbon and amino acid metabolic processes [309]. Mtb cells with disruption in genes encoding the *pstS* system were shown to be deficient in phosphate uptake, less virulent and attenuated [300]. The two proteins RegX3 and PhoR were upregulated in lineage 7 strains. These proteins have been shown to be involved in an important mechanism for Mtb survival under phosphate starvation [300, 301].

Components of the Mtb type 7 secretion system proteins, particularly Esx-1 (EspB) and Esx-3 (EccA3, EccC3, EccD3, EccB3, EccE3, MycP3), were downregulated in lineage 7 strains. ESX-3 and ESX-1are crucial for virulence or viability in Mtb [302]. MmpL5, a membrane transporter protein downregulated in lineage 7 strains, is involved in the export of siderophores for the bacterial acquisition of iron and Mtb drug resistance [303].

4.1.3. Differential abundance of proteins involved in cell wall and lipid biosynthesis

The two enzymes involved in peptidoglycan biosynthesis [310], MurE and MurF, were downregulated in lineage 7 strains. These enzymes are potential targets for discovery of novel antibiotics [311]. The methyl transferase Rv2952 is another downregulated protein involved in the biosynthesis of complex cell wall lipids, phenolglycolipid (PGL) and PDIM, which are unique to pathogenic mycobacteria [304]. These molecules are Mtb virulence factors playing a key role in permeability barrier and in bacterial pathogenicity [305]. Conversely, phenolpthiocerol synthesis type-I polyketide synthase (PpsC and PpsE) and mycolic acid synthesis enzymes (AccD6 and KasA) were upregulated in lineage 7 strains. These proteins are involved in the biosynthesis of Mtb cell wall components [312-314]. Cell envelope integrity is crucial for protection of Mtb against environmental factors. Therefore, the differential regulation of proteins involved in this process may have an effect on MTBC growth and survival.

4.1.4. Differential abundance of proteins involved in bioenergetics

A number of proteins involved in the tricarboxylic acid (TCA) cycle, including CitA, BkdA and BkdB, were downregulated in lineage 7 strains. TCA cycle generates NADH, a key metabolite involved in redox homeostasis, energy generation and other metabolic processes in mycobacteria [306]. ABC transporters CydC and CydD are involved in energy metabolism (aerobic respiration) and combats nitrosative stress [307].

Proteins mainly involved in energy generation during anaerobic condition were upregulated in lineage 7 strains. These include proteins FrdA, LpdA and components of the

cytochrome bc1-aa3 respiratory network. FrdA and LpdA play a role in energy generation during anaerobic condition [315, 316]. In addition, the AceaA and AceaB enzymes involved in glyoxylate cycle that serves as an alternative to the TCA cycle, were upregulated in lineage 7 strains. The upregulation of proteins involved in anaerobic respiration in lineage 7 strains, may be considered as a compensatory mechanism for generating energy to maintain basic physiological functions.

4.2. Protein acetylation in the MTBC

4.2.1. First combined N_{ε}- and O-acetylome map of the MTBC

Acetylome analysis identified 2490 class-I acetylation sites matched to 953 unique proteins, harboring up to 25 acetylation sites per protein. The acetylated proteins identified accounted for 23.87% of the total proteins annotated in Mtb. Highly acetylated proteins are known to be involved in stress responses in Mtb and other bacteria [255, 317]. The abundance of acetylation was highest on T residues (39.32%), followed by S (36.39%), Y (18.80%), and K residues (5.66%). The proportion O-acetylation is similar to the phosphorylation profile on the S, T and Y residues [318]. The high number of acetylated proteins identified in MTBC cells may reflect the vast diversity of metabolic processes regulated by acetylation, as shown in the Gene Ontology analysis and protein-protein interaction network analysis.

Nε-acetylation is the most commonly studied acetylation both in eukaryotes and prokaryotes [260]. A recent study on the Mtb acetylome, using an anti-acetyllysine antibody-enriched sample, reported a total of 1128 Nε-acetylation sites on 658 proteins [256]. O-acetylation is a common modification of bacterial peptidoglycan and other glycoconjugates [270, 271]. Furthermore, the bacterial O-acetyltransferase YopJ has been reported to regulate host signaling via acetylation of serine and threonine residues in the activation loop of mitogen-activated protein kinase (MAPK) and thereby blocking phosphorylation [267]. Furthermore, N-hydroxyarylamine O-acetyltransferase (NhoA) from *Salmonella typhimurium* has been shown to exert bacterial O-acetyltransferase activity [319]. This study presented the first report on O-acetylation of Mtb proteins. On the basis of the evidence discussed so far, including the YopJ O-acetyltransferase, we propose that the mechanism of O-acetylation and deacetylation of S, T and Y residues in MTBC probably follows a similar pattern as N_{ε}-acetylation. Further investigations are required to elucidate the responsible mechanisms.

The acetylome analysis in Paper II was based on the same 162 MS raw files that were used in the primary proteomic study in Paper I , involving two lineage 7 strains and the lineage 4

reference strain H37Rv [320]. Considering the differences in the MaxQuant output files and the project goals between the two parts of the analysis, the MS raw data was deposited with separate PRIDE/ProteomeXchange dataset identifiers.

4.2.2. Regulatory role of N_ε- and O-acetylation

The polar amino acid residues including lysine (K), tyrosine (Y), serine (S) and threonine (T) are the most frequent catalytic amino acid residues [321, 322]. Acetylation of these major active site residues may affect protein activity [267, 323]. Furthermore, S, T and Y are the major substrates for protein kinases that are hubs for complex regulatory networks and are involved in blocking phagosome-lysosome fusion, the hallmark of Mtb pathogenesis [244, 268]. The competitive inhibition between acetylation and phosphorylation for the same residues may have implications in fine-tuning biological processes in Mtb. Acetylation of K residues within enzyme active site increases the net negative charge on lysine residues [324, 325], which results in concomitant alteration in protein activity [326], protein-protein and protein-DNA interactions [327], local protein conformation [328], and protein localization [325, 329], reviewed in [261].

Pyruvate kinase (Pyk) active site signature was one of the PROSITE motifs identified, with acetylation at position K221. In accordance with this finding, mutagenesis of the active site K221 of Pyk was shown to reduce the enzyme activity by a factor of 10^4 to 10^5 [330]. Even though data is lacking on the role of O-acetylation in bacteria, it is possible that direct acetylation of such active site residues may modulate, abolish or induce the enzyme activity [331] or interfere with the phosphorylation event [332].

Nε-acetylation regulates cellular metabolism via various mechanisms such as enzymatic activation or inhibition, and by influencing protein stability [255, 333, 334]. Acetyl-CoA synthetase (ACS) is an acetylated protein involved in the synthesis of acetyl coenzyme A (AcCoA), a key intermediate in energy metabolism and an acetyl group donor in protein acetylation. Once generated, it is used in ATP synthesis via the TCA cycle, synthesis of fatty acids, amino acids and other metabolites, reviewed in [335]. ACS was the first enzyme in Mtb shown to be regulated by reversible post-translational acetylation via cAMP-dependent protein acetyltransferase [262, 263, 336]. Thus, acetylation of enzymes involved in AcCoA synthesis, ACS, may also modulate all the processes involving AcCoA. The acetylation status of ACS determines its activity that may influence the availability of the acetyl donor, AcCoA, and the metabolic state of the cell [335].

A number of acetylated enzymes involved in the TCA cycle and two enzymes, ICL and GlcB, involved in the glyoxylate cycle were identified [333, 337]. GlcB is also a virulence factor involved in Mtb adherence to lung epithelial cells [338]. Both copies of the ICL genes (*aceaA* and *aceaB*) are essential for Mtb survival *in vivo* [339]. His356 is a catalytic active site residue of ICL in *E. coli* [340]. AceAb was acetylated at S355, near the catalytic residue H356, which may affect the enzyme activity. Acetylation of AceA has been associated with a reduction in AceA activity, and enzyme activity was restored by deacetylation [333]. When the TCA cycle is down, the glyoxylate cycle takes the role of replenishment of TCA cycle intermediates using AcCoA from fatty acid β-oxidation [341, 342]. Protein acetylation has been shown to regulate the activity of enzymes controlling the direction of glycolysis versus gluconeogenesis, and the branching between TCA cycle and glyoxylate cycle by acetyltransferase (Pat) and deacetylase [333].

Fatty acid metabolism is among the pathways identified by GO enrichment analysis. Besides being a source of AcCoA, fatty acids are major Mtb cell wall component, which is associated with Mtb pathogenicity [343]. Reversible protein acetylation has been shown to regulate the activity of a number of fatty-acid-CoA ligases in Mtb [257, 334]. Further functional studies are required to validate these claims.

4.2.3. Gene Ontology analysis of acetylated proteins

GO analysis showed that Nε- and O-acetylated proteins are involved in similar metabolic processes as previously reported for Nε-acetylated proteins [255, 256]. However, a separate GO analysis of exclusively O-acetylated proteins and N_{ε}-acetylated proteins provided a distinct profile of biological processes and KEGG pathways associated with each of these forms of acetylation. O-acetylated proteins are involved in a broad range of KEGG pathway and biological processes than the Nε-acetylated proteins. O-acetylation occurs in a higher stoichiometry than N_{ε}-acetylation. These frequent occurrences together with their competitive inhibition of kinase substrates may contribute to the broader role of O-acetylated proteins in regulating bacterial physiology [331, 332]. Further enrichment-dependent methods for each of the four residues are necessary to complete the catalog profiling of Nε- and O-acetylated proteins in MTBC.

4.2.4. Acetylated MTBC proteins with roles in antimicrobial resistance (AMR)

Mtb modifies proteins involved in AMR or drug targets to decrease or block the affinity for drug binding without affecting normal activity [331]. A number of proteins, including KatG, InhA, NdhA KasA, AhpC, FadE24, and AcpM, which are all associated with isoniazid (INH) resistance

were acetylated [331, 344]. KatG is responsible for peroxidative activation of the prodrug INH and acts as a virulence factor to protect Mtb against oxidative stress [345]. Mutations at KatG positions S315T and R463L have been shown to diminish its capacity to activate INH and confer INH resistance [346]. Succinylation of KatG at K310 near the S315T mutation assists the enzyme in retaining its native antioxidant activity, while the INH activating property was reduced by almost 30% with the minimum inhibitory concentration of bacteria increased up to 200-fold [331]. Acetylation on KatG S465 was found near the natural mutation R463L, which may affect the protein activity.

Serine/threonine protein kinases (STPKs) and two-component signal transduction systems (TCS) are key regulators of metabolic processes, including drug resistance [347, 348]. Four out of the eleven Mtb STPKs, namely PknD, PknK, PknG and PknH, were acetylated at various residues. OpcA and Wag31 have been shown to be up-regulated in INH-resistant Mtb strains [286, 331]. OpcA and Wag31 are involved in peptidoglycan biosynthesis and oxidative stress responses [349, 350]. MurF is another acetylated protein involved in cell wall synthesis and implicated in vancomycin resistance in Mtb [310]. Acetylated protein PpiA is involved in cationic antimicrobial peptide (CAMP) resistance [351].

InhA is a major enzyme involved in the biosynthesis of mycolic acids. Mutation at InhA I74T has been associated with resistance to ethambutol (EMB), INH, rifampicin (RMP), and streptomycin (SM) [352]. We identified an acetylation site on InhA at position T79 which is only five amino acids away from the natural mutation I74T. Other acetylated proteins with a role in resistance to first-line anti-TB drugs include RpoB, EmbR, PhoP, FabG1 and RpsL [344, 353].

Mutations in the genes encoding DNA gyrase subunits, *gyrA* and *gyrB*, are the most common mechanisms for acquiring fluoroquinolone (FQ) resistance in Mtb [354]. The quinolone resistance-determining region resides between codons 74 to 113 in *gyrA* and between codons 461 to 538 in *gyrB* [354-356]. The two acetylation sites identified, T500 and S473, were located within the quinolone resistance-determining region of GyrB, and may play a role in drug resistance, DNA replication, and Mtb survival. Another acetylated protein involved in drug resistance is enhanced intracellular survival (Eis), an acetyltransferase, which confers resistance to kanamycin by modifying the drug [357, 358]. Thus, acetylation of a protein in an active site residue or anywhere in the protein sequence may alter the protein function in various modes.

4.2.5. Quantitative acetylome analysis between MTBC lineage 7 and H37Rv strains

Quantitative analysis was performed by considering those hits having at least five valid values from a total of nine technical replicates per sample, which resulted in 1085 acetylation sites (Paper II). The missing values were imputed from the normal distribution to perform a statistical test. Student T-test at FDR 0.05 resulted in 261 differentially acetylated sites on 165 proteins, out of which 257 acetylation sites on 160 proteins were hypo-acetylated in lineage 7 strains compared to the lineage 4 H37Rv strain. The lineage 7 strain is the most recently identified lineage of Mtb, characterized by slow-growth and reduced virulence phenotypes [41, 320]. Proteins involved in carbon and fatty acid metabolism, stress response, growth, virulence and the Esx-3 secretion system were hypo-acetylated in lineage 7 strains. Our former proteomic study using the same sample revealed that proteins involved in growth and virulence were differentially abundant between lineage 7 strain and H37Rv [320].

The two TCA cycle enzymes, DlaT and AcnA, were hypo-acetylated at lysine residues in lineage 7 strains. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and enolase (Eno) are two enzymes involved in the glycolytic pathway exhibiting reduced acetylation. Energy metabolism is associated with growth and virulence in intracellular bacteria, including Mtb [359, 360]. Enzymes involved in *Salmonella enterica* central carbon metabolism are regulated by reversible lysine acetylation [333]. In *S. enterica*, acetylation of GAPDH has been shown to favor the glycolytic pathway while inhibiting gluconeogenesis by more than 30%, while deacetylation of GAPDH by the sirtuin CobB stimulates gluconeogenesis and inhibits the glycolytic pathway [333]. Furthermore, it has been shown that *S. enterica* with deficient in the deacetylase *cobB* (with high acetylation) grew faster than the wild-type cells in minimal glucose medium, but grew slower than the wild-type in minimal citrate medium. This shows the importance of protein acetylation in regulating bacterial growth [333]. The activity of pyruvate dehydrogenase, one of the hypo-acetylated enzymes involved in the TCA cycle, is reduced by lysine acetylation in eukaryotes [361, 362].

ACS exhibited reduced acetylation in MTBC lineage 7 strains. AcCoA, produced by ACS, is the substrate for the glyoxylate cycle [263]. The glyoxylate cycle has been shown to be up-regulated during Mtb dormancy and persistence [363, 364]. Moreover, it has been shown that ACS is involved in controlling the expression and/or activity of the glyoxylate cycle in *E. coli* [365, 366]. The acetylation status of Mtb ACS is crucial in modulating Mtb physiology and persistence [262, 263] (as discussed above in section 4.2.2). LldD, the enzyme that catalyzes the

conversion of L-lactate to pyruvate, was hypo-acetylated in MTBC lineage 7 strains compared to H37Rv. Mtb utilizes lactate as a potential carbon and energy source inside MØs [367].

Furthermore, a number of long-chain-fatty-acid-CoA ligases associated with fatty acid metabolism were hypo-acetylated in MTBC lineage 7 strains. Fatty acids and their derivatives are an integral component of the cell wall complex and implicated in Mtb pathogenicity, fitness and survival [255, 343, 368]. Mtb utilizes fatty acids as a principal source of energy during dormancy and reactivation [360, 369]. Survival of Mtb inside the phagolysosome depends on the pathogen's ability to synthesize virulence factors and other biomolecules in a glucose-limited stress environment [342]. Reversible acetylation has been shown to modulate the activity of several fatty-acid-CoA ligases [334]. PKS13, an enzyme involved in the final steps of mycolic acid biosynthesis, was hypo-acetylated in lineage 7 strains. Mycolic acids are an integral component of Mtb cell wall and known to be related to its pathogenicity [343].

Other groups of clinically important hypo-acetylated proteins include ESX-1 and ESX-3 secretion system components, Pst system proteins, chaperone proteins, oxidoreductases, Mce proteins, STPKs, and proteins associated with drug resistance. These proteins are indispensable for Mtb in signal transduction mechanisms that lead to bacterial adaptation to its environment [348], detoxification and drug resistance [346, 370], or involved in entry and survival of the pathogen inside MØs [371]. Esx-3 is implicated in divalent metal homeostasis, immune modulation and is crucial for Mtb growth *in vivo* and *in vitro* [372]. PstS 1 is involved in inorganic phosphate uptake and its disruption has been shown to be associated with decreased virulence and attenuated growth [300]. Although the exact mechanism remains unknown, acetylation of enzymes involved in Mtb fitness and survival may lead to a change in the net charge of the protein, alter stability and compartmentalization, brings a conformational change and/or block kinase substrates and may modulate enzyme activity. In conclusion, acetylation of a number of clinically important proteins involved in diverse metabolic functions with a significant qualitative and quantitative the difference across the two MTBC lineages may indicate the impeccable role of this PTM in regulating the bacilli physiology.

4.3. Protein glycosylation in the MTBC

4.3.1. MTBC clinical strains exhibit both O- and N-glycosylation

A broad array of 57 target glycan residues were included in the analysis (Paper III, Supplementary Table S1). The glycoproteomic analysis of MTBC strains representing lineages 3, 4, 5 and 7 led to the identification of 2944 glycosylation events on 1325 unique proteins (Figure

8A, 8C). The term "glycosylation event" is used to avoid confusion since a single site can be glycosylated by more than one type of glycan residue. O-glycosylation constituted 2455 (83.4%) of the events identified (1311 events at T and 1144 events at S residues) and the remaining 489 sites (17%) were glycosylated at N residues (Figure 8B). To our knowledge, the discovery of such a high number of glycosylation sites in these four lineages of MTBC is unprecedented. Glycoproteomic analysis revealed that 945 (32.1%) of the total glycosylation events identified on 587 (44.2%) of the glycoproteins were shared amongst the four MTBC lineages. Analysis of the proteome abundance among the four lineages of MTBC showed that 94.7% of the proteins were shared between the four lineages (Figure 8D). The very high versatility at the level of PTMs may indicate the power of PTMs in explaining the variability among MTBC, much more than the proteomic profiles alone. The study also provides the first report on N-glycosylation of MTBC proteins.

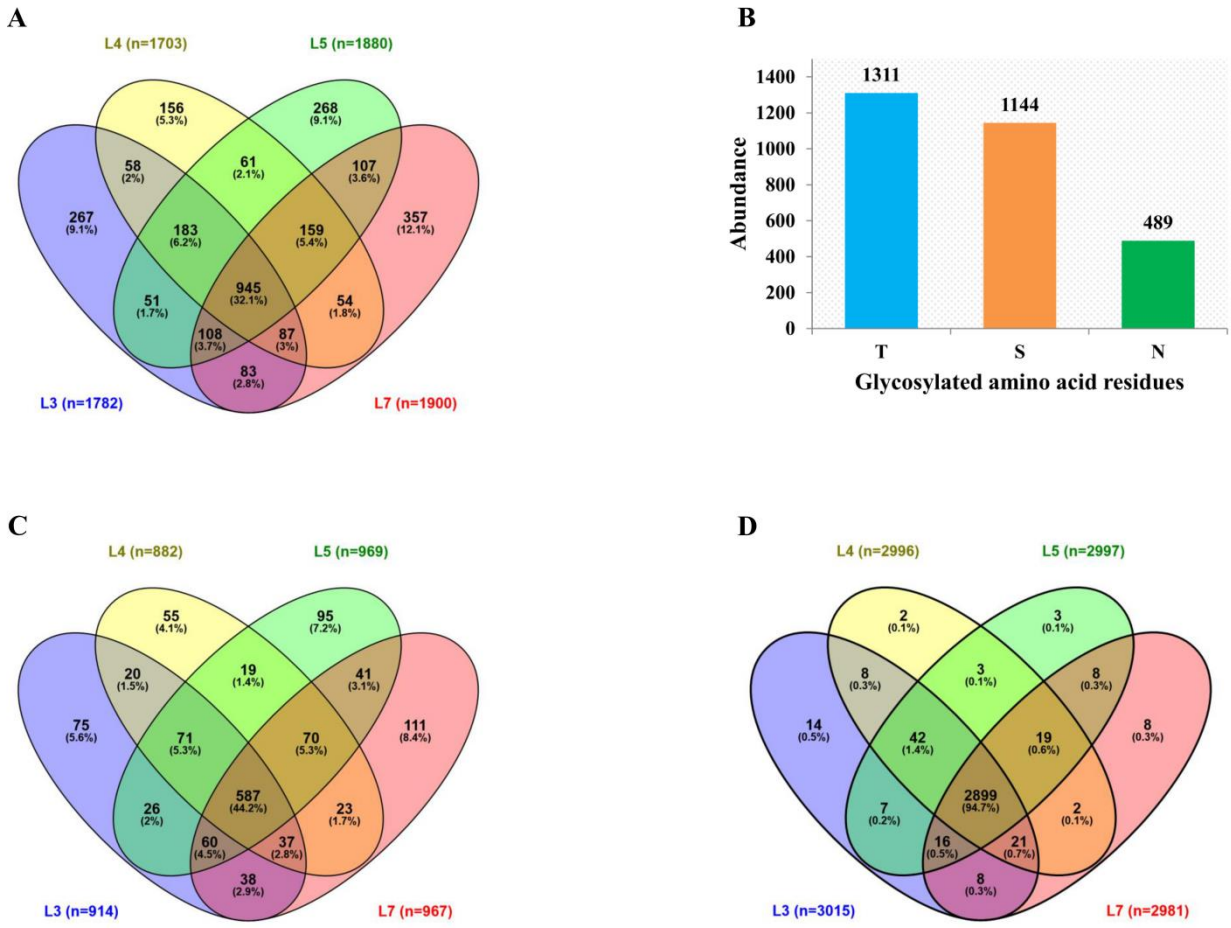


Figure 8. Abundance of glycosylation events and glycoproteins in the *M. tuberculosis* complex (MTBC). Number of glycosylation events identified among the four lineages (N=2944) (A), Number of O- and N-glycosylation in MTBC (N=2944) (B), number of glycoproteins among the four lineages (N=1325) (C) and the number of proteins identified in the four lineages of MTBC (N=3060) (D).

Protein O-glycosylation occurred at numerous sites on surface-exposed proteins with no apparent amino acid sequence specificity [272]. As previously reported, there is a relatively high propensity for R, A, P, L, G, V, S and T flanking the modified sites in a significant portion of the O-glycosylation sites mapped [273, 275, 373]. No sequence specificity was recorded for the N-glycosylation, and the MTBC genomes do not harbour homologs of the pgl genes encoding enzymes required for N-glycosylation in Gram-negative bacteria [374]. A number of suggested signature motifs were identified where R was enriched between the -8 and +8 positions in all the cases in contrast to the D/E-Y-N-X-S/T (Y, X #P) motif proposed for N-glycosylation [375]. The fact that N-linked glycosylation in MTBC occurs without sequence specificity, at sites other than those shown in other species, may indicate that mycobacteria might have a system for N-linked glycosylation which is different from those known in other organisms. This versatility in N-glycosylation motif might be partly explained by the broad diversity of the glycan residues included in the analysis and the high degree of specificity in glycosyltransferases [376]. The occurrence of the majority of glycosylation events on S and T residues indicate a possible interplay between glycosylation and phosphorylation. Glycosylation events may occur on the same S and T residues of the protein or competitively at adjacent or residues in close proximity, and hence potentially allow control of cellular signaling [373].

4.3.2. Gene Ontology analysis of glycoproteins

Based on the Gene Ontology (GO) analysis, fatty acid metabolism and lipid homeostasis, growth of symbiont in host cell and responses to host immune response were the most highly enriched biological processes. The cell wall and the plasma membrane were the two most enriched cellular components (Figure 9A). The molecular functions include ATP binding, oxidoreductase activity, acyl-CoA dehydrogenase activity, fatty-acyl-CoA binding, helicase activity, DNA binding, electron carrier activity and ligase activity (Figure 9A). The GO analysis showed that the majority of the glycoproteins identified were localized in the cell wall and plasma membrane while lipid homeostasis, fatty acid metabolism, and responses to the host immune response were among the biological processes enriched, which was in agreement with former reports [272]. The non-glycosylated complement, however, encompassed proteins with functions and localization primarily in the cytoplasm. A separate GO analysis of O- and N-glycosylated proteins revealed that O-glycosylated proteins are involved in a broader range of biological processes than the N-glycosylated proteins (Figure 9B). Furthermore, GO analysis of proteins with uniquely identified glycosylation sites showed strain specific enrichment of biological processes (Figure 9C).

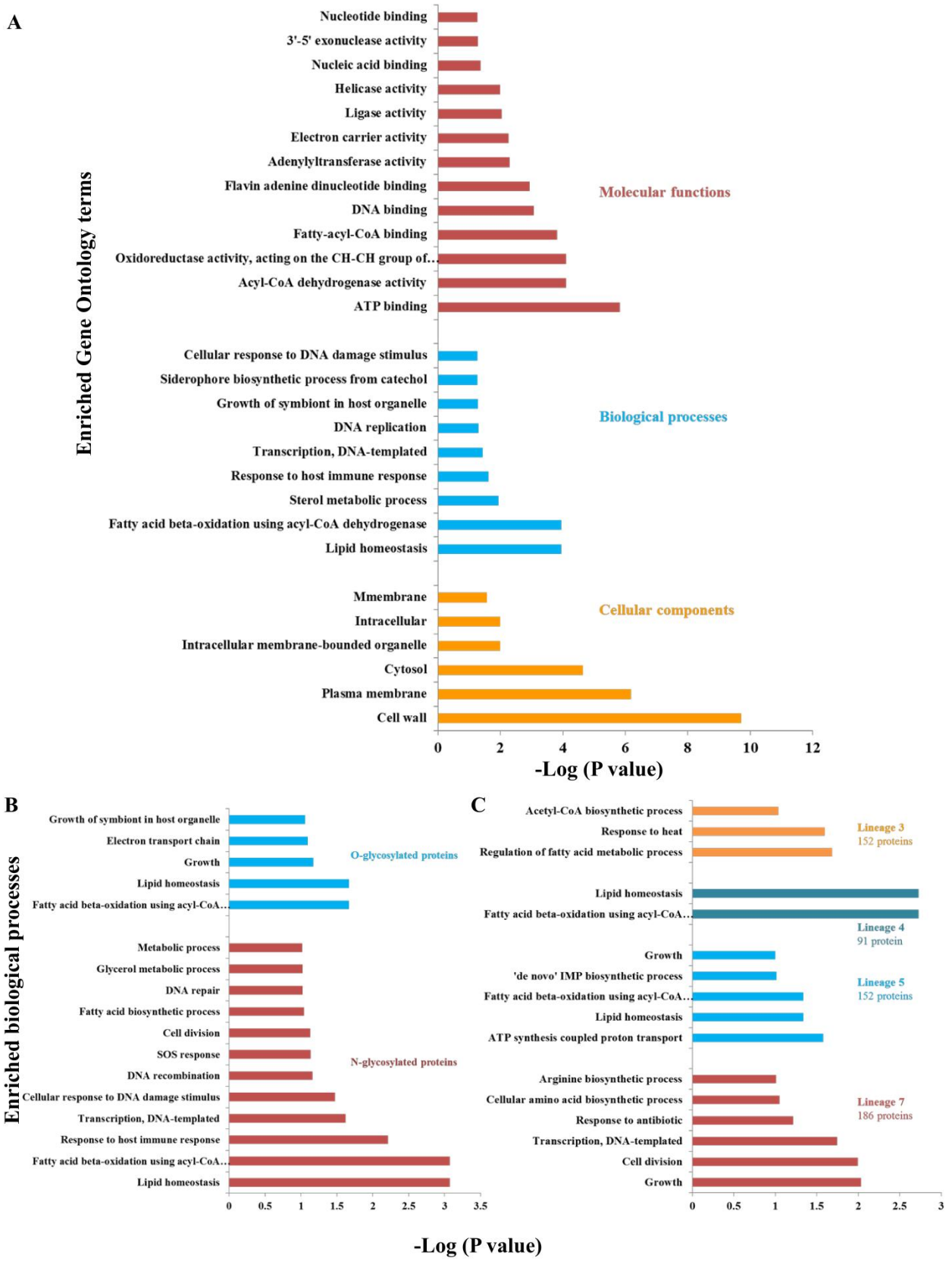


Figure 9. Gene Ontology (GO) analysis. GO analysis of identified MTBC glycoproteins (A). A separate GO analysis of O- and N-glycosylated proteins (B) and enriched biological processes from uniquely identified glycosylation sites among the four lineages (C).

4.3.3. Glycoproteins involved in pathogen-host interaction

The outermost layer of the Mtb cell envelope is a major determinant of virulence and pathogenicity and is mainly composed of proteins, polysaccharides and lipids [113, 377]. Furthermore, the cell envelope acts as a permeability barrier and play a role in promoting the phagocytosis of MTBC [378], maintaining cell integrity, regulating phagosome maturation [379] and in pathogen-host interactions [377, 378].

Lipoproteins were amongst the highly glycosylated Mtb proteins identified in this study. They are known to be involved in colonization, invasion, evasion of host defence and immunomodulation, cell envelope biogenesis, transport across membrane, nutrient acquisition, adhesion, cell invasion and initiation of inflammatory processes [159, 276, 279, 380, 381]. The lipoprotein LpqH was the most heavily glycosylated lipoprotein among all identified glycoproteins, with N-terminally clustered 33 O-mannosylation events. Some of the glycosylation sites have previously been reported as part of the mannose receptors (MR) binding domain of LpqH [273, 382]. Three of the sites, T41, S43 and S48, are part of a mature protein fragment (residues 41-60), that was reported to prevent uptake of Mtb by MØ-like U937 cell line [383]. Mannosylated LpqH is the major adhesin for the MØ MR and dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) and thereby stimulating phagocytosis and subsequent production of IL-1, IL-12 and TNF-α [280]. LpqH, an immunodominant TLR2 agonist, is crucial for Mtb growth and replication in mice, even in IFN-γ-deficient mice [384]. Furthermore, prolonged exposure to LpqH inhibits IFN-γ production and MHC class II expression, leading to persistence of Mtb infection [159]. Altering the glycosylated S residues in LpqH has been shown to affect binding affinity and exposure to proteolytic cleavage [274].

Another major group of identified glycolipoproteins involved in pathogen-host interactions are the MCE-family of proteins [380]. These glycoproteins play an active role in TB disease development and in-host virulence [113]. A total of 14 glycosylation events were identified on proteins expressed from the four Mtb *mce* operons (*mce1*, *mce2*, *mce3* and *mce4*). The invasion-/adhesin-like MCE family glycolipoproteins encoded by *mce*s are located at the cell surface of Mtb and possibly involved in entry and survival inside MØs [385].

Many other clinically important glycoproteins were identified. The cell surface glycoprotein Apa binds to DC-SIGN and surfactant protein, facilitates colonization and invasion of host cells [386]. Changes in the glycosylation pattern of Apa lead to a reduced stimulatory T-

lymphocyte response, exhibiting the biological role of the glycan moiety [387]. The immunogenic glycoprotein MPT64 is a virulence factor involved in Mtb infection of human cells and is a promising candidate for a subunit-based anti-TB vaccine [388]. Heparin-binding hemagglutinin (HbhA) glycoprotein mediates adherence to epithelial cells and is required for extrapulmonary dissemination of Mtb [389]. LprG is another glycolipoprotein that blocks host cell phagosome-lysosome fusion, and is required for full Mtb virulence [390]. LprG also controls cell wall permeability and efflux of drugs, and therefore plays a role in Mtb susceptibility to first-line anti-TB drugs [391].

4.3.4. Glycoproteins involved in MTBC cell envelope biogenesis

Membrane-associated proteins involved in lipid and fatty acid metabolism, cell envelope biosynthesis, pathogen-host interaction, transport, transcriptional regulation, and chaperone functions were also glycosylated. After LpqH, the meromycolate extension acyl carrier protein AcpM was the second most highly glycosylated protein identified with 21 glycosylation events. AcpM is involved in mycolic acid biosynthesis, a major component of the Mtb cell wall. Glycosylation sites S41, S43 and T51 were detected within the AcpM carrier protein (CP) domain profile. Importantly, one of the glycosylation sites identified (S41) is the binding site for 4'-phosphopantetheine, an activator of AcpM [392]. Other glycoproteins involved in mycolic acid synthesis include methoxy mycolic acid synthases, mycolic acid synthase PcaA, polyketide synthase Pks13, beta-ketoacyl-ACP synthases (KasA, KasB) and mycolic acid biosynthesis protein FabG1. Bacilli lacking all mycolic acid methyltransferases are viable but highly attenuated and hyperinflammatory in mice [148]. For example, Pks13 catalyzes the last condensation step of mycolic acid biosynthesis and is essential for the mycobacterial survival [393]. Glycoproteins FbpB and FbpC also possess a mycolyltransferase activity [118]. These mycolic acid synthase help to maintain the Mtb cell wall integrity by catalyzing the transfer of mycolic acids to cell wall AG, and through the synthesis of the virulence factor TDM [118]. FbpB and FbpC are T- and B-cell antigens, and may have an application in sero-diagnostics [394].

Penicillin-binding glycoproteins (PBPs) PbpA, PbpB, PonA1 and LdtA are transpeptidases involved in the synthesis of cross-linked PG that is part of the cell wall biogenesis [395]. Glycoproteins involved in PG biosynthesis include MurA, MurE, MurF, LprQ, FtsW, MviN, GlmS, GlmM, DacB1 and Wag31 [396]. Mur enzymes and PBPs are potential antibiotic targets [396]. Alpha-(1→3)-arabinofuranosyltransferase (AftD) is involved in the biosynthesis of the AG region of the mAGP complex, an essential component of the mycobacterial cell wall

[115]. EmbC is involved in the polymerization of arabinose into the arabinan of the mycobacterial cell wall AG and is linked to resistance to EMB [397].

A number of polyketide synthases, PapA1, Rv2951c, FadD26 and FadD28 are involved in the synthesis of Mtb cell wall component, PDIM and other lipids [312], while glycolipoproteins LppX, DrrC and MmpL7 are required for the translocation and localization of PDIM in the cell wall [398]. PDIMs are comprised of a number of virulence-enhancing lipids that act as defensive, offensive, or adaptive effectors of virulence [399]. Inactivation of mycobacterial *pps* and *drr* operons has been linked to defects in PDIM synthesis and secretion, respectively [400]. PknD, a regulator of MmpL7, has been proposed to be a potential anti-TB drug target [401].

Glycoproteins such as GlgM, GlgB, GlgE, TreZ and MalQ are enzymes involved in the biosynthesis of alpha-D-glucan, a constituent of Mtb capsular polysaccharides with D-arabino-D-mannan (AM) and D-mannan [110, 377]. These enzymes are required for Mtb virulence [402]. GlgE-mediated 1,4 α-glucan synthesis has been implicated in *in vitro* lysosomal stress and can potentially be exploited for killing intracellular Mtb [403]. The T10 glycosylation site in GlgE has been shown to be a regulatory kinase substrate and a validated anti-TB drug target [404]. GlgB is a potential target for inhibitors [405]. Glycosylated mannosyltransferases PimB and Rv2181 are involved in the biosynthesis of lipoglycans LM, LAM and phosphatidylinositol (PI) [406]. Mannosyltransferases MgtA, Ppm1 and Rv1459c are involved in the synthesis of immunomodulatory LM and LAM via alpha-(1→6)-mannopyranosyltransferase activity [407]. A number of glycosylated fatty acyl-AMP ligases that have been shown to play a role in cell wall biosynthesis, production of complex lipids and growth were identified [408].

4.3.5. Other clinically important glycoproteins identified

MTBC BlaC was glycosylated at 12 sites. BlaC hydrolyzes nitrocefin and other β-lactams, thereby increasing Mtb resistance towards the various classes of β-lactam antibiotics [409]. Chaperone and TCS proteins harbor many glycosylation sites. For example, the chaperones GroEL2 and GroS were found to have 18 and 10 glycosylation sites, respectively. The differential expression of chaperone glycoproteins in response to heat shock has previously been reported [113]. TCS regulate various aspects of mycobacterial physiology, including virulence, dormancy, persistence, and drug resistance [410]. The glycoprotein PhoPR regulates multiple virulence-associated processes in Mtb, including the biosynthesis of polyketide-derived lipids and acyltrehaloses. The inactivation of acyltrehaloses attenuates Mtb sufficiently to make it a possible

live vaccine candidate [112, 113]. The DosR/WhiB3 regulon is associated with hypoxia and redox adaptation, while WhiB3/PhoP is involved in cell wall lipid biosynthesis [411]. The DevS/DosR regulon is required for full Mtb virulence and is involved in regulating stress, dormancy and hypoxia [412].

A number of ESX secretion system glycoproteins, including the crucial T-cell antigen ESAT-6, were identified. The ESX secretion system is essential for full Mtb virulence (ESX-1) and physiological processes (ESX-3) [413]. The ABC transporter DrrC, Rv0194, Rv2994, Rv1273c and a number of MmpL glycoproteins are efflux pumps for anti-TB drugs, contributing to AMR [414, 415]. Apart from having roles in AMR, MmpLs are involved in the export of cell wall associated lipids and siderophores, and thus are attractive pharmacological targets [416, 417]. Proteins involved in the general secretion (Sec) pathway and twin-arginine translocation pathways (TatB) were also glycosylated. These proteins are essential for bulk export of proteins in Mtb [413]. CpnT, the first autotransporter-like protein to be identified in Mtb, was glycosylated at a domain that is required for the membrane localization of this protein [418].

4.3.6. Quantitative glycoproteomic analysis in the MTBC

Mapping the exclusive presence and/or differential abundance of Mtb glycoproteins naturally or during exposure to environmental stress or infection contributes to elucidating the selective advantages and survival strategies adopted by a specific pathogen. This information is fundamental for drug or vaccine discovery process [113]. Quantitative analysis revealed that 101 events on 67 proteins were differentially glycosylated between the four MTBC lineages (Supplementary Table S2). Notably, more than 64% of the differentially glycosylated proteins were found to be cell envelope-associated proteins involved in cell wall synthesis, protein secretion, virulence, detoxification and adaptation. The MTBC strains exhibited lineage-specific glycoproteomic profiles. The hierarchical clustering of the differentially glycosylated proteins coincided with the phylogeny of MTBC, where the modern lineages (lineage 3 and lineage 4 strains) clustered together, separated from the ancient lineages (lineage 5 and lineage 7 strains) [42].

Clinically important proteins such as FadE35, LppW, LdtA, PurK, PPE42 and UvrA were hyper-glycosylated in lineage 3 and lineage 4 strains than in the ancient lineages [395]. PurK has been identified to be a high-confidence drug target [419] while antigen PPE42 is known to elicit humoral immune response against Mtb [420]. Glycoproteins including LprG, Mce2D, DagK, UvrC and VapC45 were hyper-glycosylated in all lineages except lineage 7 strains. LprG plays a

role in transport and localization of the TLR2 agonists, LAM, PIM, LM and triacylglycerides to the cell surface, maintaining cell envelope integrity, and inhibition of phagosome-lysosome fusion, thereby enhancing Mtb survival inside MØs [390, 391]. DagK is involved in the biosynthesis of Mtb virulence factors PI and PIMs [421]. The membrane proteins DsbF and Rv0412c, iron-sulfur cluster carrier protein Mrp, ATPase MoxR3 and PknA were hyper-glycosylated in the ancient lineages. Proteins involved in lipid metabolism such as FadD9, FadD34 PG synthesizing enzyme like DacB1, an oxidoreductases BpoB and Rv2766c were hyper-glycosylated in lineage 7 strains.

