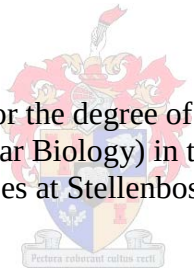
The evolution of the *Mycobacterium tuberculosis* proteome in response to the development of drug resistance

by
Suereta Fortuin



Dissertation presented for the degree of Doctor of Philosophy in Medical Sciences (Molecular Biology) in the Faculty of Medicine and Health Sciences at Stellenbosch University

Supervisor: Prof. Robin M Warren
Prof Nicolaas C gey van Pittius
Prof Haraald G Wiker
Dr. Gustavo A de Souza

March 2013

Declaration

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2013

SUMMARY

This study is the first of its kind to highlight the importance of using the latest state of the art technology available in the field of proteomics as a complementary tool to characterize the proteome of members of the *Mycobacterium tuberculosis* Beijing lineage which have been linked to outbreaks and drug resistance of Tuberculosis (TB).

Our label-free comparative analysis of two closely related *M. tuberculosis* strains with different transmission patterns and levels of virulence highlighted numerous factors that may alter metabolic pathways leading to hyper-virulence whereby the strain was able to rapidly replicate in the host and cause extensive disease. This comparative analysis clearly demonstrated that both instrumentation and analysis software impacts on the number of proteins identified and thereby the interpretation of the proteomic data. These proteomes also served as substrates for the discovery of phosphorylation sites, a field of research that reflects a significant knowledge gap in the field of *M. tuberculosis*. By using differential separation techniques in combination with the state of the art mass spectrometry we described the phosphorylation sites on 286 proteins. This was the first study to document phosphorylation of tyrosine residues in *M. tuberculosis*. By this means, our data set further extend and complement previous knowledge regarding phosphorylated peptides and phosphorylation sites in *M. tuberculosis.*

Using advanced mass spectrometry methods we further investigated the impact of the *in vivo* evolution of rifampicin resistance on the proteome of a rifampicin-resistant strain containing a S531L *rpoB* mutation. We identified the presence of over-abundant proteins which could provide novel insight into potential compensatory

mechanisms that the bacillus uses to reduce susceptibility to anti-TB drugs. Our findings suggest that proteins involved in a stress response may relate to an altered physiology enabling the pathogen to tolerate and persist when exposed to anti-TB drugs. Together this suggests that structural changes in the RNA polymerase precipitated a cascade of events leading to alterations of metabolic pathways. In addition, we present the first comprehensive analysis of the effect of rifampicin on the proteome of a rifampicin resistant *M. tuberculosis* isolate suggesting that rifampicin continues to influence the biology of *M. tuberculosis* despite the presence of an *rpoB* mutation. Our analysis showed alterations in the cell envelope composition and allowing the bacterium to survive in a metabolically dormant/persistent growth state.

The results presented in this study illustrate the full potential of using a proteomic approach as a complementary molecular technique to select promising candidate molecules and genes for further characterization using the tools of molecular biology.

OPSOMMING

Die huidige studie is ‘n eerste van sy soort, deur die nuutste gevorderde tegnologie in die proteomika veld te gebruik. Die proteoom van lede van die *Mycobacterium tuberculosis* Beijing stam, wat die oorsaak is van tuberkulose (TB) uitbrake en ook weerstandige TB, is gekarakteriseer.

Ons merkervrye vergelykende analise van twee naby verwante *M. tuberculosis* stamme met verskillende vlakke van oordraagbaarheid en virulensie, beklemtoon verskeie faktore wat metaboliese paaie mag verander, wat kan ly tot hiper-virulensie, wat die TB-stam in staat stel om vinniger te repliseer in die gasheer en ‘n uitgebreide siektetoestand kan veroorsaak. Die analise het duidelik gewys dat die toerusting wat gebruik word, sowel as die sagteware ‘n invloed kan hê op die hoeveelheid proteïne wat geïdentifiseer kan word en daardeur intrepretasie van proteomika data kan beïnvloed. Hierdie proteome dien as substrate vir die ondekking van fosforilasie setels, ‘n veld van navorsing wat dui op ‘n gaping in ons kennis van *M. tuberculosis*. Deur gebruik te maak van differensiële skeidingstegnieke en moderne spektrometrie beskryf ons fosforileringsetels in 286 proteine. Hierdie is die eerste studie wat fosforilasie van tirosien residue in *M. tuberculosis* beskryf. Hierdeur komplimenteer en brei ons data die huidige kennis oor gefosforileerde peptiede en fosforilasie setels in *M. tuberculosis* uit.

Deur gebruik te maak van gevorderde massa spektrometriese tegnieke het ons verder ook die impak van *in vivo* evolusie van rifampicin weerstandigheid op die proteoom van ‘n rifampicin weerstandige TB-stam met die algemene S531L *rpoB* mutasie ondersoek. Ons het proteïne geïdentifiseer wat in groot hoeveelhede voorkom en kan

nuwe insigte gee tot potensiele kompenserende meganismes wat deur die bacillus gebruik word om vatbaarheid vir anti-TB middels te verminder. Ons bevindings dui daarop dat proteïene betrokke in ‘n stresreaksie mag lei tot ‘n verandering in fisologie wat die patogeen in staat stel om anti-TB middels te verdra en te volhard in die teenwoordigheid van sulke middels. Saam impliseer dit dat ‘n ketting van gebeure wat lei tot veranderinge in metaboliese paaie, word vooraf gegaan deur strukturele veranderinge in die RNS polimerase. Tesame hiermee bied ons ook die eerste omvattende analise aan van die effek wat rifampicin op die proteoom van ‘n rifampicin weerstandige *M. tuberculosis* isolaat het, en wat aan die hand doen dat rifampicin voordurend die biologie van *M. tuberculosis* beïnvloed, ten spyte van die teenwoordigheid van ‘n *rpoB* mutasie. Ons analise dui op veranderinge in die samestelling van die selomhulsel wat die bakterie toelaat om te oorleef in ‘n metabolies dormante staat.

Die resultate wat in hierdie studie aangebied word illustreer die volle potensiaal van ‘n proteomiese benadering as komplementêre molekulêre tegniek om belowende kandidaat molekules en gene te kies vir verdere karakterisering, deur gebruik te maak van molekulêre tegnieke.

LIST OF AWARDS, PRESENTATION AND PUBLICATIONS

Awards

2012

April 2012- March 2013: Medical Research Council, South Africa, PhD Internship

2011

April- Sept 2011: Columbia University- Southern African Fogarty AITRP Traineeship
April 2011-March 2012: Medical Research Council South Africa, PhD Internship
April 2011: Welcome Trust-funded SACORE PhD Scholarship

2010

Jan 2010 – June 2010: Additional funding under South Africa-Norway Programme on research cooperation (PHASEII)
Oct 2007 – Sept 2010: South Africa-Norway Programme on research Co-operation

2009

March 2009: Europe-Africa Frontier Research Conference Series on Infectious Diseases Grant
Aug 2009: Best Poster Presentation, Infectious disease Session, Academic Year Day, Stellenbosch University
Sept 2009: HUPO Young investigator award (Conference attendance scholarship)

Conferences

2012

SACORE Annual General Meeting, Gaborone, Botswana, 2012
Oral presentation: The evolution of the *Mycobacterium tuberculosis* proteome

Stellenbosch University, 56th Academic Year Day, Faculty of Health Sciences, 2012
Oral presentation: The proteome and phosphoproteome of a hypo- and hyper-virulent clinical *Mycobacterium tuberculosis* Beijing strains

Keystone Symposia (Proteomics, Interactomes), Stockholm, Sweden
Poster Presentation: A phosphoproteomic approach to characterize mechanisms of virulence in clinical *M. tuberculosis* Beijing strains

2010

Analitika 2010 Conference, Stellenbosch, South Africa
Oral Presentation: Using a label-free proteomic method to identify differentially abundant proteins in closely related hypo-and hyper-virulent clinical mycobacterium tuberculosis Beijing isolates

Norway programme on research cooperation. Concluding conference. Pretoria, South Africa
Oral Presentation: Collaborative Research between South Africa, Norway and the rest of the world

ESF 4th conference on Functional Genomics and Disease, Dresden, Germany
Poster Presentation: A label-free method identified differentially abundant proteins in related *M.tuberculosis* Beijing strains

2009

ESF Infectious disease conference 2009, Cape Town, South Africa
Poster presentation: The identification and characterization of phospho-proteins in *Mycobacterium tuberculosis*

Stellenbosch University, 53rd Academic Year Day, Faculty of Health Sciences, 2009
Poster Presentation: A label-free method identified differentially abundant Proteins in Related M. tuberculosis Beijing strains (**Best Poster Presentation**: Infectious diseases)

SA Medical Research Council, Research Day, 2009
Poster Presentation: A label-free method identified differentially abundant Proteins in Related M. tuberculosis Beijing strains

HUPO 2009 Toronto, Canada (Young Investigator Award)
Poster Presentation: A label-free method identified differentially abundant Proteins in Related *M. tuberculosis* Beijing strains

2008

Stellenbosch University, 52nd Academic year day, Faculty of Health, 2008
Poster presentation: Identification and characterization of Phospho-proteins in *Mycobacterium tuberculosis*

MRC research day 2008
Poster Presentation: Identification and characterization of Phospho-proteins in *Mycobacterium tuberculosis*

Published manuscripts

1. Using a label-free proteomic method to identify differentially abundant proteins in closely related hypo- and hyper-virulent clinical *Mycobacterium tuberculosis* Beijing isolates. Gustavo A. de Souza, **Suereta Fortuin**, Diana Aguilar, Rogelio Hernandez Pando, Christopher R. E. McEvoy, Paul D. van Helden, Christian J. Koehler, Bernd Thiede, Robin M. Warren and Harald G. Wiker. ***Mol and Cell Proteomics***, 2010 Nov; 9(11):2414-23.