Four glycosylation sites on GroeL1, a chaperone involved in mycolic acid biosynthesis during biofilm formation [422], were uniformly hyper-glycosylated in lineage 4 strains. PbpB, another hyper-glycosylated lineage 4 strains, is an essential enzyme involved in peptidoglycan biosynthesis and has been predicted to be an important drug target [423]. These proteins are essential virulence factors involved in cell wall biosynthesis, stress response, immunomodulation, efficient host cell invasion, survival, growth and other physiological processes [1]. The relative abundance of these essential glycoproteins across the different lineages of Mtb may lead to a specific phenotype with better adaptability to the host.

Table 2. Summary of novel findings discovered in this study.

Discovery	**Potential implications/ inferences**	**Reference**
Differential abundance of a number of proteins involved in bacterial fitness and survival	• Explains the potential basis for phenotypic variation among H37Rv (lineage 4) and lineage 7 strains	**Paper I**
O-acetylation in MTBC	• First report on the presence and regulatory potential of O-acetylation in MTBC	**Paper II**
First combined Nε- and O-acetylation in MTBC	• Improved understanding of the role of acetylation in MTBC physiology particularly in regulating the function of metabolic enzymes and carbon metabolism	**Paper II**
N- glycosylation in MTBC	• First report on N-glycosylation in MTBC	**Paper III**
Combined N- and O-acetylation and glycosylation in MTBC	• The first comparative study in MTBC at the level of PTMs • First discovery of such an abundance of acetylated proteins and glycoproteins in MTBC clinical strains	**Papers II, III**

4.4. Methodological considerations

We employed an in-gel based discovery proteomics in all the experiments. In-gel digestion of proteins isolated by gel electrophoresis, originally introduced in 1996 [424], is robust, reproducible and effective method [425]. Sequencing of sharp, molecular weight-separated protein bands increases the dynamic range of protein mixtures thereby enabling the analysis of entire proteomes of cell lysates [426]. In-gel digestion has been shown to provide a significantly higher number of eukaryotic plasma membrane protein identifications than the in-solution and on-filter digestion methods [425]. However, it is laborious and time-consuming protein sequencing method [425]. Other drawbacks of the method include the low recovery of large and/or hydrophobic digested peptides from the gel and limited accessibility of the enzyme or chemical cleavage to the proteins embedded into the gel [427]. We think that overnight digestion in a thermos-shaker can improve digestion efficiency of trypsin while the multistep protein extraction step will ensure efficient protein recovery.

In-solution digestion-based method requires less processing time, but it introduces sample losses due to a low re-solubilization of aggregated proteins [428]. On-filter-based digestion has been shown to provide better digestion efficiency especially for identification of membrane proteins where large and hydrophobic peptides are present [427, 429]. The universality and reproducibility of this method and the loss of protein on the filter are still the concerns [430, 431].

Despite all the achievements, characterizing the vast array of modified protein forms in the proteome continues to be the challenge [229]. Another challenge in PTM studies is their low abundance in the proteome and therefore the requirement of specific enrichment strategies to target only a specific type of PTMs prior to MS analysis [248]. The enrichment-dependent techniques, however, do not provide a comprehensive PTM profile of a proteins unless a separate enrichment is used for each PTM types [432, 433] It has also been reported that PTM site identification can be influenced by the antibodies used for immunoaffinity enrichment of the modified peptides [261]. Furthermore, the enrichment-dependent technique does not allow direct quantification of PTM fractional occupancy, the percentage of protein copies modified with specific PTM event [434-436]. The use of sequential enrichment strategy has been described for the analysis of multiple PTMs from the same biological sample [433], but peptides with multiple PTMs will be underestimated with this strategy.

However, we think that further advancements in the field of MS-based proteomics and the in-gel protein fractionation will enable us to identify several thousands of Mtb proteins and PTMs with improved analytical depth, reproducibility and coverage [437]. For example, previously the detection of PTMs, acetylation in particular, was challenging due to the requirement of relatively massive material and the difficulties of analysis in distinguishing acetylation modification from tri-methylation. Most of the problems have been solved with the availability of more sensitive analytical proteomic instrumentation, such as the tandem mass spectrometer [438]. Finally, modification-specific antibodies could be generated to permit the use of immunoaffinity-enriched samples for nano LC-MS/MS analysis, which may enable the identification and quantification of a complete set of a specific PTM and to establish the global PTM map of MTBC.

Our experiments emphasize on the sources of strain variation in natural conditions using cellular proteins, but the real impact of the differentially regulated proteins and PTMs will be examined when the bacilli is challenges with stressors mimicking the natural host responses, MØ based assays or via the use of other infection models. Furthermore, the analysis of the proteome and PTM contents of MTBC secretome and membrane proteins, which constitute the major set of virulence factors, will lead to improved understanding of the pathogen and the development of effective methods to diagnose and tackle TB disease, some of which are currently underway.

5. LIMITATIONS OF THE STUDY

Due to limited capacity, the study focuses on selected representative clinical and reference strains of MTBC lineages. We employed the classic reference laboratory strain H37Rv from lineage 4 and clinical lineage 7 strains for the comparative expression proteomics and acetylation studies. For the glycoproteomic study we included clinical strains representing lineages 3, 4, 5 and 7. The use of all representative strains from each lineage of MTBC, including lineages 1, 2 and 6, would have provided an even better overview on the potential of proteomics and PTMs in explaining the variability in phenotype and geographical distributions associated to each lineage. The use of multiple strains per lineage would also have contributed to deeper characterization of the core proteome and PTM of each lineage. Due to capacity and time limits for the slow-growing MTBC strains, the study only focused on the MS data and bioinformatics analysis. The search algorithms used in the MS data analysis may sometimes result in incorrect localization of PTM sites [230, 231]. The use of genetic and other molecular biology tools are required to validate and further elucidate the real impacts of these (modified) proteins in MTBC fitness and survival.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

Through shotgun-based proteomic and PTM analysis between MTBC strains of varying phenotype, we have discovered previously unknown, statistically significant quantitative differences in the expression of numerous proteins and PTMs. These findings shed new light on differential virulence in MTBC strains. In particular, the proteomics data suggests more strain-specific bacterial fitness in the H37Rv strains than in lineage 7 strains, including increased abundance of proteins involved in iron scavenging in the host, bioenergetics, stress responses and a number of other virulence factors (**Paper I**).

Our acetylome result provides the first extensive data on the combined Nε- and O-acetylation in MTBC H37Rv and lineage 7 strains. The frequent occurrences of O-acetylation in MTBC may lead to the hypothesis that it is involved in regulating a wider range of metabolic processes than Nε-acetylation, as evidenced in the bioinformatics analysis. Quantitative acetylome analysis revealed differential acetylation of proteins involved in growth, virulence, bioenergetics, pathogen-host interactions, and stress responses (**Paper II**).

MTBC face the host environment with a coat of many glycans, which are predominant in the cell envelope. Identification of glycoproteins and elucidating their contribution to bacterial physiology is an important aspect of understanding host-microbe interactions, which is fundamental for diagnostic, drug or vaccine discovery process. This study identified a number of glycoproteins, most of which are cell envelope- and membrane-associated. These glycoproteins are involved in cell wall and mycomembrane biosynthesis, pathogen-host interaction, membrane transport, transcriptional regulation and chaperone functions. Quantitative analysis showed differential glycosylation profile of proteins in a lineage specific manner. The quantitative analysis revealed a specific clustering of the modern and ancient lineages, which is in agreement with the phylogenetic relationship of the strains (**Paper III**).

The study provides novel insight regarding the nature and diversity of the modified proteins in MTBC and opens a new avenue of research to elucidate the range of functions regulated by protein acetylation and glycosylation. Furthermore, results from the proteomic and PTM analysis showed the presence of significant quantitative differences virulence-associated proteins across the various MTBC lineages. Differential regulation of proteins involved in virulence, energy metabolism, growth and stress responses among members of MTBC may lead to a metabolic state that makes one strain better adapted to a certain environment than the other.

Due to the limited number of functional studies on the impact of protein acetylation and glycosylation conducted to date, the contribution of these PTMs in MTBC physiology is not fully understood. The findings from the proteomics and PTM analysis combined with genetic manipulation can explain the mechanism of MTBC gene regulation and generate new insights into the molecular basis for TB pathogenesis. This study provides new knowledge on post-translationally modified proteins and sites on a number of clinically important proteins, which is the basis for future functional studies. Furthermore, the use of immunoaffinity-enriched samples for MS analysis may enable the identification of large numbers of acetylated and glycosylated proteins to establish the complete PTM catalog.

REFERENCES

1. Forrellad, M.A., et al., Virulence factors of the Mycobacterium tuberculosis complex. Virulence, 2013. **4**(1): p. 3-66.
2. Koch, R., The etiology of tuberculosis. Reviews of infectious diseases, 1982. **4**(6): p. 1270-1274.
3. WHO, Global Tuberculosis Report 2018. http://www.who.int/tb/publications/global_report/en/ (Accessed on November 27, 2018).
4. Hershkovitz, I., et al., Detection and molecular characterization of 9000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. PloS one, 2008. **3**(10): p. e3426.
5. Huard, R.C., et al., Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. Journal of bacteriology, 2006. **188**(12): p. 4271-4287.
6. van Ingen, J., et al., Characterization of Mycobacterium orygis as M. tuberculosis complex subspecies. Emerg Infect Dis, 2012. **18**(4): p. 653-5.
7. Bayraktar, B., et al., Species distribution of the Mycobacterium tuberculosis complex in clinical isolates from 2007 to 2010 in Turkey: a prospective study. J Clin Microbiol, 2011. **49**(11): p. 3837-41.
8. Murray, P.R., K.S. Rosenthal, and M.A. Pfaller, Medical microbiology. 2015: Elsevier Health Sciences.
9. Peters, J.S., et al., Identification of quantitative proteomic differences between Mycobacterium tuberculosis lineages with altered virulence. Frontiers in microbiology, 2016. **7**.
10. Abel, L., et al., Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. Lancet Infect Dis, 2018. **18**(3): p. e64-e75.
11. O'Garra, A., et al., The immune response in tuberculosis. Annu Rev Immunol, 2013. **31**: p. 475-527.
12. Donoghue, H.D., Human tuberculosis--an ancient disease, as elucidated by ancient microbial biomolecules. Microbes Infect, 2009. **11**(14-15): p. 1156-62.
13. Dheda, K., C.E. Barry, and G. Maartens, Tuberculosis. The Lancet, 2016. **387**(10024): p. 1211-1226.
14. Queirós, J., et al., Genome-wide associations identify novel candidate loci associated with genetic susceptibility to tuberculosis in wild boar. Scientific Reports, 2018. **8**(1): p. 1980.
15. Caws, M., et al., The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoS Pathog, 2008. **4**(3): p. e1000034.
16. 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.
17. Coscolla, M. and S. Gagneux. Consequences of genomic diversity in Mycobacterium tuberculosis. in Seminars in immunology. 2014. Elsevier.
18. Manabe, Y.C., et al., Different strains of Mycobacterium tuberculosis cause various spectrums of disease in the rabbit model of tuberculosis. Infection and immunity, 2003. **71**(10): p. 6004-6011.
19. Nicol, M.P. and R.J. Wilkinson, The clinical consequences of strain diversity in Mycobacterium tuberculosis. Transactions of the Royal Society of Tropical Medicine and Hygiene, 2008. **102**(10): p. 955-965.
20. Lienhardt, C., From exposure to disease: the role of environmental factors in susceptibility to and development of tuberculosis. Epidemiologic reviews, 2001. **23**(2): p. 288-301.
21. Narasimhan, P., et al., Risk factors for tuberculosis. Pulmonary medicine, 2013. **2013**.
22. Donoghue, H.D., Insights gained from palaeomicrobiology into ancient and modern tuberculosis. Clinical Microbiology and Infection, 2011. **17**(6): p. 821-829.
23. Van Embden, J., et al., Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. Journal of clinical microbiology, 1993. **31**(2): p. 406-409.

24. Brudey, K., et al., Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC microbiology, 2006. **6**(1): p. 23.
25. 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.
26. Supply, P., et al., Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. Journal of clinical microbiology, 2006. **44**(12): p. 4498-4510.
27. Oelemann, M.C., et al., Assessment of an optimized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. Journal of clinical microbiology, 2007. **45**(3): p. 691-697.
28. Maes, M., et al., 24-locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. Tuberculosis, 2008. **88**(5): p. 490-494.
29. Gagneux, S., Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control. Vol. 1019. 2017: Springer.
30. Alland, D., et al., Role of large sequence polymorphisms (LSPs) in generating genomic diversity among clinical isolates of Mycobacterium tuberculosis and the utility of LSPs in phylogenetic analysis. Journal of clinical microbiology, 2007. **45**(1): p. 39-46.
31. Mathema, B., et al., Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev, 2006. **19**(4): p. 658-85.
32. Brosch, R., et al., A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proceedings of the national academy of Sciences, 2002. **99**(6): p. 3684-3689.
33. Richter, E., et al., Usefulness of the GenoType MTBC assay for differentiating species of the Mycobacterium tuberculosis complex in cultures obtained from clinical specimens. Journal of clinical microbiology, 2004. **42**(9): p. 4303-4306.
34. Richter, E., et al., Evaluation of genotype MTBC assay for differentiation of clinical Mycobacterium tuberculosis complex isolates. Journal of Clinical Microbiology, 2003. **41**(6): p. 2672-2675.
35. Neonakis, I., et al., Evaluation of the GenoType MTBC assay for differentiating 120 clinical Mycobacterium tuberculosis complex isolates. European Journal of Clinical Microbiology & Infectious Diseases, 2007. **26**(2): p. 151-152.
36. Bouakaze, C., et al., Identification and genotyping of Mycobacterium tuberculosis complex species by use of a SNaPshot Minisequencing-based assay. Journal of clinical microbiology, 2010. **48**(5): p. 1758-1766.
37. Cole, S., et al., Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature, 1998. **393**(6685): p. 537-544.
38. Garnier, T., et al., The complete genome sequence of Mycobacterium bovis. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7877-82.
39. Smith, N.H., et al., Ecotypes of the Mycobacterium tuberculosis complex. J Theor Biol, 2006. **239**(2): p. 220-5.
40. Brites, D. and S. Gagneux, The Nature and Evolution of Genomic Diversity in the Mycobacterium tuberculosis Complex, in Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control. 2017, Springer. p. 1-26.
41. Yimer, S.A., et al., Mycobacterium tuberculosis lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. J Clin Microbiol, 2015. **53**(4): p. 1301-9.
42. Comas, I., et al., Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet, 2013. **45**(10): p. 1176-82.
43. Bos, K.I., et al., Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature, 2014. **514**(7523): p. 494-7.

44. Gagneux, S., Ecology and evolution of Mycobacterium tuberculosis. Nature Reviews Microbiology, 2018. **16**: p. 202.
45. Homolka, S., et al., High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. BMC Microbiol, 2008. **8**: p. 103.
46. Balcells, M.E., et al., A first insight on the population structure of Mycobacterium tuberculosis complex as studied by spoligotyping and MIRU-VNTRs in Santiago, Chile. PLoS One, 2015. **10**(2): p. e0118007.
47. Bifani, P.J., et al., Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol, 2002. **10**(1): p. 45-52.
48. 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.
49. 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. Clinical & Experimental Immunology, 2004. **137**(3): p. 460-468.
50. Parwati, I., R. van Crevel, and D. van Soolingen, Possible underlying mechanisms for successful emergence of the Mycobacterium tuberculosis Beijing genotype strains. Lancet Infect Dis, 2010. **10**(2): p. 103-11.
51. Huyen, M.N., et al., Tuberculosis relapse in Vietnam is significantly associated with Mycobacterium tuberculosis Beijing genotype infections. J Infect Dis, 2013. **207**(10): p. 1516-24.
52. Cowley, D., et al., Recent and rapid emergence of W-Beijing strains of Mycobacterium tuberculosis in Cape Town, South Africa. Clin Infect Dis, 2008. **47**(10): p. 1252-9.
53. Yang, C., et al., Mycobacterium tuberculosis Beijing strains favor transmission but not drug resistance in China. Clin Infect Dis, 2012. **55**(9): p. 1179-87.
54. 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.
55. Kong, Y., et al., Association between Mycobacterium tuberculosis Beijing/W lineage strain infection and extrathoracic tuberculosis: Insights from epidemiologic and clinical characterization of the three principal genetic groups of M. tuberculosis clinical isolates. J Clin Microbiol, 2007. **45**(2): p. 409-14.
56. Firmani, M.A. and L.W. Riley, Mycobacterium tuberculosis CDC1551 is resistant to reactive nitrogen and oxygen intermediates in vitro. Infect Immun, 2002. **70**(7): p. 3965-8.
57. Ribeiro, S.C., et al., Mycobacterium tuberculosis strains of the modern sublineage of the Beijing family are more likely to display increased virulence than strains of the ancient sublineage. J Clin Microbiol, 2014. **52**(7): p. 2615-24.
58. Merker, M., et al., Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet, 2015. **47**(3): p. 242-9.
59. Warner, D.F., A. Koch, and V. Mizrahi, Diversity and disease pathogenesis in Mycobacterium tuberculosis. Trends Microbiol, 2015. **23**(1): p. 14-21.
60. Hanekom, M., et al., Mycobacterium tuberculosis Beijing genotype: a template for success. Tuberculosis (Edinb), 2011. **91**(6): p. 510-23.
61. Zhang, W., et al., Post-translational modifications are enriched within protein functional groups important to bacterial adaptation within a deep-sea hydrothermal vent environment. Microbiome, 2016. **4**(1): p. 49.
62. Sutherland, I., E. Svandova, and S. Radhakrishna, Alternative models for the development of tuberculosis disease following infection with tubercle bacilli. Bull Int Union Tuberc, 1976. **51**(1): p. 171-9.
63. Hernandez-Pando, R., et al., Persistence of DNA from Mycobacterium tuberculosis in superficially normal lung tissue during latent infection. The Lancet, 2000. **356**(9248): p. 2133-2138.
64. Comstock, G.W., Epidemiology of tuberculosis. Am Rev Respir Dis, 1982. **125**(3 Pt 2): p. 8-15.
65. Knechel, N.A., Tuberculosis: pathophysiology, clinical features, and diagnosis. Critical care nurse, 2009. **29**(2): p. 34-43.

66. Pai, M., et al., Tuberculosis. Nat Rev Dis Primers, 2016. **2**: p. 16076.
67. Barnes, H.L. and L.R. Barnes, The Duration of Life in Pulmonary Tuberculosis with Cavity. Trans Am Climatol Clin Assoc, 1928. **44**: p. 39-55.
68. Andrews, J.R., et al., Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. Clin Infect Dis, 2012. **54**(6): p. 784-91.
69. Lahey, T., et al., Recurrent tuberculosis risk among HIV-infected adults in Tanzania with prior active tuberculosis. Clin Infect Dis, 2013. **56**(1): p. 151-8.
70. Verver, S., et al., Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. Am J Respir Crit Care Med, 2005. **171**(12): p. 1430-5.
71. Sia, J.K., M. Georgieva, and J. Rengarajan, Innate Immune Defenses in Human Tuberculosis: An Overview of the Interactions between Mycobacterium tuberculosis and Innate Immune Cells. J Immunol Res, 2015. **2015**: p. 747543.
72. Khan, N., et al., Innate Immunity Holding the Flanks until Reinforced by Adaptive Immunity against Mycobacterium tuberculosis Infection. Front Microbiol, 2016. **7**: p. 328.
73. Wawrocki, S. and M. Druszczynska, Inflammasomes in Mycobacterium tuberculosis-Driven Immunity. Can J Infect Dis Med Microbiol, 2017. **2017**: p. 2309478.
74. Scriba, T.J., A.K. Coussens, and H.A. Fletcher, Human Immunology of Tuberculosis. Microbiology spectrum, 2016. **4**(5).
75. Frieden, T.R., et al., Tuberculosis. The Lancet, 2003. **362**(9387): p. 887-899.
76. Smith, I., Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. Clinical microbiology reviews, 2003. **16**(3): p. 463-496.
77. Bermudez, L.E., et al., The efficiency of the translocation of Mycobacterium tuberculosis across a bilayer of epithelial and endothelial cells as a model of the alveolar wall is a consequence of transport within mononuclear phagocytes and invasion of alveolar epithelial cells. Infection and immunity, 2002. **70**(1): p. 140-146.
78. Lawn, S.D. and A.I. Zumla, Tuberculosis. The Lancet, 2011. **378**(9785): p. 57-72.
79. Weiss, G. and U.E. Schaible, Macrophage defense mechanisms against intracellular bacteria. Immunol Rev, 2015. **264**(1): p. 182-203.
80. Jo, E.-K., Mycobacterial interaction with innate receptors: TLRs, C-type lectins, and NLRs. Current opinion in infectious diseases, 2008. **21**(3): p. 279-286.
81. Harding, C.V. and W.H. Boom, Regulation of antigen presentation by Mycobacterium tuberculosis: a role for Toll-like receptors. Nature Reviews Microbiology, 2010. **8**(4): p. 296-307.
82. El-Etr, S.H. and J.D. Cirillo, Entry mechanisms of mycobacteria. Front Biosci, 2001. **6**: p. D737-D747.
83. 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. The Journal of Immunology, 2001. **167**(2): p. 910-918.
84. Jo, E.K., et al., Intracellular signalling cascades regulating innate immune responses to Mycobacteria: branching out from Toll‐like receptors. Cellular microbiology, 2007. **9**(5): p. 1087-1098.
85. Verrall, A.J., et al., Early clearance of Mycobacterium tuberculosis: a new frontier in prevention. Immunology, 2014. **141**(4): p. 506-513.
86. Pieters, J., Mycobacterium tuberculosis and the macrophage: maintaining a balance. Cell host & microbe, 2008. **3**(6): p. 399-407.
87. Aderem, A. and R.J. Ulevitch, Toll-like receptors in the induction of the innate immune response. Nature, 2000. **406**(6797): p. 782-787.
88. Shi, J., et al., Role of IL-10 and TNF-alpha during Mycobacterium tuberculosis infection in murine alveolar macrophages. Genet Mol Res, 2016. **15**(3).
89. Flynn, J.L. and J. Chan, Tuberculosis: latency and reactivation. Infection and immunity, 2001. **69**(7): p. 4195-4201.
90. Flynn, J.L. and J. Chan, Immunology of tuberculosis. Annual review of immunology, 2001. **19**(1): p. 93-129.

91. Martinez, F.O., L. Helming, and S. Gordon, Alternative activation of macrophages: an immunologic functional perspective. Annual review of immunology, 2009. **27**: p. 451-483.
92. Flynn, J.L. and J. Chan, Immune evasion by Mycobacterium tuberculosis: living with the enemy. Curr Opin Immunol, 2003. **15**(4): p. 450-5.
93. Mootoo, A., et al., TNF-alpha in tuberculosis: a cytokine with a split personality. Inflamm Allergy Drug Targets, 2009. **8**(1): p. 53-62.
94. Marino, S., et al., TNF and IL-10 are major factors in modulation of the phagocytic cell environment in lung and lymph node in tuberculosis: a next-generation two-compartmental model. Journal of theoretical biology, 2010. **265**(4): p. 586-598.
95. Cavalcanti, Y.V.N., et al., Role of TNF-alpha, IFN-gamma, and IL-10 in the development of pulmonary tuberculosis. Pulmonary medicine, 2012. **2012**.
96. Cooper, A., et al., The role of interleukin-12 in acquired immunity to Mycobacterium tuberculosis infection. Immunology, 1995. **84**(3): p. 423.
97. Bertagnolli, M., et al., IL-12 augments antigen-dependent proliferation of activated T lymphocytes. The Journal of Immunology, 1992. **149**(12): p. 3778-3783.
98. Lockhart, E., A.M. Green, and J.L. Flynn, IL-17 production is dominated by γδ T cells rather than CD4 T cells during Mycobacterium tuberculosis infection. The Journal of Immunology, 2006. **177**(7): p. 4662-4669.
99. Ferrero, E., et al., Macrophages exposed to Mycobacterium tuberculosis release chemokines able to recruit selected leucocyte subpopulations: focus on γδ cells. Immunology, 2003. **108**(3): p. 365-374.
100. Schluger, N.W. and W.N. Rom, The host immune response to tuberculosis. American journal of respiratory and critical care medicine, 1998. **157**(3): p. 679-691.
101. Piccini, P., et al., Clinical peculiarities of tuberculosis. BMC infectious diseases, 2014. **14**(1): p. S4.
102. Woodworth, J.S. and S.M. Behar, Mycobacterium tuberculosis-specific CD8+ T cells and their role in immunity. Critical Reviews™ in Immunology, 2006. **26**(4).
103. Weerdenburg, E.M., P.J. Peters, and N.N. van der Wel, How do mycobacteria activate CD8+ T cells? Trends in microbiology, 2010. **18**(1): p. 1-10.
104. Lefford, M.J., Transfer of adoptive immunity to tuberculosis in mice. Infect Immun, 1975. **11**(6): p. 1174-81.
105. Stenger, S., et al., An antimicrobial activity of cytolytic T cells mediated by granulysin. Science, 1998. **282**(5386): p. 121-5.
106. Ngai, P., et al., Gamma interferon responses of CD4 and CD8 T-cell subsets are quantitatively different and independent of each other during pulmonary Mycobacterium bovis BCG infection. Infect Immun, 2007. **75**(5): p. 2244-52.
107. Kaufmann, S.H. and A.J. McMichael, Annulling a dangerous liaison: vaccination strategies against AIDS and tuberculosis. Nat Med, 2005. **11**(4 Suppl): p. S33-44.
108. Kozakiewicz, L., et al., The role of B cells and humoral immunity in Mycobacterium tuberculosis infection, in The New Paradigm of Immunity to Tuberculosis. 2013, Springer. p. 225-250.
109. Lu, L.L., et al., A Functional Role for Antibodies in Tuberculosis. Cell, 2016. **167**(2): p. 433-443.e14.
110. Angala, S.K., et al., The cell envelope glycoconjugates of Mycobacterium tuberculosis. Crit Rev Biochem Mol Biol, 2014. **49**(5): p. 361-99.
111. Brennan, P.J., Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. Tuberculosis, 2003. **83**(1): p. 91-97.
112. Jackson, M., M.R. McNeil, and P.J. Brennan, Progress in targeting cell envelope biogenesis in Mycobacterium tuberculosis. Future Microbiol, 2013. **8**(7): p. 855-75.
113. Sonawane, A., et al., Role of glycans and glycoproteins in disease development by Mycobacterium tuberculosis. Critical reviews in microbiology, 2012. **38**(3): p. 250-266.
114. Chiaradia, L., et al., Dissecting the mycobacterial cell envelope and defining the composition of the native mycomembrane. Sci Rep, 2017. **7**(1): p. 12807.
115. Alderwick, L.J., et al., The Mycobacterial cell wall—peptidoglycan and Arabinogalactan. Cold Spring Harbor perspectives in medicine, 2015. **5**(8): p. a021113.

116. Mahapatra, S., et al., Mycobacterial lipid II is composed of a complex mixture of modified muramyl and peptide moieties linked to decaprenyl phosphate. Journal of bacteriology, 2005. **187**(8): p. 2747-2757.
117. Welin, A., Survival strategies of Mycobacterium tuberculosis inside the human macrophage. 2011, Linköping University Electronic Press.
118. Puech, V., et al., Evidence for a partial redundancy of the fibronectin‐binding proteins for the transfer of mycoloyl residues onto the cell wall arabinogalactan termini of Mycobacterium tuberculosis. Molecular microbiology, 2002. **44**(4): p. 1109-1122.
119. Ryll, R., Y. Kumazawa, and I. Yano, Immunological properties of trehalose dimycolate (cord factor) and other mycolic acid-containing glycolipids--a review. Microbiol Immunol, 2001. **45**(12): p. 801-11.
120. Ciaramella, A., et al., Mycobacterial 19-kDa lipoprotein mediates Mycobacterium tuberculosis-induced apoptosis in monocytes/macrophages at early stages of infection. Cell death and differentiation, 2000. **7**(12): p. 1270.
121. Torrelles, J.B. and L.S. Schlesinger, Diversity in Mycobacterium tuberculosis mannosylated cell wall determinants impacts adaptation to the host. Tuberculosis (Edinb), 2010. **90**(2): p. 84-93.
122. Daffe, M. and G. Etienne, The capsule of Mycobacterium tuberculosis and its implications for pathogenicity. Tuber Lung Dis, 1999. **79**(3): p. 153-69.
123. Casadevall, A. and L.A. Pirofski, Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. Infect Immun, 1999. **67**(8): p. 3703-13.
124. Smith, H., Microbial surfaces in relation to pathogenicity. Bacteriological Reviews, 1977. **41**(2): p. 475.
125. Watson, D. and C. Brandly, Virulence and pathogenicity. Annual Reviews in Microbiology, 1949. **3**(1): p. 195-220.
126. Calder, B., et al., Mycobacterial proteomics: analysis of expressed proteomes and post-translational modifications to identify candidate virulence factors. Expert Rev Proteomics, 2015. **12**(1): p. 21-35.
127. Wu, H.J., A.H. Wang, and M.P. Jennings, Discovery of virulence factors of pathogenic bacteria. Curr Opin Chem Biol, 2008. **12**(1): p. 93-101.
128. Philips, J.A. and J.D. Ernst, Tuberculosis pathogenesis and immunity. Annu Rev Pathol, 2012. **7**: p. 353-84.
129. Welin, A. and M. Lerm, Inside or outside the phagosome? The controversy of the intracellular localization of Mycobacterium tuberculosis. Tuberculosis, 2012. **92**(2): p. 113-120.
130. Dey, B. and W.R. Bishai. Crosstalk between Mycobacterium tuberculosis and the host cell. in Seminars in immunology. 2014. Elsevier.
131. Chan, J., et al., Lipoarabinomannan, a possible virulence factor involved in persistence of Mycobacterium tuberculosis within macrophages. Infection and immunity, 1991. **59**(5): p. 1755-1761.
132. Yuan, Y., et al., Identification of a gene involved in the biosynthesis of cyclopropanated mycolic acids in Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences, 1995. **92**(14): p. 6630-6634.
133. Sequeira, P.C., R.H. Senaratne, and L.W. Riley, Inhibition of toll-like receptor 2 (TLR-2)-mediated response in human alveolar epithelial cells by mycolic acids and Mycobacterium tuberculosis mce1 operon mutant. Pathogens and disease, 2014. **70**(2): p. 132-140.
134. Wolf, A.J., et al., Mycobacterium tuberculosis infects dendritic cells with high frequency and impairs their function in vivo. J Immunol, 2007. **179**(4): p. 2509-19.
135. Schlesinger, L., et al., Phagocytosis of Mycobacterium tuberculosis is mediated by human monocyte complement receptors and complement component C3. The Journal of Immunology, 1990. **144**(7): p. 2771-2780.
136. Schlesinger, L., Macrophage phagocytosis of virulent but not attenuated strains of Mycobacterium tuberculosis is mediated by mannose receptors in addition to complement receptors. The Journal of Immunology, 1993. **150**(7): p. 2920-2930.

137. Xu, S., et al., Intracellular trafficking in Mycobacterium tuberculosis and Mycobacterium avium-infected macrophages. J Immunol, 1994. **153**(6): p. 2568-78.
138. Sturgill-Koszycki, S., et al., Lack of acidification in Mycobacterium phagosomes produced by exclusion of the vesicular proton-ATPase. Science, 1994. **263**(5147): p. 678-81.
139. 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.
140. Bach, H., et al., Mycobacterium tuberculosis virulence is mediated by PtpA dephosphorylation of human vacuolar protein sorting 33B. Cell Host Microbe, 2008. **3**(5): p. 316-22.
141. Sun, J., et al., Mycobacterial nucleoside diphosphate kinase blocks phagosome maturation in murine RAW 264.7 macrophages. PLoS One, 2010. **5**(1): p. e8769.
142. Middlebrook, G., R.J. Dubos, and C. Pierce, VIRULENCE AND MORPHOLOGICAL CHARACTERISTICS OF MAMMALIAN TUBERCLE BACILLI. J Exp Med, 1947. **86**(2): p. 175-84.
143. Bloch, H., Studies on the virulence of tubercle bacilli; isolation and biological properties of a constituent of virulent organisms. J Exp Med, 1950. **91**(2): p. 197-218, pl.
144. Noll, H., et al., The chemical structure of the cord factor of Mycobacterium tuberculosis. Biochim Biophys Acta, 1956. **20**(2): p. 299-309.
145. Glickman, M.S., J.S. Cox, and W.R. Jacobs, Jr., A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis. Mol Cell, 2000. **5**(4): p. 717-27.
146. Rao, V., et al., Trans-cyclopropanation of mycolic acids on trehalose dimycolate suppresses Mycobacterium tuberculosis -induced inflammation and virulence. J Clin Invest, 2006. **116**(6): p. 1660-7.
147. Hunter, R.L., et al., Multiple roles of cord factor in the pathogenesis of primary, secondary, and cavitary tuberculosis, including a revised description of the pathology of secondary disease. Ann Clin Lab Sci, 2006. **36**(4): p. 371-86.
148. Barkan, D., et al., Mycobacterium tuberculosis lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice. Infect Immun, 2012. **80**(6): p. 1958-68.
149. Astarie-Dequeker, C., et al., Phthiocerol Dimycocerosates of M. tuberculosis Participate in Macrophage Invasion by Inducing Changes in the Organization of Plasma Membrane Lipids. PLOS Pathogens, 2009. **5**(2): p. e1000289.
150. Cox, J.S., et al., Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. Nature, 1999. **402**(6757): p. 79-83.
151. Chen, L., Q.W. Xie, and C. Nathan, Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. Mol Cell, 1998. **1**(6): p. 795-805.
152. Ojha, A.K., et al., Growth of Mycobacterium tuberculosis biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol Microbiol, 2008. **69**(1): p. 164-74.
153. Ahmad, S., Pathogenesis, immunology, and diagnosis of latent Mycobacterium tuberculosis infection. Clinical and Developmental Immunology, 2010. **2011**.
154. Nigou, J., et al., Mannosylated lipoarabinomannans inhibit IL-12 production by human dendritic cells: evidence for a negative signal delivered through the mannose receptor. The Journal of Immunology, 2001. **166**(12): p. 7477-7485.
155. Kang, P.B., et al., The human macrophage mannose receptor directs Mycobacterium tuberculosis lipoarabinomannan-mediated phagosome biogenesis. Journal of Experimental Medicine, 2005. **202**(7): p. 987-999.
156. Pancholi, P., et al., Presentation of mycobacterial antigens by human dendritic cells: lack of transfer from infected macrophages. Infect Immun, 1993. **61**(12): p. 5326-32.
157. Repique, C.J., et al., Susceptibility of mice deficient in the MHC class II transactivator to infection with Mycobacterium tuberculosis. Scand J Immunol, 2003. **58**(1): p. 15-22.
158. Scanga, C.A., et al., Depletion of CD4(+) T cells causes reactivation of murine persistent tuberculosis despite continued expression of interferon gamma and nitric oxide synthase 2. J Exp Med, 2000. **192**(3): p. 347-58.

159. Rezwan, M., et al., Lipoprotein synthesis in mycobacteria. Microbiology, 2007. **153**(3): p. 652-658.
160. Uchida, Y., et al., Accelerated immunopathological response of mice infected with Mycobacterium tuberculosis disrupted in the mce1 operon negative transcriptional regulator. Cell Microbiol, 2007. **9**(5): p. 1275-83.
161. Queiroz, A., et al., Comparative metabolic profiling of mce1 operon mutant vs wild-type Mycobacterium tuberculosis strains. Pathog Dis, 2015. **73**(8): p. ftv066.
162. Cantrell, S.A., et al., Free mycolic acid accumulation in the cell wall of the mce1 operon mutant strain of Mycobacterium tuberculosis. J Microbiol, 2013. **51**(5): p. 619-26.
163. 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.
164. Constant, P., et al., Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the Mycobacterium tuberculosis complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methyl esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. J Biol Chem, 2002. **277**(41): p. 38148-58.
165. Reed, M.B., et al., A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature, 2004. **431**(7004): p. 84-7.
166. Kuehn, M.J. and N.C. Kesty, Bacterial outer membrane vesicles and the host–pathogen interaction. Genes & development, 2005. **19**(22): p. 2645-2655.
167. Lee, J., et al., Proteomic analysis of extracellular vesicles derived from Mycobacterium tuberculosis. Proteomics, 2015. **15**(19): p. 3331-3337.
168. Russell, D.G., et al., Mycobacterium tuberculosis and the four-minute phagosome. ASM NEWS-AMERICAN SOCIETY FOR MICROBIOLOGY, 2005. **71**(10): p. 459.
169. Berney, M. and L. Berney-Meyer, Mycobacterium tuberculosis in the Face of Host-Imposed Nutrient Limitation. Microbiol Spectr, 2017. **5**(3).
170. McKinney, J.D., et al., Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature, 2000. **406**(6797): p. 735-8.
171. Vromman, F. and A. Subtil, Exploitation of host lipids by bacteria. Curr Opin Microbiol, 2014. **17**: p. 38-45.
172. 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.
173. van der Wel, N., et al., M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. Cell, 2007. **129**(7): p. 1287-98.
174. Smith, J., et al., Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in Mycobacterium marinum escape from the vacuole. Infect Immun, 2008. **76**(12): p. 5478-87.
175. de Jonge, M.I., et al., ESAT-6 from Mycobacterium tuberculosis dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. J Bacteriol, 2007. **189**(16): p. 6028-34.
176. Brodin, P., et al., Dissection of ESAT-6 system 1 of Mycobacterium tuberculosis and impact on immunogenicity and virulence. Infect Immun, 2006. **74**(1): p. 88-98.
177. Stanley, S.A., et al., Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. Proc Natl Acad Sci U S A, 2003. **100**(22): p. 13001-6.
178. Groenen, P.M., et al., Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. Mol Microbiol, 1993. **10**(5): p. 1057-65.
179. Ross, B.C., et al., Molecular cloning of a highly repeated DNA element from Mycobacterium tuberculosis and its use as an epidemiological tool. J Clin Microbiol, 1992. **30**(4): p. 942-6.
180. Ramakrishnan, L., N.A. Federspiel, and S. Falkow, Granuloma-specific expression of Mycobacterium virulence proteins from the glycine-rich PE-PGRS family. Science, 2000. **288**(5470): p. 1436-9.
181. Sassetti, C.M. and E.J. Rubin, Genetic requirements for mycobacterial survival during infection. Proceedings of the National Academy of Sciences, 2003. **100**(22): p. 12989-12994.

182. Talaat, A.M., et al., The temporal expression profile of Mycobacterium tuberculosis infection in mice. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(13): p. 4602-4607.
183. Boldrin, F., et al., The phosphatidyl-myo-inositol mannosyltransferase PimA is essential for Mycobacterium tuberculosis growth in vitro and in vivo. Journal of bacteriology, 2014. **196**(19): p. 3441-3451.
184. Reddy, P.V., et al., Iron storage proteins are essential for the survival and pathogenesis of Mycobacterium tuberculosis in THP-1 macrophages and the guinea pig model of infection. Journal of bacteriology, 2012. **194**(3): p. 567-575.
185. Kumar, A., et al., Mycobacterium tuberculosis DosS is a redox sensor and DosT is a hypoxia sensor. Proceedings of the National Academy of Sciences, 2007. **104**(28): p. 11568-11573.
186. Hu, Y., et al., Deletion of the Mycobacterium tuberculosis α-crystallin-like hspX gene causes increased bacterial growth in vivo. Infection and immunity, 2006. **74**(2): p. 861-868.
187. Ducati, R.G., et al., The resumption of consumption: a review on tuberculosis. Memórias do Instituto Oswaldo Cruz, 2006. **101**(7): p. 697-714.
188. Rickman, L., et al., A two-component signal transduction system with a PAS domain-containing sensor is required for virulence of Mycobacterium tuberculosis in mice. Biochemical and biophysical research communications, 2004. **314**(1): p. 259-267.
189. Fleischmann, R.D., et al., Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. J Bacteriol, 2002. **184**(19): p. 5479-90.
190. Brosch, R., et al., The evolution of mycobacterial pathogenicity: clues from comparative genomics. Trends in microbiology, 2001. **9**(9): p. 452-458.
191. Cosma, C.L., D.R. Sherman, and L. Ramakrishnan, The secret lives of the pathogenic mycobacteria. Annual Reviews in Microbiology, 2003. **57**(1): p. 641-676.
192. Kelkar, D.S., et al., Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry. Mol Cell Proteomics, 2011. **10**(12): p. M111 011627.
193. Bespyatykh, J.A., E.A. Shitikov, and E.N. Ilina, Proteomics for the Investigation of Mycobacteria. Acta Naturae, 2017. **9**(1): p. 15-25.
194. Zhang, Y., et al., Protein analysis by shotgun/bottom-up proteomics. Chemical reviews, 2013. **113**(4): p. 2343-2394.
195. Cordwell, S.J., Exploring and exploiting bacterial proteomes. Methods Mol Biol, 2004. **266**: p. 115-35.
196. Bell, C., et al., Characterization of the Mycobacterium tuberculosis proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. Journal of proteome research, 2011. **11**(1): p. 119-130.
197. Coiras, M., et al., Application of proteomics technology for analyzing the interactions between host cells and intracellular infectious agents. Proteomics, 2008. **8**(4): p. 852-873.
198. Lindon, J.C., G.E. Tranter, and D. Koppenaal, Encyclopedia of spectroscopy and spectrometry. 2016: Academic Press.
199. Liang, K.-H., Bioinformatics for biomedical science and clinical applications. 2013: Elsevier.
200. Gengenbacher, M., et al., Mycobacterium tuberculosis in the Proteomics Era, in Molecular Genetics of Mycobacteria, Second Edition. 2014, American Society of Microbiology. p. 241-260.
201. Payne, S.H., The utility of protein and mRNA correlation. Trends in biochemical sciences, 2015. **40**(1): p. 1-3.
202. Maier, T., M. Güell, and L. Serrano, Correlation of mRNA and protein in complex biological samples. FEBS letters, 2009. **583**(24): p. 3966-3973.
203. Matthiesen, R. and J. Bunkenborg, Introduction to mass spectrometry-based proteomics, in Mass Spectrometry Data Analysis in Proteomics. 2013, Springer. p. 1-45.
204. Gross, J.H., Mass spectrometry: a textbook. 3rd ed. ed. 2017: Springer Science & Business Media.
205. Aebersold, R. and M. Mann, Mass spectrometry-based proteomics. Nature, 2003. **422**(6928): p. 198-207.

206. Angel, T.E., et al., Mass spectrometry-based proteomics: existing capabilities and future directions. Chemical Society Reviews, 2012. **41**(10): p. 3912-3928.
207. Otto, A., D. Becher, and F. Schmidt, Quantitative proteomics in the field of microbiology. Proteomics, 2014. **14**(4-5): p. 547-65.
208. Cañas, B., et al., Mass spectrometry technologies for proteomics. Briefings in Functional Genomics, 2006. **4**(4): p. 295-320.
209. Chait, B.T., Mass spectrometry: bottom-up or top-down? Science, 2006. **314**(5796): p. 65-66.
210. Washburn, M.P., D. Wolters, and J.R. Yates III, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature biotechnology, 2001. **19**(3): p. 242.
211. Kelleher, N.L., Peer reviewed: Top-down proteomics. 2004, ACS Publications.
212. Xie, F., et al., Liquid chromatography-mass spectrometry-based quantitative proteomics. Journal of Biological Chemistry, 2011: p. jbc. R110. 199703.
213. Bantscheff, M., et al., Quantitative mass spectrometry in proteomics: a critical review. Analytical and bioanalytical chemistry, 2007. **389**(4): p. 1017-1031.
214. Hsieh, E.J., et al., Comparison of database search strategies for high precursor mass accuracy MS/MS data. J Proteome Res, 2010. **9**(2): p. 1138-43.
215. Nesvizhskii, A.I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. Journal of proteomics, 2010. **73**(11): p. 2092-2123.
216. Nesvizhskii, A.I., A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. J Proteomics, 2010. **73**(11): p. 2092-123.
217. Nesvizhskii, A.I., O. Vitek, and R. Aebersold, Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods, 2007. **4**(10): p. 787-97.
218. Elias, J.E. and S.P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods, 2007. **4**(3): p. 207-14.
219. Perkins, D.N., et al., Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis, 1999. **20**(18): p. 3551-67.
220. Eng, J.K., A.L. McCormack, and J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.
221. Geer, L.Y., et al., Open mass spectrometry search algorithm. J Proteome Res, 2004. **3**(5): p. 958-64.
222. Craig, R. and R.C. Beavis, TANDEM: matching proteins with tandem mass spectra. Bioinformatics, 2004. **20**(9): p. 1466-7.
223. Cox, J., et al., Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res, 2011. **10**(4): p. 1794-805.
224. Cox, J. and M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol, 2008. **26**(12): p. 1367-72.
225. Eng, J.K., et al., A face in the crowd: recognizing peptides through database search. Mol Cell Proteomics, 2011. **10**(11): p. R111.009522.
226. Houel, S., et al., Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. J Proteome Res, 2010. **9**(8): p. 4152-60.
227. Shteynberg, D., et al., Combining results of multiple search engines in proteomics. Mol Cell Proteomics, 2013. **12**(9): p. 2383-93.
228. Aggarwal, S. and A.K. Yadav, False Discovery Rate Estimation in Proteomics. Methods Mol Biol, 2016. **1362**: p. 119-28.
229. Jensen, O.N., Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr Opin Chem Biol, 2004. **8**(1): p. 33-41.

230. Huang, X., et al., ISPTM: an iterative search algorithm for systematic identification of post-translational modifications from complex proteome mixtures. J Proteome Res, 2013. **12**(9): p. 3831-42.
231. Griffiths, J.R., Analysis of Protein Post-Translational Modifications by Mass Spectrometry. 2016: John Wiley & Sons.
232. Prabakaran, S., et al., Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med, 2012. **4**(6): p. 565-83.
233. Zhang, J., et al., MS/MS/MS reveals false positive identification of histone serine methylation. J Proteome Res, 2010. **9**(1): p. 585-94.
234. Daugelat, S., et al., Secreted antigens of Mycobacterium tuberculosis: characterization with T lymphocytes from patients and contacts after two-dimensional separation. Journal of Infectious Diseases, 1992. **166**(1): p. 186-190.
235. 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. Infection and Immunity, 1997. **65**(11): p. 4515-4524.
236. Wallis, R., R. Paranjape, and M. Phillips, Identification by two-dimensional gel electrophoresis of a 58-kilodalton tumor necrosis factor-inducing protein of Mycobacterium tuberculosis. Infection and immunity, 1993. **61**(2): p. 627-632.
237. 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. Infection and immunity, 1999. **67**(1): p. 327-336.
238. Schubert, O.T., et al., The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. Cell host & microbe, 2013. **13**(5): p. 602-612.
239. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng, 2009. **11**: p. 49-79.
240. Galassie, A.C. and A.J. Link, Proteomic contributions to our understanding of vaccine and immune responses. Proteomics Clinical Applications, 2015. **9**(11-12): p. 972-989.
241. Phillips, C.I. and M. Bogyo, Proteomics meets microbiology: technical advances in the global mapping of protein expression and function. Cell Microbiol, 2005. **7**(8): p. 1061-76.
242. Ribet, D. and P. Cossart, Pathogen-mediated posttranslational modifications: A re-emerging field. Cell, 2010. **143**(5): p. 694-702.
243. van Els, C.A., et al., Toward Understanding the Essence of Post-Translational Modifications for the Mycobacterium tuberculosis Immunoproteome. Front Immunol, 2014. **5**: p. 361.
244. Chao, J., et al., Protein kinase and phosphatase signaling in Mycobacterium tuberculosis physiology and pathogenesis. Biochim Biophys Acta, 2010. **1804**(3): p. 620-7.
245. Canova, M.J. and V. Molle, Bacterial serine/threonine protein kinases in host-pathogen interactions. J Biol Chem, 2014. **289**(14): p. 9473-9.
246. Cain, J.A., N. Solis, and S.J. Cordwell, Beyond gene expression: the impact of protein post-translational modifications in bacteria. J Proteomics, 2014. **97**: p. 265-86.
247. Mann, M. and O.N. Jensen, Proteomic analysis of post-translational modifications. Nat Biotechnol, 2003. **21**(3): p. 255-61.
248. Soufi, B., et al., Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. Curr Opin Microbiol, 2012. **15**(3): p. 357-63.
249. van Noort, V., et al., Cross‐talk between phosphorylation and lysine acetylation in a genome‐reduced bacterium. Molecular systems biology, 2012. **8**(1): p. 571.
250. Ribet, D. and P. Cossart, Post-translational modifications in host cells during bacterial infection. FEBS Lett, 2010. **584**(13): p. 2748-58.
251. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl, 2005. **44**(45): p. 7342-72.

252. Bastos, P.A.D., J.P. da Costa, and R. Vitorino, A glimpse into the modulation of post-translational modifications of human-colonizing bacteria. Journal of proteomics, 2017. **152**: p. 254-275.
253. Olsen, J.V. and M. Mann, Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol Cell Proteomics, 2013. **12**(12): p. 3444-52.
254. Chicooree, N., R.D. Unwin, and J.R. Griffiths, The application of targeted mass spectrometry-based strategies to the detection and localization of post-translational modifications. Mass Spectrom Rev, 2015. **34**(6): p. 595-626.
255. Liu, F., et al., Acetylome analysis reveals diverse functions of lysine acetylation in Mycobacterium tuberculosis. Mol Cell Proteomics, 2014. **13**(12): p. 3352-66.
256. Xie, L., et al., Proteome-wide lysine acetylation profiling of the human pathogen Mycobacterium tuberculosis. Int J Biochem Cell Biol, 2015. **59**: p. 193-202.
257. Nambi, S., et al., Cyclic AMP-dependent protein lysine acylation in mycobacteria regulates fatty acid and propionate metabolism. J Biol Chem, 2013. **288**(20): p. 14114-24.
258. Hayden, J.D., et al., Reversible acetylation regulates acetate and propionate metabolism in Mycobacterium smegmatis. Microbiology, 2013. **159**(Pt 9): p. 1986-99.
259. Graham, R.L. and S. Hess, Mass spectrometry in the elucidation of the glycoproteome of bacterial pathogens. Current Proteomics, 2010. **7**(1): p. 57-81.
260. Ouidir, T., T. Kentache, and J. Hardouin, Protein lysine acetylation in bacteria: Current state of the art. Proteomics, 2016. **16**(2): p. 301-9.
261. Carabetta, V.J. and I.M. Cristea, The regulation, function, and detection of protein acetylation in bacteria. J Bacteriol, 2017: p. JB. 00107-17.
262. Xu, H., S.S. Hegde, and J.S. Blanchard, Reversible Acetylation and Inactivation of Mycobacterium tuberculosis Acetyl-CoA Synthetase Is Dependent on cAMP. Biochemistry, 2011. **50**(26): p. 5883-5892.
263. Li, R., et al., Purification and characterization of the acetyl-CoA synthetase from Mycobacterium tuberculosis. Acta Biochim Biophys Sin, 2011. **43**(11): p. 891-899.
264. Okkels, L.M., et al., CFP10 discriminates between nonacetylated and acetylated ESAT‐6 of Mycobacterium tuberculosis by differential interaction. Proteomics, 2004. **4**(10): p. 2954-2960.
265. Ghosh, S., et al., Lysine acetylation of the Mycobacterium tuberculosis HU protein modulates its DNA binding and genome organization. Mol Microbiol, 2016. **100**(4): p. 577-88.
266. Kosono, S., et al., Changes in the Acetylome and Succinylome of Bacillus subtilis in Response to Carbon Source. PLoS One, 2015. **10**(6): p. e0131169.
267. Mukherjee, S., et al., Yersinia YopJ acetylates and inhibits kinase activation by blocking phosphorylation. Science, 2006. **312**(5777): p. 1211-4.
268. Cousin, C., et al., Protein-serine/threonine/tyrosine kinases in bacterial signaling and regulation. FEMS Microbiol Lett, 2013. **346**(1): p. 11-9.
269. Denk, D. and A. Bock, L-cysteine biosynthesis in Escherichia coli: nucleotide sequence and expression of the serine acetyltransferase (cysE) gene from the wild-type and a cysteine-excreting mutant. J Gen Microbiol, 1987. **133**(3): p. 515-25.
270. Anonsen, J.H., et al., Structural and genetic analyses of glycan O-acetylation in a bacterial protein glycosylation system: evidence for differential effects on glycan chain length. Glycobiology, 2017: p. 1-12.
271. Moynihan, P.J. and A.J. Clarke, O-Acetylated peptidoglycan: controlling the activity of bacterial autolysins and lytic enzymes of innate immune systems. Int J Biochem Cell Biol, 2011. **43**(12): p. 1655-9.
272. Daubenspeck, J.M., et al., General N-and O-Linked Glycosylation of Lipoproteins in Mycoplasmas and Role of Exogenous Oligosaccharide. PLOS ONE, 2015. **10**(11): p. e0143362.
273. Smith, G.T., M.J. Sweredoski, and S. Hess, O-linked glycosylation sites profiling in Mycobacterium tuberculosis culture filtrate proteins. Journal of proteomics, 2014. **97**: p. 296-306.
274. Herrmann, J., et al., Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from Mycobacterium tuberculosis. The EMBO journal, 1996. **15**(14): p. 3547.

275. Herrmann, J.L., et al., Analysis of post‐translational modification of mycobacterial proteins using a cassette expression system. Febs Letters, 2000. **473**(3): p. 358-362.
276. Becker, K., et al., Lipoprotein Glycosylation by Protein-O-Mannosyltransferase (MAB_1122c) Contributes to Low Cell Envelope Permeability and Antibiotic Resistance of Mycobacterium abscessus. Frontiers in microbiology, 2017. **8**: p. 2123.
277. Brülle, J.K., et al., Cloning, expression and characterization of Mycobacterium tuberculosis lipoprotein LprF. Biochemical and biophysical research communications, 2010. **391**(1): p. 679-684.
278. Sartain, M.J. and J.T. Belisle, N-Terminal clustering of the O-glycosylation sites in the Mycobacterium tuberculosis lipoprotein SodC. Glycobiology, 2008. **19**(1): p. 38-51.
279. Kovacs-Simon, A., R. Titball, and S.L. Michell, Lipoproteins of bacterial pathogens. Infection and immunity, 2011. **79**(2): p. 548-561.
280. Diaz-Silvestre, H., et al., The 19-kDa antigen of Mycobacterium tuberculosis is a major adhesin that binds the mannose receptor of THP-1 monocytic cells and promotes phagocytosis of mycobacteria. Microbial pathogenesis, 2005. **39**(3): p. 97-107.
281. Malik, A.N. and P. Godfrey-Faussett, Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease. Lancet Infect Dis, 2005. **5**(3): p. 174-83.
282. Pérez-Llarena, F.J. and G.J.F.i.M. Bou, Proteomics as a tool for studying bacterial virulence and antimicrobial resistance. 2016. **7**: p. 410.
283. Li, P., et al., Comparative Proteomics Analysis of Human Macrophages Infected with Virulent Mycobacterium bovis. Front Cell Infect Microbiol, 2017. **7**: p. 65.
284. de Souza, G.A. and H.G. Wiker, A proteomic view of mycobacteria. Proteomics, 2011. **11**(15): p. 3118-3127.
285. Gunawardena, H.P., et al., Comparison of the membrane proteome of virulent Mycobacterium tuberculosis and the attenuated Mycobacterium bovis BCG vaccine strain by label-free quantitative proteomics. Journal of proteome research, 2013. **12**(12): p. 5463-5474.
286. Jhingan, G.D., et al., Comparative Proteomic Analyses of Avirulent, Virulent, and Clinical Strains of Mycobacterium tuberculosis Identify Strain-specific Patterns. J Biol Chem, 2016. **291**(27): p. 14257-73.
287. Jena, L., et al., Comparative proteomic analysis of Mycobacterium tuberculosis strain H37Rv versus H37Ra. Int J Mycobacteriol, 2013. **2**(4): p. 220-6.
288. Cordwell, S.J., A.S. Nouwens, and B.J. Walsh, Comparative proteomics of bacterial pathogens. Proteomics, 2001. **1**(4): p. 461-472.
289. González-Zamorano, M., et al., Mycobacterium tuberculosis glycoproteomics based on ConA-lectin affinity capture of mannosylated proteins. 2009. **8**(2): p. 721-733.
290. Sinha, S., et al., Proteome analysis of the plasma membrane of Mycobacterium tuberculosis. Comp Funct Genomics, 2002. **3**(6): p. 470-83.
291. Bell, C., et al., Characterization of the Mycobacterium tuberculosis proteome by liquid chromatography mass spectrometry-based proteomics techniques: a comprehensive resource for tuberculosis research. J Proteome Res, 2012. **11**(1): p. 119-30.
292. 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.
293. Målen, H., et al., Definition of novel cell envelope associated proteins in Triton X-114 extracts of Mycobacterium tuberculosis H37Rv. 2010. **10**(1): p. 132.
294. Schubert, O.T., et al., Absolute proteome composition and dynamics during dormancy and resuscitation of Mycobacterium tuberculosis. Cell host & microbe, 2015. **18**(1): p. 96-108.
295. de Souza, G.A., et al., Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo-and hypervirulent clinical Mycobacterium tuberculosis Beijing isolates. 2010. **9**(11): p. 2414-2423.
296. Verma, R., et al., Quantitative proteomic and phosphoproteomic analysis of H37Ra and H37Rv strains of Mycobacterium tuberculosis. 2017. **16**(4): p. 1632-1645.

297. Bespyatykh, J., et al., Proteome analysis of the Mycobacterium tuberculosis Beijing B0/W148 cluster. Scientific reports, 2016. **6**: p. 28985.
298. Malen, H., et al., Comparison of membrane proteins of Mycobacterium tuberculosis H37Rv and H37Ra strains. BMC Microbiol, 2011. **11**: p. 18.
299. Gunawardena, H.P., et al., Comparison of the membrane proteome of virulent Mycobacterium tuberculosis and the attenuated Mycobacterium bovis BCG vaccine strain by label-free quantitative proteomics. J Proteome Res, 2013. **12**(12): p. 5463-74.
300. Peirs, P., et al., Mycobacterium tuberculosis with disruption in genes encoding the phosphate binding proteins PstS1 and PstS2 is deficient in phosphate uptake and demonstrates reduced in vivo virulence. Infection and immunity, 2005. **73**(3): p. 1898-1902.
301. Rifat, D., D.A. Belchis, and P.C. Karakousis, senX3-independent contribution of regX3 to Mycobacterium tuberculosis virulence. BMC microbiology, 2014. **14**(1): p. 265.
302. van Winden, V.J., et al., Mycosins are required for the stabilization of the ESX-1 and ESX-5 Type VII secretion membrane complexes. MBio, 2016. **7**(5): p. e01471-16.
303. Hartkoorn, R.C., S. Uplekar, and S.T. Cole, Cross-resistance between clofazimine and bedaquiline through upregulation of MmpL5 in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy, 2014. **58**(5): p. 2979-2981.
304. Yu, J., et al., Both phthiocerol dimycocerosates and phenolic glycolipids are required for virulence of Mycobacterium marinum. Infection and immunity, 2012. **80**(4): p. 1381-1389.
305. Simeone, R., et al., Molecular dissection of the biosynthetic relationship between phthiocerol and phthiodiolone dimycocerosates and their critical role in the virulence and permeability of Mycobacterium tuberculosis. Febs j, 2007. **274**(8): p. 1957-69.
306. Kumar, A., et al., Redox homeostasis in mycobacteria: the key to tuberculosis control? Expert reviews in molecular medicine, 2011. **13**.
307. Holyoake, L.V., et al., CydDC-mediated reductant export in Escherichia coli controls the transcriptional wiring of energy metabolism and combats nitrosative stress. Biochem J, 2016. **473**(6): p. 693-701.
308. Rifat, D., W.R. Bishai, and P.C. Karakousis, Phosphate depletion: a novel trigger for Mycobacterium tuberculosis persistence. The Journal of infectious diseases, 2009. **200**(7): p. 1126-1135.
309. Tischler, A.D., et al., Mycobacterium tuberculosis requires phosphate-responsive gene regulation to resist host immunity. Infection and immunity, 2013. **81**(1): p. 317-328.
310. Munshi, T., et al., Characterisation of ATP-dependent Mur ligases involved in the biogenesis of cell wall peptidoglycan in Mycobacterium tuberculosis. PLoS One, 2013. **8**(3): p. e60143.
311. Moraes, G.L., et al., Structural and functional features of enzymes of Mycobacterium tuberculosis peptidoglycan biosynthesis as targets for drug development. Tuberculosis (Edinb), 2015. **95**(2): p. 95-111.
312. Bisson, G.P., et al., Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, rpoB mutant Mycobacterium tuberculosis. Journal of bacteriology, 2012. **194**(23): p. 6441-6452.
313. Pawelczyk, J., et al., AccD6, a key carboxyltransferase essential for mycolic acid synthesis in Mycobacterium tuberculosis, is dispensable in a nonpathogenic strain. Journal of bacteriology, 2011. **193**(24): p. 6960-6972.
314. Bhatt, A., et al., The Mycobacterium tuberculosis FAS‐II condensing enzymes: their role in mycolic acid biosynthesis, acid‐fastness, pathogenesis and in future drug development. Molecular microbiology, 2007. **64**(6): p. 1442-1454.
315. Watanabe, S., et al., Fumarate reductase activity maintains an energized membrane in anaerobic Mycobacterium tuberculosis. PLoS pathogens, 2011. **7**(10): p. e1002287.
316. Zheng, H., et al., Genetic basis of virulence attenuation revealed by comparative genomic analysis of Mycobacterium tuberculosis strain H37Ra versus H37Rv. PloS one, 2008. **3**(6): p. e2375.

317. Ma, Q. and T.K. Wood, Protein acetylation in prokaryotes increases stress resistance. Biochem Biophys Res Commun, 2011. **410**(4): p. 846-51.
318. Fortuin, S., et al., Phosphoproteomics analysis of a clinical Mycobacterium tuberculosis Beijing isolate: expanding the mycobacterial phosphoproteome catalog. Front Microbiol, 2015. **6**: p. 6.
319. Saito, K., et al., Mechanism of activation of proximate mutagens in Ames' tester strains: The acetyl-CoA dependent enzyme in Salmonella typhimurium TA98 deficient in TA981, 8-DNP6 catalyzes DNA-binding as the cause of mutagenicity. Biochemical and biophysical research communications, 1983. **116**(1): p. 141-147.
320. Yimer, S.A., et al., Comparative Proteomic Analysis of Mycobacterium tuberculosis Lineage 7 and Lineage 4 Strains Reveals Differentially Abundant Proteins Linked to Slow Growth and Virulence. Front Microbiol, 2017. **8**: p. 795.
321. Holliday, G.L., J.B. Mitchell, and J.M. Thornton, Understanding the functional roles of amino acid residues in enzyme catalysis. J Mol Biol, 2009. **390**(3): p. 560-77.
322. Bartlett, G.J., et al., Analysis of catalytic residues in enzyme active sites. J Mol Biol, 2002. **324**(1): p. 105-21.
323. Thao, S. and J.C. Escalante-Semerena, Control of protein function by reversible Nvarepsilon-lysine acetylation in bacteria. Curr Opin Microbiol, 2011. **14**(2): p. 200-4.
324. Shaw, B.F., et al., Lysine acetylation can generate highly charged enzymes with increased resistance toward irreversible inactivation. Protein Science, 2008. **17**(8): p. 1446-1455.
325. Li, T., et al., Acetylation modulates cellular distribution and DNA sensing ability of interferon-inducible protein IFI16. Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10558-63.
326. Zhang, T., et al., Acetylation negatively regulates glycogen phosphorylase by recruiting protein phosphatase 1. Cell Metab, 2012. **15**(1): p. 75-87.
327. Thao, S., et al., Nε− lysine acetylation of a bacterial transcription factor inhibits its DNA-binding activity. PloS one, 2010. **5**(12): p. e15123.
328. Howe, F.S., et al., Lysine acetylation controls local protein conformation by influencing proline isomerization. Mol Cell, 2014. **55**(5): p. 733-44.
329. Ishfaq, M., et al., Acetylation regulates subcellular localization of eukaryotic translation initiation factor 5A (eIF5A). FEBS Lett, 2012. **586**(19): p. 3236-41.
330. Sakai, H., Mutagenesis of the active site lysine 221 of the pyruvate kinase from Bacillus stearothermophilus. J Biochem, 2005. **137**(2): p. 141-5.
331. Xie, L., et al., First succinyl-proteome profiling of extensively drug-resistant Mycobacterium tuberculosis revealed involvement of succinylation in cellular physiology. J Proteome Res, 2015. **14**(1): p. 107-19.
332. Mukherjee, S., Y.H. Hao, and K. Orth, A newly discovered post-translational modification--the acetylation of serine and threonine residues. Trends Biochem Sci, 2007. **32**(5): p. 210-6.
333. Wang, Q., et al., Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. Science, 2010. **327**(5968): p. 1004-7.
334. Guan, K.L. and Y. Xiong, Regulation of intermediary metabolism by protein acetylation. Trends Biochem Sci, 2011. **36**(2): p. 108-16.
335. Shi, L. and B.P. Tu, Protein acetylation as a means to regulate protein function in tune with metabolic state. 2014, Portland Press Limited.
336. You, D., et al., Acetyl coenzyme A synthetase is acetylated on multiple lysine residues by a protein acetyltransferase with a single Gcn5-type N-acetyltransferase (GNAT) domain in Saccharopolyspora erythraea. Journal of bacteriology, 2014. **196**(17): p. 3169-3178.
337. Zhao, S., et al., Regulation of cellular metabolism by protein lysine acetylation. Science, 2010. **327**(5968): p. 1000-4.
338. Kinhikar, A.G., et al., Mycobacterium tuberculosis malate synthase is a laminin ‐ binding adhesin. Molecular microbiology, 2006. **60**(4): p. 999-1013.
339. 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.

340. Rehman, A. and B.A. Mcfadden, The Consequences of Replacing Histidine 356 in Isocitrate Lyase fromEscherichia coli. Archives of biochemistry and biophysics, 1996. **336**(2): p. 309-315.
341. Wayne, L.G. and K.Y. Lin, Glyoxylate metabolism and adaptation of Mycobacterium tuberculosis to survival under anaerobic conditions. Infect Immun, 1982. **37**(3): p. 1042-9.
342. Lorenz, M.C. and G.R. Fink, Life and death in a macrophage: role of the glyoxylate cycle in virulence. Eukaryot Cell, 2002. **1**(5): p. 657-62.
343. Guenin-Mace, L., R. Simeone, and C. Demangel, Lipids of pathogenic Mycobacteria: contributions to virulence and host immune suppression. Transbound Emerg Dis, 2009. **56**(6-7): p. 255-68.
344. Smith, T., K.A. Wolff, and L. Nguyen, Molecular biology of drug resistance in Mycobacterium tuberculosis, in Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism. 2012, Springer. p. 53-80.
345. Ng, V.H., et al., Role of KatG catalase-peroxidase in mycobacterial pathogenesis: countering the phagocyte oxidative burst. Mol Microbiol, 2004. **52**(5): p. 1291-302.
346. Cohen, T., M.C. Becerra, and M.B. Murray, Isoniazid resistance and the future of drug-resistant tuberculosis. Microb Drug Resist, 2004. **10**(4): p. 280-5.
347. Gotoh, Y., et al., Two-component signal transduction as potential drug targets in pathogenic bacteria. Curr Opin Microbiol, 2010. **13**(2): p. 232-9.
348. Prisic, S. and R.N. Husson, Mycobacterium tuberculosis Serine/Threonine Protein Kinases. Microbiol Spectr, 2014. **2**(5).
349. Mukherjee, P., et al., Novel role of Wag31 in protection of mycobacteria under oxidative stress. Mol Microbiol, 2009. **73**(1): p. 103-19.
350. Kang, C.M., et al., Wag31, a homologue of the cell division protein DivIVA, regulates growth, morphology and polar cell wall synthesis in mycobacteria. Microbiology, 2008. **154**(Pt 3): p. 725-35.
351. Unal, C.M. and M. Steinert, Microbial peptidyl-prolyl cis/trans isomerases (PPIases): virulence factors and potential alternative drug targets. Microbiol Mol Biol Rev, 2014. **78**(3): p. 544-71.
352. Tseng, S.T., et al., The mutations of katG and inhA genes of isoniazid-resistant Mycobacterium tuberculosis isolates in Taiwan. J Microbiol Immunol Infect, 2015. **48**(3): p. 249-55.
353. Ramaswamy, S.V., et al., Single nucleotide polymorphisms in genes associated with isoniazid resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother, 2003. **47**(4): p. 1241-50.
354. Chien, J.Y., et al., Mutations in gyrA and gyrB among Fluoroquinolone- and Multidrug-Resistant Mycobacterium tuberculosis Isolates. Antimicrob Agents Chemother, 2016. **60**(4): p. 2090-6.
355. Wang, J.Y., et al., Fluoroquinolone resistance in Mycobacterium tuberculosis isolates: associated genetic mutations and relationship to antimicrobial exposure. J Antimicrob Chemother, 2007. **59**(5): p. 860-5.
356. Pitaksajjakul, P., et al., Mutations in the gyrA and gyrB genes of fluoroquinolone-resistant Mycobacterium tuberculosis from TB patients in Thailand. Southeast Asian J Trop Med Public Health, 2005. **36 Suppl 4**: p. 228-37.
357. Chen, W., et al., Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB. Proc Natl Acad Sci U S A, 2011. **108**(24): p. 9804-8.
358. Tsodikov, O.V., K.D. Green, and S. Garneau-Tsodikova, A random sequential mechanism of aminoglycoside acetylation by Mycobacterium tuberculosis Eis protein. PLoS One, 2014. **9**(4): p. e92370.
359. Eisenreich, W., et al., Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. Nat Rev Microbiol, 2010. **8**(6): p. 401-12.
360. Munoz-Elias, E.J. and J.D. McKinney, Carbon metabolism of intracellular bacteria. Cell Microbiol, 2006. **8**(1): p. 10-22.
361. Zhang, X., et al., Lysine Acetylation of Pyruvate Dehydrogenase Reduces Enzymatic Activity and Contributes to Impaired Substrate Metabolism in the Failing Myocardium. 2014, Am Heart Assoc.
362. Ozden, O., et al., SIRT3 deacetylates and increases pyruvate dehydrogenase activity in cancer cells. Free Radic Biol Med, 2014. **76**: p. 163-172.

363. McKinney, J.D., et al., Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. Nature, 2000. **406**(6797): p. 735.
364. Shi, L., et al., Carbon flux rerouting during Mycobacterium tuberculosis growth arrest. Mol Microbiol, 2010. **78**(5): p. 1199-215.
365. Wolfe, A.J., The acetate switch. Microbiol Mol Biol Rev, 2005. **69**(1): p. 12-50.
366. Contiero, J., et al., Effects of mutations in acetate metabolism on high-cell-density growth of Escherichia coli. Journal of Industrial Microbiology & Biotechnology, 2000. **24**(6): p. 421-430.
367. Billig, S., et al., Lactate oxidation facilitates growth of Mycobacterium tuberculosis in human macrophages. Scientific reports, 2017. **7**(1): p. 6484.
368. Banerjee, R., et al., Comparative genomics of cell envelope components in mycobacteria. PLoS One, 2011. **6**(5): p. e19280.
369. Daniel, J., T. Sirakova, and P. Kolattukudy, An acyl-CoA synthetase in Mycobacterium tuberculosis involved in triacylglycerol accumulation during dormancy. PLoS One, 2014. **9**(12): p. e114877.
370. Colangeli, R., et al., The Mycobacterium tuberculosis iniA gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. Mol Microbiol, 2005. **55**(6): p. 1829-40.
371. Zhang, F. and J.P. Xie, Mammalian cell entry gene family of Mycobacterium tuberculosis. Mol Cell Biochem, 2011. **352**(1-2): p. 1-10.
372. Zondervan, N.A., et al., Regulation of Three Virulence Strategies of Mycobacterium tuberculosis: A Success Story. International journal of molecular sciences, 2018. **19**(2): p. 347.
373. Gloster, T.M. and D.J. Vocadlo, Mechanism, structure, and inhibition of O-GlcNAc processing enzymes. Current signal transduction therapy, 2010. **5**(1): p. 74-91.
374. Cain, J.A., et al., Proteomics reveals multiple phenotypes associated with N-linked glycosylation in Campylobacter jejuni. Mol Cell Proteomics, 2019.
375. Kowarik, M., et al., Definition of the bacterial N-glycosylation site consensus sequence. Embo j, 2006. **25**(9): p. 1957-66.
376. Rini, J.M. and J.D. Esko, Glycosyltransferases and Glycan-Processing Enzymes, in Essentials of Glycobiology, rd, et al., Editors. 2015, Cold Spring Harbor Laboratory Press. Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved.: Cold Spring Harbor (NY). p. 65-75.
377. Sambou, T., et al., Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: biosynthesis and impact on the persistence in mice. Molecular microbiology, 2008. **70**(3): p. 762-774.
378. Torrelles, J.B., A.K. Azad, and L.S. Schlesinger, Fine discrimination in the recognition of individual species of phosphatidyl-myo-inositol mannosides from Mycobacterium tuberculosis by C-type lectin pattern recognition receptors. The Journal of immunology, 2006. **177**(3): p. 1805-1816.
379. Fenton, M.J., L.W. Riley, and L.S. Schlesinger, Receptor-mediated recognition of Mycobacterium tuberculosis by host cells, in Tuberculosis and the tubercle bacillus. 2005, American Society of Microbiology. p. 405-426.
380. Sutcliffe, I.C. and D.J. Harrington, Lipoproteins of Mycobacterium tuberculosis: an abundant and functionally diverse class of cell envelope components. FEMS microbiology reviews, 2004. **28**(5): p. 645-659.
381. Nguyen, H.T., et al., A lipoprotein modulates activity of the MtrAB two-component system to provide intrinsic multidrug resistance, cytokinetic control and cell wall homeostasis in Mycobacterium. Mol Microbiol, 2010. **76**(2): p. 348-64.
382. Parra, J., et al., Scrutiny of Mycobacterium tuberculosis 19 kDa antigen proteoforms provides new insights in the lipoglycoprotein biogenesis paradigm. Sci Rep, 2017. **7**: p. 43682.
383. Ocampo, M., et al., Specific interaction between Mycobacterium tuberculosis lipoprotein-derived peptides and target cells inhibits mycobacterial entry in vitro. Chem Biol Drug Des, 2014. **84**(6): p. 626-41.

384. 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. 7153-7159.
385. Ahmad, S., et al., The six mammalian cell entry proteins (Mce3A–F) encoded by the mce3 operon are expressed during in vitro growth of Mycobacterium tuberculosis. Scandinavian journal of immunology, 2005. **62**(1): p. 16-24.
386. Ragas, A., et al., The Mycobacterium tuberculosis cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A. Journal of biological chemistry, 2007. **282**(8): p. 5133-5142.
387. Satchidanandam, V., et al., The glycosylated Rv1860 protein of Mycobacterium tuberculosis inhibits dendritic cell mediated TH1 and TH17 polarization of T cells and abrogates protective immunity conferred by BCG. PLoS Pathog, 2014. **10**(6): p. e1004176.
388. Rodríguez, D., et al., Peptides from the Mycobacterium tuberculosis Rv1980c protein involved in human cell infection: insights into new synthetic subunit vaccine candidates. Biological chemistry, 2010. **391**(2/3): p. 207-217.
389. Pethe, K., et al., The heparin-binding haemagglutinin of M. tuberculosis is required for extrapulmonary dissemination. Nature, 2001. **412**(6843): p. 190-194.
390. Becker, K. and P. Sander, Mycobacterium tuberculosis lipoproteins in virulence and immunity–fighting with a double‐edged sword. FEBS letters, 2016. **590**(21): p. 3800-3819.
391. Bianco, M.V., et al., Role of P27-P55 operon from Mycobacterium tuberculosis in the resistance to toxic compounds. BMC infectious diseases, 2011. **11**(1): p. 195.
392. Zimhony, O., et al., AcpM, the Meromycolate Extension Acyl Carrier Protein of Mycobacterium tuberculosis, Is Activated by the 4′-Phosphopantetheinyl Transferase PptT, a Potential Target of the Multistep Mycolic Acid Biosynthesis. Biochemistry, 2015. **54**(14): p. 2360-2371.
393. Gavalda, S., et al., The Polyketide Synthase Pks13 Catalyzes a Novel Mechanism of Lipid Transfer in Mycobacteria. Chemistry & Biology, 2014. **21**(12): p. 1660-1669.
394. Steingart, K.R., et al., Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: a meta-analysis. Clinical and vaccine immunology, 2009. **16**(2): p. 260-276.
395. Kieser, K.J., et al., Peptidoglycan synthesis in Mycobacterium tuberculosis is organized into networks with varying drug susceptibility. Proceedings of the National Academy of Sciences, 2015. **112**(42): p. 13087-13092.
396. Lovering, A.L., S.S. Safadi, and N.C. Strynadka, Structural perspective of peptidoglycan biosynthesis and assembly. Annual review of biochemistry, 2012. **81**: p. 451-478.
397. Goude, R., et al., The arabinosyltransferase EmbC is inhibited by ethambutol in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy, 2009. **53**(10): p. 4138-4146.
398. Goude, R. and T. Parish, The genetics of cell wall biosynthesis in Mycobacterium tuberculosis. 2008.
399. Wei, J., et al., Genome-wide transcription analyses in Mycobacterium tuberculosis treated with lupulone. Brazilian Journal of Microbiology, 2014. **45**(1): p. 333-342.
400. Waddell, S., et al., Inactivation of polyketide synthase and related genes results in the loss of complex lipids in Mycobacterium tuberculosis H37Rv. Letters in applied microbiology, 2005. **40**(3): p. 201-206.
401. Pérez, J., et al., Mycobacterium tuberculosis transporter MmpL7 is a potential substrate for kinase PknD. Biochemical and biophysical research communications, 2006. **348**(1): p. 6-12.
402. Koliwer-Brandl, H., et al., Metabolic Network for the Biosynthesis of Intra- and Extracellular alpha-Glucans Required for Virulence of Mycobacterium tuberculosis. PLoS Pathog, 2016. **12**(8): p. e1005768.
403. Lin, W., et al., Transcriptional Profiling of Mycobacterium tuberculosis Exposed to In Vitro Lysosomal Stress. Infection and immunity, 2016. **84**(9): p. 2505-2523.
404. Leiba, J., et al., Mycobacterium tuberculosis maltosyltransferase GlgE, a genetically validated antituberculosis target, is negatively regulated by Ser/Thr phosphorylation. Journal of Biological Chemistry, 2013. **288**(23): p. 16546-16556.