2. Proteogenomic analysis of polymorphisms and gene annotation divergences in prokaryotes using a clustered mass spectrometry-friendly database. GA de Souza, MØ Arntzen, **S Fortuin**, AC Schürch, H Målen, CRE McEvoy, D van Soolingen, B Thiede, RM. Warren, HG Wiker. ***Mol and Cell Proteomics***, 2011 Jan;10(1):M110.002527.

3. Multiplexed Activity-Based Protein Profiling of the Human Pathogen *Aspergillus fumigatus* Reveals Large Functional Changes Upon Exposure to Human Serum. Wiedner SD; Burnum KE; Pederson LM; Anderson LN; **Fortuin S**; Chauvigne'-Hines LM; Shukla AK; Ansong C; Panisko EA; Smith RD; Wright AT. ***J Bio Chem.,*** 2012 Sep 28;287(40):33447-59.

4. Novel and Widespread Adenosine Nucleotide-Binding in Mycobacterium tuberculosis Ansong CK, Payne SH, Haft DH, Corrie O, , Chauvigné-Hines LM; Purvine SO; Shukla AK; **Fortuin S**, Smith RD, Adkins JN, Grundner C; Wright AT ***J Chem Biol***., 2013 Jan 24;20(1):123-33.

ACKNOWLEDGEMENTS

My appreciation to the following people, organizations and intuitions for their contribution and support during this study:

- Lord God Almighty, creator of the Universe for providing me with Godly wisdom and using me in his service to do his will.

- My sincere appreciation for the opportunity to work together with Prof. Robin Warren (promoter), Prof. Harald Wiker (co-promoter), Prof. Nicolaas Gey van Pittius (co-promoter) and Dr Gustavo de Souza (co-promoter) on this unique and interesting project. The success of the project would not have been possible if it was not for your patience, guidance suggestions and support.

- My mother Rose- and (late) father Jakobus Fortuin, for instilling me the value of education, for your unconditional love, prayers and constant motivation.

- My brothers Whesly, Winston, Irwin, Franklin and only sister Rozanne for your patience, motivation and support.

- Prof. Herald Wiker for hosting me at the Gades Institute, Bergen University, Bergen, Norway.

- Dr. Gustavo de Souza and Dr. Gisele Tomazella for your hospitality, availability, support and patience in whenever I asked for your help, Oslo University, Oslo, Norway.

- Prof. Matthias Mann and Dr. Nagarjuna Nagaraj for hosting me at the Max Plank institute and training in phosphoproteomics, Martinsried, Munich, Germany.

- Dr. Ljiljana Pasa-Tolic, for hosting me at the Pacific Northwest laboratories. Dr. Aaron Wright for making provision in your lab to be trained in proteomics related techniques and Mr. Sam Purvine training in proteomics data management and analysis. Battelle, Richland, WA, USA.

- Financial support: The National Research Foundation (RSA), Norwegian Research Council (Norway), National Institute of Health –Forgarty (USA),

Medical Research Council (RSA), Southern Africa Consortium for Research Excellence-Welcome Trust (SACORE) (United Kingdom), Kwazulu-Natal Research Institute for Tuberculosis and HIV (K-RITH) (USA)

- The division of Molecular Biology and Human Genetics; Head of the department, Prof. Paul van Helden and colleagues for their support and words of encouragement.

- To amazing scientific discussions with Gail Louw, Lynthia Paul, Lizma Streicher, Dominique Andersen and Monique Williams. I hope that we will continue to have these discussions.

- My friends Natalie Bruiners, Maragretha de Vos, Laurianne Loebenberg and Danielle Stanley-Josephs for their constant support and words of motivation.

- Special and long term friends around the world for their support and encouragement via emails, post cards, letters and phone calls; Shabana Bharoocha (Singapore), Christern family (Rotterdam, The Netherlands), Henk and Edith van Oostrum (Amsterdam, The Netherlands), Maider Marin (Mayagüez, Puerto Rico).

Philippians 4: 13

"For I can do everything through Jahweh-Elohim, who gives me strength"

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

°C	:	Degree Celsius
µl	:	microliters
ABC	:	ATP binding cassette
ADC	:	Albumin dextrose catalase
Ambic	:	Ammonium bicarbonate
Amp	:	Ampere
ACN	:	Acetonitrile
DC	:	Dextrose catalase
ATP	:	Adenosine triphosphate
bp	:	base pairs
BSA	:	Bovine serum albumin
cDNA	:	Complementary DNA
CFUs	:	Colony forming units
dH_2O	:	Distilled water
ddH_2O	:	Double distilled water
DNA	:	Deoxyribonucleic acid
dNTP	:	Deoxyribonecleotide triphosphate
DTT	:	Dithiolthreitol
FA	:	Formic acid
FDR	:	False discovery rate
g	:	Grams
GITC	:	Guanidine-thiocyanate
HCl	:	Hydrochloric acid
HCD	:	Higher-energy collisional dissociation
HIV	:	Human immunodeficiency virus

hr(s)	:	Hour (s)
IAA	;	Iodoacetamide
IS	:	Insertion sequence
LAM	:	Latin-American and Mediterranean
LJ	:	Løwenstein-Jensen
LTQ	:	Linear Trap Quadrupole
MeOH	:	Methanol
M. tuberculosis	:	*Mycobacterium tuberculosis*
MDR	:	Multi Drug Resistant
MFS	:	The Major Facilitator Super family
MGIT	:	Mycobacterial Growth Indicator Tube
MIC	:	Minimum Inhibitory Concentration
µl	:	microliter
µM	:	micro molar
mg	:	milligram
ml	:	milliliters
mM	:	milli Molar
mRNA	:	Messenger RNA
m.s	:	meter per second
MS	:	Mass spectrometry
ng	:	nano grams
OADC	:	Oleic Albumin Dextrose Catalase
OD	:	Optical density
PCR	:	Polymerase chain reaction
PDIM	:	Phthiocerol dimycocerosate
PGL	:	phenoglycolipids
QRT-PCR	:	Quantitative REAL-TIME PCR
RFLP	:	Restriction Fragment Length Polymorphism

RIF	:	Rifampicin
rpm	:	Revolutions per minute
RNA	:	Ribonucleic acid
RRDR	:	RIF Resistance Determining Region
rRNA	:	Ribosomal RNA
SA	:	South Africa
SCX	:	Strong Cation Exchange
SDS	:	Sodium dodecyl sulphate
SMR	:	Small Multidrug Resistance
SNP	:	Single nucleotide polymorphism
TB	:	Tuberculosis
TDR	:	Total drug resistant
TFA	:	Trifluoroacetic acid
Tm	:	Melting temperature
U	:	Units
V	:	Volt
WCL	:	Whole cell lysate
XDR	:	Extreme Drug Resistant
ZN	:	Ziehl-Neelsen

CHAPTER 1

INTRODUCTION

Background

The pathogen *Mycobacterium tuberculosis* (*M. tuberculosis*), the causative agent of tuberculosis (TB), remains a global public health concern. According to the World Health Organisation (WHO) it is estimated that approximately 2 million people die from TB per year, while 9 million new cases develop ("WHO | Global tuberculosis control 2011," 2012). This places an enormous burden on the South African National TB Control Programme since they are required to diagnose an estimated 540,000 cases of TB every year. This is further constrained by the fact that 70 % of the TB treatment budget is allocated to the treatment of multi-drug resistance TB (MDR).

A major landmark in the history of TB research was the deciphering of the *M. tuberculosis* genome sequence which has provided a blue print for research into the physiology of this pathogen with the goal of developing new drugs, diagnostics and vaccines (Cole *et al*., 1998; Starck *et al*., 2004). Molecular epidemiological studies have shown that the current epidemic is driven by numerous different genotypes and that most TB patients develop disease through recent infection (transmission) (Van der Spuy *et al*., 2009). Furthermore, patients are at risk of reinfection either before or after developing disease, suggesting an extremely high infection pressure in high burden settings. The frequency at which the different *M. tuberculosis* genotypes appear in different settings suggests the evolution of different pathogenic characteristics (Van der Spuy et al., 2009). Recently, animal studies have shown that certain sub-types of the *M. tuberculosis* Beijing strains are more virulent than others (Parwati *et al*., 2010), further emphasising the idea that not all strains (or sub-strains) possess the same virulence characteristics. However, the mechanisms underlying these differences remain to be elucidated. Comparative genomics suggest that the genome of *M. tuberculosis* has evolved through single nucleotide polymorphisms

(SNP's), insertions and deletions (Filliol *et al*., 2006). Despite an ever increasing number of genomes being sequenced the mechanisms defining pathogenesis remain largely unknown. Furthermore, our knowledge of how mutations may impact on structural changes and protein function, regulatory mechanisms or Post Translational Modifications (PTM's) remains largely unknown. Studies of these changes could be crucial in predicting the function of specific proteins involved in virulence and the design of new anti-tubercular vaccines and drugs. Protein phosphorylation, one of the most studied PTM's, could provide a lead in identifying the cause of phenotypic differences of genetically similar clinical *M. tuberculosis* strains. Protein phosphorylation is involved in the regulation of numerous essential biological processes described in both prokaryotes and eukaryotes and it is speculated that perturbations in the equilibrium of bacterial kinases and phosphatases may be involved in infectious diseases. In order to understand the biology and pathogenic mechanisms of *M. tuberculosis* it is relevant to establish which proteins are phosphorylated at a whole proteome level. Characterising and identifying phosphorylated proteins is a challenging analytical task due to the low abundance and stoichiometry of phosphopeptides/proteins in complex protein samples and the presence of active phosphatases in the cells lysed during protein extractions (Oda *et al*., 2001). Thus, we have optimised a recently developed method using combining strong cation exchange (SCX) with titanium dioxide (TiO_2) to enrich for phosphorylated peptides (Wiśniewski *et al*., 2010; Nagaraj *et al*., 2012) in a complex *M. tuberculosis* protein sample.