405. Dkhar, H.K., et al., Discovery of Mycobacterium tuberculosis α-1, 4-glucan branching enzyme (GlgB) inhibitors by structure-and ligand-based virtual screening. Journal of Biological Chemistry, 2015. **290**(1): p. 76-89.
406. Torrelles, J.B., et al., Inactivation of Mycobacterium tuberculosis mannosyltransferase pimB reduces the cell wall lipoarabinomannan and lipomannan content and increases the rate of bacterial-induced human macrophage cell death. Glycobiology, 2009. **19**(7): p. 743-755.
407. Mishra, A.K., et al., Identification of a novel α (1→ 6) mannopyranosyltransferase MptB from Corynebacterium glutamicum by deletion of a conserved gene, NCgl1505, affords a lipomannan‐and lipoarabinomannan‐deficient mutant. Molecular microbiology, 2008. **68**(6): p. 1595-1613.
408. Portevin, D., et al., The Acyl-AMP Ligase FadD32 and AccD4-containing Acyl-CoA Carboxylase Are Required for the Synthesis of Mycolic Acids and Essential for Mycobacterial Growth IDENTIFICATION OF THE CARBOXYLATION PRODUCT AND DETERMINATION OF THE ACYL-CoA CARBOXYLASE COMPONENTS. Journal of Biological Chemistry, 2005. **280**(10): p. 8862-8874.
409. Nampoothiri, K., et al., Molecular cloning, overexpression and biochemical characterization of hypothetical β‐lactamases of Mycobacterium tuberculosis H37Rv. Journal of applied microbiology, 2008. **105**(1): p. 59-67.
410. Zhou, P., et al., Mycobacterium tuberculosis two-component systems and implications in novel vaccines and drugs. Crit Rev Eukaryot Gene Expr, 2012. **22**(1): p. 37-52.
411. Domenech, P., et al., Unique Regulation of the DosR Regulon in the Beijing Lineage of Mycobacterium tuberculosis. J Bacteriol, 2017. **199**(2).
412. Converse, P.J., et al., Role of the dosR-dosS two-component regulatory system in Mycobacterium tuberculosis virulence in three animal models. Infection and immunity, 2009. **77**(3): p. 1230-1237.
413. Feltcher, M.E., J.T. Sullivan, and M. Braunstein, Protein export systems of Mycobacterium tuberculosis: novel targets for drug development? Future microbiology, 2010. **5**(10): p. 1581-1597.
414. da Silva, P.E.A., et al., Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis. FEMS Immunology & Medical Microbiology, 2011. **63**(1): p. 1-9.
415. Gupta, A.K., et al., Microarray analysis of efflux pump genes in multidrug-resistant Mycobacterium tuberculosis during stress induced by common anti-tuberculous drugs. Microbial drug resistance, 2010. **16**(1): p. 21-28.
416. Chalut, C., MmpL transporter-mediated export of cell-wall associated lipids and siderophores in mycobacteria. Tuberculosis (Edinb), 2016. **100**: p. 32-45.
417. Viljoen, A., et al., The diverse family of MmpL transporters in mycobacteria: from regulation to antimicrobial developments. Molecular microbiology, 2017. **104**(6): p. 889-904.
418. Danilchanka, O., et al., An outer membrane channel protein of Mycobacterium tuberculosis with exotoxin activity. Proc Natl Acad Sci U S A, 2014. **111**(18): p. 6750-5.
419. Raman, K., K. Yeturu, and N. Chandra, targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis. BMC Syst Biol, 2008. **2**: p. 109.
420. Chakhaiyar, P., et al., Regions of high antigenicity within the hypothetical PPE major polymorphic tandem repeat open-reading frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. J Infect Dis, 2004. **190**(7): p. 1237-44.
421. Owens, R.M., et al., M. tuberculosis Rv2252 encodes a diacylglycerol kinase involved in the biosynthesis of phosphatidylinositol mannosides (PIMs). Mol Microbiol, 2006. **60**(5): p. 1152-63.
422. Ojha, A., et al., GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. Cell, 2005. **123**(5): p. 861-873.
423. Amir, A., et al., Mycobacterium tuberculosis H37Rv: In Silico Drug Targets Identification by Metabolic Pathways Analysis. Int J Evol Biol, 2014. **2014**: p. 284170.
424. Shevchenko, A., et al., Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Analytical chemistry, 1996. **68**(5): p. 850-858.

425. Choksawangkarn, W., et al., Comparative study of workflows optimized for in-gel, in-solution, and on-filter proteolysis in the analysis of plasma membrane proteins. Journal of proteome research, 2012. **11**(5): p. 3030-3034.
426. Shevchenko, A., et al., In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature protocols, 2006. **1**(6): p. 2856.
427. Luque-Garcia, J.L. and T.A. Neubert, On-membrane tryptic digestion of proteins for mass spectrometry analysis, in Protein Blotting and Detection. 2009, Springer. p. 331-341.
428. Fic, E., et al., Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. Electrophoresis, 2010. **31**(21): p. 3573-3579.
429. Wiśniewski, J.R., et al., Universal sample preparation method for proteome analysis. Nature methods, 2009. **6**(5): p. 359.
430. Wisniewski, J.R. and M. Mann, Spin filter-based sample preparation for shotgun proteomics Reply. Nature Methods, 2009. **6**(11): p. 785-786.
431. Bereman, M.S., J.D. Egertson, and M.J. MacCoss, Comparison between procedures using SDS for shotgun proteomic analyses of complex samples. Proteomics, 2011. **11**(14): p. 2931-2935.
432. Swaney, D.L., et al., Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat Methods, 2013. **10**(7): p. 676-82.
433. Mertins, P., et al., Integrated proteomic analysis of post-translational modifications by serial enrichment. Nature methods, 2013. **10**(7): p. 634.
434. Wu, R., et al., A large-scale method to measure absolute protein phosphorylation stoichiometries. Nat Methods, 2011. **8**(8): p. 677-83.
435. Olsen, J.V., et al., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal, 2010. **3**(104): p. ra3.
436. Li, Z., et al., Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. Nat Commun, 2014. **5**: p. 4405.
437. Yang, Y., et al., Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. Protein Cell, 2015. **6**(4): p. 265-74.
438. Kim, S.C., et al., Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Mol Cell, 2006. **23**(4): p. 607-18.

SUPPORTING INFORMATION

Ample glycosylation in the membrane and cell envelope explain the phenotypic diversity and virulence in the *Mycobacterium tuberculosis* complex

Alemayehu Godana Birhanu, Solomon Abebe Yimer, Shewit Kalayou, Tahira Riaz, Ephrem Debebe Zegeye, Carol Holm-Hansen, Gunnstein Norheim, Abraham Aseffa, Markos Abebe, Tone Tønjum

Supplementary Table S1. A list of glycan residues analyzed with their respective occurrences at multiplicities of _1, _2 and _3.

Glycan residue	Formula	Neutral loss / Monoisotopic mass	M+H+ ion	Number of glycosylation sites				Remark
				-1	-2	-3	Total sites	
DeoxyHex	C6H10O4	146.0579	147.0652	98	94	67	259	Fuc, Rha
Hept	C7H12O6	192.0634	193.0707	95	97	42	234	Heptoses
Pent	C5H8O4	132.0422	133.0495	95	93	43	231	Pentoses (Ara, Xyl, Rib)
Hex	C6H10O5	162.0528	163.0601	67	83	48	198	Hexoses
HexNac	C8H13O5N	203.0794	204.0866	57	61	29	147	N-Acetylhexosamine
HexN	C6H11O4N	161.0688	162.0761	35	54	57	146	Hexosamines
MurNGlyc	C11H17O8N	291.0954	292.1027	64	48	15	127	N-Glycolyl-Muramic Acid
MurNac	C11H17O7N	275.1005	276.1078	50	60	13	123	N-acetylmuramic acid
Mur	C9H15O6N	233.0899	234.0972	34	50	26	110	Muraminic acid
HexA	C6H8O6	176.0321	177.0394	34	46	19	99	Hexuronic acid
HexHex	C12H20O10	324.1056	325.1129	35	42	14	91	Hexobiose /Trehalose
HexNacMurNac	C19H30O12N2	478.1799	479.1872	51	18		69	PG repeating units
HexHexNac	C14H23O10N	365.1322	366.1395	39	29		68	
OAcMurNac	C13H19O8N	317.1111	318.1183	43	20	4	67	2,6-N,O-diacetyl muramic acid
HexNacOAc_MurNac	C21H32O13N2	520.1904	521.1977	47	13		60	PG repeating units (O-acetylated)
OAcMurNGlyc	C13H19O9N	333.106	334.1133	37	21	1	59	2-N-Glycoly 6-O-acetyl muramic acid
HexHexHex	C18H30O15	486.1585	487.1657	40	11	6	57	Hexotriose
HexNacMurNGlyc	C19H30O13N2	494.1748	495.1821	31	17	3	51	PG repeating units
HexNac(2)	C16H26O10N2	406.1587	407.166	26	18	5	49	
HexOAc_MurNac	C19H29O13N	479.164	480.1712	35	13		48	
OAcHexHex	C14H22O11	366.1162	367.1235	26	18		44	O-acetylated trehalose
Hex(1)HexNAc(2)	C22H36O15N2	568.2116	569.2188	22	20		42	
HexMurNGlyc	C17H 27O13N	453.1482	454.1555	32	10		42	

AraGlc	C20H34O13	482.1999	483.2072	21	14	6	41	Arabinogalactan
HexMurNac	C17H27O12N	437.1533	438.1606	32	9		41	
HexOAc_MurNGlyc	C19H 29O14N	495.1588	496.1661	38	2		40	
HexNacOAc_MurNGlyc	C21H32O14N2	536.1854	537.1926	30	6		36	PG repeating units (O-acetylated)
HexNac(1)MurNac(1)MurNGlyc(1)	C30H47O20N3	769.2753	770.2826	17	6		23	PG repeating units
HexNac(1)OAc_MurNac(1)OAc_MurNGlyc(1)	C34H51O22N3	853.2964	854.3037	21	2		23	PG repeating units (O-acetylated)
HexNac(2)MurNac	C27H43O17N3	681.2592	682.2665	21			21	PG repeating units
HexNac(1)MurNac(2)	C30H47O19N3	753.2804	754.2877	19	1		20	PG repeating units
HexNac(1)MurNGlyc(2)	C30H47O21N3	785.2702	786.2775	15	5		20	PG repeating units
HexNac(2)OAc_MurNGlyc	C29H45O19N3	739.2647	740.272	20			20	PG repeating units (O-acetylated)
HexNac(1)(OAc)2_MurNac(2)	C34H51O21N3	837.3015	838.3088	18	1		19	PG repeating units (O-acetylated)
HexNac(2)MurNGlyc	C27H43O18N3	697.2541	698.2614	14	4		18	PG repeating units
HexNacOAc_MurNacMurNGlyc	C32H49O21N3	811.2859	812.2931	18			18	PG repeating units (O-acetylated)
HexNac(2)OAc_MurNac	C29H45O18N3	723.2698	724.2771	14	2		16	PG repeating units (O-acetylated)
HexNac(1)OAc_MurNac(2)	C32H49O20N3	795.2909	796.2982	14			14	PG repeating units (O-acetylated)
Hex(2)HexNAc(2)	C28H46O20N2	730.2644	731.2717	11	2		13	
UDP-MurNGly	C20H29O19N3P2	677.087	678.0943	12			12	Activated sugar
ADP-Hept	C17H25O15N5P2	601.0822	602.0895	11			11	
UDP-Pent	C14H20O15N2P2	518.0339	519.0412	11			11	Activated sugar
ADP-HexN	C16H24O13N6P2	570.0877	571.0949	10			10	
ADP-HexNac	C18H26O14N6P2	612.0982	613.1055	10			10	
HexNac(1)OAc_MurNGlyc(2)	C32H49O22N3	827.2808	828.288	8	2		10	PG repeating units (O-acetylated)
HexNac(1)(OAc)2_MurNGlyc(2)	C34H51O23N3	869.2913	870.2986	9			9	PG repeating units (O-acetylated)
UDP-HexNac	C17H25O16N3P2	598.071	590.0783	9			9	Activated sugar
ADP-DeoxyHex	C16H23O13N5P2	555.0768	556.084	8			8	
UDP-DeoxyHex	C15H22O15N2P2	532.0495	533.0568	7			7	Activated sugar
UDP-MurNac	C20H29O20N3P2	661.0921	662.0994	7			7	Activated sugar
ADP-Hex	C16H23O14N5P2	571.0717	572.0789	6			6	
UDP-HexN	C15H23O15N3P2	547.0604	548.0677	6			6	Activated sugar
ADP-MurNac	C21H30O16N6P2	684.1193	685.1266	5			5	
ADP-MurNGly	C21H30O17N6P2	700.1143	701.1215	5			5	
ADP-Pent	C15H21O13N5P2	541.0611	542.0684	5			5	
UDP-Hept	C16H24O17N2P2	578.055	579.0623	5			5	Activated sugar
UDP-Hex	C15H22O16N2P2	548.0445	549.0517	4			4	Activated sugar
Total							**2944**	

Supplementary Table S2. A list of differentially glycosylated proteins in MTBC, P≤0.05, values normalized to Z-scores.

	Protein Name	Gene ID	Position	Amino acid	Charge	Glycan residue	L3	L4	L5	L7	-Log ANOVA p value
1	ATPase	moxR3	306	S	5	ADP-DeoxyHex	-0.432	-1.334	0.595	0.090	1.762
2	Uncharacterized HTH-type transcriptional regulator Rv1828	Rv1828	29	T	2	ADP-Hept	0.180	0.638	0.795	-1.840	3.270
3	Nicotinate-nucleotide pyrophosphorylase [carboxylating]	nadC	141	T	2	ADP-HexN	1.356	-0.676	0.337	-0.334	1.490
4	Conserved protein	Rv3683	7	T	2	ADP-HexNac	0.121	0.698	0.922	-1.372	1.923
5	Uncharacterized protein Rv2567	Rv2567	565	T	2	ADP-Pent	0.579	0.457	1.034	-1.726	1.537
6	Acyl-CoA dehydrogenase	fadE35	106	S	3	DeoxyHex	1.012	0.330	-0.527	-0.845	1.414
7	Ribonuclease VapC10	vapC10	6	S	3	DeoxyHex	-0.308	1.547	-0.570	-0.114	1.636
8	Diacylglycerol kinase	dagK	2	S	2	DeoxyHex	0.354	0.855	0.145	-1.877	2.697
9	Fatty acyl-AMP ligase	fadD34	383	T	2	DeoxyHex	-0.663	-0.581	-0.472	1.608	2.880
10	Membrane protein	Rv0412c	394	S	2	DeoxyHex	-0.781	-0.772	1.328	0.436	3.596
11	Iron-sulfur cluster carrier protein	mrp	373	S	2	DeoxyHex	-0.998	-0.853	0.399	1.272	2.529
12	Allophanate hydrolase	Rv0263c	2	T	2	DeoxyHex	-0.811	-0.828	-0.113	1.246	2.211
13	Allophanate hydrolase	Rv0263c	3	T	2	DeoxyHex	-0.822	-0.509	-0.190	1.247	2.181
14	Membrane protein	Rv0412c	388	T	2	DeoxyHex	-0.916	-0.998	1.383	0.396	4.548
15	Iron-sulfur cluster carrier protein	mrp	379	T	2	DeoxyHex	-0.854	-0.835	0.367	1.299	2.703
16	Type III pantothenate kinase	coaX	59	T	2	DeoxyHex	0.331	0.588	0.325	-1.708	1.433
17	Type III pantothenate kinase	coaX	61	T	2	DeoxyHex	0.334	0.580	0.328	-1.720	1.453
18	Acyl-CoA dehydrogenase	fadE35	110	S	3	Hep	1.031	0.300	-0.619	-0.763	1.362
19	L,D-transpeptidase 1	ldtA	147	T	2	Hep	0.339	1.020	-0.713	-0.183	1.320
20	Uncharacterized protein	Rv1868	9	T	2	Hep	-0.799	-1.053	0.608	1.081	3.322
21	Penicillin-binding protein PbpB	pbpB	83	S	3	Hex	-0.793	1.313	-0.080	-0.788	1.506
22	Uncharacterized protein	Rv3090	289	T	2	Hex	-1.237	0.097	0.209	0.733	2.493
23	Uncharacterized protein	Rv0690c	156	N	2	Hex	-0.513	-0.045	-0.166	1.776	2.174
24	Probable conserved lipoprotein DsbF	dsbF	47	T	3	Hex	0.923	0.507	0.183	-1.278	1.611
25	Serine/threonine-protein kinase PknA	pknA	252	T	2	Hex	-1.003	-0.877	0.692	1.212	4.338
26	UvrABC system protein A	uvrA	628	T	2	Hex	1.349	0.622	-0.769	-0.975	2.016
27	Serine/threonine-protein kinase PknA	pknA	256	N	2	Hex	-1.008	-0.887	0.685	1.206	4.204
28	UvrABC system protein A	uvrA	636	N	2	Hex	1.404	0.518	-0.652	-0.743	1.744
29	Possible exported protein	Rv3483c	78	S	4	Hex	0.697	1.179	-0.374	-1.219	1.864
30	Probable conserved lipoprotein DsbF	dsbF	26	S	3	Hex	-0.395	-0.955	0.598	1.324	1.787

31	Lipoarabinomannan carrier protein LprG	lprG	161	N	3	Hex(1)HexNAc(2)	0.255	0.438	0.755	-1.499	2.161
32	N5-carboxyaminoimidazole ribonucleotide synthase	purK	204	S	3	Hex(1)HexNAc(2)	0.626	0.536	0.059	-0.874	1.698
33	Cobyric acid synthase	cobQ	2	S	2	HexA	-0.115	0.603	0.525	-1.410	1.969
34	Uncharacterized protein Rv1303	Rv1303	85	S	2	HexA	0.164	0.684	0.737	-1.619	2.159
35	Antitoxin VapB11	vapB11	7	S	2	HexHex	-0.584	0.397	-1.145	0.758	3.506
36	Membrane protein	Rv3737	114	T	3	HexHex	-0.070	0.463	0.793	-1.069	1.365
37	Membrane protein	Rv3737	116	S	3	HexHex	0.038	0.585	0.923	-1.351	1.706
38	Probable conserved lipoprotein DsbF	dsbF	33	T	3	HexHex	-0.756	-0.793	0.599	1.341	1.915
39	Probable conserved lipoprotein DsbF	dsbF	40	T	3	HexHex	-0.583	-0.874	0.622	1.329	2.028
40	Membrane protein	Rv3737	114	T	3	HexHex	0.047	0.583	0.915	-1.372	1.707
41	Uncharacterized oxidoreductase Rv1144	Rv1144	110	N	2	HexHex	0.884	-0.624	0.948	-1.068	1.552
42	Uncharacterized oxidoreductase Rv1144	Rv1144	116	N	2	HexHex	0.884	-0.624	0.948	-1.068	1.552
43	Probable transcriptional regulatory protein NarL	narL	113	S	3	HexHex	0.993	0.322	-0.707	-0.632	1.550
44	Probable transcriptional regulatory protein NarL	narL	114	T	3	HexHex	1.085	-0.263	-1.160	-0.850	2.443
45	Putative amidase AmiD	amiD	211	S	3	HexHexNac	-0.607	0.216	1.070	-1.418	2.477
46	Acetyl-CoA acetyltransferase	fadA3	12	S	3	HexN	0.583	-1.101	-1.069	1.236	1.896
47	ABC transporter substrate-binding protein	Rv0987	215	S	2	HexN	-0.135	0.287	0.742	-1.174	1.762
48	Uncharacterized protein Rv1303	Rv1303	80	S	2	HexN	0.178	0.669	0.720	-1.646	2.117
49	Cobyric acid synthase	cobQ	11	S	2	HexN	-0.153	0.560	0.538	-1.384	1.860
50	Cobyric acid synthase	cobQ	9	T	2	HexN	-0.153	0.560	0.538	-1.384	1.860
51	Cobyric acid synthase	cobQ	10	T	2	HexN	-0.153	0.560	0.538	-1.384	1.860
52	Uncharacterized protein	Rv3098c	130	S	2	HexNac	-0.406	-0.725	1.637	-0.813	3.328
53	Mammalian cell entry protein	mce2D	270	S	2	HexNac	0.025	0.491	1.298	-1.047	1.383
54	D-alanyl-D-alanine carboxypeptidase	dacB1	306	T	2	HexNac	-0.507	-0.338	-0.605	1.513	2.245
55	D-alanyl-D-alanine carboxypeptidase	dacB1	312	T	2	HexNac	-0.427	-0.364	-0.512	1.548	2.761
56	60 kDa chaperonin 1	groEL1	150	S	2	HexNac	-0.281	1.434	-0.355	-0.661	1.936
57	60 kDa chaperonin 1	groEL1	151	S	2	HexNac	-0.256	1.437	-0.709	-0.552	1.869
58	60 kDa chaperonin 1	groEL1	141	T	2	HexNac	-0.248	1.415	-0.442	-0.607	1.805
59	RNA polymerase-binding protein RbpA	rbpA	9	S	2	HexNac(1)MurNac(2)	-0.765	0.921	1.027	-1.143	1.336
60	Transcriptional regulatory protein	Rv0818	65	S	2	HexNac(2)	-0.457	1.371	-0.496	-1.012	2.119
61	Possible GDP-mannose 4,6-dehydratase Gca (GDP-D-mannose dehydratase)	Rv0112; gca	211	T	3	HexNac(2)OAc_MurNGly	-0.776	0.643	0.980	-1.077	1.449

62	Multifunctional 2-oxoglutarate metabolism enzyme	kgd	1137	N	3	HexNac(2)OAc_MurNGly	0.117	0.521	0.899	-1.304	1.429
63	Uncharacterized protein	Rv1765c Rv2015c	7;7	S	3	HexNacMurNac	1.033	-1.312	-0.010	-0.802	1.847
64	Probable short-chain type dehydrogenase/reductase	Rv2766c	174	S	3	HexNacOAc_MurNac	-0.562	-0.338	-1.025	1.506	1.658
65	DNA-binding protein HU homolog	hup	2	N	3	HexNacOAc_MurNacMurNGly	-0.784	0.279	0.948	1.114	1.409
66	Polyketide synthetase MbtD (Polyketide synthase)	mbtD	863	S	3	HexNacOAc_MurNGly	-0.160	-0.925	1.164	-0.402	1.775
67	Probable fatty-acid-CoA ligase FadD9 (Fatty-acid-CoA synthetase) (Fatty-acid-CoA synthase)	fadD9	97	T	2	HexOAc_MurNGly	-0.003	-0.308	-0.857	1.850	2.378
68	Glutamate--tRNA ligase	gltX	6	T	2	Mur	1.024	-0.672	0.497	-1.628	2.036
69	Lipoprotein	Rv3693	145	S	2	Mur	-1.132	-0.085	0.357	0.905	1.320
70	Lipoprotein	Rv3693	146	S	2	Mur	-1.183	-0.050	0.374	0.899	1.478
71	Lipoprotein	Rv3693	148	S	2	Mur	-1.482	-0.092	0.352	0.901	1.473
72	Uncharacterized PPE family protein PPE42	PPE42	12	S	2	MurNac	0.930	0.941	-0.830	-1.069	2.913
73	Uncharacterized HTH-type transcriptional regulator Rv1395	Rv1395	338	S	3	MurNac	0.008	0.821	0.974	-1.732	2.138
74	Probable transposase	Rv1035c	38	T	3	MurNac	-0.844	-0.601	-0.514	1.293	1.547
75	60 kDa chaperonin 1	groEL1	148	T	2	MurNac	-0.251	1.452	-0.557	-0.428	1.846
76	Possible transposase	Rv3638; Rv2944	187;176	S	2	MurNac	-0.123	0.353	0.475	-1.362	1.557
77	Uncharacterized protein	Rv0059	9	S	2	MurNac	1.403	-0.777	-0.543	-0.430	1.359
78	Possible transposase	Rv3638; Rv2944	191;180	T	2	MurNac	-0.156	0.336	0.462	-1.332	1.442
79	Possible roxidase BpoB (Non-haem roxidase)	bpoB	159	N	2	MurNGly	-0.405	-0.598	-0.731	1.502	1.438
80	Probable transposase	Rv1035c	39	N	3	MurNGly	-0.418	-0.467	-0.482	1.257	1.468
81	Uncharacterized protein	Rv3098c	132	S	2	MurNGly	-0.675	-0.427	1.601	-0.634	2.426
82	Possible transcriptional regulatory protein (Probably TetR-family)	Rv0238	69	S	3	MurNGly	0.564	-0.137	0.871	-0.848	1.588
83	UvrABC system protein C	uvrC	469	T	2	MurNGly	0.623	0.332	0.590	-1.448	1.362
84	Putative lipoprotein LppW	lppW	69	N	3	MurNGly	0.808	0.817	0.012	-1.143	1.501
85	Putative lipoprotein LppW	lppW	67	S	3	MurNGly	0.808	0.818	-0.001	-1.375	1.829
86	Putative lipoprotein LppW	lppW	77	T	3	MurNGly	0.805	0.814	0.029	-1.310	1.689

87	4-hydroxyacetophenone monooxygenase	Rv3049c	281	N	3	OAcHexHex	0.438	0.756	0.319	-1.825	2.927
88	Possible methyltransferase/methylase	Rv0521	45	S	2	OAcMurNGly	0.091	-1.039	0.309	1.209	1.302
89	Probable conserved integral membrane leucine and alanine rich protein	Rv2508c	145	S	2	OAcMurNGly	0.842	-1.250	-0.869	0.923	1.305
90	Beta-lactamase	blaC	35	S	2	Pentoses	0.967	-1.020	0.553	-0.553	1.421
91	Putative ribonuclease VapC45	vapC45	83	T	2	Pentoses	0.621	0.774	0.426	-1.494	1.681
92	L,D-transpeptidase 1	ldtA	156	S	2	Pentoses	0.339	1.020	-0.713	-0.183	1.320
93	L,D-transpeptidase 1	ldtA	151	S	2	Pentoses	0.339	1.020	-0.713	-0.183	1.320
94	Phytoene synthase	crtB	13	S	2	Pentoses	-0.617	0.596	1.141	-1.327	2.364
95	Phytoene synthase	crtB	15	T	2	Pentoses	-0.958	0.605	1.143	-0.867	1.621
96	Phytoene synthase	crtB	17	T	2	Pentoses	-0.470	0.596	1.078	-1.135	1.554
97	Uncharacterized ABC transporter ATP-binding protein Rv2326c	Rv2326c	591	T	2	UDP-MurNac	-0.361	-1.192	1.467	-0.388	2.002
98	30S ribosomal protein S13	rpsM	63	N	3	UDP-MurNGly	-1.198	0.645	0.549	0.459	1.769
99	Possible transcriptional regulatory protein (Probably MarR-family)	Rv0042c	132	S	2	UDP-MurNGly	-0.050	0.427	0.875	-1.121	1.496
100	Uncharacterized protein	Rv3863	11	S	2	UDP-MurNGly	0.068	-0.912	1.366	-0.925	1.889
101	Possible exported protein	Rv3483c	77	T	4	Hex	0.693	1.186	-0.406	-1.016	1.543

I

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 May 2017
doi: 10.3389/fmicb.2017.00795

Comparative Proteomic Analysis of *Mycobacterium tuberculosis* Lineage 7 and Lineage 4 Strains Reveals Differentially Abundant Proteins Linked to Slow Growth and Virulence

Solomon A. Yimer[1,2], Alemayehu G. Birhanu[2,3], Shewit Kalayou[1], Tahira Riaz[2], Ephrem D. Zegeye[4], Getachew T. Beyene[1], Carol Holm-Hansen[5], Gunnstein Norheim[5], Markos Abebe[6], Abraham Aseffa[6] and Tone Tønjum[1,2]*

[1] Department of Microbiology, Oslo University Hospital, Oslo, Norway, [2] Department of Microbiology, University of Oslo, Oslo, Norway, [3] Department of Medical Biotechnology, Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia, [4] Centre for Applied Biotechnology, Uni Research Environment, Bergen, Norway, [5] Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway, [6] Department of Research and Innovation, Armauer Hansen Research Institute, Addis Ababa, Ethiopia

OPEN ACCESS

Edited by:
Biswarup Mukhopadhyay, Virginia Tech, USA

Reviewed by:
Thomas Dick, Rutgers University, USA
Nomakorinte Gcebe, Agricultural Research Council, South Africa
Jonathan M. Blackburn, University of Cape Town, South Africa

***Correspondence:**
Solomon A. Yimer
s.a.yimer@medisin.uio.no

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *28 December 2016*
Accepted: *18 April 2017*
Published: *09 May 2017*

Citation:
Yimer SA, Birhanu AG, Kalayou S, Riaz T, Zegeye ED, Beyene GT, Holm-Hansen C, Norheim G, Abebe M, Aseffa A and Tønjum T (2017) Comparative Proteomic Analysis of Mycobacterium tuberculosis Lineage 7 and Lineage 4 Strains Reveals Differentially Abundant Proteins Linked to Slow Growth and Virulence. Front. Microbiol. 8:795. doi: 10.3389/fmicb.2017.00795

In order to decipher the nature of the slowly growing *Mycobacterium tuberculosis (M.* tuberculosis) lineage 7, the differentially abundant proteins in strains of *M. tuberculosis* lineage 7 and lineage 4 were defined. Comparative proteomic analysis by mass spectrometry was employed to identify, quantitate and compare the protein profiles of strains from the two *M. tuberculosis* lineages. Label-free peptide quantification of whole cells from *M. tuberculosis* lineage 7 and 4 yielded the identification of 2825 and 2541 proteins, respectively. A combined total of 2867 protein groups covering 71% of the predicted *M. tuberculosis* proteome were identified. The abundance of 125 proteins in *M. tuberculosis* lineage 7 and 4 strains was significantly altered. Notably, the analysis showed that a number of *M. tuberculosis* proteins involved in growth and virulence were less abundant in lineage 7 strains compared to lineage 4. Five ABC transporter proteins, three phosphate binding proteins essential for inorganic phosphate uptake, and six components of the type 7 secretion system ESX-3 involved in iron acquisition were less abundant in *M. tuberculosis* lineage 7. This proteogenomic analysis provided an insight into the lineage 7-specific protein profile which may provide clues to understanding the differential properties of lineage 7 strains in terms of slow growth, survival fitness, and pathogenesis.

Keywords: ***Mycobacterium tuberculosis*, tuberculosis, lineage 7, proteomics, Ethiopia, mass spectrometry, type 7 secretion**

INTRODUCTION

Tuberculosis (TB) has claimed an uncountable number of lives over centuries. One third of the global population is infected with the causative agent *Mycobacterium tuberculosis (M. tuberculosis)*, which is the main cause of TB. Each year, ~10.4 million people contract TB and 1.8 million die from the disease (WHO, 2016). In line with the World Health Organization (WHO) vision, the world

is united in the quest to eliminate TB by 2050 (WHO, 2016). A holistic approach to understand host, environmental and bacterial factors to TB susceptibility and disease is crucial to achieve the global TB elimination target. There is a large body of evidence on host and environmental attributes for TB (Rieder, 1999). However, given the potential implications for severity of illness and transmission, there is limited knowledge regarding the basic mechanisms underlying the physiology and pathogenesis of the different *M. tuberculosis* lineages.

Large sequence polymorphisms classify *M. tuberculosis* into 7 main lineages. These are lineage 1 (Indo-Oceanic), lineage 2 (East Asian including "Beijing"), lineage 3 (CAS/Delhi), lineage 4 (Euro-American including Latin American Mediterranean (LAM), Haarlem, X type and T families), lineage 5 and lineage 6 (West African 1 and 2, respectively), and lineage 7 (Comas et al., 2013). Lineage 7 was recently detected in Ethiopia and among Ethiopian immigrants in Djibouti (Blouin et al., 2012; Firdessa et al., 2013; Yimer et al., 2013, 2015). We have shown that cells of the *M. tuberculosis* lineage 7 grow slowly *in vitro* and that lineage 7 infections are associated with prolonged delay in seeking health care among patients compared to other lineages (Yimer et al., 2015). Whole genome sequencing (WGS) demonstrated that *M. tuberculosis* lineage 7 cells host a high number of mutations in genes involved in carbohydrate transport and metabolism, transcription, energy production and conversion, all of which contribute to the slow-growth phenotype (Yimer et al., 2016). Former WGS studies have described characteristics and attributes of the other *M. tuberculosis* lineages that may play important roles in the pathogenesis of TB (Coscolla and Gagneux, 2014). For example, the Beijing genotype (lineage 2) is associated with high bacillary load in acid-fast bacillus (AFB) smears and frequently acquires drug resistance (Coscolla and Gagneux, 2014). Lineages 2, 3, and 4 exhibited a lower early inflammatory response compared to lineage 1 and lineage 6 (Chacon-Salinas et al., 2005). Lineage 3 showed a higher anti-inflammatory phenotype compared to Lineage 4 (Portevin et al., 2011). *Mycobacterium africanum* (lineage 6) acquires drug resistance at a lower rate than the Euro-American (lineage 4) in Ghana (Albanna et al., 2011). This shows that even though they are genetically closely related, different strains of *M. tuberculosis* present very diverse clinical phenotypes in terms of virulence. We therefore hypothesized that there may be a mechanism underlying the variation in pathogenicity observed between strains of *M. tuberculosis* detectable at the proteomic level. Global proteomic characterization by use of mass spectrometry represents a powerful tool and is an important supplement to genomics in defining the protein expression patterns (indicating which genes are expressed and down- or upregulated; Peirs et al., 2005; Kelkar et al., 2011; Schubert et al., 2015; Bespyatykh et al., 2016; Jhingan et al., 2016; Peters et al., 2016). To date, no studies have addressed the relationship between the *M. tuberculosis* lineage 7 slow-growth phenotype and its proteomic signatures.

The objective of this study was to characterize the differentially abundant protein profile of *M. tuberculosis* lineage 7 (L7-35 and L7-28) and lineage 4 H37Rv strains and define the proteomic profiles relevant for growth and pathogenicity. This proteomic study generated novel insight into the differentially abundant (DA) proteins in *M. tuberculosis* lineage 7 vs. lineage 4 strains. A total of 2,867 proteins covering 71% of the predicted *M. tuberculosis* proteome were identified. The analysis of DA proteins by pathway clustering provided an overview on the lineage-specific protein interaction profile. This information may be used to explain the differential behavior of lineage 7 strains in terms of slow growth, survival fitness, and pathogenesis.

MATERIALS AND METHODS

Mycobacterial Strains and Growth Conditions

Whole genome sequenced *M. tuberculosis* lineage 7 strains (L7-35 and L7-28) collected form the Amhara Region of Ethiopia and lineage 4 strain H37Rv were streaked onto Middlebrook 7H10 plates in triplicates from freezer stocks, and incubated in a humidified 37°C, 5% CO_2 incubator. After 32 days the cells were carefully scrapped off the agar plates and put into 50 mL Falcon® tubes. The cell pellets were gently resuspended in 30 mL PBS, pH 7.4, and centrifuged at 3,900 rpm for 20 min at 4°C. The cell pellets were subsequently transferred into 2 mL screw capped tubes (Sarstedt, Nümbrecht, Germany) and resuspended in 1 mL PBS, and heat inactivated at 80°C for 90 min. Culturing and processing of the *M. tuberculosis* samples up until the heat inactivation step were conducted in a biosafety level 3 facility at Oslo University Hospital, Norway. The heat-inactivated *M. tuberculosis* samples were stored at −80°C until lysed for mass spectrometry analysis.

Proteomic Analysis

Cell lysis

The heat-inactivated cell pellets were resuspended in lysis buffer containing 2% SDS, 10 mM Tris-HCl (pH 7.5), 1 tablet per 50 mL EDTA-free Protease Inhibitor Cocktail (Sigma-Aldrich, Cleveland, US) and 1 tablet per 10 mL PhosSTOP Phosphatase Inhibitor Cocktail (Roche). The samples were subsequently transferred into Lysing Matrix B tubes (Roche) and disrupted mechanically by bead beating using MagNa Lyser (Roche Diognostics, GmbH, Mannheim, Germany) for 90 s, speed 6.0. The lysis procedure followed by 1 min cooling on ice was repeated six times. The lysate was clarified by centrifugation (15,000 × g for 15 min) at 21°C, and the supernatant containing the whole cell lysate proteins was transferred in to new 2 mL screw cap micro tubes (Sarstedt, Nümbrecht, Germany).

In-Gel Trypsin Digestion of Cellular Proteins

Hundred μg of protein sample dissolved in NuPAGE LDS sample buffer (4x) and NuPAGE Sample Reducing Agent (10X) (Life Technologies, USA) were incubated for 10 min at 70°C and pre-fractionated by 1.0 mm, 4–12% NuPAGE Novex Bis-Tris SDS-PAGE gel (Life Technologies), at 80 V for 5 min followed by 20 min at 200 V. Gels were Coomassie-stained using a Colloidal Blue Staining kit for NuPAGE as per manufacturer's instructions. After staining, each gel lane was divided into 6 fractions, and each fraction was subjected to in-gel reduction, alkylation, and tryptic digestion (Shevchenko et al., 2006). In brief, proteins were reduced using 10 mM DTT for 1 h at

56°C and alkylated with 55 mM iodoacetamide for 1 h at room temperature (Sigma-Aldrich, Cleveland, US). The reduced and alkylated peptides were digested with sequencing-grade trypsin (Promega, WI, USA, 1:100; w/w) for 16 h at 37°C in 50 mM NH_4HCO_3. The trypsin-digested protein samples were extracted from the gel using sequential (50 and 100%) acetonitrile (ACN), dried by SpeedVac concentrator (Eppendorf, concentrator 5301) and re-suspended using 0.05% trifluoroaceticacid (TFA). For desalting, the peptide samples were loaded on to C_{18} stage tips activated and equilibrated with 95% ACN/0.1% FA and 0.1% formic acid (FA), respectively. The loaded samples were washed with 0.05% TFA and eluted with 95% ACN/0.1% FA. The eluent was dried using a SpeedVac concentrator, re-suspended in 0.1% FA, transferred to auto-sampler nano-liquid chromatography (LC) vials and stored at −20°C. The proteomic work flow is depicted in Supplementary file 1.

Mass spectrometry (MS) analysis

Peptide identification and quantitation were performed by label-free quantification (LFQ) LC-MS/MS using a Q Exactive hybrid quadropole-orbitrap instrument interfaced with an EASY 1000-nano-LC electrospray ion source (Thermo-Fisher Scientific, Biberach, Germany). Peptides were injected in triplicates into a pre-analytic column (Acclaim PepMap 100, 75 μm × 2 cm, nanoviper, C18, 3 μm, 100 Å, Thermo Fisher Scientific) and separated on an analytical column (PepMap RSLC, C18, 2 μm, 100 Å, 50 μm × 15 cm, Thermo Fisher Scientific) with a 75 min solvent gradient and flow rate of 0.3 μL/min at 60°C. The gradient used was from 2 to 30% solvent B for 30 min followed by 30–75% solvent B from 30 to 35 min and 75 to 90% solvent B from 35 to 70 min. Thereafter the gradient was kept at 90% solvent B from 70 to 75 min, using 0.1% formic acid (FA) in 3% acetonitrile (ACN) as solvent A and 0.1% FA in 97% ACN as solvent B (FA: LC-MS grade, Fluka; ACN: LC-MS grade, Merck Laboratories). The MS instrument was operated in the data-dependent acquisition mode with automatic switching between MS and MS/MS scans. The full MS scans were acquired at 70 K resolution, with automatic gain control target of 1×10^6 ions, maximum injection time of 200 ms and MS scan range 300–1,800 m/z. Higher energy collision dissociation (HCD) was used for peptide fragmentation with normalized collision energy set to 28. The MS/MS scans were performed using a data-dependent top10 method at a resolution of 17.5 K with an automatic gain control target of 5×10^4 ions at maximum injection time of 100 ms and isolation window of 2.0 m/z units. An under fill ratio of 10% and dynamic exclusion duration of 30 s were applied. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [1] with the dataset identifier PXD006117.

Database search

The MS/MS data analysis was performed using the MaxQuant (MQ) software package (version 1.4.0.5; Cox and Matthias, 2008) for analyzing large MS/MS data sets, employing its integrated Andromeda search algorithms (Cox et al., 2011). The raw spectral data were searched against the *M. tuberculosis* H37Rv reference proteome UP000001584 (UniProt-proteome) using reverse decoy databases and a selection of known contaminants provided by MQ. The following parameters were applied for the database search: Enzyme specificity was set as Trypsin/P, and a maximum of two missed cleavages and a mass tolerance of 0.5 Da for fragment ion were applied. The "re-quantify" and "match between runs" options were utilized with a retention time alignment window of 3 min. Carbamidomethylation of cysteine was set as a fixed modification and acetylation of the protein N-terminus, conversion of N-terminal glutamine and glutamic acid to pyroglutamic acid and oxidation of methionine were set as variable modifications for database searches. The first search for precursor ions was performed with a mass tolerance of 20 ppm for calibration, while 6 ppm was applied for the main search. For protein identification, at least 1 unique peptide was required per protein group (Cox and Matthias, 2008; Zhao and Lin, 2010). Minimum peptide length of 7 amino acids was required for identification. The maximum false discovery rate (FDR) cutoff of 0.01 (1%) was set at both the peptide spectra matches and the protein group levels. For all other parameters, the default setting was applied. Following protein identification by a database search, validation for multiple comparisons was corrected using Benjamini-Hochberg correction (Benjamini et al., 2001). To aid in the control of false positives, the database was supplemented with additional sequences for common contaminants and reversed sequences of each entry.

Bioinformatics analysis

Bioinformatics analysis was performed using the Perseus software (version 1.5.1.6) as previously described (Tyanova et al., 2016). The protein groups output from MQ was used as the basis for all the subsequent statistical and ontology enrichment analysis. LFQ intensities were used to assess differences in the abundance of proteins between the two *M. tuberculosis* lineages. Abundance estimation of the proteins identified was performed using intensity-based absolute quantification (iBAQ) values. Briefly, the protein groups output was filtered by removing matches to the reverse database, matches only identified by site, and common contaminants. Subsequently, LFQ intensities were transformed to $\log_2$. For quantitative comparisons, three technical replicates of each biological experiment ($n = 3$) were averaged based on their median values. An LFQ intensity category was then created that consisted of the intensities from two clinical isolates (*M. tuberculosis* L728 and L735) of lineage 7. This merger served as a core proteome for lineage 7, which was later used to compare against the H37Rv lineage 4 reference strain. For statistical analysis, at least two valid LFQ intensities out of the three biological experiments were required. Signals that originally were zero (missing values) were imputed with random numbers from a normal distribution. The mean and standard deviation were chosen to best simulate low abundance values below the noise level (width = 0.3; shift = 1.8; Hubner et al., 2010). A two-tailed unpaired *t*-test with an FDR value of 0.05 and $S_0 = 2$ was applied to identify proteins for which the abundance was significantly changed between the two *M. tuberculosis* lineages (Tusher et al., 2001). The resulting *t*-test -significant proteins for each *M. tuberculosis* lineage were analyzed for

annotation enrichments. A two-tailed Fisher's Exact Test was used to assess the significance of enrichment terms. Proteins assigned to enriched term categories ($p < 0.05$) were grouped according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification. The GO and KEGG categories of proteins identified were added using the Uniprot annotation for the *M. tuberculosis* reference proteome database (Gene ontology[1]).

Protein Interaction and Network Analysis

The Search Tool for the Retrieval of Interacting Genes version 10.0 (Franceschini et al., 2013; STRING, http://string-db.org/) was used to interpret the biological significance of DA proteins in terms of predicted protein-protein interaction networks. The required minimum interaction score of at least 0.4 was used as the cut-off point criterion. The Cytoscape software (Shannon et al., 2003) was used to visualize the interaction network predicted. A list of DA proteins, the official gene identifier, and the corresponding relative abundance value were separately uploaded to Cytoscape (http://www.cytoscape.org/). Properties of the network including node degree and edge attributes were then analyzed. Nodes represent proteins and edges represent the interactions/connections between the proteins. The degree represents the number of interactions associated with the protein. Proteins with a large degree are known as hub proteins (Azuaje et al., 2010) and are considered to be the essential or key proteins in the network (Ideker and Sharan, 2008). The Network Analyzer option in Cytoscape 3.4.1 was used to compute the degree and *between-ness* centrality of the network (Assenov et al., 2008). The MCODE program was used to identify the most inter-connected nodes (Bader and Hogue, 2003).

Ethics Approval

The study obtained ethics approval from the Regional Committee for Medical Research Ethics in Eastern Norway (REK Øst) and the Ethiopian Science and Technology Ministry in Addis Ababa, Ethiopia. Written informed consent was obtained from the study participants before the study was conducted.

RESULTS

Comprehensive Proteome Analysis of *M. tuberculosis* Lineage 7 and Lineage 4 Strains

A total of 2867 *M. tuberculosis* proteins were identified with 99% confidence at the peptide and protein levels (Supplemental file 2) representing 71% protein coverage of the predicted *M. tuberculosis* proteome. The total number of proteins identified in *M. tuberculosis* lineage 7 (L7-35 and L7-28) and in lineage 4 were 2,825 and 2,541, respectively. Among the 2,825 identified proteins in lineage 7 strains, 2499 (87%) proteins were shared in all the biological experiments (**Figures 1A–C**). Mutual exclusivity analysis revealed 326 and 42 strain-specific protein groups in lineage 7 and lineage 4, respectively (Supplemental file 2), and the main DA component pathways are addressed below. The overlap in protein identification in the different biological replicates and *M. tuberculosis* lineages is shown in the Venn diagram (**Figures 1A–C**). Of the annotated components, 1,783 (62%) have an assigned molecular function, 1,110 (38.7%) are involved in known biological processes, 948 (33%) are assigned by cellular compartment, and 829 (28.9%) have an assigned KEGG function. The complete list of protein groups is presented in Supplemental file 3.

Protein Profile of *M. tuberculosis* Lineage 7 (L7-35 and L7-28) and Lineage 4 Strains

The abundance of the *M. tuberculosis* proteins identified was quantified by iBAQ. This technique takes into account the normalization and summation of MS/MS signals in relation to peptide size, length, and number of theoretical peptides considered acceptable for all the proteins that are defined in a specific proteome run. The 10 most abundant protein classes identified in lineage 7 strain were chaperones, hydrolase isomerase, ligase, oxidoreductase, transfer carrier protein, and nucleic acid-binding proteins. The least abundant proteins identified include the ESAT-6-like protein (EsxT), MCE-family protein (Mce2F), PE-PGRS family protein and PE family protein. The iBAQ intensity in the aggregate proteome covered a dynamic range of six orders of magnitude between the most abundant and least abundant proteins (**Figures 2A,B**; Supplemental file 3).

After identifying the 2,867 proteins in the composite *M. tuberculosis* proteome, the reproducibility of our label-free quantification workflow was assessed (Supplemental file 1). The Pearson correlations (R-values) of biological replicates using normalized protein LFQ intensities were computed. The analysis showed that the R-value between normalized intensities was high (**Figures 3A,B**, Supplemental file 4) and thus was suitable for accurate comparisons of protein abundance differences.

For DA protein comparisons, criteria were set that fulfill two valid LFQ intensity values from each biological triplicate. This resulted in a total of 1,946 proteins. Using a set of statistical criteria, T-test p-value 0.05, $S_0 = 2$ and fold change cutoff point ± 2, we found the abundances of 125 proteins to be significantly changed (**Figure 3C**, Supplemental file 5).

The Proteomes of the *M. tuberculosis* Lineage 7 Strains Are Significantly Different from Lineage 4

Among the 125 differentially regulated proteins, 64 were downregulated and 61 proteins were upregulated (**Figure 3C**, Supplemental file 5). The 125 DA protein groups were further subjected to unsupervised hierarchical cluster analysis. The resulting cluster-gram is shown in **Figure 4A**. KEGG analysis was then conducted to investigate whether these proteins were enriched for any particular pathway. The pathway sub-clusters that were significantly enriched are shown in **Figures 4B–D** (Supplementary file 6). The DA proteins were also categorized into their functional categories as defined by TubercuList. A majority of the proteins detected belong to categories of intermediary metabolism and respiration (39.2%), lipid metabolism (20%), cell wall- and cell processes-related (24.8%), conserved hypotheticals (2.4%) and unknowns (0.8%)

[1]www.uniprot.org/help/gene_ontology

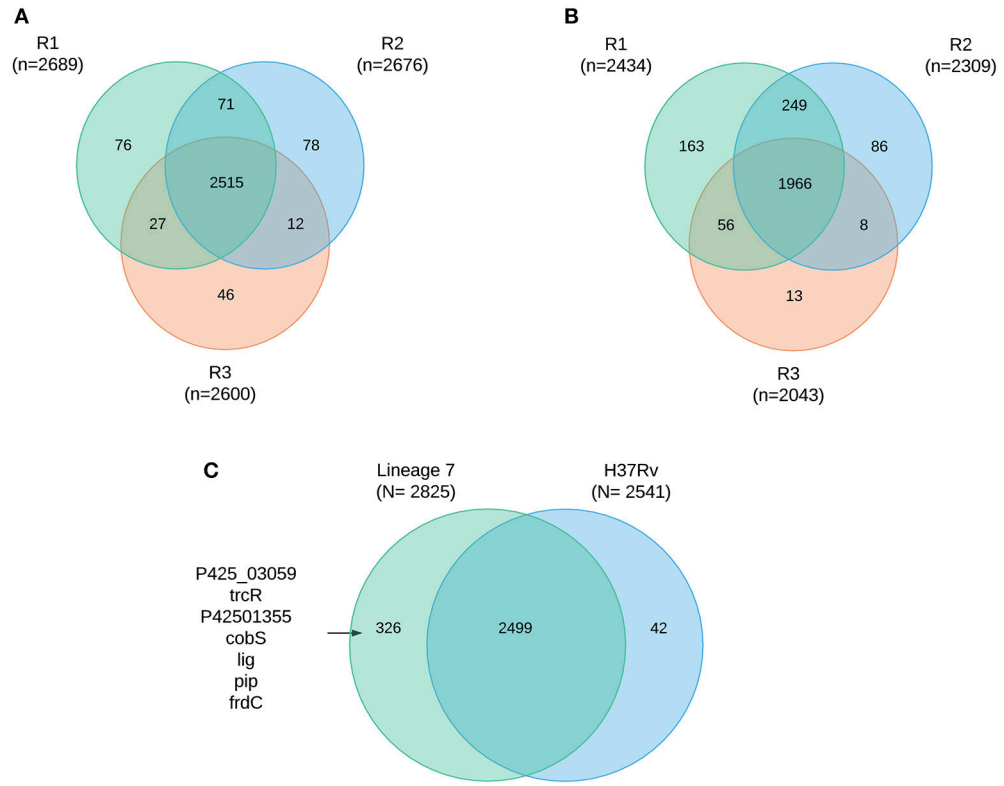


FIGURE 1 | Protein coverage Venn plot illustrating protein identification overlaps among the three biological replicates (R1–R3) in *M. tuberculosis* lineage 7 (A) and in lineage 4 **(B)** isolates, **(C)** overlap of protein identification between lineages and some of the mutually exclusively identified proteins.

(Supplemental file 7). As shown in **Figure 5**, the number of less abundant proteins involved in cell wall and cell processes is greater in lineage 7 strains than in lineage 4. In addition, the number of more abundant proteins involved in intermediary metabolism and respiration is greater in lineage 7 strains than in lineage 4.

To visualize the functional and molecular interaction network, a total of 125 DA proteins were searched into the online STRING protein query database. Eighty eight of the 125 proteins were present in recognized and predicted networks with total interaction edges of 174 interaction networks. **Figure 6** suggests that LpdA, FrdA, EccB3, EccC3, and PstB2 represent a significant protein hub. Three network sub clusters of proteins were highly correlated to functions of Pst system, Esx-3 secretion machinery, energy metabolism, and oxidative stress responses (**Figure 6**).

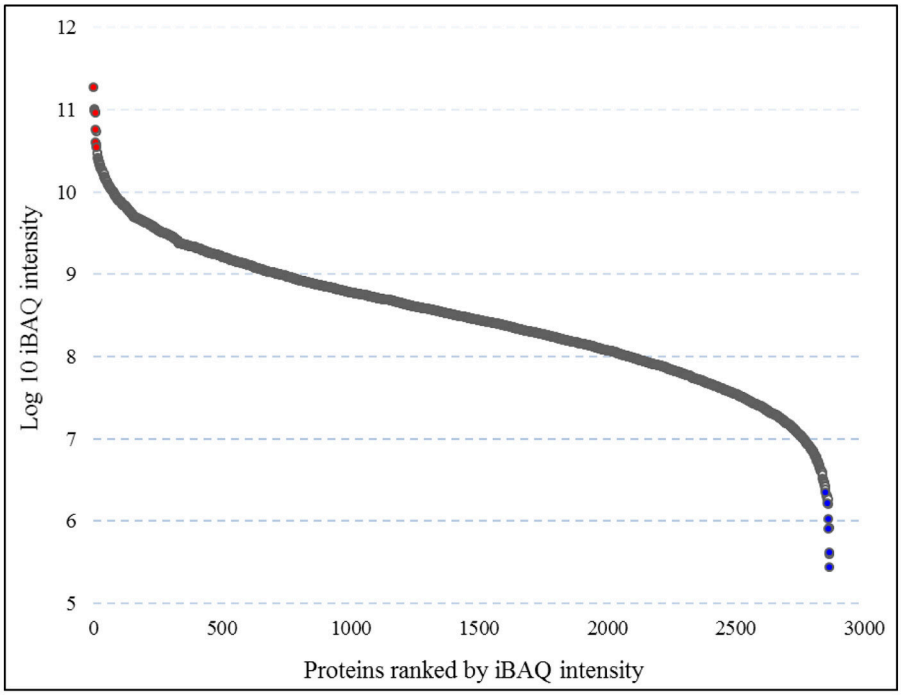
Cluster 1 in **Figure 6** shows interactions of ABC transporter proteins Rv0932c (Pst1), (PstS2), Rv0933 (PstB), and Rv0292 which are essential for the high-affinity capture of periplasmic inorganic phosphate (Pi) in *M. tuberculosis*. These proteins were significantly downregulated with a fold change ranging from 31 to 262. In addition, Rv0491 (RegX3) and Rv0758 (PhoR) part of the two-component regulatory system were upregulated (**Table 1**). The Lineage 7 strains also exhibited downregulation of Rv3881c (EspB), Rv1196 (PPE18), Rv1743 (PknE), and Rv1980c (Mpt64), which play important roles in the ESX-1 secretion system and apoptosis (**Table 2**).

Cluster 3 in **Figure 6** depicts interactions of the type 7 secretion system (Esx-3) components that are crucial for iron acquisition in *M. tuberculosis*. The ESX-3 secretion machinery proteins (Rv2083 Rv0282 (EccA3), (EccB3), Rv0284 (EccC3), (Rv0290/EccD3), and Rv0292 (EccE3) were downregulated with a fold change ranging from 19 to 55. In addition, Rv0291 (MycP3), Rv0676c (MmpL5), and Rv0036 proteins were downregulated in lineage 7 strains. Rv1884c (RpfC), a resuscitation-promoting factor protein, was upregulated in lineage 7 strains compared to lineage 4 (**Table 3**).

Several proteins, Rv2952 (AhpD), Rv0969 (CtpV), Rv0968 (MshD), Rv1161 (NarG), Rv2394 (GgtB), CysD, CysN, CysH, and SirA that are involved in counteracting the effect of reactive oxygen intermediates (ROI) and/or reactive nitrogen intermediates (RNI) were less abundant in lineage 7 (L7-35 and L7-28) compared to lineage 4. Lineage 7 strains also showed downregulation of ABC transporter proteins Rv1620c (CydC) and Rv1621c (CydD). In contrast, a number of proteins that counteract the detrimental effects of ROI and RNI were upregulated in lineage 7 (L7-35 and L7-28) isolates compared to lineage 4. These included GnD1, ZwF1, Rv3119 (MoaE1), RV3324C (MoaC), RV3323C, and (MoaX) proteins (**Table 4**).

The DA proteins CitA, BkdB BkdA, FrdA, and LpdA play important roles in the citrate cycle (TCA) pathway were differentially regulated. While CitA, BkdB, BkdA proteins

A



B

Protein IDs	Gene	Protein name	iBAQ intensity*
I6Y7T8	groL	60 kDa chaperonin	11.27
I6X3P5	acpP	Acyl carrier protein	11.01
I6X5Y3	P425_03110	DNA-binding protein	11.00
I6X969	P425_00491	Iron-regulated heparin binding hemagglutinin	11.00
I6X2I8	P425_01936	L-lactate dehydrogenase	10.99
I6Y4F5	tuf	Elongation factor	10.98
I6WY60	P425_00354	Isoniazid inducible protein	10.96
I6X626	P425_03155	Electron transfer flavoprotein alpha subunit	10.76
I6Y3F9	groS	10 kDa chaperonin	10.74
I6YCV1	P425_02299	Glutamine synthetase	10.60
I6YC53	esxT	ESAT-6 like protein EsxT	6.27
O07784	mce2F	MCE-family protein	6.27
I6Y6W4	P425_00104	Fatty-acyl-CoA thioesterase type III	6.22
I6YBX3	Rv3361c	Conserved protein	6.21
O53333	Rv3183	Possible transcriptional regulatory protein	6.02
O53500	Rv2102	Uncharacterized protein	5.92
Q79FL8	PE_PGRS30	PE-PGRS family protein PE_PGRS30	5.90
P95210	Rv0401	Probable conserved transmembrane protein	5.61
I6XZT7	P425_02174	PE family protein	5.60
O06605	Rv1585c	Possible phage PhiRv1 protein	5.44

FIGURE 2 | Protein dynamic range estimation. (A) Combined intensity based absolute quantification (iBAQ) values for the 2,825 proteins were plotted with $\log_{10}$ iBAQ intensity on the y axis, and proteins were ranked by iBAQ intensity on the x axis. The plot shows a dynamic range of 6 orders of magnitude. **(B)** List of the ten most (colored red) and 10 least (colored blue) abundant proteins based on iBAQ intensity.

were less abundant, the FrdA and Rv3303c (LpdA) proteins were upregulated in lineage 7 strains compared to lineage 4. Rv3043c (CtaD), Rv2200c (CtaC), QcrA (Rv1446c), QcrC (Rv2194), QcrB (Rv2196), which are the major respiratory route in mycobacteria, were also upregulated. In addition, the Rv1915 (AceaA) and Rv1916 (AceaB) GnD1 and PpgK proteins were upregulated in lineage 7 strains (**Table 5**).

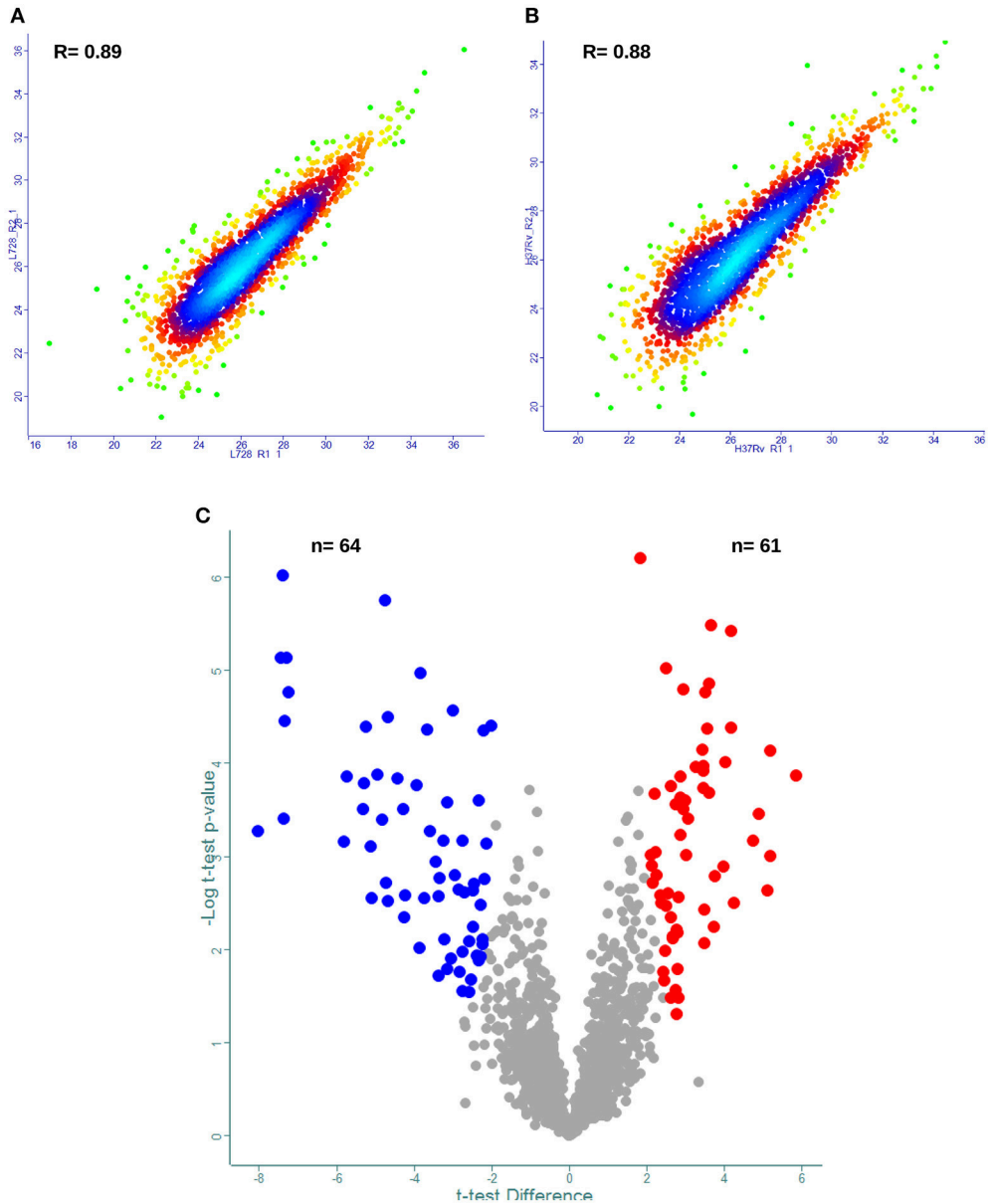


FIGURE 3 | Label free proteome quantification of *Mycobacterium tuberculosis* lineages. Quantitative analysis was performed using the MaxQuant and Perseus software environments as described in the method section. Reproducibility of the analytic workflow for biological replicates was assessed by Pearson correlation coefficients (*R*-values). **(A,B)** Represent density scatter plot of peptide intensities for biological replicates of *M. tuberculosis* lineage 7 and lineage 4, respectively. **(C)** volcano plot of protein abundance differences as a function of statistical significance (*t*-test $p \leq 0.05$ and fold change cutoff point ±2) between lineage 7 and lineage 4 isolates. Y-axis indicates *p*-value ($-\log_{10}$). X-axis shows protein ratio (x-axis) in lineage 7 vs. lineage 4 stains. The color code indicates upregulation (red) and downregulation (blue). Proteins with no statistically significant difference in abundances between the two lineages are shown in gray.

Several proteins involved in cell wall/lipid biosynthesis were less abundant in lineage 7 than in lineage 4. The RV2952 was markedly downregulated with a 164-fold change in lineage 7 strains. Furthermore, the MmaA3, IniC, Pks4, MurE, and MurF proteins were less abundant in lineage 7 strains. The PpsE PpsC, AccD6, and KasA proteins were upregulated in lineage 7 than in lineage 4 (**Table 6**).

DISCUSSION

This is the first study to generate information on the DA proteomic profile of *M. tuberculosis* lineage 7 vs. lineage 4 strains. We compared the proteomes in the two lineages and obtained 2867 protein groups that cover 71% of the predicted *M. tuberculosis* proteome ($n = 4023$). Former studies have

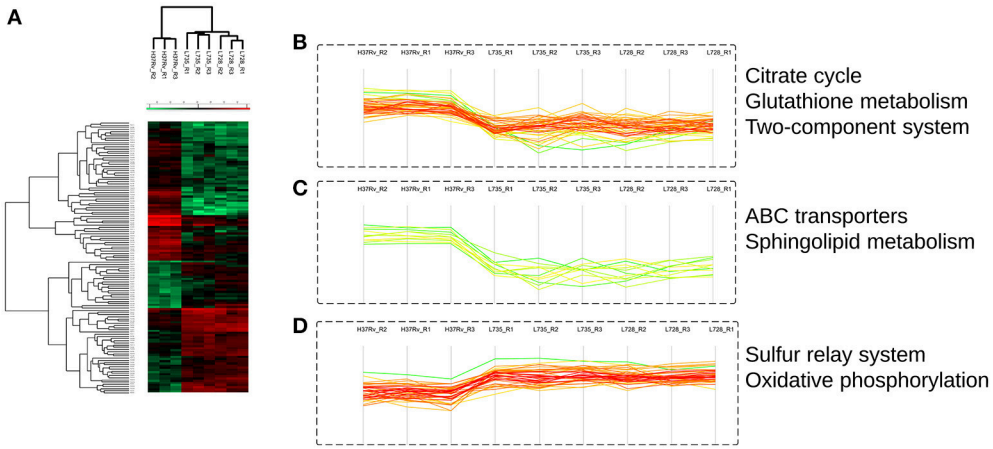


FIGURE 4 | Quantitative proteomics analysis reveals distinct pattern of *M. tuberculosis* differentially abundant proteins that are involved in several important pathways. (A) Unsupervised hierarchical clustering representing the *T*-test significant proteins ($n = 125$). Color code indicates the normalized median abundance of the proteins belonging to the category (red more abundant; green less abundant). **(B–D)** Representative clusters of significantly enriched KEGG pathways.

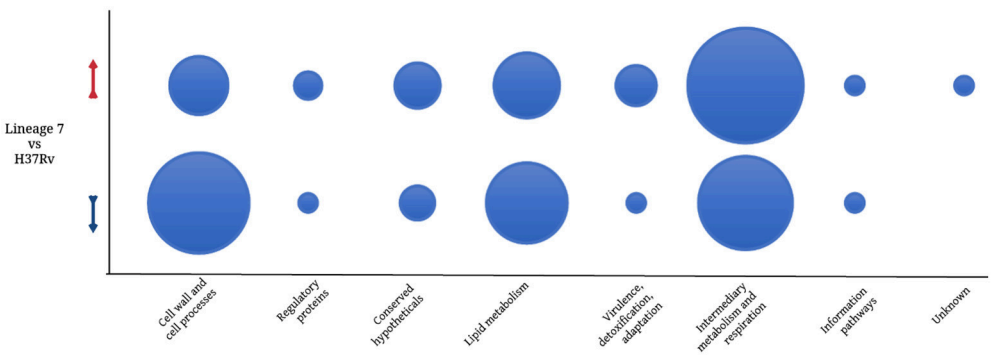


FIGURE 5 | Bubble plot comparing the number of *M. tuberculosis* differentially abundant proteins in each functional category. Red arrow denotes upregulation and the blue arrow indicates downregulation. Functional categorization was performed according to TubercuList v 2.6 (http://tuberculist.epfl.ch/).

documented even higher protein coverages with maximal ranges of 77–82% of the predicted *M. tuberculosis* proteome (Kelkar et al., 2011; Peters et al., 2016). Compared to recent publications that reported total protein coverages of 46% (Bespyatykh et al., 2016), 54% (Jhingan et al., 2016) and 62% (Schubert et al., 2015), the number of proteins identified in our study is relatively high.

This study showed considerable differences in the levels of protein abundance between *M. tuberculosis* lineage 7 and lineage 4 strains. Transporter proteins of the Pst system Rv0932 (PstS2), RV0933 (PstB), and RV0934 (PstS1) were less abundant in the two lineage 7 strains. PstA, PstC, PstS, and PstB form an ABC transporter essential for the capture of periplasmic inorganic phosphate (P_i) in *M. tuberculosis* (Rifat et al., 2009). While PstS binds P_i with high affinity, PstB provides the energy required for P_i transport from the periplasm to the cytosol. A former study demonstrated that *M. tuberculosis* strains with a disruption in genes encoding the pst system were deficient in phosphate uptake and exhibited decreased virulence and attenuated growth, and this was attributed to the absence of the PstS2 protein (Peirs et al., 2005). In our study, PstS2 in lineage 7 strains was 261-fold lower than that in lineage 4. The phosphate starvation response (PSR) is an important mechanism for *M. tuberculosis* survival under phosphate-depleted conditions (Rifat et al., 2009). In this regard, the *M. tuberculosis* genes *regX3* and *phoR* are known to encode components that are essential for bacillary survival during phosphate limitation and for regulation of PSR (Rifat et al., 2009, 2014). In the two lineage 7 strains, we observed upregulation of the RegX3 and PhoR proteins that may suggest enhanced expression of PSR. P_i is an essential component of DNA, RNA, ATP, phospholipids, and proteins, and is crucial for energy transfer, protein activation, and carbon and amino acid metabolic processes (Tischler et al., 2013). The significant downregulation of essential proteins for P_i uptake in lineage 7 (L7-35 and L7-28) strains suggests that these proteins may contribute to the relatively slow *in vitro* growth described in our former study (Yimer et al., 2015).

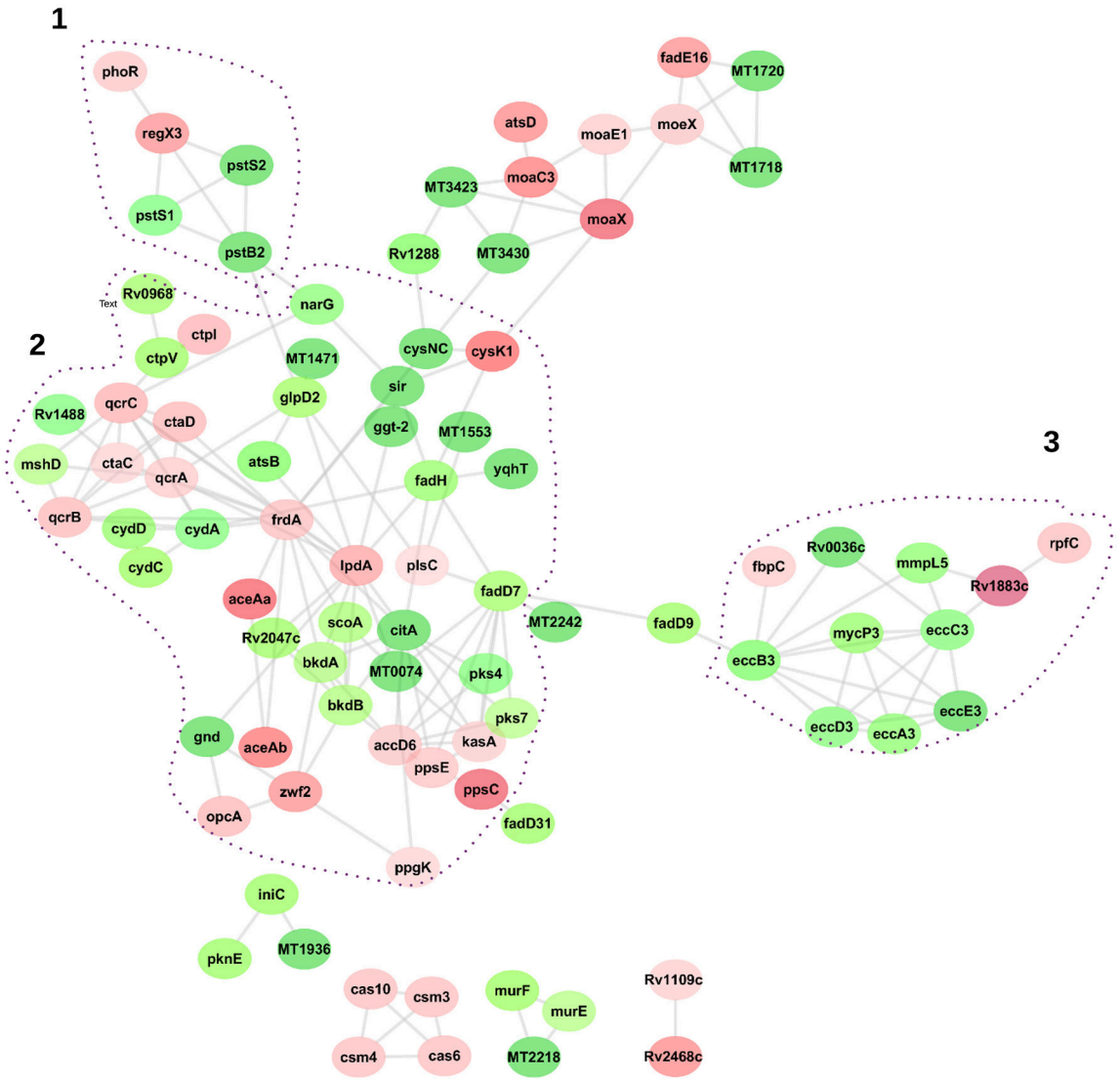


FIGURE 6 | Network interaction analysis of differentially abundant proteins related to *M. tuberculosis* lineages 7 and 4. Graphical representation of the interaction was generated by the Cytoscape software version 3.4. Proteins are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). The intensity of the node color indicates the increased (red) or decreased (green) abundance according to fold changes. Dot-lined circles indicate sub-networks of proteins linked to specific KEGG pathways.

This study showed downregulation of proteins involved in the type 7 secretion systems ESX-3 and ESX-1 in lineage 7 (L7-35 and L7-28) strains. ESX-1, ESX-3, and ESX-5 have been shown to be crucial for virulence or viability in *M. tuberculosis* (van Winden et al., 2016). The ESX-3 secretion machinery is encoded by an operon, including the genes *eccA3*, *eccB3*, *eccC3*, *eccD3*, *rv0291*, and *eccE3* (**Table 1**). All proteins encoded by these genes were downregulated in lineage 7 strains. The Esx-3 system is essential for siderophore-mediated iron acquisition, and is critical for *in vitro* growth of *M. tuberculosis* (Serafini et al., 2013). *M. tuberculosis* uses siderophore molecules (mycobactin and carboxymycobactin) to acquire iron. MmpS4/MmpL4 and MmpS5/MmpL5 proteins are essential for the biosynthesis and transport of siderophores in *M. tuberculosis* (Wells et al., 2013). In the current study, MmpL5 was downregulated in lineage 7 strains. A non-synonymous mutation in the *mmpL4* gene was identified in our recent WGS study (Yimer et al., 2016). The Rv0291 (MycP3) protein downregulated in lineage 7 strains is an essential protein for *M. tuberculosis* growth *in vitro* (Sassetti et al., 2003). The downregulation of several components of the essential type 7 secretion system ESX-3 may expose lineage 7 strains to iron starvation that potentially affects its growth *in vitro*.

A number of proteins that are involved in counteracting the effect of ROI and RNI were downregulated in the two lineage 7 strains compared to lineage 4 (**Table 3**). The protein CtpV is responsible for copper transport in *M. tuberculosis* and copper

TABLE 1 | List of the differentially abundant proteins involved in the Pst system in *M. tuberculosis* lineage 7 (L7-35 and L7-28) vs. lineage 4 (H37Rv) strains.

Protein name	Protein IDs	Gene name	Rv number	Fold change
Phosphate-binding protein PstS	I6Y569	pstS2	Rv0932c	−261.84
Phosphate import ATP-binding protein PstB	I6XWL3	pstB	Rv0933	34.46
Phosphate-binding protein PstS	I6XA55	pstS1	Rv0934	−31.02
Two-component system sensory transduction protein	I6Y7Y4	regX3	Rv0491	12.09
Two component system response sensor kinase PhoR	P71815	phoR	Rv0758	5.9

TABLE 2 | List of the *M. tuberculosis* differentially abundant PPE proteins in *M. tuberculosis* lineage 7 (L7-35 and L7-28) vs. lineage 4 (H37Rv) strains.

Protein name	Protein IDs	Gene symbol	Rv number	Fold change
PPE family protein PPE18	L7N675	PPE18	Rv1196	−34.87
Secreted ESX-1 substrate protein B EspB	I6YHD6	espB	Rv3881c	−8.29
Immunogenic protein MPT64	I6YC43	mpt64	Rv1980c	−10.35
PPE family protein PPE51	I6YBA9	PPE51	Rv3136	−10.15
Conserved protein	O06216	Rv2161c	Rv2161c	−156.57

TABLE 3 | List of the *M. tuberculosis* differentially abundant proteins and corresponding fold changes involved in type 7 secretion system.