The use of proteomics as a complementary molecular tool is not only important for defining the biology of drug sensitive TB but is also essential for understanding the persistence and successful spread of drug resistant strains. Classical dogma has incorrectly assumed that the acquisition of resistance compromised fitness and thus transmissibility (Cohen *et al*., 2003). This dogma has been challenged by molecular epidemiological studies which have clearly demonstrated that drug resistant strains are actively transmitted and that the drug resistance epidemic is driven by transmission (Cohen *et al*., 2003). *In vitro* studies analysing spontaneously generated drug resistant TB mutants have shown a direct correlation between mutations in a specific target gene and resistance to the anti-TB drug used to select the resistant mutants (Meier *et al*., 1996). Consequently, the relationship between a defined mutation and resistance to a specific anti-TB drug has become a dogma and now forms the basis for the development of molecular-based drug-resistance diagnostic assays (Ramaswamy and Musser, 1998; Johnson *et al*., 2006). However, the above studies have failed to exclude the possibility that additional modulatory mutations may occur concurrently in association with the evolution of drug-resistance. This hypothesis is supported by molecular epidemiological studies and *in vitro* velocity assays which provide evidence to suggest that compensatory mutations may ameliorate the fitness cost associated with acquisition of drug resistance (Comas *et al*., 2011). This has lead scientists to focus their research efforts on the discovery of compensatory mutations which may explain restoration of fitness and transmissibility of drug resistant *M. tuberculosis* strains (Comas *et al*., 2011).

The accumulation of mutations occurring during the evolution of drug resistance may be more complex than only restoring fitness as these mutations may be associated

with an altered physiology. This is supported by the observation that the level of Rifampicin (RIF) resistance (as measured by the Minimum Inhibitory Concentration (MIC)) was highly variable among isolates with the same genetic background and identical mutation conferring resistance to RIF (Louw *et al*., 2011). However, the genetic and/or proteomic basis of these phenotypes remain unknown. Our limited knowledge of the physiology of drug resistant TB can partly be ascribed to current dogma (one mutation one resistance phenotype), which polarised scientific thinking that drug resistant strains are physiologically identical to drug sensitive strains with the exception of the drug resistance mutations. Thus, the interplay between antibiotics/metabolic inhibitors and gene expression is limited to studies on drug sensitive strains of *M. tuberculosis*. We do not know how drug resistant strains respond to the drugs to which they are resistant. The importance of this is underscored by the fact that all patients diagnosed with TB are placed onto a standard treatment regimen until the time that either treatment failure or drug susceptibility testing suggests the need for an alternative regimen.

Over the last decade the tremendous advances in proteomics now provide the opportunity to identify proteins involved in novel pathways which is associated with virulence, transmissibility and resistance to antibiotics. By using a novel proteomic approach we showed an over-abundance of cell wall proteins as well as regulatory proteins in a highly-transmissible strain as compared to a non-transmitted strain (de Souza *et al*., 2010). Using a proteomics approach may enable the identification of compensatory mechanisms that the mycobacterium develops in response to antibiotics. Thus, if drug resistant strains demonstrate a unique physiology it is conceivable that this physiology influences the propensity of a given drug resistant

strain to develop resistance to specific additional drugs or to adapt in a manner which induces cross-resistance to subsequent treatment.

Hypothesis: The phenotype of a *M. tuberculosis* strain is defined by changes in the proteome as a result of altered gene expression which may or may not be derived from genomic mutation.

Overall aim: To determine the proteomes of different clinical *M. tuberculosis* isolates which demonstrate different phenotypes (virulence and drug resistance).

Ethical approval: Ethical approval for the above projects has been granted by the Faculty of Medicine and Health Sciences under the project titles:

1. Characterisation of the Mycobacterium tuberculosis phosphorylome and its role in pathogenesis (N06/10/204).
2. An investigation into the evolutionary history and biological characteristics of the members of the genus *Mycobacterium*, with specific focus on the different strains of *Mycobacterium tuberculosis*, other members of the *M. tuberculosis* complex and non-tuberculous mycobacteria (NTM) (92/008, 96/093, 2000/c056, 2000/C061, 2002/C118, N04/08/135).

CHAPTER 2

Literature review

My contribution to this chapter: Writing of Manuscript

Introduction

All biological and physiological processes are directly or indirectly regulated by protein phosphorylation, a signal transduction cascade which links the external- to the internal cellular environment (Kobir *et al*., 2011). This process is considered to be the most important known reversible PTM found in nature, and is involved in multiple cellular processes ranging from metabolism and homeostasis to cellular signaling (Graves and Krebs, 1999). Protein phosphorylation relies on amino acid specific protein kinases which phosphorylate the free hydroxyl group of serine, threonine and tyrosine (Cozzone, 1998). The protein thereby gains a γ-phosphoester residue at the expense of adenosine triphosphate (ATP) which acts as the phosphoryl donor. A non-phosphorylated protein can be phosphorylated at multiple sites by the same kinase in a metabolic pathway or kinases belonging to different pathways (Fiuza *et al*., 2008). The phosphorylation of a single protein at multiple sites increases the extent to which its activity can be regulated (Graves and Krebs, 1999). Phosphorylation is also coupled to the reverse reaction (dephosphorylation) which is mediated by protein phosphatases. These enzymes catalyze the reaction whereby the phosphate is removed from the phosphorylated protein (Graves and Krebs, 1999). This restoration of the dephosphorylated state, and *vice versa* (Hunter, 1995), is capable of rapid and efficient activation or inactivation of candidate proteins and their subsequent functions. This cascade represents a bidirectional and reusable molecular “on/off” switch (Cohen, 2002).

Although protein phosphorylation has been extensively investigated in eukaryotes, it was only discovered in prokaryotes in the late 1970’s (Garnak and Reeves, 1979a, 1979b; Manai and Cozzone, 1979). In prokaryotes, 9 amino acids have been found to

be phosphorylated and include; serine, threonine, tyrosine, histidine, lysine, arginine, aspartate, glutamate, and cysteine (Klumpp *et al*., 2002).

It is now well recognised that protein phosphorylation is a key process in the regulation of cell homeostasis which has been implicated in disease and pathogenesis (Tan *et al*., 2009). As such, phosphoproteomics have become a major focus in the research of infectious diseases. To date, the phosphoproteomes of 9 pathogenic bacteria (Mijakovic and Macek, 2012), including *Mycobacterium tuberculosis* (*M. tuberculosis*) H37Rv have been described and summarized in table 2.1.

Table 2.1: Published phosphorylomes of pathogenic bacteria

Pathogenic bacteria	# Proteins	# Phosphosites	Reference
Corynebacterium *glutamicum*	78	103	(Bendt *et al*., 2003)
Campylobacter jejuni	36	0	(Voisin *et al*., 2007)
Bacillus subtilis	78	78	(Macek *et al*., 2007)
Klebsiella pneumonia	81	93	(Lin *et al*., 2009)
Pseudomonas aruginosa	57	55	(Ravichandran *et al*., 2009)
Pseudomonas putida	56	53	(Ravichandran *et al*., 2009)
Streptococcus pneumonia	84	163	(Sun *et al*., 2010)
Mycoplasma pneumonia	63	16	(Schmidl *et al*., 2010)
Mycobacterium tuberculosis	301	516	(Prisic *et al*., 2010)

Mycobacterium tuberculosis

Unravelling the *M. tuberculosis* genome sequence has provided a blue print for research into the physiology of this pathogen with the goal of developing new drugs, diagnostics and vaccines (Cole *et al*., 1998; Starck *et al*., 2004). According to Tuberculist (http://tuberculist.epfl.ch/) a total of 4018 open reading frames (ORFs) were identified in the genome and these were grouped in 12 functional categories (Cole *et al*., 1998). Previous studies using monoclonal antibodies showed that

'eukaryotic-like' protein phosphorylation occurs in the pathogen, *M. tuberculosis* (Chow *et al*., 1994). Furthermore, the use of molecular techniques such as Southern Blot analysis, PCR and whole genome sequencing demonstrated that *M. tuberculosis* encodes 11 'eukaryotic-like' protein kinases (Av-Gay and Everett, 2000) as well as 11 complete two-component regulatory systems (Stock *et al*., 2000). The existence of these 'eukaryotic-like' Ser/Thr protein kinases (STPK) genes in the *M. tuberculosis* genome indicates that protein phosphorylation plays a central role in regulating various biological functions, ranging from adaptative responses to bacterial pathogenicity (Av-Gay and Everett, 2000; Ravichandran *et al*., 2009).

M. tuberculosis STPKs and their substrates

STPKs are single transmembrane receptors, which can be autophosphorylated at the serine- and threonine residues on the intracellular domains of the receptor. STPKs' are known to act as sensors of the external environment, which in turn signal to the intracellular environment allowing for regulation of developmental changes and host pathogen interactions (Tyagi and Sharma, 2004) .

In an attempt to classify the mycobacterial STPKs', a phylogenetic tree was generated by analysing the full length gene sequences of the 11 STPKs' from 6 completely sequenced mycobacterial genomes (Narayan *et al*., 2007). This phylogenetic construction grouped the STPK's into 5 clades (Table 2.2).