Protein name	Protein IDs	Gene symbol	Rv number	Fold change
ESX-3 secretion system protein EccA3	I6Y3E7	eccA3	Rv0282	−19.12
ESX-3 secretion system protein EccC3	I6X8X9	eccC3	Rv0284	−25.51
ESX-3 secretion system protein EccD3	I6Y7G0	eccD3	Rv0290	−25.68
ESX-3 secretion system protein EccB3	I6XUX6	eccB3	Rv0283	−37.78
ESX-3 secretion system protein EccE3	I6Y3F4	eccE3	Rv0292	−55.98
Uncharacterized protein	I6X8C7	Rv0036c	Rv0036c	−161.42
Membrane-anchored mycosin MycP	O53695	mycp3	Rv0291	−13.40
Fatty-acid–CoA ligase FadD9	Q50631	fadH9	Rv2590	−7.14
Resuscitation-promoting factor RpfC	I6XZ79	rpfC	Rv1884c	6.82
Diacylglycerol acyltransferase	I6Y2U9	fbpC	Rv0129c	5.47
Transmembrane transport protein MmpL5	I6XVY5	mmpl5	Rv0676c	−19.5
Uncharacterized protein	O07748	Rv1883c	Rv1883c	57.35

levels increase during hypoxia (White et al., 2009; Ward et al., 2010). Excess copper is toxic and must be regulated in *M. tuberculosis*. Several proteins, including CysD, CysN, CysH, and SirA involved in cysteine biosynthesis and sulfur metabolism, were downregulated in lineage 7. Cysteine is a precursor for glutathione biosynthesis (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/). The MshD and GgtB proteins are involved in mycothiol biosynthesis (Buchmeier et al., 2006) and glutathione metabolism (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/), respectively. Lineage 7 strains showed downregulation of ABC transporter proteins Rv1620c (CydC) and Rv1621c (CydD) that are involved in cytochrome biosynthesis (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/). CydDC contributes to nitric oxide tolerance (Holyoake et al., 2016). In contrast, several proteins that counteract the lethal effects of ROI and RNI in phagocytic cells were upregulated in the two lineage 7 strains (**Table 3**). GnD1 and ZwF1 proteins are required for glutathione metabolic process (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/). Rv3119 (MoaE1), Rv3324c (MoaC), and Rv3323c (MoaX) proteins are important in sulfur relay system (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/) and redox reaction in various compounds and are crucial for molybdenum cofactor (MoCo) biosynthesis (Williams et al., 2014). Molybdenum enzymes are involved in the sulfur, carbon, and nitrogen metabolism (Williams et al., 2014). In general, the downregulation of several proteins that counteract the effect of ROI and RNI highlights the exposure of lineage 7 cells to hostile environments. Conversely, the upregulation of specific proteins points to a compensatory mechanism adapted by lineage 7 strains to escape from the detrimental effect of ROI and RNI stress exposures.

Lineage 7 (L7-35 and L7-28) strains expressed various proteins involved in metabolic pathways that generate energy for survival. A number of proteins that play an important role in the TCA pathway were differentially expressed. While CitA, BkdB, and BkdA were downregulated, FrdA and LpdA proteins were upregulated in lineage 7 strains compared to lineage 4. The TCA is an aerobic core metabolic pathway which is crucial for the final steps of the oxidation of carbohydrates and fatty acids (Wikipedia[2]). The TCA pathway supplies NADH for use in oxidative phosphorylation and other metabolic processes. One study showed that in dormant cultivatable *M. tuberculosis*, the TCA was suppressed (Converse et al., 2009). When *M. tuberculosis* experiences hypoxia due to oxygen depletion, increased levels of NADH and inability to close the respiratory cycle, the bacteria shifts the direction of TCA enzymes from an oxidative direction to a reductive direction and starts to actively secrete succinic acid to complete the respiratory cycle. This process is facilitated by an enzyme called fumarate reductase (FrdA). We found an upregulation of FrdA that may signal a shift in metabolic pathway by lineage 7 strains. A former study showed that the FrdA protein was upregulated under hypoxic conditions in *M. tuberculosis* (Watanabe et al., 2011). Lineage 7 strains exhibited upregulation of LpdA protein; LpdA has quinone reductase activity and catalyzes the formation of NADH, which is important for energy production under anaerobic conditions

[2]https://en.wikipedia.org/wiki/Citric_acid_cycle

TABLE 4 | List of the differentially abundant proteins involved in ROI and RNI stress exposure responses in *M. tuberculosis* lineage 7 (L7-35 and L7-28) vs. lineage 4 (H37Rv) strains.

Protein name	Protein IDs	Gene symbol	Rv number	Fold change
Uncharacterized protein	O07748	Rv1883c	Rv1883c	57.35
MoaD-MoaE fusion protein MoaX	Q6MWY3	moaX	Rv3323c	36.35
Cysteine synthase	I6Y910	cysK1	Rv2334	18.13
Isocitrate lyase AceAb	O07717	aceAb	Rv1916	16.33
Cyclic pyranopterin monophosphate synthase accessory protein	I6YBT7	moaC3	Rv3324c	15.71
Glucose-6-phosphate 1-dehydrogenase	I6XBH9	zwf2	Rv1447c	12.15
Arylsulfatase AtsD	I6XVW9	atsD	Rv0663	13.45
Glucose-6-phosphate 1-dehydrogenase	I6XBH9	zwf2	Rv1447c	12.15
Fumarate reductase FrdA	I6YAW6	frdA	Rv1552	6.83
Molybdenum cofactor biosynthesis protein E1 MoaE1	I6X6B1	moaE1	Rv3119	5.36
Probable component linked with the assembly of cytochrome transport transmembrane ATP-binding protein ABC transporter CydC	O06137	cydC	Rv1620c	−6.00
Respiratory nitrate reductase alpha chain NarG	I6Y9T4	narG	Rv1161	−21.51
Arylsulfatase AtsB	O65931	atsB	Rv3299c	−39.47
Conserved protein	O06216	Rv2161c	Rv2161c	−156.57
Methyltransferase	I6XFR4	Rv2952	Rv2952	−164.28

TABLE 5 | List of the differentially abundant proteins involved in energy metabolism in *M. tuberculosis* lineage 7 (L7-35 and L7-28) vs. lineage 4 (H37Rv) strains.

Protein name	Protein IDs	Gene symbol	Rv number	Fold change
6-phosphogluconate dehydrogenase, decarboxylating	Q79FJ2	gnd1	Rv1844c	17.85
Isocitrate lyase AceAa	O07718	aceAa	Rv1915	29.61
NAD(P)H quinone reductase LpdA	I6XGU5	lpdA	Rv3303C	9.52
Isocitrate lyase AceAb	O07717	aceAb	Rv1916	16.33
Arylsulfatase AtsD	I6XVW9	atsD	Rv0663	13.45
Ubiquinol-cytochrome C reductase QcrC	I6Y059	qcrC	Rv2194	8.09
Cytochrome C oxidase polypeptide I CtaD	I6YAZ7	ctaD	Rv3043c	7.61
Ubiquinol-cytochrome C reductase QcrB	I6YCT0	qcrB	Rv2196	7.25
Fumarate reductase FrdA	I6YAW6	frdA	Rv1552	6.83
Rieske iron-sulfur protein QcrA	I6XDR2	qcrA	Rv2195	5.62
Transmembrane cytochrome C oxidase subunit II CtaC	I6XDR7	ctaC	Rv2200c	4.57
Polyphosphate glucokinase PpgK	I6YE62	ppgK	Rv2702	4.42
Gamma-glutamyl transpeptidase GgtB	P71750	ggtB	Rv2394	−4.76
Branched-chain keto acid dehydrogenase E1 component alpha subunit BkdA	I6YDK3	bkdA	Rv2497c	−4.93
Glycerol-3-phosphate dehydrogenase	I6Y352	glpD2	Rv3302c	−5.84
Branched-chain keto acid dehydrogenase E1 component beta subunit BkdB	I6XEG1	bkdB	Rv2496c	−5.02
Citrate synthase II CitA	I6Y908	citA	Rv0889c	−150.96

(Zheng et al., 2008). As shown in **Figure 6**, FrdA and LpdA exhibit interactions with a number of proteins indicating the central role of these enzymes in energy metabolism in lineage 7 strains.

We also found upregulation of the CtaD, CtaC, QcrA, QcrC, and QcrB proteins that are components of the bc1-aa3 pathway. The bc1-aa3 cytochrome pathway is essential for growth *in vitro* and is the major respiratory route in mycobacteria (Matsoso et al., 2005). These proteins are also upregulated under hypoxia and poor-energy environments (Cook et al., 2014). In addition, the AceaA and AceaB enzymes involved in glyoxylate cycle that serves as an alternative to the TCA cycle, were upregulated in lineage 7. The GnD1 protein upregulated in lineage 7 strains functions in the pentose phosphate pathway as the main generator of cellular NADPH (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/). In addition, the PpgK protein involved in glucose phosphorylation (Kyoto Encyclopedia of Genes and Genomes; http://www.genome.jp/kegg/) was upregulated in lineage 7 strains.

M. tuberculosis is adapted to inhabit a wide range of intracellular and extracellular environments. A typical feature

TABLE 6 | List of the differentially abundant proteins involved in cell wall/lipid biosynthesis in *M. tuberculosis* lineage 7 (L7-35 and L7-28) vs. lineage 4 (H37Rv) strains.

Protein name	Protein IDs	Gene symbol	Rv number	Fold change
UDP-N-acetylmuramoyl-tripeptide–D-alanyl-D-alanine ligase	I6YCL0	murF	Rv2157c	−6.01
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate–2,6-diaminopimelate ligase	I6X3E1	murE	Rv2158c	−4.64
Isoniazid inducible protein IniC	I6XV19	iniC	Rv0343	−7.74
Acyl-CoA ligase FadD31	I6Y7V6	fadD31	Rv1925	−7.97
Fatty-acid-CoA ligase FadD9	Q50631	fadD9	Rv2590	−7.15
Polyketide synthase Pks7	P94996	pks7	Rv1661	−4.42
Polyketide beta-ketoacyl synthase Pks4	I6Y9V4	pks4	Rv1181	−26.47
Conserved protein	O06216	Rv2161c	Rv2161c	−156.57
Methyltransferase	I6XFR4	Rv2952	Rv2952	−164.28
Methoxy mycolic acid synthase 3 MmaA3	I6XVV3	mmaA3	Rv0643c	−10.88
Phenolpthiocerol synthesis type-I polyketide synthase PpsC	I6X5S4	ppsC	Rv2933	36.43
Phenolpthiocerol synthesis type-I polyketide synthase PpsE	I6Y228	ppsE	Rv2935	6.30
Propionyl-CoA carboxylase beta chain 6 AccD6	I6XDV6	accD6	Rv2247	6.08
3-oxoacyl-[acyl-carrier-protein] synthase 1 KasA	I6Y8T4	kasA	Rv2245	5.18

of this adaptation is the ability to respire and regenerate ATP via various energy-generating metabolic pathways (Cook et al., 2014). We observed downregulation of essential enzymes involved in core metabolic pathway of the TCA cycle. Therefore, the upregulation of the proteins specified above, which are involved in aerobic and anaerobic respiration in lineage 7 strains, may be considered as a compensatory mechanism for generating energy to maintain basic physiological functions *M. tuberculosis* lineage 7 cells.

Several proteins involved in cell wall/lipid biosynthesis were differentially expressed between the two lineage 7 strains and lineage 4. The Rv2952 protein, which is required for the biosynthesis of phenolglycolipid (PGL) and production of dimycocerosates of phthiocerol (DIM), was markedly downregulated in lineage 7 cells. Rv2952 is a methyltransferase that catalyzes the transfer of a methyl group for the production of DIM and PGL (Pérez et al., 2004). A study indicated that a Pks12 mutant strain was deficient in the synthesis of DIM, and that that the growth and virulence of this strain were reduced (Sirakova et al., 2003). Our recent study exhibited non-synonymous mutations in the *pkS12* gene (Yimer et al., 2016), suggesting a possible effect of this gene on the efficiency of PDIM synthesis. Similarly, the IniC protein downregulated in lineage 7 strains is suggested to participate in the regulation of cell wall growth (Alland et al., 2000). The downregulated MurE and MurF proteins are key enzymes of peptidoglycan biosynthetic pathway (Munshi et al., 2013). AccD6, KasA, PpsE, PpsC proteins were upregulated in lineage 7 strains compared to lineage 4. AccD6 and KasA proteins are key components in the mycolate biosynthesis (Bhatt et al., 2007; Pawelczyk et al., 2011). Mycolic acids are considered major virulence factors. PpsE and PpsC proteins are involved in phenolpthiocerol and phthiocerol dimycocerosate and important for virulence in *M. tuberculosis* (Bisson et al., 2012). The integrity of cell wall envelope is crucial for protection of *M. tuberculosis* against environmental stress factors. Therefore, the downregulation of a number of proteins involved in cell wall/lipid biosynthesis may have an effect on the growth and survival of *M. tuberculosis* lineage 7 cells.

This study addresses in detail the DA proteomic profile of *M. tuberculosis* lineage 7 (L7-35 and L7-28) vs. lineage 4 (H37Rv) strains. The analysis provides new insight into the lineage-specific protein profile variations that may explain the particular character of lineage 7 cells in terms of growth/survival and pathogenesis as compared to lineage 4. A number of proteins involved in *M. tuberculosis* growth and virulence fitness are less abundant in the two lineage 7 isolates in contrast to lineage 4. This may suggest that the *in vitro* slow-growth of *M. tuberculosis* lineage 7 bacilli and delayed health seeking among patients infected with lineage 7 strains observed in our earlier study may be phonotypic characteristics of lineage 7 cells. However, the most preferred method of linking phenotypic characteristics with a particular protein is to disrupt the gene encoding that protein and assess phenotypic alterations that the unavailability of that protein vests on the phenotype *in vitro* and in animal models. Therefore, further study by knocking out genes encoding the pst system and ESX-3 secretion system proteins, and observing phenotypic changes that the absence the corresponding proteins confers on the phenotype of lineage 7 vs. other lineages in *in vitro* and *in vivo* models, are warranted. In addition, further work examining the immune response of patients infected with lineage 7 vs. other lineages is imperative.

AUTHOR CONTRIBUTIONS

TT and SY conceived the study and study design. SY collected the lineage 7 isolates. SY and EZ performed specimen handling and cultivation. AB, TR, and SK performed the MS analysis. SK, SY, and AB performed the bioinformatics analysis. SY, SK, AB, and TT evaluated and interpreted the data and drafted the paper. All authors edited and approved the final manuscript.

FUNDING

Funding was received from the Research Council of Norway (RCN) FRIMEDBIO project 204747 and RCN GLOBVAC projects 234506 to TT and 192468 to CH, and Norwegian South-Eastern Health Authority project 2013080 to SY and TT.

ACKNOWLEDGMENTS

The authors thank the patients for consenting to participate in the study. We also thank the selected health care facilities in the Amhara Region, Ethiopia, for facilitating the study. We are grateful to the Armauer Hansen Research Institute (AHRI), Addis Ababa, Ethiopia, and the Norwegian Institute of Public Health for facilitating the transfer of lineage 7 strains for WGS at Oslo University Hospital.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00795/full#supplementary-material

REFERENCES

Albanna, A. S., Reed, M. B., Kotar, K. V., Fallow, A., McIntosh, F. A., Behr, M. A., et al. (2011). Reduced transmissibility of East African Indian strains of Mtb. *PLoS ONE* 6:*e*25075. doi: 10.1371/journal.pone.0025075

Alland, D., Steyn, A. J., Weisbrod, T., Aldrich, K., and Jacobs, W. R. (2000). Characterization of the *Mycobacterium tuberculosis* iniBAC promoter, a promoter that responds to cell wall biosynthesis inhibition. *J. Bacteriol.* 182, 1802–1811. doi: 10.1128/JB.182.7.1802-1811.2000

Assenov, Y., Ramírez, F., Schelhorn, S. E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics* 24, 282–284. doi: 10.1093/bioinformatics/btm554

Azuaje, F., Devaux, Y., and Wagner, D. R. (2010). Coordinated modular functionality and prognostic potential of a heart failure biomarker-driven interaction network. *BMC Syst. Biol.* 4:60. doi: 10.1186/1752-0509-4-60.

Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* 4:2. doi: 10.1186/1471-2105-4-2

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125, 279–284. doi: 10.1016/S0166-4328(01)00297-2

Bespyatykh, J., Shitikov, E., Butenko, I., Altukhov, I., Alexeev, D., and Mokrousov, I. (2016). Proteome analysis of the *Mycobacterium tuberculosis* Beijing B0/W148 cluster. *Sci. Rep.* 6:28985. doi: 10.1038/srep28985

Bhatt, A., Molle, V., Besra, G. S., Jacobs, W. R. Jr., and Kremer, L. (2007). The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Mol. Microbiol.* 64, 1442–1454. doi: 10.1111/j.1365-2958.2007.05761.x

Bisson, G. P., Mehaffy, C., Broeckling, C., Prenni, J., Rifat, D., and Lun, D. (2012). Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, *rpoB* mutant *Mycobacterium tuberculosis*. *J. Bacteriol.* 194, 6441–6452. doi: 10.1128/JB.01013-12

Blouin, Y., Hauck, Y., Soler, C., Fabre, M., Vong, R., and Dehan, C. (2012). Significance of the identification in the Horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. *PLoS ONE* 7:12. doi: 10.1371/journal.pone.0052841

Buchmeier, N. A., Newton, G. L., and Fahey, R. C. (2006). A mycothiol synthase mutant of *Mycobacterium tuberculosis* has an altered thiol-disulfide content and limited tolerance to stress. *J. Bacteriol.* 188, 6245–6252. doi: 10.1128/JB.00393-06

Chacon-Salinas, R., Serafín-López, J., Ramos-Payán, R., Méndez-Aragón, P., Hernández-Pando, R., Van Soolingen, D., et al. (2005). Differential pattern of cytokine expression by macrophages infected *in vitro* with different *Mycobacterium tuberculosis* genotypes. *Clin. Exp. Immunol.* 140, 443–449. doi: 10.1111/j.1365-2249.2005.02797.x

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* 45, 1176–1182. doi: 10.1038/ng.2744

Converse, P. J., Karakousis, P. C., Klinkenberg, L. G., Kesavan, A. K., Ly, L. H., and Allen, S. S. (2009). Role of the dosR-dosS two-component regulatory system in *Mycobacterium tuberculosis* virulence in three animal models. *Infect. Immun.* 77, 1230–1237. doi: 10.1128/IAI.01117-08

Cook, G. M., Hards, K., Vilchèze, C., Hartman, T., and Berney, M. (2014). Energetics of respiration and oxidative phosphorylation in mycobacteria. *Microbiol. Spectr.* 2:3. doi: 10.1128/microbiolspec.MGM2-0015-2013

Coscolla, M., and Gagneux, S. (2014). Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin. Immunol.* 26, 441–444. doi: 10.1016/j.smim.2014.09.012

Cox, J., and Matthias, M. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi: 10.1038/nbt.1511

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805. doi: 10.1021/pr101065j

Firdessa, R., Berg, S., Hailu, E., Schelling, E., Gumi, B., Erenso, G., et al. (2013). Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerging Infect. Dis.* 19, 460–463. doi: 10.3201/eid1903.120256

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., and Roth, A. (2013). STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094

Holyoake, L. V., Hunt, S., Sanguinetti, G., Cook, G. M., Howard, M. J., Rowe, M. L., et al. (2016). CydDC-mediated reductant export in Escherichia coli controls the transcriptional wiring of energy metabolism and combats nitrosative stress. *Biochem. J.* 473, 693–701. doi: 10.1042/BJ20150536

Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., et al. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions. *J. Cell Biol.* 17, 739–754. doi: 10.1083/jcb.200911091

Ideker, T., and Sharan, R. (2008). Protein networks in disease. *Genome Res.* 18, 644–652. doi: 10.1101/gr.071852.107

Jhingan, G. D., Kumari, S., Jamwal, S. V., Kalam, H., Arora, D., Jain, N., et al. (2016). Comparative proteomic analyses of avirulent, virulent, and clinical strains of *Mycobacterium tuberculosis* identify strain-specific patterns. *J. Biol. Chem.* 291, 14257–14273. doi: 10.1074/jbc.M115.666123

Kelkar, D. S., Kumar, D., Kumar, P., Balakrishnan, L., Muthusamy, B., Yadav, A. K., et al. (2011). Proteogenomic analysis of *Mycobacterium tuberculosis* by high resolution mass spectrometry. *Mol. Cell. Proteomics.* 10:M111.011627. doi: 10.1074/mcp.M111.011627

Matsoso, L. G., Kana, B. D., Crellin, P. K., Lea-Smith, D. J., Pelosi, A., Powell, D., et al. (2005). Function of the cytochrome bc_1-aa_3 branch of the respiratory network in mycobacteria and network adaptation occurring in response to its disruption. *J. Bacteriol.* 187, 6300–6308. doi: 10.1128/JB.187.18.6300-6308.2005

Munshi, T., Gupta, A., Evangelopoulos, D., Guzman, J. D., Gibbons, S., Nicholas, H., et al. (2013). Characterization of ATP-Dependent Mur Ligases involved in the biogenesis of cell wall peptidoglycan in *Mycobacterium tuberculosis*. *PLoS ONE* 8:e60143. doi: 10.1371/journal.pone.0060143

Pawelczyk, J., Brzostek, A., Kremer, L., Dziadek, B., Rumijowska-Galewicz, A., Fiolka, M., et al. (2011). AccD6, a key carboxyltransferase essential for mycolic acid synthesis in *Mycobacterium tuberculosis*, is dispensable in a nonpathogenic strain. *J. Bacteriol.* 193, 6960–6972. doi: 10.1128/JB.05638-11

Peirs,. P., Lefèvre, P., Boarbi, S., Wang X. M., Denis, O., Braibant, M., et al (2005). *Mycobacterium tuberculosis* with disruption in genes encoding the phosphate binding proteins PstS1 and PstS2 is deficient in phosphate uptake and demonstrates reduced *in vivo* virulence. *Infect. Immun.* 73, 1898–1902. doi: 10.1128/IAI.73.3.1898-1902.2005

Pérez, E., Constant, P., Laval, F., Lemassu, A., Lanéelle, M. A., Daffé, M., et al. (2004). Molecular dissection of the role of two methyltransferases in the biosynthesis of phenolglycolipids and phthiocerol dimycoserosate in the *Mycobacterium tuberculosis* complex. *J. Biol. Chem.* 279, 42584–42592. doi: 10.1074/jbc.M406134200

Peters, J. S., Calder, B., Gonnelli, G., Degroeve, S., Rajaonarifara, E., Mulder, N., et al. (2016). Identification of quantitative proteomic differences between *Mycobacterium tuberculosis* lineages with altered virulence. *Front. Microbiol.* 7:813. doi: 10.3389/fmicb.2016.00813

Portevin, D., Gagneux, S., Comas, I., and Young, D. (2011). Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog.* 7:e1001307. doi: 10.1371/journal.ppat.1001307

Rieder, H. L. (1999). *Epidemiologic Basis of Tuberculosis Control*. Paris: International Union Against Tuberculosis and Lung Disease.

Rifat, D., Belchis, D. A., and Karakousis, P. C. (2014). *senX3*-independent contribution of *regX3* to *Mycobacterium tuberculosis* virulence. *BMC Microbiol.* 14:265. doi: 10.1186/s12866-014-0265-8

Rifat, D., Bishai, W. R., and Karakousis, P. C. (2009). Phosphate depletion: a novel trigger for *Mycobacterium tuberculosis* persistence. *J. Infect. Dis.* 200, 1126–1135. doi: 10.1086/605700

Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high-density mutagenesis. *Mol. Microbiol.* 48, 77–84. doi: 10.1046/j.1365-2958.2003.03425.x

Schubert, O. T., Ludwig, C., Kogadeeva, M., Zimmermann, M., Rosenberger, G., Gengenbacher, M., et al. (2015). Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*. *Cell Host Microbe* 18, 96–108. doi: 10.1016/j.chom.2015.06.001

Serafini, A., Pisu, D., Palù, G., Rodriguez, G. M., and Manganelli, R. (2013). The ESX-3 secretion system is necessary for iron and zinc homeostasis in *Mycobacterium tuberculosis*. *PLoS ONE* 8:e78351. doi: 10.1371/journal.pone.0078351

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860. doi: 10.1038/nprot.2006.468

Sirakova, T. D., Dubey, V. S., Kim, H. J., Cynamon, M. H., and Kolattukudy, P. E. (2003). The largest open reading frame *(pks12)* in the *Mycobacterium tuberculosis* genome is involved in pathogenesis and dimycocerosyl phthiocerol synthesis. *Infect. Immun.* 71, 3794–3801. doi: 10.1128/IAI.71.7.3794-3801.2003

Tischler, A. D., Leistikow, R. L., Kirksey, M. A., Voskuil, M. I., and McKinney, J. D. (2013). *Mycobacterium tuberculosis* requires phosphate-responsive gene regulation to resist host immunity. *Infect. Immun.* 81, 317–328. doi: 10.1128/IAI.01136-12

Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.* 98, 5116–5121. doi: 10.1073/pnas.091062498

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., et al. (2016). The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nat. Methods* 13, 731–740. doi: 10.1038/nmeth.3901

van Winden, V. J., Ummels, R., Piersma, S. R., Jiménez, C. R., Korotkov, K. V., Bitter, W., et al. (2016). Mycosins are required for the stabilization of the ESX-1 and ESX-5 Type VII secretion membrane complexes. *MBio* 7:e01471-16. doi: 10.1128/mbio.01471-16

Ward, S. K., Abomoelak, B., Hoye, E. A., Steinberg, H., and Talaat, A. M. (2010). CtpV: a putative copper exporter required for full virulence of *Mycobacterium tuberculosis*. *Mol. Microbiol.* 77, 1096–1110. doi: 10.1111/j.1365-2958.2010.07273.x

Watanabe, S., Zimmermann, M., Goodwin, M. B., Sauer, U., Barry, C. F., and Boshoff, H. I. (2011). Fumarate Reductase Activity Maintains an Energized Membrane in Anaerobic *Mycobacterium tuberculosis*. *PLoS Pathog.* 7:e1002287. doi: 10.1371/journal.ppat.1002287

Wells, R. M., Jones, C. M., Xi, Z., Speer, A., Danilchanka, O., Doornbos, K. S., et al. (2013). Discovery of a siderophore export system essential for virulence of *Mycobacterium tuberculosis*. *PLoS Pathog.* 9:e1003120. doi: 10.1371/journal.ppat.1003120

White, C., Lee, J., Kambe, T., Fritsche, K., and Petris, M. J. (2009). A role for the ATP7A copper-transporting ATPase in macrophage bactericidal activity. *J. Biol. Chem.* 284, 33949–33956. doi: 10.1074/jbc.M109.070201

WHO (2016). World Health Organization and Geneva Switzerland. Available online at: http://www.who.int/tb/publications/global_report/en/www.who.int/tb/publications/global_report/en/ (Accessed Dec 14, 2016).

Williams, M., Mizrahi, V., and Kana, B. D. (2014). Molybdenum cofactor: a key component of *Mycobacterium tuberculosis* pathogenesis? *Crit. Rev. Microbiol.* 40, 18–29. doi: 10.3109/1040841X.2012.749211

Yimer, S. A., Hailu, E., Derese, Y., Bjune, G. A., and Holm-Hansen, C. (2013). Spoligotyping of *Mycobacterium tuberculosis* isolates among pulmonary tuberculosis patients in Amhara Region, Ethiopia. *APMIS.* 121, 878–885. doi: 10.1111/apm.12046

Yimer, S. A., Namouchi, A., Zegeye, E. D., Holm-Hansen, C., Norheim, G., Abebe, M., et al. (2016). Deciphering the recent phylogenetic expansion of the originally deeply rooted *Mycobacterium tuberculosis* lineage 7. *BMC Evol Biol.* 16:146. doi: 10.1186/s12862-016-0715-z

Yimer, S. A., Norheim, G., Namouchi, A., Zegeye, E. D., Kinander, W., Tønjum, T., et al. (2015). *Mycobacterium tuberculosis* lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. *Clin. Microbiol.* 53, 1301–1309. doi: 10.1128/JCM.03566-14

Zhao, Y., and Lin, Y. H. (2010). Whole-cell protein identification using the concept of unique peptides. *Genomics Proteomics Bioinform.* 8, 33–41. doi: 10.1016/S1672-0229(10)60004-6

Zheng, H., Lu, L., Wang, B., Pu, S., Zhang, X., Zhu, G., et al. (2008). Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PLoS ONE* 11:e2375. doi: 10.1371/journal.pone.0002375

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

III

SCIENTIFIC REPORTS

OPEN

Ample glycosylation in membrane and cell envelope proteins may explain the phenotypic diversity and virulence in the *Mycobacterium tuberculosis* complex

Received: 11 May 2018
Accepted: 24 January 2019
Published online: 27 February 2019

Alemayehu Godana Birhanu[1,2], Solomon Abebe Yimer[1,5], Shewit Kalayou[3], Tahira Riaz[1], Ephrem Debebe Zegeye[4], Carol Holm-Hansen[6], Gunnstein Norheim[5,6], Abraham Aseffa[7], Markos Abebe[7] & Tone Tønjum[1,3]

Multiple regulatory mechanisms including post-translational modifications (PTMs) confer complexity to the simpler genomes and proteomes of *Mycobacterium tuberculosis* (Mtb). PTMs such as glycosylation play a significant role in Mtb adaptive processes. The glycoproteomic patterns of clinical isolates of the *Mycobacterium tuberculosis* complex (MTBC) representing the lineages 3, 4, 5 and 7 were characterized by mass spectrometry. A total of 2944 glycosylation events were discovered in 1325 proteins. This data set represents the highest number of glycosylated proteins identified in Mtb to date. O-glycosylation constituted 83% of the events identified, while 17% of the sites were N-glycosylated. This is the first report on N-linked protein glycosylation in Mtb and in Gram-positive bacteria. Collectively, the bulk of Mtb glycoproteins are involved in cell envelope biosynthesis, fatty acid and lipid metabolism, two-component systems, and pathogen-host interaction that are either surface exposed or located in the cell wall. Quantitative glycoproteomic analysis revealed that 101 sites on 67 proteins involved in Mtb fitness and survival were differentially glycosylated between the four lineages, among which 64% were cell envelope and membrane proteins. The differential glycosylation pattern may contribute to phenotypic variabilities across Mtb lineages. The study identified several clinically important membrane-associated glycolipoproteins that are relevant for diagnostics as well as for drug and vaccine discovery.

Tuberculosis (TB) is a major threat to public health, causing more than three deaths per minute globally. The causative agent is *Mycobacterium tuberculosis* (Mtb), and the TB crisis is exacerbated by the emergence of multi-drug-resistant (MDR) and extensively drug-resistant (XDR) Mtb strains. This situation highlights the urgent need for a comprehensive understanding of virulence and pathogenicity determinants of the *M. tuberculosis* complex (MTBC) to pave the way for the development of alternative TB control and prevention. Post-translational modifications (PTMs) including protein glycosylation are crucial in this regard. The unique Mtb glycoconjugates in the cell envelope are the predominant basis for host–pathogen interactions, antigenicity, and virulence determination and constitute one of the major components causing antimicrobial resistance

[1]Department of Microbiology, University of Oslo, PO Box 4950, Nydalen, NO-0424, Oslo, Norway. [2]Addis Ababa University, Institute of Biotechnology, PO Box 1176, Addis Ababa, Ethiopia. [3]International Center of Insect Physiology and Ecology (ICIPE), P.O. Box 30772-00100, Nairobi, Kenya. [4]Centre for Applied Biotechnology, Uni Research Environment, Bergen, Norway. [5]Coalition for Epidemic Preparedness Innovations (CEPI), P.O. BOX 123, Torshov, 0412, Oslo, Norway. [6]Infection Control and Environmental Health, Norwegian Institute of Public Health, PO Box 4404, Nydalen, NO-0403, Oslo, Norway. [7]Armauer Hansen Research Institute, Jimma Road, PO Box 1005, Addis Ababa, Ethiopia. Solomon Abebe Yimer and Shewit Kalayou contributed equally. Correspondence and requests for materials should be addressed to A.G.B. (email: alexbiology97@yahoo.com) or T.T. (email: tone.tonjum@medisin.uio.no)

(AMR) in Mtb[1–6]. Glycosylation in Mtb has mainly been detected in surface-exposed proteins and in some other membrane proteins[7]. The unique structure, antigenicity and essentiality of Mtb cell envelope glycoconjugates for mycobacterial growth provide opportunities for the development of novel drugs, vaccines, diagnostics and biomarkers[1].

Mtb glycoproteins play a critical role in a number of biological activities including cell adhesion and invasion, protein stability, localization, and maintenance of protein conformation[8–12]. Other functions influenced by glycosylation are cellular signaling, AMR, immunomodulation, intracellular bacterial survival, biofilm formation, protein complex formation, antigenicity, pathogenicity and virulence[8–12]. Recently, it has been shown that protein glycosylation was associated with low cell envelope permeability and AMR in the multi-resistant *Mycobacterium abscessus*[13]. Furthermore, glycosylation protects proteolytically-sensitive cleavage sites, thereby maintaining the membrane-association of the protein by its lipid anchor, and may also be linked to protein export[10,14].

The Mtb cell envelope is composed of an inner plasma membrane, a cell wall core with an outer mycomembrane, and an outermost layer, known as the capsule composed of polysaccharides, lipids and proteins[1]. The cell wall core is composed of peptidoglycan (PG) covalently linked via phosphoryl-N-acetylglucosaminosylrhamnosyl to arabinogalactan (AG), which in turn is esterified to α-alkyl, β-hydroxy long-chain mycolic acids, forming the mycolyl arabinogalactan-peptidoglycan (mAGP) complex[1,15]. This complex is essential for bacterial viability and is the basis of susceptibility and resistance to many anti-TB drugs including ethambutol (EMB) and ethionamide (ETH)[12,16]. Mannose-capped lipoarabinomannan (LAM), one of the key Mtb virulence factors, is a surface-exposed lipoglycan anchored to the inner and outer membranes via a mannosyl phosphate inositol[17]. The soluble components of the cell envelope include free lipids, proteins, LAM, and phosphatidylinositol mannosides (PIMs), which are signaling effector molecules in bacterial pathogenesis and disease processes[15]. The nature and amounts of the mycomembrane and capsular material vary among Mtb isolates and is likely to impact significantly on the pathogen phenotype and outcome of the pathogen-host interaction[9,18].

PG is a polymer of alternating N-acylated muramic acid (MurNac) and N-acetylglucosamine (GlcNac) residues linked in a β (1 → 4) configuration with cross-linked peptides of varying composition attached to the muramyl moieties[19,20]. PG glycosyltransferases use the lipid-linked donor precursor for the synthesis of oligo-β-(1 → 4)-[GlcNAc-β-(1 → 4)-MurNAc(peptide)] glycan strands[21]. In contrast to most other bacteria, muramic acid moieties are N-glycolylated (oxidized) in mycobacteria[15]. In Mtb, MurNGly, MurNAc and Mur residues are present in the precursor pool and in the PG[20].

The membrane-bound glycosyltransferases or oligosaccharyltransferases (OSTs) catalyze the transfer of the monosaccharide moiety of an activated nucleotide-sugar substrate from lipid carriers to acceptor substrates, such as monosaccharides, oligosaccharides, proteins, lipids, small organic molecules, and DNA, and hence produce a wide variety of biomolecules[8,22]. Glycosidases are enzymes involved in both the degradation of glycans and the removal of monosaccharides to form intermediates that are acted upon by glycosyltransferases for the biosynthesis of glycans[22].

Campylobacter jejuni and *Neisseria meningitidis* have well-characterized bacterial N-linked and O-linked glycosylation systems, respectively[23]. In N-linked protein glycosylation, an oligosaccharide is transferred by N-OST from a lipid donor to asparagines (N) located within the well-recognized consensus sequence D/E-Y-N-X-S/T and N-X-S/T (Y≠P, X≠P) of proteins[24,25]. Bacterial O-OSTs are responsible for the reversible attachment of glycans to hydroxyl groups of serine (S), threonine (T) and tyrosine (Y) residues, with no apparent sequence specificity[26,27].

Glycoproteomics is likely to identify Mtb virulence factors because glycoproteins on the bacterial cell envelope are used by mycobacteria to enable their entry into the primary human host cell, the macrophage[28]. It has been proposed that Mtb interacts with mannose receptors (MRs) on host cells via mannosylated proteins to enter the macrophages[29]. Despite the vital importance of these proteins in Mtb pathogenesis, our current knowledge of Mtb glycoproteins is still limited, and only a few secreted and cell wall-associated glycoproteins have been described to date[8,28,30,31]. Previous studies have used laboratory strains as model systems to study glycosylation in Mtb. However, only a few sub-groups within the genetically conserved MTBC appear to cause extensive outbreaks with different clinical presentation and AMR[32–35]. In this study, we employed qualitative and quantitative mass spectrometry and bioinformatics to explore the glycoproteomic patterns of clinical isolates from four lineages of the MTBC, lineages 3, 4, 5 and 7, to investigate the role of protein glycosylation in Mtb adaptation, survival and AMR.

Our study reveals the presence of a number of glycoproteins that play roles in MTBC virulence and pathogenesis. These include proteins involved in pathogen-host interaction, transport and biosynthesis of MTBC cell envelope components, and drug efflux pumps, which are attractive pharmacological targets. Furthermore, we found quantitative differences in glycosylation patterns among the different lineages of MTBC that may potentially contribute to explaining their phenotypic characteristics.

Results

Abundance of both O- and N-glycosylation profile among members of the MTBC. After filtering the data for potential contaminants and hits to the reverse database, the peptide intensities were log2-transformed. For protein identification, the data was further filtered using localization probability of 0.7, PEP of 0.05 and having valid values in at least one sample resulted in 2944 class-I glycosylation events derived from 1325 unique proteins in MTBC strains representing lineages 3, 4, 5 and 7. The term "glycosylation event" is used to avoid confusion when a single glycosylation site is glycosylated by more than one type of glycan residues. O-glycosylation constituted 2455 (83%) of the events identified (1311 events at T and 1144 events at S residues) and the remaining 489 sites (17%) were glycosylated at N residues (Fig. 1A,C, Supplementary Table S1). Comparative glycoproteomic analysis among the four MTBC lineages revealed that 945 (32.1%) of the total glycosylation events identified were shared amongst the four lineages (Fig. 1A). Comparison at the level of unique

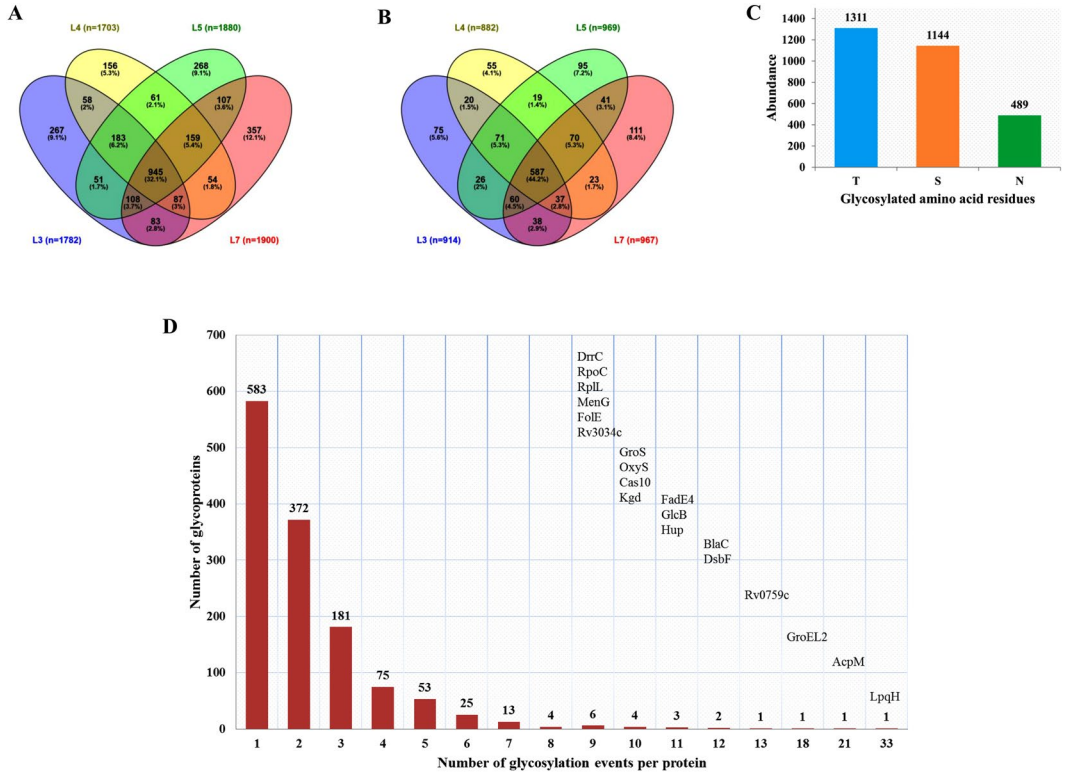


Figure 1. Abundance of glycosylation and glycoproteins in MTBC. Venn diagram showing the number of glycosylation events identified among the four lineages (N = 2944) (**A**), and the number of glycoproteins identified among the four lineages (N = 1325) (**B**), the number of N- and O-glycosylation events identified (N = 2944) (**C**), and number of glycosylation events identified per individual glycoprotein (**D**).

glycoproteins revealed that 44.2% of the glycoproteins were shared among the four lineages, irrespective of the glycosylation site and glycan residues (Fig. 1B).

Among the 57 most common naturally occurring glycan residues, deoxyhexoses (DeoxyHex) was the most frequently identified glycan residue in our search, followed by Hept, pent, Hex, HexN, HexNac/GlcNac, MurNGly and MurNac (Supplementary Table S3). We also identified sugar molecules attached to both ADP and UDP in comparable proportions, which may contribute to an activated nucleotide-sugar substrates for OST (Supplementary Table S3).

The glycosylation events identified per protein ranged from 1 to 33 (Fig. 1D, Supplementary Table S2). The lipoprotein LpqH was found to be the most highly glycosylated protein with no less than 33 events, hosting 24 events on T and 9 events on S residues (Fig. 1D). Notably, all glycan residues found on LpqH were composed of solely hexoses, while a cocktail of glycan residues were observed in other proteins harboring many glycosylation events (Supplementary Table S1). Among the 33 glycosylation events detected on LpqH, 20 events were common across all four lineages. Other highly glycosylated proteins included AcpM, GroeL2, BlaC, DsbF, FadE4 and HupB (Fig. 1D, Supplementary Table S2). Several hypothetical proteins were also glycosylated.

LpqH and AcpM have glycosylation sites concentrated in the interacting domains. The glycosylation sites in LpqH clustered in the N-terminus, densely located between residues 27–48 of the 159 amino acid protein (Fig. 2A). The amino acid residues from 41–60 are known to be involved in the binding of LpqH with the host MR. The four sites, T40, T41, S43 and Ser48 are located in this binding domain region of LpqH. In AcpM, three glycosylation sites in the helix S41, S43 and T51 are found in the carrier protein (CP) domain profile (Fig. 2C).

No apparent amino acid sequence specificity for Mtb glycosyltransferases. Comparing the 31-mer unique sequences of all peptides containing a glycosylation site by WebLogo yielded a "consensus" sequence atlas (Fig. 3). The distribution of the amino acids flanking the modified site showed a relatively high propensity for R, L, A, V, P and G residues (Fig. 3).

The glycolipoproteins and glycoproteins identified are involved in diverse biological functions. Based on the Gene Ontology (GO) analysis, fatty acid metabolism and lipid homeostasis, growth of symbiont in the host cell and responses to the host immune system were highly enriched biological processes. The cell wall and the plasma membrane were the two highly enriched cellular components of the glycoproteins identified. The molecular functions include ATP binding, oxidoreductase activity, acyl-CoA dehydrogenase activity,

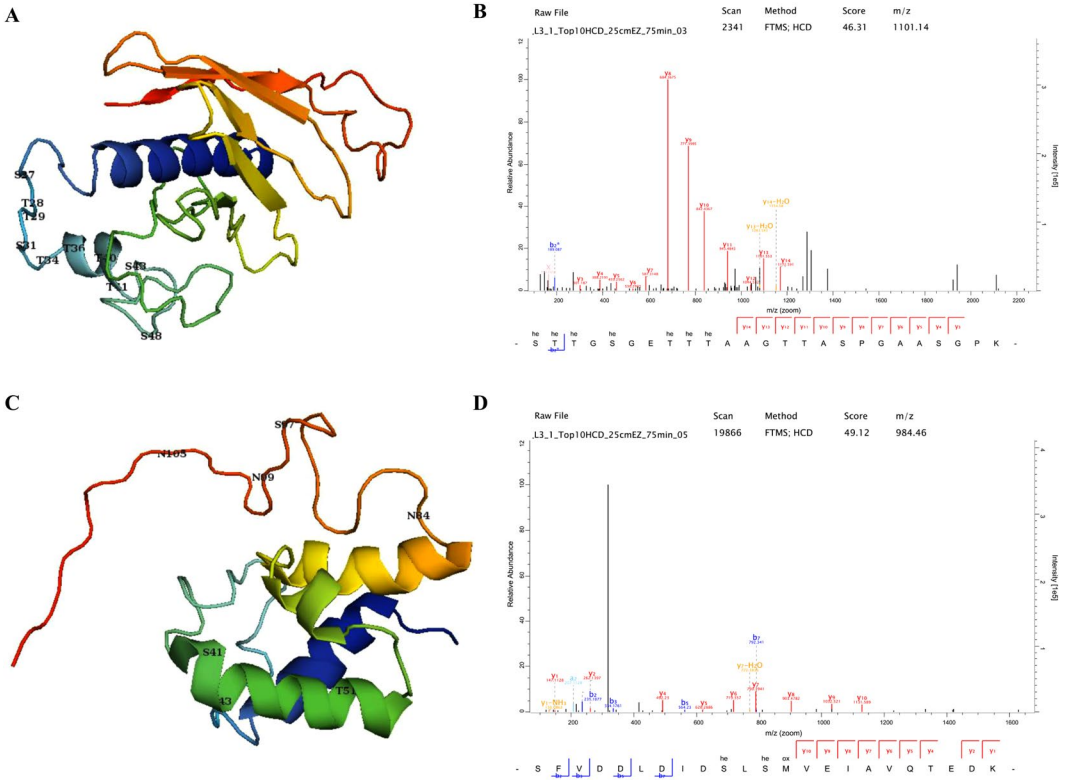


Figure 2. The 3D models, acetylation sites and representative spectra of LpqH (**A**,**B**) and AcpM (**C**,**D**). The glycosylation sites in LpqH are clustered in the N-terminus between residues 27–48 of the 159 amino acid protein (**A**). The four sites, T40, T41, S43 and S48 were located in this binding domain of LpqH. In AcpM, three sites in the helix, S41, S43 and T51, were found in the carrier protein (CP) domain profile (**C**). (he = hexose).

fatty-acyl-CoA binding, helicase activity, DNA binding, electron carrier activity and ligase activity (Fig. 4). The roles of the Mtb glycoproteins identified are summarized in Table 1. The non-glycosylated complement, however, encompassed proteins with functions and localization primarily in the cytoplasm. Furthermore, the GO analysis of uniquely identified glycoproteins provided strain-specific enrichment of biological processes, molecular functions (Supplementary Table S6).

Glycosylated proteins are involved in 14 specific metabolic pathways. Through protein-protein interaction (PPI) network analysis, 14 highly interconnected clusters were identified (Fig. 5). Most of the interacting glycoproteins identified were part of common pathways involved in fatty acid and lipid metabolism, protein synthesis, pathogen-host interaction, PG, AG, mycolic acid and capsule biosynthesis, stress responses, two-component systems (TCS), energy metabolism, and DNA replication repair and recombination (3R) (Fig. 5).

Glycolipoproteins involved in pathogen-host interaction. Our GO analysis a showed that most of the glycoproteins identified were localized in the cell wall and plasma membrane while lipid homeostasis, fatty acid metabolism, and response to the host immune system were among the biological processes enriched (Fig. 4). Lipoproteins were amongst the highly glycosylated Mtb proteins identified in this study, and are known to be involved in colonization, invasion, evasion of host defence and immunomodulation, cell envelope biogenesis, transport across membrane, nutrient acquisition, adhesion, cell invasion and initiation of inflammatory processes (Table 2)[13,14,36–38]. These glycoproteins include the LpqH, the MCE-family proteins, Apa, Heparin-binding hemagglutinin (HbhA) and LprG.

Glycosylation of proteins involved in MTBC cell envelope biogenesis. Membrane-associated proteins involved in lipid and fatty acid metabolism, cell envelope biosynthesis, pathogen-host interaction, transport, transcriptional regulation, and chaperone functions were also glycosylated (Fig. 5). After LpqH, the meromycolate extension acyl carrier protein AcpM was the second highly glycosylated protein identified with 21 glycosylation events. AcpM is involved in mycolic acid biosynthesis, a major component of the Mtb cell wall. Other glycoproteins involved in mycolic acid synthesis include methoxy mycolic acid synthases (MmaA1, MmaA2, MmaA3), mycolic acid synthase PcaA, polyketide synthases Pks13, enzymes involved in the synthesis of the Mtb cell wall components (PpsA, PpsC, PpsD, PpsE, PapA1, Rv2951c, FadD26, FadD28, LppX, DrrC and MmpL7), beta-ketoacyl-ACP synthases (KasA, KasB), mycolyltransferases (FbpB and FbpC), mycolic acid biosynthesis a protein FabG1, penicillin-binding glycoproteins (PBPs) (PbpA, PbpB, PonA1 and LdtA), proteins involved in PG biosynthesis (MurA, MurE, MurF, LprQ, FtsW, MviN, GlmS, GlmM, DacB1 and Wag31), AftD, EmbC, enzymes

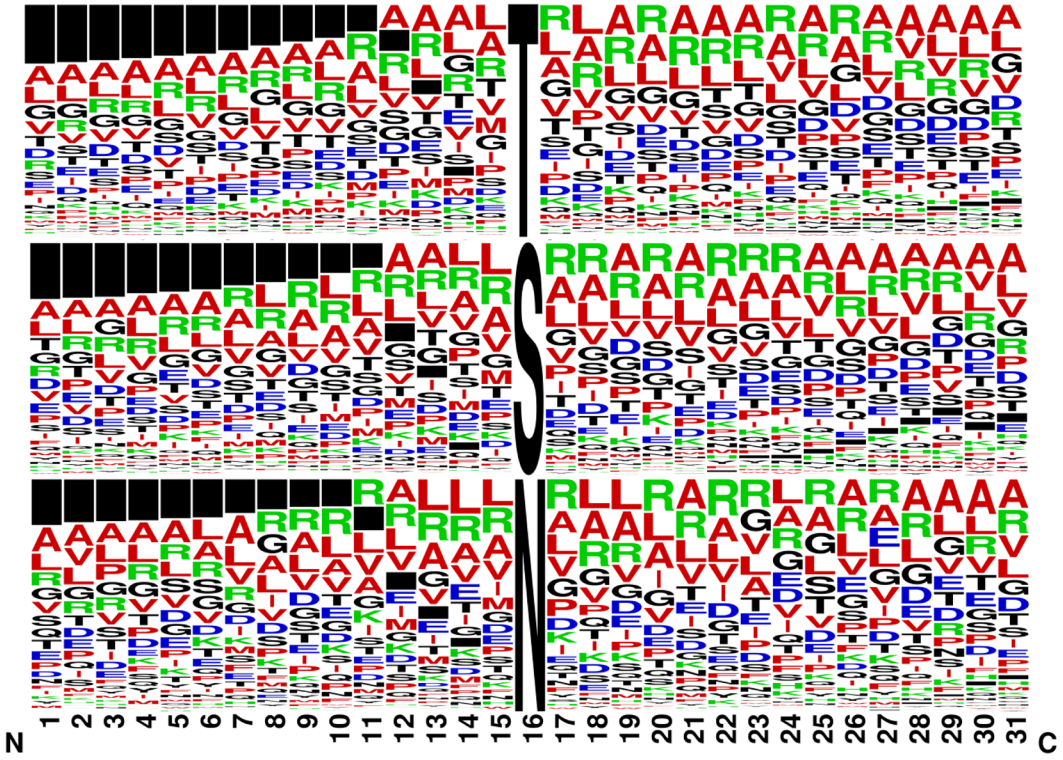


Figure 3. Glycosylation motif analysis. The N- and O-glycosylation motif generated from the high confidence identification indicates a higher likelihood of basic R, hydrophobic P, A, V and L, interspersed with Polar G; and some hydrophilic S and T around the O-glycosylation site. In addition, the glycosylation motifs seem to cluster predominantly at the N-terminus as indicated by the black boxes on the left. The height of each amino acid indicates its relative frequency at that specific position.

involved in the biosynthesis of alpha-D-glucan (GlgM, GlgB, GlgE, TreZ and MalQ), enzymes involved in biosynthesis of lipoglycans (PimB, Rv2181, MgtA, Ppm1 and Rv1459c).

Other clinically important glycoproteins identified. Clinically important glycoproteins include BlaC, chaperone proteins, TCS proteins, ESX secretion system proteins and other transporter proteins. Mtb BlaC was glycosylated at 12 sites, while the chaperones GroEL2 and GroS were found to have 18 and 10 glycosylation sites, respectively.

Glycosylation of cytoplasmic proteins in MTBC. Cytoplasmic proteins involved in translation and DNA metabolism were glycosylated (Fig. 5A,B). The cytochrome P450 proteins were also glycosylated.

MTBC strains exhibit lineage-specific glycoproteomic profiles. The GO analyses of exclusively identified glycoproteins provided strain-specific enrichment of biological processes and molecular functions (Supplementary Table S6). Among the 2944 glycosylation events detected, 1010 had valid values in at least six valid LFQ intensity values from the total of 12 biological replicates (50%) and were thus subjected to further quantitative analysis. The missing values were imputed from the normal distribution and the log2-transformed data was normalized to Z-scores for further statistical testing. Multiple sample test analysis at a $P < 0.05$ level of significance revealed that 101 sites on 67 proteins were differentially glycosylated (differential abundance of a peptide/protein glycosylated at a specific site) between the four MTBC lineages studied (Fig. 6A, Supplementary Tables S4 and S5). Notably, most of the differentially glycosylated proteins (43/67: 64%) were located in the cell wall and cell membrane or possess a membrane component. These proteins belonged to different functional categories including lipid metabolism, cell wall and cell processes, virulence, detoxification and adaptation, and hypothetical proteins (Supplementary Table S5). Clinically important differentially glycosylated proteins include the lipoarabinomannan carrier protein LprG, chaperone proteins GroEL1, class A β-lactamases BlaC, mammalian cell entry (Mce)-family protein Mce2D, peroxidase BpoB, penicillin-binding glycoprotein PbpB, and a number of proteins involved in fatty acid and lipid metabolism (Supplementary Table S5).

In the hierarchical clustering of the differentially glycosylated proteins, the modern lineages (lineage 3 and lineage 4 strains) clustered together, separated from the ancient lineages (lineage 5 and lineage 7 strains) (Fig. 6A). Four separate clusters of proteins with a particular glycosylation profile were generated. The first cluster included 14 hyper-glycosylated proteins (having a significantly higher number of glycosylation events at a specific position in a protein than the average) in lineage 3 and lineage 4 strains, encompassing LppW (S67, N69, T77), UvrA (T628, N636), PurK (S204), PPE42 (S12), FadE35 (S106, S110), LdtA (T147, S151, S156), DsbF (T47), NarL

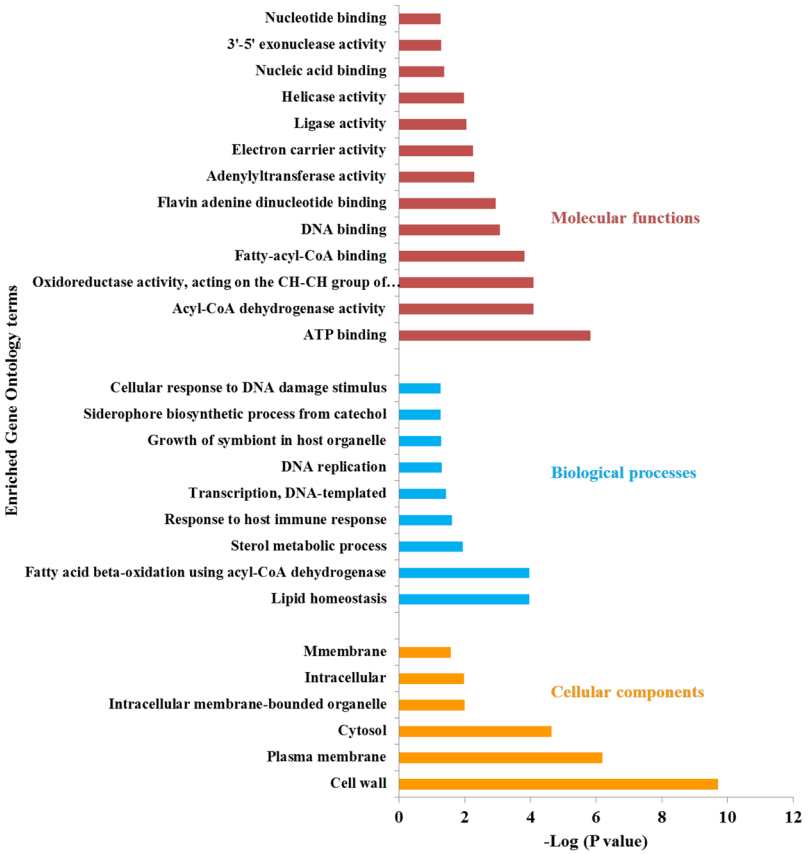


Figure 4. Gene Ontology analysis of Mtb glycoproteins. The gene ontology analysis showed that the majority of the glycoproteins identified were localized in the cell wall and plasma membrane while lipid homeostasis, fatty acid metabolism, and response to the host immune system were among the enriched biological processes.

(S113, T114) and Rv3483c (T77, S78) (Fig. 6B). Cluster two included 26 glycoproteins such as Mce2D (S270), DagK (S2), LprG (N161), UvrC (T469) and VapC45 (T83) that were hyper-glycosylated in lineage 3, lineage 4 and lineage 5 strains (Fig. 6C). Membrane proteins DsbF (S26, T33, T40) and Rv0412c (388, 394), Mrp (S373, T379), ATPase MoxR3 (S306) and PknA (T252, N256) were hyper-glycosylated in ancient lineages (lineage 5 and lineage 7 strains) (Fig. 6D). Cluster four included 11 glycoproteins, such as FadD9 (T97), DacB1 (T306, T312), FadD34 (T383), short chain type dehydrogenase/reductase Rv2766c (174) and peroxidase BpoB (N159), that were hyper-glycosylated in lineage 7 strains (Fig. 6E). Virulence factors such as PbpB (S83), GroeL1 (T141, T148, S150, S151), VapC10 (S6), VapB11 (S7) and the transcriptional regulatory protein Rv0818 (S65) were hyper-glycosylated in lineage 4 strains. The alanine rich protein Rv3863 (S11), the transmembrane ATP-binding protein ABC transporter protein Rv2326c and a polyketide synthase involved in sidrophore biosynthesis MbtD (T591) were hyper-glycosylated in lineage 5 strains (Supplementary Table S4).

Discussion

Our analysis identified with high confidence a total of 2944 glycosylation events on 1325 Mtb unique proteins. To our knowledge, the discovery of such a large number of glycosylation sites in these four clinical strains from different MTBC lineages is unprecedented. About 83.4% of the glycosylation events were localized on S and T residues, indicating a possible interplay with phosphorylation. It has been reported that different glycosylation events may occur on the same S and T residues of the protein or competitively at adjacent or residues in close proximity, and hence potentially allow control of cellular signaling[27]. The study provides the first evidence on N-linked protein glycosylation in Mtb and Gram-positive bacteria. Protein glycosylation occurred at numerous sites on surface-exposed proteins with no apparent amino acid sequence specificity (Fig. 3)[7]. As previously reported, there is a relatively high propensity for R, A, P, L, G, V, S and T flanking the modified sites in a significant portion of the glycosylation sites mapped[11,27,30] (Fig. 3). However, a number of suggested signature motifs were identified in nearly 17% of the events and R was enriched between the −8 and +8 positions in contrast to the D/E-Y-N-X-S/T (Y, X #P) motif proposed for N-glycosylation[24]. This difference might partly be due to the diversity of the glycan residues analyzed and the high degree of specificity for both their donor and acceptor substrates in the glycosyltransferases[39]. Comparative analysis revealed that only 32.1% of the glycosylation events and 44.2% of the glycoproteins were shared among the four lineages. The higher versatility at the level of PTMs may indicate the power of PTMs in explaining the phenotypic variability among MTBC than the proteomic studies.

Function	Process	Glycoproteins involved
Cell envelope synthesis	Mycolic acid synthesis	AcpM, MmaA1, MmaA2, MmaA3, PcaA, Pks13, FbpB and FbpC
	PDIM synthesis and transport	PpsA, PpsC, PpsD, PpsE, PapA1, FadD26, FadD28, LppX, DrrC and MmpL7
	PG synthesis	PBPs) PbpA (Rv0016c), PbpB (Rv2163c), PonA1 (Rv0050) and LdtA (Rv0116c), MurA, MurE, MurF, LprQ, FtsW, MviN, GlmS, GlmM, DacB1 and Wag31
	Capsule biosynthesis	GlgM (Rv1212c), GlgB (Rv1326c), GlgE (Rv1327c), TreZ (Rv1562c) and MalQ (Rv1781c)
	AG	AftD (Rv0236c), DprE1, EmbC and EmbR
	Lipoglycans (LM, LAM and PI)	PimB (Rv2188c), EmbC, Rv1459c and Rv2181
Membrane transport proteins	Sec	SecA1, SecY, SecD, SecE2, SecF
	Tat	TatB
	MmpL lipid transporters	MmpL1, MmpL3, MmpL4, MmpL5, MmpL6, MmpL8, MmpL9, MmpL10, MmpL11, MmpL12 and MmpL13b
	ATP-binding cassette (ABC)	DrrC, DppA, DppC, DppD, FecB, UgpC, UgpE, ProZ, CydD, MalQ, Rv2326c, Rv2041c, Rv1680, Rv3197, Rv0987, Rv1281c, Rv3092c, Rv1747, Rv1273c, Rv1739c, Rv2564, Rv0073
	Type-VII secretion	EccA1, EccB1, EccCb1, EccA2, EccB2, EccC2, EccD3, EccB3, EccC4, EccB4, EccA5, EccC5,
	Others	CpnT, NanT, IrtA, IrtB, ArsC
MCE family proteins		Mce2A, Mce2D, Mce1E/LprK, Mce3R, Mce1A, Mce4C, Mce1C, Mce3C, Mce2R, Mce1R, Mce1B, Mce2F, Mce2B, Mce3D and Apa*
Regulatory proteins		DevS-DosT/DosR, PhoR, WhiB3, WhiB4, WhiB5, WhiB7, TcrA, PrrA/PrrB, MtrA/MtrB, KdpD, KdpC, MoxR3, NarL, EmbR, PdtaR, GlnB, Mce1R, KstR, BlaR, BlaI and OxyS, Rv1353c, Rv0890c, Rv3095, Rv0494, Rv0043c, RamB, Rv0081, Rv0339c
Chaperones		GroS, DnaK, GroeL1, GroeL2, ClpB, ClpX, Hsp
Role in AMR		BlaC, KatG, RpoC, KasA, AhpD, FadE24, AcpM, IniB, IniC, EthA, OpcA, Wag31, RpoB, EmbR, EmbC, FabG1, RpsL, Mdh, Ndh, Alr, MtrAB, Rv2994, Rv0194, LprG, GyrA and GyrB
Potential drug targets		Mur enzymes, DrrC, PknD, MmpL3, GlgB, GlgE, Hpt, PbpA, PbpB, PonA1 and LdtA

Table 1. Virulence-associated membrane-bound glycoproteins, proteins involved in regulation, antimicrobial resistance (AMR) and chaperone proteins identified in MTBC.

DeoxyHex was the most frequently identified glycan residue in our search, followed by Hept, pent, Hex, HexN, HexNac/GlcNac, MurNGly and MurNac (Supplementary Table S3). In bacteria, 6-deoxy-hexoses, like fucose and rhamnose, are important components of cell surface glycans[40]. The pentose sugars arabinose and galactose are components of the heteropolysaccharide, AG, which serves to connect PG with the outer mycolic acid layer[19]. Bacterial heptosyltransferases are reported to be involved in O-glycosylation of autotransporters using ADP-heptose[41]. The presence of frequently occurring glycan residues attached to lipoproteins, extracellular polysaccharides (EPSs) and glycoproteins might alter the structure and function of these biomolecules in particular and bacterial physiology in general[40]. We identified both ADP- and UDP linked to different sugar molecules to form an activated nucleotide-sugar substrates for OST (Supplementary Tables S1 and S3). Most publications reported that only UDP-linked sugars were the substrates for OST[8,22], while other reports showed that a particular OST, heptosyltransferase, used ADP-heptose as an activated nucleotide-sugar substrate[41].

The outermost layer of the Mtb cell envelope is a major determinant of virulence and pathogenicity, and is mainly composed of proteins, polysaccharides and small amount of lipids[12,42]. It acts as a permeability barrier of the cell envelope, promoting the phagocytosis of Mtb[43], maintaining cell integrity, regulating phagosome maturation[44] and playing diverse roles in the pathogen-host interactions[42,43]. The gene ontology analysis showed that the majority of the glycoproteins identified were localized in the cell wall and plasma membrane while lipid homeostasis, fatty acid metabolism, and response to the host immune response were among the biological processes enriched. Besides, the PPI network analysis showed that most of these cell-envelope associated glycoproteins are involved in pathogen-host interaction and fatty acid/lipid metabolism. These cell envelope-associated glycoproteins have been shown to have a vital role in Mtb virulence and pathogenesis (reviewed in[12]).

Lipoproteins were amongst the highly glycosylated Mtb proteins identified in this study. Lipoproteins are a functionally diverse class of membrane-bound proteins involved in colonization, invasion, evasion of host defence and immunomodulation, cell envelope biogenesis, transport across the membrane, nutrient acquisition, adhesion, cell invasion and initiation of inflammatory processes (Table 2)[13,14,36]. The lipoprotein LpqH was the most densely glycosylated lipoprotein detected, with 33 N-terminally clustered O-glycosylation events, where all glycan residues were hexoses (Supplementary Table S1). Notably, these sites were densely located between residues 27–48 within the 159 amino acid protein. Some of the glycosylation sites have previously been reported as part of the MR binding domain of LpqH, as shown in the 3D model (Fig. 2)[30,45]. Three of the sites, T41, S43 and S48, were part of a mature protein fragment (residues 41–60) that was reported to prevent uptake of Mtb by macrophage-like U937 cells[46]. Altering the glycosylated Ser residues in LpqH have been shown to affect binding affinity and exposure to proteolytic cleavage[10]. LpqH, an immunodominant TLR2 agonist, is crucial for Mtb growth and multiplication in IFN-γ-activated macrophages as well as in IFN-γ-deficient mice[47]. Mannosylated LpqH is the major adhesin for the macrophage MR and DC-SIGN, and the mannose residue serves as an adhesin for binding to the host MR[29].

Other groups of identified glycolipoproteins involved in pathogen-host interaction are the MCE-family of proteins[36]. These glycoproteins have an active role in disease development and in-host virulence[12]. A total of 14 glycosylation events were identified on proteins expressed from the four Mtb *mce* operons (*mce1*, *mce2*, *mce3* and

mce4). The invasion-/adhesin-like MCE family glycolipoproteins encoded by *mce*s are located at the cell surface of Mtb and possibly involved in entry and survival inside macrophages[48].

A number of other clinically important glycoproteins were identified. The cell surface glycoprotein Apa binds to DC-SIGN and surfactant protein, facilitates colonization and invasion of host cells[49]. Changes in the glycosylation pattern of Apa lead to a reduced stimulatory T-lymphocyte response, exhibiting the biological role of the glycan moiety[50]. Glycosylation is also required for proper localization of superoxide dismutases (SodB)[51]. The immunogenic glycoproteins MPT64 and Apa are virulence factors involved in Mtb infection of human cells and is a promising candidate for a subunit-based anti-TB vaccine[12,52]. Heparin-binding hemagglutinin (HbhA) glycoprotein mediates adherence to epithelial cells and is required for extrapulmonary dissemination of Mtb[53]. The lipoprotein LprG is another glycolipoprotein that blocks host cell phagosome-lysosome fusion, and is required for full Mtb virulence[54].

Glycoproteins associated with drug efflux pumps, drug-hydrolyzing enzymes, or capable of altering Mtb cell wall permeability mediates the development of AMR (reviewed in[12]). These include proteins like the mycobacterial membrane protein large (MmpL) proteins, daunorubicin-dim-transport integral membrane protein ABC transporter (DrrC), class a beta-lactamase (BlaC) and LprG (Table 1). DrrC, Rv0194, Rv2994, Rv1273c and a number of MmpL glycoproteins are efflux pumps for anti-TB drugs, contributing to AMR[55,56]. In addition to a role in drug resistance, MmpLs are involved in the export of cell wall associated lipids and siderophores, and are attractive pharmacological targets[57,58]. BlaC hydrolyzes nitrocefin and other β-lactams, thereby increasing Mtb resistance towards different classes of β-lactam antibiotics[4]. LprG controls cell wall permeability and efflux of drugs, and therefore plays a role in Mtb susceptibility to first-line anti-TB drugs[5].

The study identified a number of membrane-associated glycoproteins involved in cell envelope biosynthesis and drug efflux pumps, which are potential Mtb drug targets (Table 1, Fig. 5). AcpM was second most densely glycosylated protein involved in mycolic acid biosynthesis, one of the major components of the Mtb cell wall. Glycosylation sites Ser41, Ser43 and Thr51 were detected within the AcpM CP domain profile. Importantly, one of the glycosylation sites identified (Ser41) is the binding site for 4′-phosphopantetheine, an activator of AcpM[59]. Other glycoproteins involved in mycolic acid synthesis include MmaA1, MmaA2, MmaA3 and PcaA, Pks13, KasA, KasB and FabG1. Bacilli lacking all mycolic acid methyltransferases are viable but highly attenuated and hyperinflammatory in mice[60]. Pks13 catalyzes the last condensation step of mycolic acid biosynthesis and is essential for the mycobacterial survival[61]. Glycoproteins FbpB (Ag85B) and FbpC (Ag85c) also possess a mycolyltransferase activity[62]. These glycoproteins help to maintain the Mtb cell wall integrity by catalyzing the transfer of mycolic acids to cell wall AG, and through the synthesis of the virulence factor cord factor (trehalose 6,6′-dimycolate, TDM)[62]. Furthermore, FbpB and FbpC are T- and B-cell antigens and may have an application in sero-diagnostics[63].

Penicillin-binding glycoproteins (PBPs) PbpA, PbpB, PonA1 and LdtA are transpeptidases involved in the synthesis of cross-linked PG that is part of the cell wall biogenesis[64]. Other essential glycoproteins involved in PG biosynthesis include MurA, MurE, MurF, LprQ, FtsW, MviN, GlmS, GlmM, DacB1 and Wag31[65]. Glycoproteins involved in PG biosynthesis, such as Mur enzymes and PBPs, are potential antibiotic targets[65]. Alpha-(1 → 3)-arabinofuranosyltransferase (AftD) is involved in the biosynthesis of the AG region of the mAGP complex, an essential component of the mycobacterial cell wall[19]. EmbC is involved in the polymerization of arabinose into the arabinan of the mycobacterial cell wall AG and is linked to resistance to EMB[66].

Polyketide synthases (PpsA, PpsC, PpsD, PpsE), PapA1, Rv2951c, FadD26 and FadD28 are multifunctional enzymes involved in the synthesis of the Mtb cell wall component, PDIM and other lipids[67], while the glycolipoproteins LppX, DrrC and MmpL7 are required for the translocation and localization of PDIM in the cell wall[68]. PDIM comprise of a number of virulence-enhancing lipids that act as defensive, offensive, or adaptive effectors of virulence[69]. Inactivation of mycobacterial *pps* and *drr* operons has been linked to defects in PDIM synthesis and secretion, respectively[70]. PknD, a regulator of MmpL7, has been proposed to be a potential anti-TB drug target[71].

Glycosyltransferases such as GlgM, GlgB, GlgE, TreZ and MalQ are enzymes involved in the biosynthesis of alpha-D-glucan, a constituent of Mtb capsular polysaccharides with D-arabino-D-mannan (AM) and D-mannan[1,42]. These enzymes are required for Mtb virulence[72]. GlgE-mediated 1,4 α-glucan synthesis has been implicated in *in vitro* lysosomal stress and can potentially be exploited for killing intracellular Mtb[73]. The Thr10 glycosylation site in GlgE has been shown to be a regulatory kinase substrate and a validated anti-TB drug target[74]. GlgB is a potential target for inhibitors[75]. Glycosylated mannosyltransferases PimB and Rv2181 are involved in the biosynthesis of lipoglycans LM, LAM and phosphatidylinositol (PI)[76]. Mannosyltransferases MgtA, Ppm1 and Rv1459c are involved in the synthesis of immunomodulatory LM and LAM via alpha-(1 → 6)-mannopyranosyltransferase activity[77]. A number of glycosylated fatty acyl-AMP ligases that have been shown to play a role in cell wall biosynthesis, production of complex lipids and growth[78] were identified. As discussed above, glycosylation is involved in regulating the activity of different enzymes. In this study, identification of glycosylated glycosyltransferases (with rare abundance) may play a role in regulating its function as an enzyme[79]. There are reports on auto-glycosylation mediated activation of glycosyltransferases in eukaryotes[80–82].

Other clinically important glycoproteins identified include chaperone and TCS proteins. The differential expression of chaperone glycoproteins, such as GroeL2 and GroS, in response to heat shock have previously been reported[12]. TCS regulate various aspects of mycobacterial physiology, including virulence, dormancy, persistence, and drug resistance[83]. The glycoprotein PhoPR regulates multiple virulence-associated processes in Mtb, including the biosynthesis of polyketide-derived lipids and acyltrehaloses. The inactivation of acyltrehaloses attenuates Mtb sufficiently to make it a possible live vaccine candidate[12,16]. The DosR/WhiB3 regulon is associated with hypoxia and redox adaptation, while WhiB3/PhoP is involved in cell wall lipid biosynthesis[84]. The DevS/DosR regulon is required for full Mtb virulence and is involved in regulating stress, dormancy and hypoxia[85].

Twenty glycosylation events on proteins belonging to the specialized ESX secretion system components, including the crucial T-cell antigen ESAT-6, were detected. The ESX secretion system is essential for full Mtb

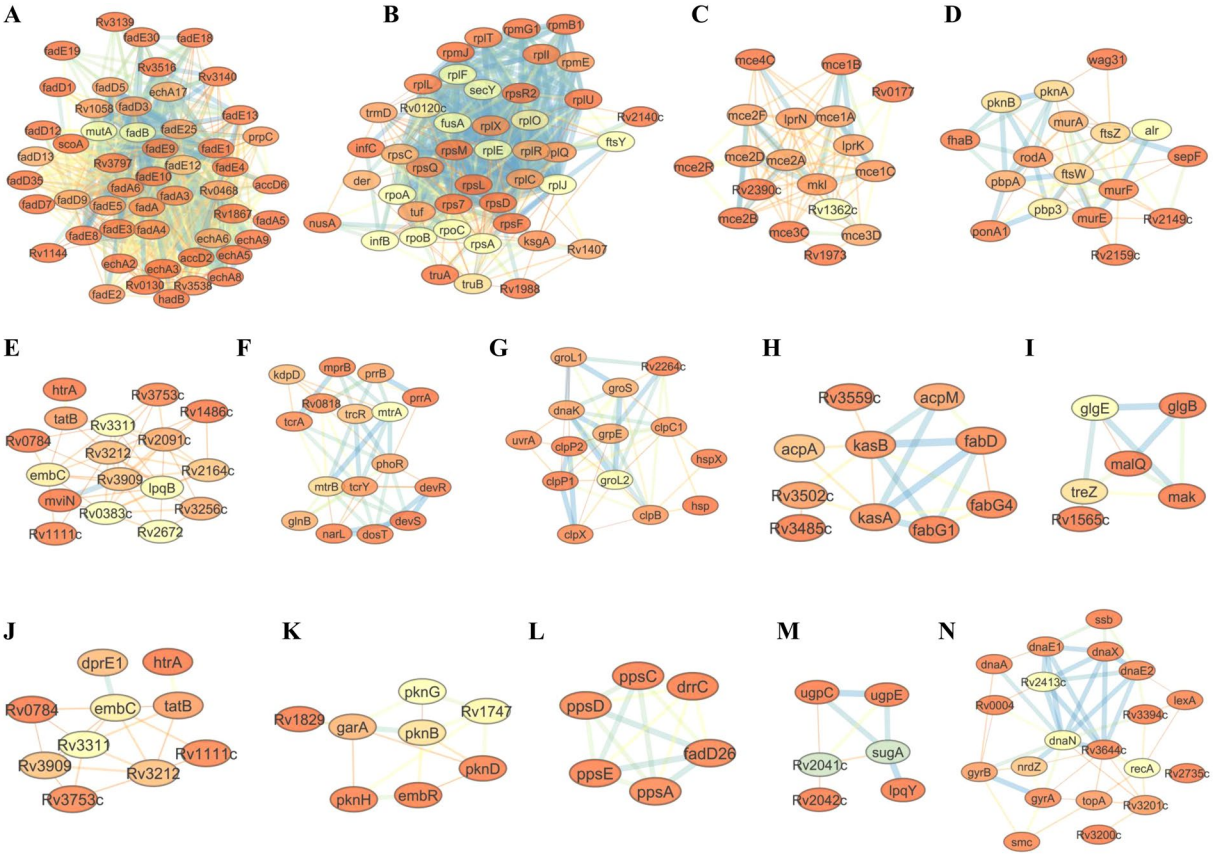


Figure 5. Protein-protein interaction networks of identified glycoproteins generated by Cytoscape. Networks are involved in lipid Metabolism & cell wall synthesis (**A**,**D**,**E**,**H**,**I**,**J**), protein synthesis (**B**), host-pathogen interaction (**C**), chaperone proteins (**G**), regulators (**F**,**K**), phthiocerol dimycocerosate (PDIM) synthesis (**L**), (sugar) transporters (**M**) and DNA replication, repair and recombination (**N**).