Table 2.2: Summary of STPK's identified in *M. tuberculosis*

Rv number	Gene name	Substrate	Function	Clade	Reference
Rv0015c	*pknA*	GroEL1, KasB, Wag31	Heat shock, Mycolic acid biosynthesis Cell divison	I	(Chaba *et al*., 2002)
Rv0014c	*pknB*	GarA,KasB, Rv0020c, Rv1422, Rv1747, Wag31	Glycogen recycling, Tricarboxylic acid cycle, Mycolic biosynthesis, FHA-containing protein, Putative ABC transporter, Cell division	I	(Av-Gay *et al*., 1999; Young *et al*., 2003)
Rv0931c	*pknD*	GarA, GroEL1, Rv1747, Rv0516c, Mmpl7	Glycogen recycling, Tricarboxylic acid cycle, Heat shock protein, Putative ABC transporter	II	(Peirs et al., 1997; Greenstein, Echols, et al., 2007)
Rv1743	*pknE*	GarA, GroEL1, KasB, Rv1747	Glycogen recycling, Tricarboxylic acid cycle, Mycolic biosynthesis, Heat shock protein, Putative ABC transporter	II	(Molle *et al*., 2008)
Rv1746	*pknF*	GarA, GroEL1, KasB, Rv0020c, Rv1747	Glycogen recycling, Tricarboxylic acid cycle, Mycolic biosynthesis, FHA-containing protein,Putative ABC transporter	III	(Koul *et al*., 2001; Molle *et al*., 2008)
Rv0410c	*pknG*	GarA	Glycogen recycling, Tricarboxylic acid cycle	V	(Koul et al., 2001; Fiuza et al., 2008)
Rv1266c	*pknH*	embR	Tricarboxylic acid cycle	II	(Sharma *et al*., 2006)
Rv2914c	*pknI*	EmbR/EmbR2	Phosphotransferase, Tricarboxylic acid cycle, Glycan biosynthesis and metabolism	III	(Narayan *et al*., 2007)
Rv2088	*pknJ*	MyBP	Tricarboxylic acid cycle, Glycan biosynthesis and metabolism, Integral membrane protein	III	(Arora *et al*., 2010)
Rv3080c	*pknK*	virS	Phosphotrasferase, Tricarboxylic acid cycle Glycan biosynthesis and metabolism	IV	(Kumar *et al*., 2009)
Rv2176	*pknL*	GroEL1, KasB Rv2175	Heat shock Protein, Mycolic acid biosynthesis	I	(Canova *et al*., 2008)

PknA, *PknB*, *PknD*, *PknE*, *PknF*, *PknH*, *PknI*, *PknJ*, *PknL* are predicted to be transmembrane receptors (Narayan *et al*., 2007) which are hypothesised to be involved in growth, stress responses and host-pathogen interactions (Grundner, Gay, *et al*., 2005). The genes *pknG* and *pknK* code for soluble protein kinases with no apparent transmembrane regions (Av-Gay and Everett, 2000).

Using transposon mutagenesis, *pknA*, *pknB* and *pknG* were shown to be genes essential for the growth of *M. tuberculosis* in culture (Sassetti and Rubin, 2003; Sassetti *et al*., 2003). *PknA*, *pknB* are *pstP* are part of an operon which encodes genes involved in cell shape control and cell wall synthesis (Av-Gay and Everett, 2000; Fernandez *et al*., 2006; Wehenkel *et al*., 2008). Both *pknA and pknB* are expressed during exponential growth, while their over expression slows growth and alters cell morphology (Kang, 2005). Expression of *pknB* and *pstP* is upregulated during exponential growth or infection (Boitel *et al*., 2003), while *pknB* is down regulated during nutrient starvation (Betts *et al*., 2000). This demonstrates their imperative influence in the survival of *M. tuberculosis* during infection. Recently it was shown that the pseudokinase and transmembrane protein MviN is a substrate for PknB and is involved in peptidoglycan synthesis in *M. tuberculosis* (Gee *et al*., 2012).

PknD alters transcription of an anti-anti-sigma factor homolog, Rv0516c which is specifically phosphorylated at a novel site, Thr2 (Greenstein, MacGurn, *et al*., 2007) upon its over expression. This residue is a conserved ser/thr phosphorylation site in the anti-anti-sigma factor family (Greenstein, MacGurn, *et al*., 2007). PknD specific activity induces a transcriptional response leading to an abnormal phosphorylation of physiological changes that is sensitive to alterations in the activity of regulatory factors other than kinases (Greenstein, Echols, *et al.*, 2007).

It has been shown that the autophosphorylated form of PknF leads to an interaction with and phosphorylation of the ABC transporter (Rv1747). The *pknF* and *Rv1747* genes are adjacently located on the *M. tuberculosis* genome to form an operon. Rv1747 has been implicated in the transport of glucose across the membrane and is required for virulent *M. tuberculosis* infection in mice (Curry *et al*., 2005). Rv1747 contain 2 of the 7 Forkhead-associated (FHA) domains encoded by the 6 FHA domain containing proteins in the *M. tuberculosis* genome. FHA domains are phosphopeptide recognition motifs spanning 80-100 amino acid residues which are folded in an 11-stranded beta sandwich (Spivey *et al*., 2011). The FHA domain mediates phosphopeptide interactions with proteins which have been phosphorylated by STPKs (Grundner, Gay, *et al*., 2005). The majority of FHA domains recognize phospho-threonine (Durocher *et al*., 2000), with specificity for residues at the C-terminal of the phospho-threonine site, particularly the +3 position. PknF binds to the phosphorylated sites at Thr-150 and Thr-208 of Rv1747 where it exhibits ATPase activity allowing the uptake of glucose by the ABC transporter (Spivey *et al*., 2011).

The physiological role of mycobacterial PknG remains unclear, however, it is hypothesised that PknG is involved in the survival of *M. tuberculosis* inside the macrophages and is a sensor for nutritional stress (Cowley *et al*., 2004). In addition, this cytosolic protein can be secreted into the macrophage providing further evidence of its important role in infection (Cowley *et al*., 2004). Autophosphorylation of PknG in the thioredoxin (Trx) domain is crucial for the survival of *M. tuberculosis* within the infected macrophage (Walburger *et al*., 2004; Scherr *et al*., 2009).

It has been shown that PknE has been implicated in intracellular survival of *M. tuberculosis* by suppressing apoptosis (Kumar and Narayanan, 2012). It is suggested that PknE could play an important role in sensing various environmental changes, in particular during infection of

the macrophages and possible activate multiple signalling processes allowing physiological adaptation in order for its survival (Kumar and Narayanan, 2012).

The PknH gene is located adjacent to the EmbR transcriptional regulator in the *M. tuberculosis* genome. Phosphorylation of EmbR enhances binding to the embCAB arabinisyltransferase promoter leading to increased transcription of these enzymes. PknH functions as a feedback regulator, altering the production of cell wall constituents in response to environmental signals. However it has been shown that the sensor domain of PknH may regulate the cell-wall production in response to environmental cues, such as compounds that are not related to the mycobacterial cell wall (Cavazos *et al*., 2012).

Based on transposon mutagenesis studies PknL was found to be non-essential for in vitro growth of both *M. tuberculosis* H37Rv and CDC1551. However subtle regulatory functions brought about by protein interactions with Rv2175c were observed (Canova *et al*., 2008).

PknK is involved in the regulation of growth whereby it regulates tRNA expression during logarithmic and stationary growth phases as a means to facilitate adaptation to changing environments (Molhotra, 2012). It has been demonstrated that PknK-mediated phosphorylation increases the affinity of the transcriptional regulator, VirS for the *mym* (mycobacterial monooxygenase) promoter and stimulates its transcription under acidic conditions (Kumar *et al*., 2009).

PknI appears to be required for balanced growth of *M. tuberculosis* inside macrophages and under the *in vitro* conditions which are used to mimic growth inside macrophages. PknI plays a pivotal role in controlling the growth of *M. tuberculosis* upon infection by slowing down the growth once inside the host macrophage. Internal signals used to activate PknI are most likely the host associated internal signals of low pH associated with limited oxygen availability (Gopalaswamy *et al*., 2009).

M. tuberculosis phosphatases and their substrates

Even though a large number of protein kinases involved in the regulatory processes of *M. tuberculosis* have been recognised, the number and variety of phosphatases has only recently been discovered (Cohen, 1997). The presence of two 'eukaryotic-like' phosphatases suggests that *M. tuberculosis* might utilise 'eukaryotic-like' signalling as alternative signalling transduction pathways (Boitel *et al*., 2003). 'Eukaryotic-like' Ser/Thr phosphatases (PstP) can be divided into two families which are categorised according to their sequence similarities, metal-ion-dependence and sensitivity to inhibitors (Barford *et al*., 1998); protein phosphatase P (PPP) (Bellinzoni *et al*., 2007) and metal-dependent protein phosphatases (PPM) (Rigden, 2011). PstP has been identified as a metalloenzyme belonging to the PPM family of Ser/Thr protein phosphatases and has been found to dephosphorylate various mycobacterial STPKs' and their substrates (Boitel *et al*., 2003; Chopra *et al*., 2003). In addition, Boitel and co-workers provided evidence that phosphorylated PknB served as a substrate for the Ser/Thr phosphatase (PstP) (Boitel *et al*., 2003).

Tyrosine Phosphorylation in *M. tuberculosis*

The protein-tyrosine kinases (PTK) and phosphatases (PTP) found in bacteria are structurally and functionally similar to those found in eukaryotes (Cozzone *et al*., 2004). The ability to reverse tyrosine phosphorylation is dependent on two putative protein-tyrosine phosphatases (PTP), PtpA and PtpB (Shi *et al*., 1998; Greenstein *et al*., 2005). According to Cole (Cole *et al*., 1998), the *M. tuberculosis* genome encodes for both protein tyrosine phosphatase A (PtpA) and protein tyrosine phosphatase B (PtpB) however, to date only one tyrosine kinases, (encoded by *Rv2232*; *ptkA*) has been identified in *M. tuberculosis* (Cowley *et al*., 2002). It has been suggested tyrosine phosphatases are secreted into the host, where they target

components of the host signalling pathway. As such, PtpA has been identified as a novel drug target and has been investigated as efficient inhibitors of PtpA (Chiaradia *et al*., 2008). Recently, Av-Gay and colleagues identified PtkA, a novel PTK that acts as substrate for mycobacterial PtpA (Bach *et al*., 2009). PtkA has been shown to phosphorylate two tyrosine residues (Tyr128 and Tyr129) located in PtpA. Furthermore, that study showed that the presence of a phosphate donor, i.e. ATP or GTP, increased the interaction between PtkA and PtpA (Bach *et al*., 2009). The function of PtpA remains unclear, although evidence suggests that it plays a role in the manipulation of the host response during *M. tuberculosis* infection (Tyagi and Sharma, 2004; Jers *et al*., 2008).

Recent studies have shown that the *M. tuberculosis* phosphatase, PtpB, is not amino acid specific but does exhibit dual specificity, implying that this enzyme can dephosphorylate proteins which have been phosphorylated on either serine or threonine residues (Greenstein *et al*., 2005). *M. tuberculosis* PtpB is routinely detected in both culture filtrates and whole cell lysates, which suggests that it is secreted into the extracellular medium (Grundner, *et al*., 2005). Interestingly, a recent study by Beresford and colleagues showed that PtpB also dephosphorylates phosphoinositides (Beresford *et al*., 2007). In the host cells phosphoinositides are involved in membrane trafficking, actin remodeling and cell survival (Grundner, *et al*., 2005). Secretion of an effector protein by the pathogen which may directly interact with these host phosphoinositides would result in the destabilization of the phosphoinositide metabolism (Silva and Tabernero, 2010). In addition, PtpB has been suggested to play an important role in the regulation of growth, development and pathogenesis of *M. tuberculosis* (Beresford *et al*., 2010). Furthermore, it was observed that the PTPs', PtpA and PtpB are secreted into the host cells and function as virulence factors (Koul *et al*., 2000, 2000; Cozzone *et al*., 2004).

Two component system in *M. tuberculosis*

Prokaryotic signal transduction generally relies on the reversible protein phosphorylation conducted by two component systems (Stock *et al*., 2000). This process of phosphorylation/dephosphorylation leads to an alteration of the proteins inherent biological activity leading to a biochemical or cellular component change (Xing *et al*., 2002) and important in the adaptive regulation of the bacteria due to environmental change or stress (Bagchi *et al*., 2005). The molecular system responsible for stimulus response in bacteria involves the so called two component system of histidine kinase sensor and associated response regulators. The two component signal transduction involves a signalling histidine kinase (HK) and an effector response regulator (RR). Here, the HK is regulated by an environmental stimuli and autophosphorylates at a histidine residue. This in effect creates a high energy phosphoryl group that is transferred to an aspartate residue in the RR. This response, in combination with the DNA-binding activity regulating gene expression (Shrivastava *et al*., 2007). Transcriptional regulatory studies of five of the mycobacterial two-component systems including devS-devR, senX3-regX3, trcR-trcS, prrA-prrB and mprA, indicates their involvement in auto-regulation where the histidine kinase phosphorylates the conserved histidine residue in the kinase domain (Haydel *et al*., 1999; Himpens *et al*., 2000; Ewann *et al*., 2004; He and Zahrt, 2005). The involvement of the remaining 7 two-component systems in pathogenesis and virulence remain to be investigated. To date only one of the 11 two-component systems MtrA/MtrB in *M. tuberculosis* was shown to be essential for growth (Zahrt and Deretic, 2000).

Table 2.3: Two-component systems of *M. tuberculosis* identified and characterised

Rv numbers	Two-component systems	References
Rv3245c/Rv3246c	MtrB-MtrA	(Zahrt and Deretic, 2000)
Rv1032c/Rv1033c	TrcS/TrcR	(Haydel *et al*., 1999)
Rv3132c/Rv3133c	DevS-DevR	(Saini *et al*., 2004)
Rv0902c/Rv0903c	PrrB/PrrA	(Ewann *et al*., 2004)
Rv0757/Rv0758	PhoP/PhoR	(Zahrt and Deretic, 2000)
Rv0844c/Rv0845c	NarL/NarS	(Tyagi and Sharma, 2004)
Rv0981/Rv0982	MprA/MprB	(Zahrt *et al*., 2003)
Rv1027c/Rv1028c	KdpE/KdpD	(Parish, Smith, Roberts, *et al*., 2003)
Rv3764c/Rv3765c	TcrY/TcrX	(Parish, Smith, Roberts, *et al*., 2003)
Rv1626/Rv3220c	PdtaR/PdtaS	(Morth *et al*., 2005)
Rv0490/Rv0491	SenX3/RegX3	(Himpens *et al*., 2000)

The PdtaR/PdtaS two-component system was first described by Morth *et al*. (Morth *et al*., 2005). PdtaR acts as a transcriptional anti-termination rather than a transcriptional initiator while the PdtaS is localised in the cytosol and is constitutively active *in vitro* (Morth *et al*., 2005). PdtaS contains sensing modules; cGMP-regulated cyclic nucleotide phosphodiesterases, adenylyl cyclases and the bacterial transcription factor FhlA (GAF) and a Per-ARNT-SiM (Drosophila Period,-Arylhydrocarbon Receptor Nuclear Transport and Drosophila Single-Minded (PAS) domains). Preu and colleagues demonstrated that the PdtaR/PdtaS two-component system is structurally similar to and has a similar function as the

EutW/EutV two-component system that is involved in the regulation of ethanolamine catabolism in gram positive-bacteria (Preu *et al*., 2012).

The two-component system MtrA/MtrB, was initially investigated in *C. glutamicum* (Möker *et al*., 2004). In *M. tuberculosis* MtrA/MtrB is involved in osmoregulation and the control of cell wall metabolism (Bott and Brocker, 2012) and has been shown to be essential (Zahrt and Deretic, 2000). The response regulator MtrA has been identified in the sera of TB patients indicating the involvement of MtrB/MtrA two-component system in pathogenesis (Singh 2001). The PhoP/PhoR two-component system has been implicated in virulence based on the observation that the *PhoP* gene is required during intracellular growth but not essential for persistence of the bacilli in the macrophages (Pérez *et al*., 2001).

The DevR-DevS, two-component systems in *M. tuberculosis* has been characterised and described in detail (Dasgupta *et al*., 2000). The DevR-DevS forms part of the DosR-regulon which is activated by hypoxic conditions and regulates expression 47 genes, induced in *M. tuberculosis* during growth under these conditions (Park, Lee, *et al*., 2003). This two-component system is found in BCG and *M. smegmatis* where it is required for long term survival during hypoxia, oxygen deprivation, heat stress, tolerance, respectively (O'Toole *et al*., 2003).

The *M. tuberculosis* PrrA-PrrB two-component system is expressed during intracellular growth and is suggested to be important in establishing infection in human macrophages (Haydel *et al*., 1999; Ewann *et al*., 2002). It has recently been shown that transcription of PrrA-PrrB was induced during nitrogen starvation and plays an important in mycobacterial growth and metabolism (Haydel *et al*., 2012). The *M. tuberculosis* SenX3-RegX3 two-component system has been implicated in virulence and phosphate-responsive control of gene expression and further alters the expression of aerobic respiration (Roberts *et al*., 2011).

M. tuberculosis phosphoproteome

To date 516 phosphorylation events which mapped to 301 proteins have been reported in *M. tuberculosis* (Prisic *et al*., 2010). The authors identified a wide range of serine/threonine phosphorylated proteins as well as previously described phosphorylated proteins including, 4 STPKs, GarA, Rv1422 and FHA thereby demonstrating the influence of the STPKs in the regulation of various cellular processes and their impact on the regulatory mechanisms, which may likely be involved in *M. tuberculosis* virulence (Prisic *et al*., 2010). They further identified a dominant motif shared by 6 *M. tuberculosis* STPKs using *motif-x* (Schwartz and Gygi, 2005). A bioinformatics approach was used to identify 215 specific phospho-acceptor residues within each phosphopeptide. The dominant motifs were validated in each phosphopeptide and lead to the identification of the +3 position relative to the phospho-threonine residue preferred by each kinase (Prisic *et al*., 2010). According to Tuberculist (Lew *et al*., 2011), the functional categorization of the phosphorylated proteins identified in the H37Rv phosphorylome analysis showed that most of the proteins were involved in cell wall and cell processes as well as intermediary metabolism and respiration (Table 2.4).