Role or function	Glycolipoprotein(s)
Antigenicity	LpqH, LprG, LppX
Adhesion and cell invasion	LpqH, LprG, LprK, LprN, LppA, LpqG, LppX, MCE
Required for growth	LpqH, LprK, SugA, LpqY, LppY, LpqB
Signal transduction	LprF, LprA, LprG, LppR, LppX, LpqB
Role in AMR	LprG, BlaC
Cell wall metabolism	PbpB, PbpA, PonA1, LprQ, LprK, LppW, LppX, LpqY, LpqB
ACB transport system	UgpE, UgpC, Rv2041c, LpqY, MalQ, DppA, FecB
Degradation	LpqP, LpqI, Rv2672, LpqL
Other enzymes and metabolic activities	GgtB, Rv0526, DsbF, LppZ, LpqD, SodB, Rv0526
Unknown function	Rv3693, Rv0679c, LppG, LpqU, LpqJ, LppO

Table 2. Representative biological activities elicited by glycolipoproteins identified from MTBC.

virulence (ESX-1) and physiological processes (ESX-3)[86]. Five proteins involved in the general secretion (Sec) pathway and a twin-arginine translocation pathways (TatB) were also found to be glycosylated. These specific proteins are essential for bulk export of proteins in Mtb[86]. CpnT, the first autotransporter-like protein to be identified in Mtb, was glycosylated at a domain that is required for the membrane localization of this protein[87]. Our former study showed that glycoproteins including LpqH, AcpM, GroEL1, GroEL2, DnaK, Pks13, KatG, LprK, SecA1 and a number of proteins involved in lipid metabolism and protein synthesis were highly acetylated in Mtb[88], which might indicate the interaction among different PTMs in fine-tuning specific cellular processes.

A recent report demonstrated a mechanism for co-regulation of Mtb cell wall synthesis and ribosome maturation (protein synthesis), and hence glycosylation of proteins involved in these two processes (Fig. 5A,B) may have a regulatory role[89]. Evidence for glycosylation of DNA-binding proteins (Dps) has been observed in *Salmonella enterica* in response to starvation and/or oxidative stress[90]. This is the first report on glycosylation of those cytosolic proteins. Glycosylation of cytochrome P450 has been demonstrated in eukaryotes (CYP2W1)[91] and in viral

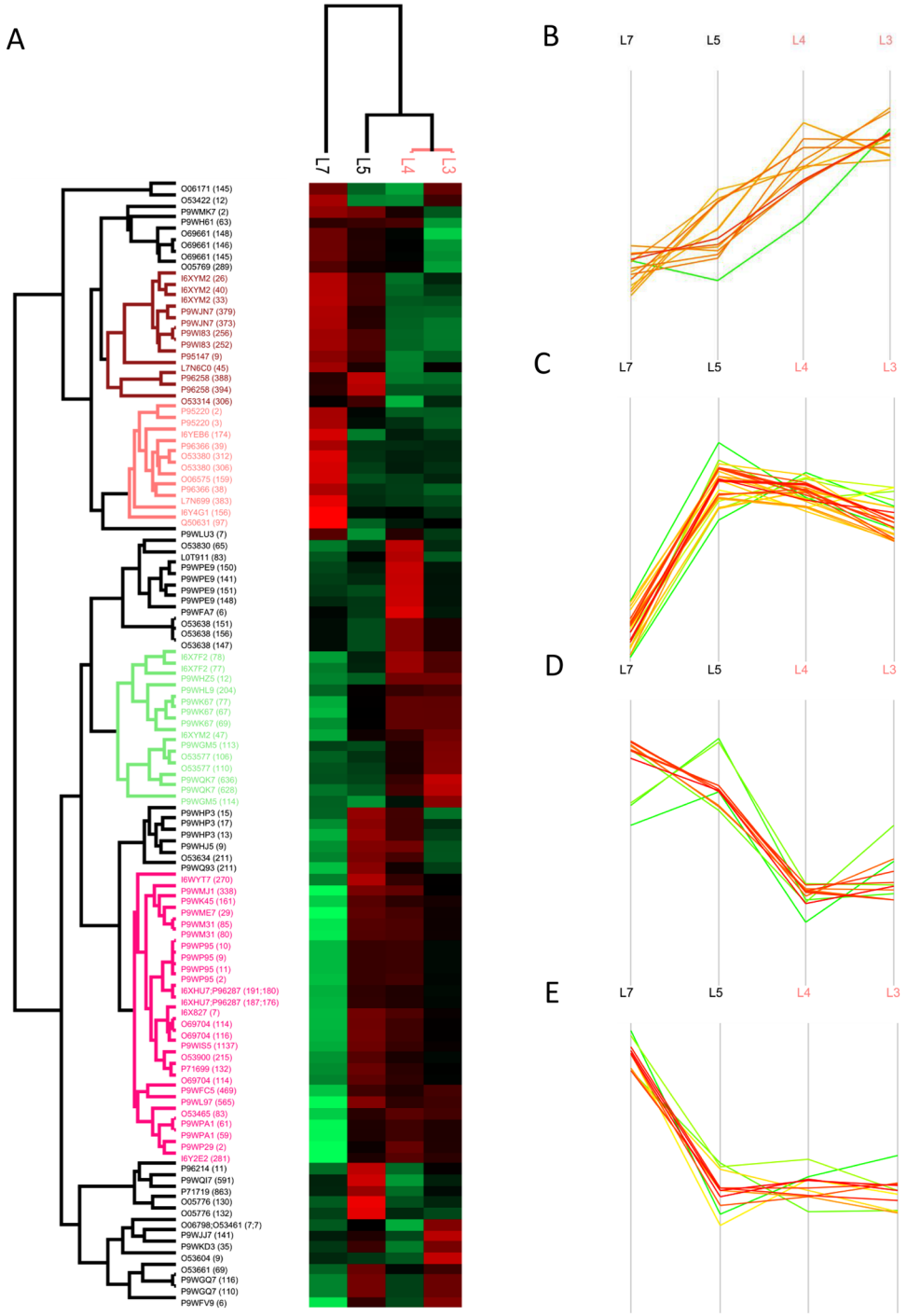


Figure 6. Profile of the 101 differentially glycosylated proteins in MTBC. Hierarchical clustering of differentially glycosylated proteins (**A**), proteins hyper-glycosylated in lineage 3 and lineage 4 strains (**B**), proteins hyper-glycosylated in lineage 3, lineage 4 and lineage 5 strains (**C**), proteins hyper-glycosylated in lineage 5 and lineage 7 strains (**D**), proteins hyper-glycosylated in lineage 7 strain (**E**).

cytochrome P450 (YP_143162)[92]. Glycosylation in this regard may enable the proper localization of cytochrome P450[14,93]. Cytochrome P450 plays a role in steroid metabolism, drug deactivation, fatty acid metabolism, xenobiotic detoxification and catabolism of exogenous compounds as a source of energy[94]. Fatty acid metabolism is a major source of carbon and energy in Mtb[95].

The PTMs identified by *in vitro* culture may only reflect the mycobacterial phenotype in the absence of stress, which may not completely if at all overlap patterns during infection. So, further mapping the exclusive presence and/or differential abundance of Mtb glycoproteins naturally or during exposure to environmental stress or infection may contribute to elucidate the selective advantages and survival strategies adopted by a specific pathogen. This information is fundamental for any drug or vaccine discovery process[12]. More than 64% of the differentially

glycosylated proteins were found to be cell envelope-associated proteins. These glycoproteins are reported to be involved in Mtb virulence and pathogenesis (reviewed in[12]). The hierarchical clustering of the differentially glycosylated proteins coincided with the phylogeny among the MTBC, where the modern lineages (lineage 3 and lineage 4 strains) clustered together, separated from the ancient lineages (lineage 5 and lineage 7 strains) (Fig. 6A)[96].

Clinically important proteins such as FadE35, LppW, LdtA, PurK, PPE42 and UvrA were hyper-glycosylated in lineage 3 and lineage 4 strains compared to the ancient lineages[64]. PurK has been identified to be a high-confidence drug target[97]. The antigen PPE42 is known to elicit humoral immune response against Mtb[98]. Glycoproteins including LprG, Mce2D, DagK, UvrC and VapC45 were hyper-glycosylated in all lineages except lineage 7 strains. LprG plays a role in transport and localization of the TLR2 agonists, LAM, PIM, LM and triacylglycerides to the cell surface, maintaining cell envelope integrity, and inhibition of phagosome-lysosome fusion, thereby enhancing Mtb survival inside macrophages[5,54]. The DagK is involved in the biosynthesis of Mtb virulence factors PI and PIMs[99]. The membrane proteins DsbF and Rv0412c, iron-sulfur cluster carrier protein Mrp, ATPase MoxR3 and PknA were hyper-glycosylated in lineage 5 and lineage 7 strains. A number of proteins involved in lipid metabolism such as FadD9, FadD34, and PG synthesis like DacB1, and oxidoreductases BpoB and Rv2766c, were hyper-glycosylated in lineage 7 strains. Four glycosylation sites on GroeL1, a chaperone involved in mycolic acid biosynthesis during biofilm formation[100], were uniformly hyper-glycosylated in lineage 4 strains. Penicillin-binding membrane protein PbpB, another hyper-glycosylated lineage 4 strains, is an essential enzyme involved in peptidoglycan biosynthesis and has been predicted to be an important drug target[101]. These proteins are essential virulence factors used by Mtb for cell wall biosynthesis, stress response, immunomodulation, efficient host cell invasion, survival, growth and other physiological processes[102]. The relative abundance of these essential glycoproteins across the different lineages of Mtb might lead to a specific phenotype with better adaptability to the host.

Identification of glycoproteins and their function contributes to a better understanding of the pathogenesis and survival strategies adopted by Mtb. This knowledge is fundamental for diagnostic, drug or vaccine discovery process. Many anti-TB drugs target the biosynthesis of PG, MA and AG, drug efflux pumps and other virulence factors used by Mtb to efficiently invade and multiply inside the host[12]. Our study has identified a large number of membrane-associated glycolipoproteins involved in Mtb pathogenesis. We present a comprehensive glycoproteome map of Mtb and show that there are significant quantitative differences across the various Mtb lineages that may directly influence phenotype. Further purification and detailed functional studies addressing selected uncharacterized glycoproteins may reveal the physiological role of protein glycosylation in defining the phenotype of a bacillus. These findings expand the current understanding of the nature and diversity of Mtb glycoproteins, open a new avenue of research for identification of potential drug targets, and create opportunities to engineer glycoproteins for their clinical applications[103].

Methods

Mtb strains and growth conditions. Four clinical strains representing four different Mtb lineages, lineage 3 (CAS-DELHI), lineage 4 (FSP471.1), lineage 5 (*M. africanum*) and lineage 7 (*Aethiops vetus*[104]) strains, were cultured on Middlebrook 7H10 agar plates for 32 days. The details of culturing, sample handling and inactivation were performed as previously described in Yimer *et al.*[105].

Proteomic analyses.

(i) **Preparation of cell lysates**. The Mtb cell pellets were mechanically disrupted by bead beating with a Mag-Na Lyser (Roche, US) as described by Yimer *et al.*[105].

(ii) **In-gel trypsin digestion**. Gel-fractionated protein samples (100 µg) from Mtb cells grown to late exponential phase were stained using a Colloidal Blue Staining kit (Invitrogen, CA) and each gel-lane was divided into six fractions. Each fraction was subjected to in-gel reduction, alkylation, and tryptic digestion as previously described[106]. Proteins were reduced using 10 mM DTT, alkylated with 55 mM iodoacetamide and digested with sequence grade trypsin (Promega, 1:100; w/w) overnight at 37 °C in 50 mM NH_4HCO_3. The in-gel digested protein samples were extracted using 50% and 100% acetonitrile (ACN), dried by SpeedVac concentrator (Eppendorf, concentrator 5301) and re-suspended using 0.05% trifluoroacetic acid (TFA). The extracted protein samples were purified using C_{18} stage tips by stacking three discs from Empore and transferred to auto-sampler nano LC vials for LC-MS/MS analysis as.

(iii) **Nano-LC-MS/MS analysis**. Peptide characterization and quantitation were performed by nano LC-MS/MS using a Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer interfaced with an EASY1000-nano-electrospray ion source (Thermo-Fisher Scientific, Biberach, Germany). The LC gradient was from 2% to 90% solvent B (0.1% FA in 97% CAN) in 50 µm × 15 cm analytical columns (PepMap RSLC, C18, 2 µm, 100 Å, Thermo Scientific) for 75 min analysis at a flow rate of 0.3 µl/min. The mass spectrometer was operated in data-dependent acquisition mode with automatic switching between MS and MS/MS scans.

The full MS scans were acquired at 70K resolution with automatic gain control (AGC) target of 1×10^6 ions between m/z = 300 to 1800 and were surveyed for a maximum injection time of 200 milliseconds (ms). Higher energy collision dissociation (HCD) was used for peptide fragmentation at normalized collision energy set to 28. The MS/MS scans were performed using a data-dependent top10 method at a resolution of 17.5K with an AGC of 5×10^4 ions at maximum injection time of 100 ms and isolation window of 2.0 m/z units. An underfill ratio of 10% and dynamic exclusion duration of 30 s was applied. For each Mtb lineage, three biological replicates were analyzed with each biological replicates fractionated into six gel bands, resulting in a total of 72 analytical runs (four lineages * three biological replicates * six fractions).

(iv) **Protein and PTM identification**. The Maxquant software (version 1.5.7.4) was employed for protein

and glycosylation site identification from the raw MS data[107]. The raw mass spectral data were searched against the Uniprot Mtb protein database containing 3993 protein sequences concatenated to reverse decoy database and protein sequences for common contaminants. Trypsin [KR].[^P] was specified as a cleavage enzyme with up to two missed cleavages. The "re-quantify" and "match between runs" options were utilized with a retention time alignment window of three min. Carbamidomethylation of cysteine residues was specified as a fixed modification and acetylation on protein N-terminal, conversion of N-terminal glutamine and glutamic acid to pyroglutamic acid, and oxidation of methionine were set as the variable modifications.

For the PTM analysis, a number of glycan residues were configured in the MaxQuant search at three different amino acid residues, N, S and T (Supplementary Table S3) and were set to variable modification. Both unique and razor peptides were used for the quantification of PTM abundance. Peptides with a minimum length of seven amino acids and detected in at least one or more of the replicates were considered for identification. For protein identification, a minimum of two peptides, of which at least one was unique, was required per protein group. All other parameters in MaxQuant were set to default values.

(v) **Bioinformatics analysis.**

Statistical analysis. Statistical significance was determined with multiple-sample test at a $P < 0.05$ level of significance using Perseus software (version 1.6.0.7). All modified peptide spectra were validated by applying stringent site localization probability of >0.70 and PEP of <0.05 prior to further analysis. PTM sites with a minimum of one valid value from the total samples were considered for PTM site identification. Modified peptides with valid values in at least 50% of the samples were considered for label-free relative quantification analysis.

Analysis of N- and O-glycosylation motifs. A sequence logo generators WebLogo (http://weblogo.berkeley.edu/logo.cgi) was used to identify the enriched amino acid motifs flanking the glycosylated sites. The sequence windows from the identification table were used to generate the sequence motifs for the three modified amino acids S, T and N.

Gene Ontology analysis of glycosylated proteins. The biological processes, cellular component and molecular function for the identified glycoproteins were analyzed using DAVID Bioinformatics Resources 6.7. The proteins were classified by GO annotation based on three terms; molecular function (MF), biological process (BP) and cellular component (CC).

Protein-protein interaction network analysis. Protein-protein interaction (PPI) networks were generated via STRING database version 10 with a high confidence threshold of 0.7 and imported into Cytoscape software (version 3.5.0) to produce the final interaction networks. Highly interconnected clusters were identified using MCODE and ClusterOne plug-in toolkits.

Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD009676.

References

1. Angala, S. K., Belardinelli, J. M., Huc-Claustre, E., Wheat, W. H. & Jackson, M. The cell envelope glycoconjugates of Mycobacterium tuberculosis. *Critical reviews in biochemistry and molecular biology* **49**, 361–399, https://doi.org/10.3109/10409238.2014.925420 (2014).
2. Jarlier, V. & Nikaido, H. Mycobacterial cell wall: structure and role in natural resistance to antibiotics. *FEMS microbiology letters* **123**, 11–18 (1994).
3. Wang, F., Cassidy, C. & Sacchettini, J. C. Crystal structure and activity studies of the Mycobacterium tuberculosis β-lactamase reveal its critical role in resistance to β-lactam antibiotics. *Antimicrobial agents and chemotherapy* **50**, 2762–2771 (2006).
4. Nampoothiri, K. *et al*. Molecular cloning, overexpression and biochemical characterization of hypothetical β-lactamases of Mycobacterium tuberculosis H37Rv. *Journal of applied microbiology* **105**, 59–67 (2008).
5. Bianco, M. V. *et al*. Role of P27–P55 operon from Mycobacterium tuberculosis in the resistance to toxic compounds. *BMC infectious diseases* **11**, 195 (2011).
6. Bigi, F. *et al*. The gene encoding P27 lipoprotein and a putative antibiotic-resistance gene form an operon in Mycobacterium tuberculosis and Mycobacterium bovis. *Microbiology* **146**, 1011–1018 (2000).
7. Daubenspeck, J. M., Jordan, D. S., Simmons, W., Renfrow, M. B. & Dybvig, K. General N-and O-Linked Glycosylation of Lipoproteins in Mycoplasmas and Role of Exogenous Oligosaccharide. *PLOS ONE* **10**, e0143362, https://doi.org/10.1371/journal.pone.0143362 (2015).
8. Calder, B., Soares, N. C., de Kock, E. & Blackburn, J. M. Mycobacterial proteomics: analysis of expressed proteomes and post-translational modifications to identify candidate virulence factors. *Expert Rev Proteomics* **12**, 21–35, https://doi.org/10.1586/14789450.2015.1007046 (2015).
9. Torrelles, J. B. & Schlesinger, L. S. Diversity in Mycobacterium tuberculosis mannosylated cell wall determinants impacts adaptation to the host. *Tuberculosis (Edinburgh, Scotland)* **90**, 84–93, https://doi.org/10.1016/j.tube.2010.02.003 (2010).
10. Herrmann, J., O'Gaora, P., Gallagher, A., Thole, J. & Young, D. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from Mycobacterium tuberculosis. *The EMBO journal* **15**, 3547 (1996).
11. Herrmann, J. L., Delahay, R., Gallagher, A., Robertson, B. & Young, D. Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. *Febs Letters* **473**, 358–362 (2000).
12. Sonawane, A., Mohanty, S., Jagannathan, L., Bekolay, A. & Banerjee, S. Role of glycans and glycoproteins in disease development by Mycobacterium tuberculosis. *Critical reviews in microbiology* **38**, 250–266 (2012).
13. Becker, K. *et al*. Lipoprotein Glycosylation by Protein-O-Mannosyltransferase (MAB_1122c) Contributes to Low Cell Envelope Permeability and Antibiotic Resistance of Mycobacterium abscessus. *Frontiers in microbiology* **8**, 2123 (2017).
14. Kovacs-Simon, A., Titball, R. & Michell, S. L. Lipoproteins of bacterial pathogens. *Infection and immunity* **79**, 548–561 (2011).

15. Brennan, P. J. Structure, function, and biogenesis of the cell wall of Mycobacterium tuberculosis. *Tuberculosis* **83**, 91–97 (2003).
16. Jackson, M., McNeil, M. R. & Brennan, P. J. Progress in targeting cell envelope biogenesis in Mycobacterium tuberculosis. *Future Microbiol* **8**, 855–875, https://doi.org/10.2217/fmb.13.52 (2013).
17. Gaur, R. L. *et al*. LprG-mediated surface expression of lipoarabinomannan is essential for virulence of Mycobacterium tuberculosis. *PLoS pathogens* **10**, e1004376 (2014).
18. Daffe, M. & Etienne, G. The capsule of Mycobacterium tuberculosis and its implications for pathogenicity. *Tubercle and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease* **79**, 153–169, https://doi.org/10.1054/tuld.1998.0200 (1999).
19. Alderwick, L. J., Harrison, J., Lloyd, G. S. & Birch, H. L. The Mycobacterial cell wall—peptidoglycan and Arabinogalactan. *Cold Spring Harbor perspectives in medicine* **5**, a021113 (2015).
20. Mahapatra, S. *et al*. Mycobacterial lipid II is composed of a complex mixture of modified muramyl and peptide moieties linked to decaprenyl phosphate. *Journal of bacteriology* **187**, 2747–2757 (2005).
21. Moran, A. P. *Microbial glycobiology: structures, relevance and applications* (Elsevier, 2009).
22. Rini, J. M. & Esko, J. D. Glycosyltransferases and glycan-processing enzymes (2017).
23. Tan, F. Y., Tang, C. M. & Exley, R. M. Sugar coating: bacterial protein glycosylation and host–microbe interactions. *Trends in biochemical sciences* **40**, 342–350 (2015).
24. Kowarik, M. *et al*. Definition of the bacterial N-glycosylation site consensus sequence. *Embo j* **25**, 1957–1966, https://doi.org/10.1038/sj.emboj.7601087 (2006).
25. Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S. & Aebi, M. The NXS/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **15**, 361–367 (2004).
26. Zarschler, K. *et al*. Protein tyrosine O-glycosylation—a rather unexplored prokaryotic glycosylation system. *Glycobiology* **20**, 787–798 (2010).
27. Gloster, T. M. & Vocadlo, D. J. Mechanism, structure, and inhibition of O-GlcNAc processing enzymes. *Current signal transduction therapy* **5**, 74–91 (2010).
28. Graham, R. L. & Hess, S. Mass spectrometry in the elucidation of the glycoproteome of bacterial pathogens. *Current Proteomics* **7**, 57–81 (2010).
29. Diaz-Silvestre, H. *et al*. The 19-kDa antigen of Mycobacterium tuberculosis is a major adhesin that binds the mannose receptor of THP-1 monocytic cells and promotes phagocytosis of mycobacteria. *Microbial pathogenesis* **39**, 97–107 (2005).
30. Smith, G. T., Sweredoski, M. J. & Hess, S. O-linked glycosylation sites profiling in Mycobacterium tuberculosis culture filtrate proteins. *Journal of proteomics* **97**, 296–306 (2014).
31. Gonzalez-Zamorano, M. *et al*. Mycobacterium tuberculosis glycoproteomics based on ConA-lectin affinity capture of mannosylated proteins. *J Proteome Res* **8**, 721–733, https://doi.org/10.1021/pr800756a (2009).
32. Peters, J. S. *et al*. Identification of Quantitative Proteomic Differences between Mycobacterium tuberculosis Lineages with AlteredVirulence. *Frontiers in Microbiology* **7**, 813, https://doi.org/10.3389/finicb.2016.00813 (2016).
33. 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. *Clinical & Experimental Immunology* **137**, 460–468 (2004).
34. Manabe, Y. C. *et al*. Different strains of Mycobacterium tuberculosis cause various spectrums of disease in the rabbit model of tuberculosis. *Infection and immunity* **71**, 6004–6011 (2003).
35. Nicol, M. P. & Wilkinson, R. J. The clinical consequences of strain diversity in Mycobacterium tuberculosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102**, 955–965 (2008).
36. Sutcliffe, I. C. & Harrington, D. J. Lipoproteins of Mycobacterium tuberculosis: an abundant and functionally diverse class of cell envelope components. *FEMS microbiology reviews* **28**, 645–659 (2004).
37. Rezwan, M., Grau, T., Tschumi, A. & Sander, P. Lipoprotein synthesis in mycobacteria. *Microbiology* **153**, 652–658 (2007).
38. Nguyen, H. T., Wolff, K. A., Cartabuke, R. H., Ogwang, S. & Nguyen, L. A lipoprotein modulates activity of the MtrAB two-component system to provide intrinsic multidrug resistance, cytokinetic control and cell wall homeostasis in Mycobacterium. *Mol Microbiol* **76**, 348–364, https://doi.org/10.1111/j.1365-2958.2010.07110.x (2010).
39. Rini, J. M. & Esko, J. D. In *Essentials of Glycobiology* (eds rd *et al*.) 65–75 (Cold Spring Harbor Laboratory Press Copyright 2015–2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved., 2015).
40. Mäki, M. & Renkonen, R. Biosynthesis of 6-deoxyhexose glycans in bacteria. *Glycobiology* **14**, 1R–15R (2003).
41. Lu, Q. *et al*. An iron-containing dodecameric heptosyltransferase family modifies bacterial autotransporters in pathogenesis. *Cell Host Microbe* **16**, 351–363, https://doi.org/10.1016/j.chom.2014.08.008 (2014).
42. Sambou, T. *et al*. Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: biosynthesis and impact on the persistence in mice. *Molecular microbiology* **70**, 762–774 (2008).
43. Torrelles, J. B., Azad, A. K. & Schlesinger, L. S. Fine discrimination in the recognition of individual species of phosphatidyl-myo-inositol mannosides from Mycobacterium tuberculosis by C-type lectin pattern recognition receptors. *The Journal of immunology* **177**, 1805–1816 (2006).
44. Fenton, M. J., Riley, L. W. & Schlesinger, L. S. In *Tuberculosis and the tubercle bacillus* 405–426 (American Society of Microbiology, 2005).
45. Parra, J. *et al*. Scrutiny of Mycobacterium tuberculosis 19 kDa antigen proteoforms provides new insights in the lipoglycoprotein biogenesis paradigm. *Sci Rep* **7**, 43682, https://doi.org/10.1038/srep43682 (2017).
46. Ocampo, M., Curtidor, H., Vanegas, M., Patarroyo, M. A. & Patarroyo, M. E. Specific interaction between Mycobacterium tuberculosis lipoprotein-derived peptides and target cells inhibits mycobacterial entry *in vitro*. *Chemical biology & drug design* **84**, 626–641, https://doi.org/10.1111/cbdd.12365 (2014).
47. 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* **25**, 7153–7159 (2007).
48. Ahmad, S., El-Shazly, S., Mustafa, A. & Al-Attiyah, R. The six mammalian cell entry proteins (Mce3A–F) encoded by the mce3 operon are expressed during *in vitro* growth of Mycobacterium tuberculosis. *Scandinavian journal of immunology* **62**, 16–24 (2005).
49. Ragas, A., Roussel, L., Puzo, G. & Rivière, M. The Mycobacterium tuberculosis cell-surface glycoprotein apa as a potential adhesin to colonize target cells via the innate immune system pulmonary C-type lectin surfactant protein A. *Journal of biological chemistry* **282**, 5133–5142 (2007).
50. Satchidanandam, V. *et al*. The glycosylated Rv1860 protein of Mycobacterium tuberculosis inhibits dendritic cell mediated TH1 and TH17 polarization of T cells and abrogates protective immunity conferred by BCG. *PLoS Pathog* **10**, e1004176, https://doi.org/10.1371/journal.ppat.1004176 (2014).
51. Ota, F., Kizuka, Y., Kitazume, S., Adachi, T. & Taniguchi, N. N-Glycosylation is essential for the secretion of extracellular superoxide dismutase. *FEBS letters* **590**, 3357–3367 (2016).
52. Rodríguez, D. *et al*. Peptides from the Mycobacterium tuberculosis Rv1980c protein involved in human cell infection: insights into new synthetic subunit vaccine candidates. *Biological chemistry* **391**, 207–217 (2010).
53. Pethe, K. *et al*. The heparin-binding haemagglutinin of M. tuberculosis is required for extrapulmonary dissemination. *Nature* **412**, 190–194 (2001).

54. Becker, K. & Sander, P. Mycobacterium tuberculosis lipoproteins in virulence and immunity–fighting with a double-edged sword. *FEBS letters* **590**, 3800–3819 (2016).
55. da Silva, P. E. A., Von Groll, A., Martin, A. & Palomino, J. C. Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis. *FEMS Immunology & Medical Microbiology* **63**, 1–9 (2011).
56. Gupta, A. K. *et al*. Microarray analysis of efflux pump genes in multidrug-resistant Mycobacterium tuberculosis during stress induced by common anti-tuberculous drugs. *Microbial drug resistance* **16**, 21–28 (2010).
57. Chalut, C. MmpL transporter-mediated export of cell-wall associated lipids and siderophores in mycobacteria. *Tuberculosis (Edinburgh, Scotland)* **100**, 32–45, https://doi.org/10.1016/j.tube.2016.06.004 (2016).
58. Viljoen, A. *et al*. The diverse family of MmpL transporters in mycobacteria: from regulation to antimicrobial developments. *Molecular microbiology* **104**, 889–904 (2017).
59. Zimhony, O. *et al*. AcpM, the Meromycolate Extension Acyl Carrier Protein of Mycobacterium tuberculosis, Is Activated by the 4′-Phosphopantetheinyl Transferase PptT, a Potential Target of the Multistep Mycolic Acid Biosynthesis. *Biochemistry* **54**, 2360–2371, https://doi.org/10.1021/bi501444e (2015).
60. Barkan, D., Hedhli, D., Yan, H.-G., Huygen, K. & Glickman, M. S. Mycobacterium tuberculosis lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice. *Infection and immunity* **80**, 1958–1968 (2012).
61. Gavalda, S. *et al*. The Polyketide Synthase Pks13 Catalyzes a Novel Mechanism of Lipid Transfer in Mycobacteria. *Chemistry & Biology* **21**, 1660–1669, https://doi.org/10.1016/j.chembiol.2014.10.011 (2014).
62. Puech, V. *et al*. Evidence for a partial redundancy of the fibronectin-binding proteins for the transfer of mycoloyl residues onto the cell wall arabinogalactan termini of Mycobacterium tuberculosis. *Molecular microbiology* **44**, 1109–1122 (2002).
63. Steingart, K. R. *et al*. Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: a meta-analysis. *Clinical and vaccine immunology* **16**, 260–276 (2009).
64. Kieser, K. J. *et al*. Peptidoglycan synthesis in Mycobacterium tuberculosis is organized into networks with varying drug susceptibility. *Proceedings of the National Academy of Sciences* **112**, 13087–13092 (2015).
65. Lovering, A. L., Safadi, S. S. & Strynadka, N. C. Structural perspective of peptidoglycan biosynthesis and assembly. *Annual review of biochemistry* **81**, 451–478 (2012).
66. Goude, R., Amin, A., Chatterjee, D. & Parish, T. The arabinosyltransferase EmbC is inhibited by ethambutol in Mycobacterium tuberculosis. *Antimicrobial agents and chemotherapy* **53**, 4138–4146 (2009).
67. Bisson, G. P. *et al*. Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, rpoB mutant Mycobacterium tuberculosis. *Journal of bacteriology* **194**, 6441–6452 (2012).
68. Goude, R. & Parish, T. The genetics of cell wall biosynthesis in Mycobacterium tuberculosis (2008).
69. Wei, J. *et al*. Genome-wide transcription analyses in Mycobacterium tuberculosis treated with lupulone. *Brazilian Journal of Microbiology* **45**, 333–342 (2014).
70. Waddell, S. *et al*. Inactivation of polyketide synthase and related genes results in the loss of complex lipids in Mycobacterium tuberculosis H37Rv. *Letters in applied microbiology* **40**, 201–206 (2005).
71. Pérez, J. *et al*. Mycobacterium tuberculosis transporter MmpL7 is a potential substrate for kinase PknD. *Biochemical and biophysical research communications* **348**, 6–12 (2006).
72. Koliwer-Brandl, H. *et al*. Metabolic Network for the Biosynthesis of Intra- and Extracellular alpha-Glucans Required for Virulence of Mycobacterium tuberculosis. *PLoS Pathog* **12**, e1005768, https://doi.org/10.1371/journal.ppat.1005768 (2016).
73. Lin, W. *et al*. Transcriptional Profiling of Mycobacterium tuberculosis Exposed to *In Vitro* Lysosomal Stress. *Infection and immunity* **84**, 2505–2523 (2016).
74. Leiba, J. *et al*. Mycobacterium tuberculosis maltosyltransferase GlgE, a genetically validated antituberculosis target, is negatively regulated by Ser/Thr phosphorylation. *Journal of Biological Chemistry* **288**, 16546–16556 (2013).
75. Dkhar, H. K. *et al*. Discovery of Mycobacterium tuberculosis α-1, 4-glucan branching enzyme (GlgB) inhibitors by structure-and ligand-based virtual screening. *Journal of Biological Chemistry* **290**, 76–89 (2015).
76. Torrelles, J. B. *et al*. Inactivation of Mycobacterium tuberculosis mannosyltransferase pimB reduces the cell wall lipoarabinomannan and lipomannan content and increases the rate of bacterial-induced human macrophage cell death. *Glycobiology* **19**, 743–755, https://doi.org/10.1093/glycob/cwp042 (2009).
77. Mishra, A. K. *et al*. Identification of a novel α (1 → 6) mannopyranosyltransferase MptB from Corynebacterium glutamicum by deletion of a conserved gene, NCgl1505, affords a lipomannan-and lipoarabinomannan-deficient mutant. *Molecular microbiology* **68**, 1595–1613 (2008).
78. Portevin, D. *et al*. The Acyl-AMP Ligase FadD32 and AccD4-containing Acyl-CoA Carboxylase Are Required for the Synthesis of Mycolic Acids and Essential for Mycobacterial Growth Identification of The Carboxylation Product and Determination of The Acyl-Coa Carboxylase Components. *Journal of Biological Chemistry* **280**, 8862–8874 (2005).
79. Varki A *et al*. editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; Chapter 17, Glycosyltransferases. Available from, https://www.ncbi.nlm.nih.gov/books/NBK20718/ (1999).
80. El-Battari, A. *et al*. Different glycosyltransferases are differentially processed for secretion, dimerization, and autoglycosylation. *Glycobiology* **13**, 941–953, https://doi.org/10.1093/glycob/cwg117 (2003).
81. Muhlenhoff, M., Eckhardt, M., Bethe, A., Frosch, M. & Gerardy-Schahn, R. Autocatalytic polysialylation of polysialyltransferase-1. *Embo j* **15**, 6943–6950 (1996).
82. Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867, https://doi.org/10.1016/j.cell.2006.08.019 (2006).
83. Zhou, P., Long, Q., Zhou, Y., Wang, H. & Xie, J. Mycobacterium tuberculosis two-component systems and implications in novel vaccines and drugs. *Critical reviews in eukaryotic gene expression* **22**, 37–52 (2012).
84. Domenech, P. *et al*. Unique Regulation of the DosR Regulon in the Beijing Lineage of Mycobacterium tuberculosis. *J Bacteriol* **199**, https://doi.org/10.1128/jb.00696-16 (2017).
85. Converse, P. J. *et al*. Role of the dosR-dosS two-component regulatory system in Mycobacterium tuberculosis virulence in three animal models. *Infection and immunity* **77**, 1230–1237 (2009).
86. Feltcher, M. E., Sullivan, J. T. & Braunstein, M. Protein export systems of Mycobacterium tuberculosis: novel targets for drug development? *Future microbiology* **5**, 1581–1597 (2010).
87. Danilchanka, O. *et al*. An outer membrane channel protein of Mycobacterium tuberculosis with exotoxin activity. *Proc Natl Acad Sci USA* **111**, 6750–6755, https://doi.org/10.1073/pnas.1400136111 (2014).
88. Birhanu, A. *et al*. Nε-and O-Acetylation in Mycobacterium tuberculosis Lineage 7 and Lineage 4 strains: Proteins Involved in Bioenergetics, Virulence and Antimicrobial Resistance are Acetylated. *Journal of proteome research* (2017).
89. Schwenk, S., Moores, A., Nobeli, I., McHugh, T. D. & Arnvig, K. B. Cell-wall synthesis and ribosome maturation are co-regulated by an RNA switch in Mycobacterium tuberculosis. *bioRxiv*, 232314 (2017).
90. Hanna, E. S. *et al*. Evidence for glycosylation on a DNA-binding protein of Salmonella enterica. *Microbial cell factories* **6**, 11 (2007).
91. Gomez, A. *et al*. Colorectal cancer-specific cytochrome P450 2W1: intracellular localization, glycosylation, and catalytic activity. *Molecular pharmacology* **78**, 1004–1011 (2010).
92. Lamb, D. C. *et al*. The first virally encoded cytochrome p450. *Journal of virology* **83**, 8266–8269 (2009).
93. Bartels, K.-M. *et al*. Glycosylation is required for outer membrane localization of the lectin LecB in Pseudomonas aeruginosa. *Journal of bacteriology* **193**, 1107–1113 (2011).

94. Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A. & Deisenhofer, J. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. *Structure* **3**, 41–62 (1995).
95. McKinney, J. D. *et al*. Persistence of Mycobacterium tuberculosis in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735–738, https://doi.org/10.1038/35021074 (2000).
96. Comas, I. *et al*. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature genetics* **45**, 1176 (2013).
97. Raman, K., Yeturu, K. & Chandra, N. targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis. *BMC systems biology* **2**, 109, https://doi.org/10.1186/1752-0509-2-109 (2008).
98. Chakhaiyar, P. *et al*. Regions of high antigenicity within the hypothetical PPE major polymorphic tandem repeat open-reading frame, Rv2608, show a differential humoral response and a low T cell response in various categories of patients with tuberculosis. *J Infect Dis* **190**, 1237–1244, https://doi.org/10.1086/423938 (2004).
99. Owens, R. M. *et al*. M. tuberculosis Rv2252 encodes a diacylglycerol kinase involved in the biosynthesis of phosphatidylinositol mannosides (PIMs). *Mol Microbiol* **60**, 1152–1163, https://doi.org/10.1111/j.1365-2958.2006.05174.x (2006).
100. Ojha, A. *et al*. GroEL1: a dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. *Cell* **123**, 861–873 (2005).
101. Amir, A. *et al*. Mycobacterium tuberculosis H37Rv: In Silico Drug Targets Identification by Metabolic PathwaysAnalysis. *International journal of evolutionary biology* **2014**, 284170, https://doi.org/10.1155/2014/284170 (2014).
102. Forrellad, M. A. *et al*. Virulence factors of the Mycobacterium tuberculosis complex. *Virulence* **4**, 3–66 (2013).
103. Facciuolo, A. & Mutharia, L. M. Mycobacterial glycoproteins: a novel subset of vaccine candidates. *Frontiers in cellular and infection microbiology* **4**, 133 (2014).
104. Nebenzahl-Guimaraes, H. *et al*. Genomic characterization of Mycobacterium tuberculosis lineage 7 and a proposed name: 'Aethiops vetus'. *Microb Genom* **2**, e000063, https://doi.org/10.1099/mgen.0.000063 (2016).
105. Yimer, S. A. *et al*. Comparative Proteomic Analysis of Mycobacterium tuberculosis Lineage 7 and Lineage 4 Strains Reveals Differentially Abundant Proteins Linked to Slow Growth and Virulence. *Front Microbiol* **8**, 795, https://doi.org/10.3389/fmicb.2017.00795 (2017).
106. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**, 2856–2860, https://doi.org/10.1038/nprot.2006.468 (2006).
107. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372, https://doi.org/10.1038/nbt.1511 (2008).

Acknowledgements

We thank the patients for consenting to participate in the study and selected health care facilities in the Amhara Region, Ethiopia, for facilitating the study. We are grateful to the Armauer Hansen Research Institute (AHRI), Addis Ababa, Ethiopia for facilitating the transfer of lineages 3, 4 and 7 strains to Oslo University Hospital. Funding was received from the Research Council of Norway (RCN) FRIMEDBIO project # 204747, NORBRAIN #197467 and GLOBVAC #234506 project to T.T. and #192468 to C.H.H., and from the Norwegian South-Eastern Health Authority project #2013080 to S.A.Y., G.N. and T.T.

Author Contributions

A.G.B. and T.T. conceived the study and study design. S.A.Y. and E.D.Z. performed specimen handling and cultivation. S.A.Y. collected the Mtb lineage 3, 4 and 7 strains. S.K. and A.G.B. performed the MS sample preparation. A.G.B. performed bioinformatics and statistical analyses. T.R. performed MS analysis. A.G.B. and T.T. evaluated and interpreted the data and drafted the paper. All authors edited and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-39654-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.