Table 2.4: Functional categorization of phosphorylated proteins identified in H37Rv phosphorylome (Prisic *et al*., 2010)

Functional group number *	Functional categories*	Number of phosphorylated proteins
F0	Virulence, detoxification and adaptation	10
F1	Lipid metabolism	22
F2	Information pathways	25
F3	Cell wall and cell processes	68
F4	Stable RNAs	0
F5	Insertion sequences and phages	4
F6	PE/PPE families	15
F7	Intermediary metabolism and respiration	72
F8	Unknown	0
F9	Regulatory proteins	20
F10	Conserved hypothetical	65
	Total number of phosphorylated proteins identified	**301**

*The functional groups from Tuberculist database, http://tuberculist.epfl.ch

6. Conclusion

The identification and characterisation of putative kinase and phosphatase genes in the *M. tuberculosis* genome suggests their participation in key biological processes such as gene expression and cell growth. The identification and characterisation of more substrates and targets of protein kinases and phosphatases in *M. tuberculosis* will help with understanding the pathogenesis and transmission of this disease causing bacteria. Phosphoproteomics can now be used to investigate phosphorylation as a potential avenue to alter bacterial growth and to discover novel proteins for potential therapeutic targets. Variation in the phosphorylation pattern may provide novel insights into regulatory pathways during various *M. tuberculosis* growth stages. Such information could enable the identification of pathways regulating the latent infection as well as characterizing novel drug targets. Unique phosphorylation patterns for each strain or sub-type may reflect strain specific mechanisms regulating pathogenicity and virulence. However, it must be acknowledged that the description of the phosphoproteome is largely limited to a single publication and it is expected that many more phosphorylated proteins are still to be discovered in *M. tuberculosis*.

CHAPTER 3

METHODOLOGY

The proteomes of different *M. tuberculosis* strains from and epidemiological sample bank demonstating varying levels of virulence and drug resistance were analysed using the latest state of the art proteomics and mass spectrometry tools. An ouline of the experimental procedures is given in figure 3.1.

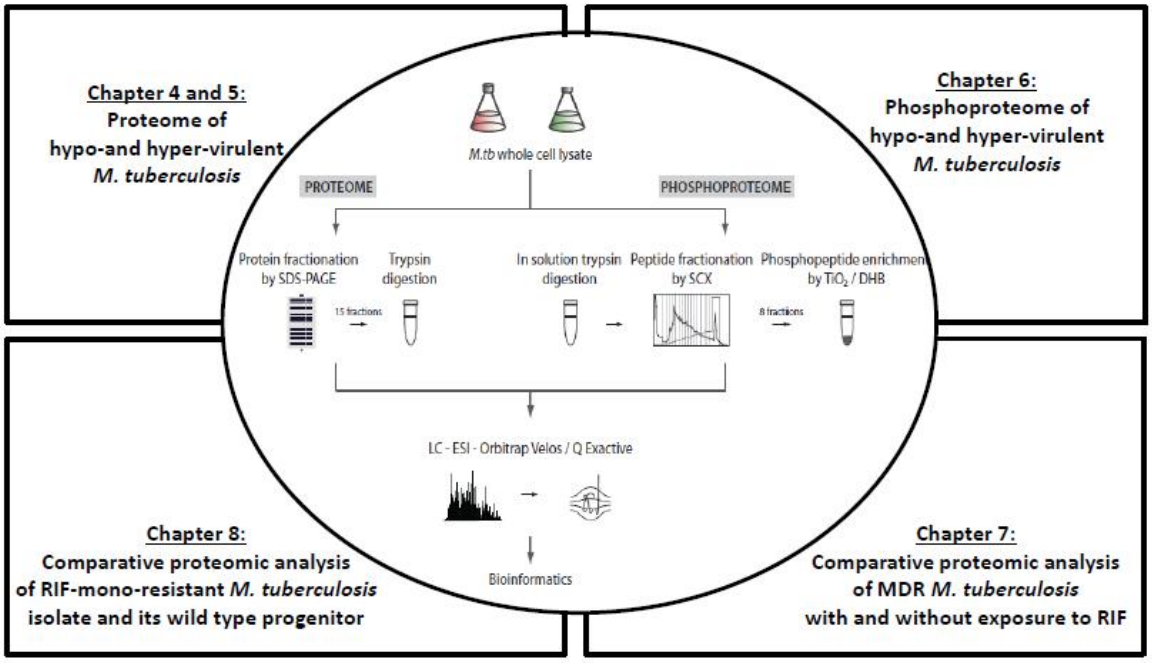


Figure 3.1: Overview of experimental study design

3.1: Genotypic classification

M. tuberculosis isolates were cultured from TB patients attending primary health care clinics in an epidemiological field site in the Western Cape, South Africa over a 12 year period (figure 3.2A). Positive *M. tuberculosis* cultures were genotyped by IS*6110* DNA fingerprinting and spoligotyping according to international standards (van Embden *et al*., 1993, 1993). The clinical isolates selected, (susceptible, rifampicin RIF-mono-resistant, and MDR) and analysed in this thesis are from the genetically distinct Beijing lineage. The presence of contaminants in *M. tuberculosis* cultures were continuously monitored with plating the culture onto a blood agar for two days (*M. tuberculosis* does not grow on blood agar within 2 days). Ziehl-Neelsen gram staining of *M. tuberculosis* smears were heat fixed (heated for 2 hours at 100 °C), stained with carbol-fuchsin (Becton, Dickinson and Company, Maryland, United States of America (USA)) and decolorized with acid alcohol for positive identification of *M. tuberculosis*. The smears were then counterstained with methylene blue (Becton, Dickson and Company, Maryland, USA) and read under the light microscope for acid-fast bacilli. *M. tuberculosis*, an acid-fast bacterium, will retain dyes when heated and treated with acidified organic compounds. Therefore the bacilli will appear pink in a contrasting background when the ZN test is done.

3.1.1. Clinical hypo- and hyper-virulent *M. tuberculosis* strains

Genotyping of *M. tuberculosis* (figure 3.2B) from our epidemiological reference bank strain collection identified 452 isolates with the Beijing genotype, of which 319 were classified as members of sub-lineage 7 and were characterised by the regions of difference (RD) 150 deletion (Hanekom *et al*., 2007). Within this sub-lineage, 288 cases were clustered into 21 transmission chains (cluster size ranged from 2 to 147), and 31 cases had unique IS6*110* DNA fingerprints. Transmission chains were defined as a series of cases having isolates with

identical IS*6110* DNA fingerprints (Van Embden *et al*., 1993) with inter-case intervals of less than 2 years and each transmission chain was assumed to be initiated by a single index case (Van der Spuy *et al*., 2009). A transmission chain unique case was defined as one having no other cases with the identical strain occurring within 2 years (Van der Spuy *et al*., 2009).

Analysis of the population structure of *M. tuberculosis* over time revealed that the Beijing genotype was increasing exponentially relative to the other dominant genotypes (figure 3.1A) (Van der Spuy *et al*., 2009). The Beijing genotype has been associated with numerous multi-drug resistant outbreaks around the world (Lan *et al*., 2003; Cox *et al*., 2005). To determine whether an actively transmitting strain showed different pathogenic properties to that of a non-transmitting strain, one isolate representative of the largest cluster (n = 147 cases) and one unique isolate were randomly selected for further analysis. The IS*6110* DNA fingerprints of these two strains with different transmission chains showed no significant variation (figure. 3.2B).

The virulence (as determined by survival, lung pathology and bacterial load) of each selected isolate was evaluated in 6 to 8 week old male BALB/c mice, in collaboration with Prof R Hernandez-Pando, Mexico. Briefly, bacteria were grown in Middlebrook 7H9 broth (Difco, Detroit, MI USA) enriched with glycerol and albumin, catalase and dextrose (Becton Dickinson, Cockysville MD, USA), and incubated with constant agitation at 37°C and 5% CO_2 for 21 days. Growth was monitored by densitometry. As soon as the culture reached stationary phase (OD_{600} =1) the bacilli were harvested, and the concentration was adjusted to 2.5 x 10^5 viable bacilli per 100µl of phosphate buffered saline (PBS) as determined by diacetate fluorescein incorporation, and 100 µl aliquots were frozen at -70°C until use. To induce progressive pulmonary TB, mice were anaesthetized with sevoflurane and inoculated

intratracheally with 2.5 x 10^5 bacilli in 100 µl PBS. Infected mice were kept in a vertical position until the effect of anaesthesia passed.

Two experiments were performed: in each experiment 2 groups of 70 mice were infected with the 2 different clinical *M.tuberculosis* strains, while an additional group served as a control. Twenty mice from each group were left undisturbed to record survival up to day 120 after infection. Six animals from each group were sacrificed by exsanguination at 1, 3, 7, 14, 21, 28, 60 and 120 days after infection. One lung lobe, right or left, was perfused with 10% formaldehyde dissolved in PBS and prepared for histopathology, determining by automated morphometry the percentage of lung surface area affected by pneumonia. The other lobe was snap-frozen in liquid nitrogen and used for the determination of bacilli loads by counting the number of colony forming units (CFU) following the method previously described (Hernández-Pando *et al*., 1996; Hernandez-Pando *et al*., 1997). All procedures were performed in a class III cabinet in a bio-safety level III facility. Infected mice were kept in cages fitted with micro-isolators connected to negative pressure. Animal work was performed in accordance with the national regulations on Animal Care and Experimentation in Mexico.

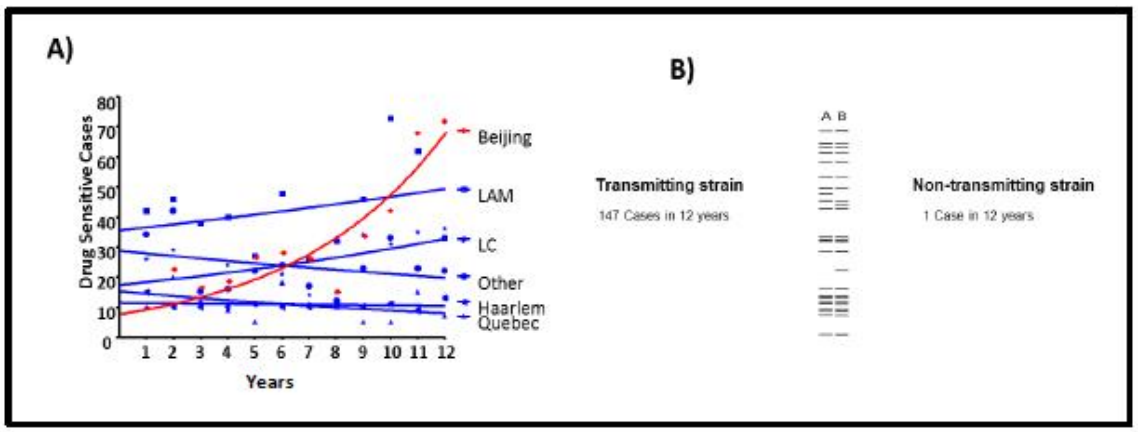


Figure 3.2: A) Twelve year monitoring of epidemiological field site of various *M. tuberculosis* genotypes. B) Two *M. tuberculosis* strains demonstrating different transmission chains (Van der Spuy *et al*., 2009)

3.1.2. Clinical hetero-resistant *M. tuberculosis* isolate

Phenotypically diagnosed drug resistant *M. tuberculosis* isolates were cultured from sputum strains and genotypically characterised by IS*6110* DNA fingerprinting (Van Embden *et al*., 1993) and spoligotyping (Kamerbeek *et al*., 1997) using internationally standardised methods. The DNA sequence chromatogram of the *rpoB* gene suggested heteroresistence by the presence of both the wild-type and mutant sequences at codon 531. The isolate were plated on selective media, 7H9 supplemented with OADC, with and without 2 µg/lm RIF. Single colonies were selected from the selective media with RIF-mono-resistant *M. tuberculosis* was characterised by the *rpoB* gene 531 TTG mutation) and without the mutation (susceptible *M. tuberculosis*) for further analysis. The *rpoB* gene in the susceptible and RIF-mono-resistant colonies was sequenced to confirm the presence and absence of the mutation, respectively.

3.1.3. Clinical MDR *M. tuberculosis* Beijing strain

Drug resistant *M. tuberculosis* bacilli were cultured from sputum samples and genotypically characterised by IS*6110* DNA fingerprinting and spoligotyping using internationally standardised methods (van Embden *et al*., 1993; Kamerbeek *et al*., 1997). The nsSNPs conferring resistance to isoniazid, RIF, ethambutol and pyrazinamide and were determined by DNA sequencing of the *katG*, *iniA* promotor, *rpoB*, *embB* and *pncA* genes, respectively (Louw *et al*., 2011). The Minimum inhibitory concentration (MIC) of RIF for the MDR isolate was determined by inoculating a 100 µl aliquot of a mid-log phase culture into enriched BACTEC 12B medium (Becton Dickinson, USA) containing between 2 and 200 µg/ml RIF for the analysis of RIF resistant isolates H37Rv (ATCC 35828) was included as a RIF susceptible control (Louw *et al*., 2011). The culture were incubated at 37^{o}C and the growth index (GI) for each isolate (at each RIF concentration) was measured daily for 9

consecutive days and compared to the GI of the corresponding isolate grown (diluted 1/100) in the absence of RIF. The MIC for the isolate was defined as the lowest concentration of RIF at which there was a complete absence of growth. The experiments were done in duplicate and were repeated on at least two separate occasions.

3.2: Freezer stock preparations of *M. tuberculosis* clinical strains

3.2.1. Hypo and hyper virulent *M. tuberculosis* clinical strains

Stock cultures of two genetically closely related *M. tuberculosis* Beijing genotype strains which demonstrated vastly different pathogenic characteristics in terms of their ability to transmit and cause disease in humans and to kill mice were inoculated into Mycobacterial Growth Indicator Tubes (MGIT) and incubated at 37^{o}C until positive growth was detected using the BACTEC 960 TB system (Becton Dickson, USA). Approximately 0.2 ml was then inoculated onto Løwenstein-Jensen (LJ) media and incubated at 37^{o} C for 6 weeks with weekly aeration to provide dissolved oxygen to the media until single colony formation was observed. A single colony was transferred from each culture and inoculated into 20 ml 7H9 Middelbrook liquid medium (Becton, Dickinson and Company, Sparks, USA) containing 0.2% (v/v) glycerol (Merck Laboratories, Saarchem, Gauteng, SA) and 0.1% Tween80 (Merck Laboratories, Saarchem, Gauteng, SA) in 250 ml screw cap tissue culture flasks (Greiner Bio-one, Germany) and supplemented with 10% -dextrose-catalase (DC) and cultured at 37^{o}C. Primary cultures were inspected for contamination by ZN staining and culture on blood agar plates. Each primary sub-culture was then sub-cultured in 20 ml Middlebrook 7H9 liquid medium supplemented with DC only and incubated at 37^{o}C in a 250 ml tissue screw cap culture flask (Greiner Bio-one, Germany). Growth of the secondary sub-cultures was monitored and re-inspected for contamination. Thereafter, 50% glycerol stocks

were prepared and stored at -80°C. Briefly, 500µl of 50% v/v glycerol solution with H_2O and 500 µl of the secondary culture with OD_{600}=1 were mixed together in a sterile 2 ml cryogenic tubes with O-rings and stored at -80°C.

3.2.2. Hetero-resistant *M. tuberculosis* clinical isolate

A single colony without the mutation (susceptible *M. tuberculosis* strain) and a single colony containing the *rpoB* mutation (mono-rifampicin resistant *M. tuberculosis* strain), with a Minimum Inhibitory Cconcentration (MIC) of between 100 and 200 µg/ml, were transferred to 10 ml 7H9 Middelbrook medium (Becton, Dickinson and Company, Sparks, USA) containing 0.2% (v/v) glycerol (Merck Laboratories, Saarchem, Gauteng, SA) and 0.1% Tween80 (Merck Laboratories, Saarchem, Gauteng, SA) and supplemented with 10% albumin-dextrose-catalase (ADC) in a 50 ml filter screw cap tissue culture flasks (Griener Bio-one, , Germany) and incubated at 37°C. The primary cultures were sub-cultured in 10 ml Middlebrook 7H9 liquid medium supplemented with DC in a 50 ml filter screw cap tissue culture flasks (Griener Bio-one, Germany) and incubated at 37°C until it reached an OD_{600} = 1.0. Secondary sub-cultures were re-inspected for contamination and 50% glycerol stocks were prepared and stored at -80°C as described above.

3.2.3. MDR resistant *M. tuberculosis* Beijing isolate with high RIF MIC

One clinical isolate from the Beijing lineage cluster 220, with a S531L mutation in the *rpoB* gene, and an MIC of 140 µg/ml was inoculated onto LJ solid medium and incubated at 37 °C for 3-4 weeks with continuous aeration. Colonies were scraped from LJ slants and incubated in 5 ml Middlebrook 7H9 medium (Becton Dickinson, Sparks, MD 21152, USA) supplemented with 0.2% (v/v) glycerol, 0.1% Tween 80 and 10% ADC in a 50 ml filter screw cap tissue culture flasks (Greiner Bio-one, Germany). After incubation for 7-10 days at 37°C these primary sub-cultures were inspected for contamination by ZN gram staining

and culture on blood agar plates. The primary sub-cultures were then sub-cultured in 10 ml Middlebrook 7H9 liquid medium supplemented with ADC and incubated at 37°C. After 10-14 days of growth (OD_{600} =0.7-0.8), the secondary sub-culture were re-inspected for contamination and a 50% glycerol stock was prepared as described above and stored at -80°C.

3.3: Culturing of *M. tuberculosis* clinical strains to mid-log growth phase

3.3.1. Culturing of *M. tuberculosis*

M. tuberculosis cultures for protein extraction was set up by inoculating 500 µl 50% glycerol stock into 20 ml 7H9 Middlebrook medium supplemented with DC in a 50 ml screw cap tissue culture flask (Greiner Bio-one, Germany) and incubated until it reached an OD_{600} of 0.9. One millilitre of the culture was then used to inoculate 2 times 50 ml 7H9 Middelbrook media supplemented with DC in a 250 ml filter screw cap tissue culture flasks (Greiner Bio-one, Germany). Both cultures were incubated at 37°C and the growth was monitored by OD_{600}. Protein extractions were done as described below when cultures reached mid-log growth phase (OD_{600} between 0.6 and 0.7).

3.3.2. RIF treated *M. tuberculosis* cultures

When each 50 ml culture in 7H9 Middelbrook medium reached an OD_{600} = 0.7 it was divided into two equal portions: one half served as control (without exposure) while the remaining half was exposed to the critical concentration (2 µg/ml) of RIF and incubated at 37°C for a further 24 hours. The critical concentration is defined as the drug concentration used to differentiate between drug susceptible and drug resistant isolates. A isolate is determined as resistant when 1% or more of the test population grows in the presence of the critical

concentration. This concentration was used as it would not result in a killing effect of high and intermediate level RIF resistant isolates. Thereafter whole cell lysate proteins were extracted as described below.

3.4: Mycobacterial whole cell lysate protein extraction

Mycobacterial cells were collected by centrifugation (10 min at 2 500 x *g*) at 4ºC and resuspended in 1 ml cold lysis buffer containing 10 mM Tris-HCl pH 7.4 (Merck Laboratories), 0.1% Tween-80 (Sigma Aldrich, St. Louis, MO), Complete Protease inhibitor cocktail (1 tablet per 25 ml) (Roche, Mannheim Germany) and Phosphatase inhibitor cocktail (1 tablet per 10 ml) (Roche, Mannheim Germany). Resuspended cells were transferred into 2 ml cryogenic tubes with O-rings and the pellet was collected after centrifugation (2 min at 14 000 x *g*; 1 min on ice; 2 min at 14 000 x *g*; 1 min ice in order to prevent the protein denaturation due to the high temperature). An equal volume of 0.1 mm glass beads (Biospec Products Inc., Bartlesville, OK) was added to the pelleted cells. In addition, 300 µl cold lysis buffer together with 10 µl RNase free DNaseI (2 U/ml) (NEB, New England Laboratories) was added and the cell walls were lysed mechanically by bead-beating for 20 seconds in a Ribolyser (Bio101 SAVANT, Vista, CA) at a speed of 4.0 followed by a 1 min cooling on ice. The lysis procedure was repeated 6 times. The lysate was clarified by centrifugation (10 000 x *g* for 5 min) at 21^{o}C and the supernatant containing the whole cell lysate proteins was retained. Thereafter the lysate was filter sterilised through a 0.22 µm pore acrodisc 25 mm PF syringe sterile filters, (Pall Life Sciences, Pall Corporation, Ann Arbour, MI) and stored at -80^{o}C until further analysis.

3.5: Determination of protein concentration

Protein concentrations were determined with the RCDC Protein assay (Bio-Rad Laboratories, Hercules, CA 94547) according to the manufacturer's instructions.

3.6: Total proteome

3.6.1. Gel electrophoresis

Fifty micrograms of whole cell lysate protein was diluted in an application buffer mixture containing 15µl HPLC grade water, 5µl NuPAGE® LDS sample buffer (Invitrogen, Carlsbad, CA) and 1ul 10 mM dithiothreitol (DTT) and heated for 5 min at 95^{o}C. Thereafter proteins were fractionated in by SDS-PAGE using a 1.0 mm 4 - 12% Nu-PAGE gradient gel, (Invitrogen, Carlsbad, CA) under reducing conditions for 40 min at 200V. SDS-PAGE gels were Coomassie stained using a Colloidal Blue staining Kit (Invitrogen, Carlsbad, CA).

3.6.2. In-gel trypsin digestion

After staining, each gel lane was divided into 12 fractions for Linear Trap Quadropole (LTQ)-Orbitrap analysis, or 10 fractions for Q Exactive analysis. Each fraction was subjected to *in gel* reduction, alkylation and tryptic digestion. Proteins were reduced in 10 mM DTT for 1 hour at 56^{o}C and alkylated with 55 mM iodoacetamide for 45 min at room temperature. The reduced and alkylated peptides were digested with sequence grade-modified trypsin 1:50 wt:wt (Promega, Madinson WI) for 16 hours at 37^{o}C in 50 mM NH4HCO3, pH 8.0. The reaction was quenched through acidification with 1% trifluoroacteric acid (TFA) (Fluka, Buchs, Switzerland). After treatment with Trypsin, peptides were extracted from gel pieces using 50% acetonitrile (ACN) followed by washing with 100% ACN and all the washes from

a single fraction were pooled together in one 1.5 eppendorf tube (Eppendorf, Hamburg 22331, Germany). The total peptide extract were dried to 50 µl and the resulting peptide mixture was desalted on RP-C18 STAGE tips (Rappsilber *et al*., 2003). Peptides were stored on the RP-C18 STAGE tips and eluted with 70% ACN and 0.1% formic acid (FA) and dried to 5 µl. The peptides diluted to a volume of 16 ul FA before mass spectrometry analysis.

3.6.3. LTQ-Orbitrap (Mass spectrometry)

All experiments were performed on a Dionex Ultimate 3000 nano-LC system (Sunnyvale CA, USA) connected to a linear quadrupole ion trap – Orbitrap (LTQ-Orbitrap) mass spectrometer (ThermoElectron, Bremen, Germany) equipped with a nanoelectrospray ion source. For liquid chromatography separation we used an Acclaim PepMap 100 column (C18, 3 µm, 100 Å) (Dionex, Sunnyvale CA, USA) capillary of 12 cm bed length 100 micron ID self packed with Reprosil_Pur C18-aq (Dr. Maisch Gmbh, Ammerbuch-Entringen, Germany). The flow rate used was 0.3 µl/min for the nano column, and the solvent gradient used was 7% B to 40% B in 87 min, then 40-80% B in 8 min. Solvent A was aqueous 2% ACN in 0.1% formic acid, whereas solvent B was aqueous 90% ACN in 0.1 % FA.

The mass spectrometer was operated in the data-dependent mode to automatically switch between Orbitrap-MS and LTQ-MS/MS acquisition. Survey full scan MS spectra (from m/z 300 to 2,000) were acquired in the Orbitrap with resolution R = 60,000 at m/z 400 (after accumulation to a target of 1,000,000 charges in the LTQ). The method used allowed sequential isolation of the most intense ions, up to six, depending on signal intensity, for fragmentation on the linear ion trap using collisionally induced dissociation at a target value of 100,000 charges.

For accurate mass measurements the lock mass option was enabled in MS mode and the polydimethylcyclosiloxane (PCM) ions generated in the electrospray process from ambient

air were used for internal recalibration during the analysis (Olsen *et al*., 2005). Target ions already selected for MS/MS were dynamically excluded for 60 seconds. General mass spectrometry conditions were: electrospray voltage, 1.5 kV; no sheath and auxiliary gas flow. Ion selection threshold was 500 counts for MS/MS, and an activation Q-value of 0.25 and activation time of 30 ms were also applied for MS/MS.

3.6.4. Q Exactive-Orbitrap (Mass spectrometry)

The proteome analysis was performed using a nanoflow Dionex Ultimate 3000 nano-LC system (Sunnyvale CA, USA) coupled online to a Q Exactive mass spectrometer (Thermo Fisher scientific, Bremen, Germany) via a Proxeon nanospray source, operated at 2.4kV. Each sample were analysed in triplicate and 4 ul of peptide mixture were injected for each replicate. The peptides were loaded in solvent A (0.1% FA in aqueous solution) in a Viper^{TM} capillary column packed with C18 resin with a bed length of 15 cm (75 µm internal diameter, 3 µm beads, 75 µm internal diameter, beads, Bremen, Germany) with a flow rate of 5 µl/min. Peptides were eluted using a linear 10-32% gradient of solvent B (90% ACN and 0.1% FA) during 120 min, with a flow of 300 nl/min.

Mass spectra were acquired using a data dependant method with an automatic switch between a full scan and choosing the 10 most abundant precursor ions from the survey scan MS/MS scans. The target value for the full scan MS spectra were 1e6 ion and resolution was 70,000 at m/z 200. MS/MS was performed by higher-energy collisional dissociation (HCD) fragmentation with a collision energy value of 30%, with target value of 1e6 ions or 120ms injection time. MS/MS spectrum was collected in the Orbitrap cell with a resolution of 35,000 at m/z 200. Ions were dynamically excluded for 45s after ion selection for MS/MS.

3.7: Data analysis

3.7.1. Exponentially Modified Protein Abundance Index (EmPAI)

3.7.1.1. Mascot search and peptide/protein validation

MS/MS peak lists from individual 48 RAW files (24 from the hypo-virulent and 24 from the hyper-virulent strains) were generated using DTA SuperCharger package, version 1.29, available at the MSQuant validation tool (see below). Protein identification was done by searching the data separately against the *M. tuberculosis* H37Rv protein database available at the Tuberculist website (version R11 tuberculist.epfl.ch). The databases were in-house modified to also contain reversed sequences of all entries as a control of false-positive identifications during analysis. Common contaminants, such as keratins, BSA and trypsin, were also added to the database. MASCOT Daemon was used for multiple searches submission on a local Mascot server v2.1 (Matrix Science). The search parameters used were: Maximum missed cleavages: 3; Carbamidomethyl (C) as fixed modification; N-acetyl (Protein), Oxidation (M), pyro-glu (Q) and pyro-glu (E) as variable modifications; Peptide mass tolerance of ± 15 parts per million; MS/MS mass tolerance of 0.5 Da. Under these criteria, Mascot indicated a minimal score of 22 for $p \leq 0.01$ and 15 for $p \leq 0.05$. All data had a mass accuracy average of 4.1 parts per million. Spectra and protein validation were performed using an open source software called MSQuant (version 1.5a61), largely used for LC-MS/MS data analysis. Proteins were validated statistically, based on the score of their individual peptides. Proteins with at least two tryptic peptides with a minimal score of 22 for each (protein false-positive probability of 0.01%), or those with only 1 peptide but a MS/MS score higher than 38 were accepted (protein false-positive probability lower than 0.25%). Using these criteria, all MS/MS identifications of peptides present in entries with reversed sequences (i.e. false-positive identifications) were not validated, since none of the reversed

proteins were identified with 2 peptides with a score higher than 21 each or 1 peptide with a score higher than 38 (the highest Mascot score for a peptide from the reversed database was 32 – data not shown). Identifications with only one unique peptide were accepted only after manual validation. Quality criteria for manual validation were the assignment of major peaks, the occurrence of uninterrupted y- or b-ion series of at least 3 consecutive amino acids, the preferred cleavages N-terminal to proline bonds and C-terminal to Asp or Glu bonds, and the possible presence of a2/b2 ion pairs.

3.7.1.2. Protein abundance estimation

The peptide list was submitted to the emPAI calculation tool, and emPAI values for individual proteins identified in each sample were obtained (Ishihama *et al*., 2005). Before the submission of the data, genes *Rv1198*, *Rv1793* and *Rv2346c* were manually merged, and renamed the complex to ESATx-like. This was done because it is impossible to determine which of those proteins are contributing the tryptic peptides that are shared between them (these 94 amino acids long proteins share 86 identical amino acid residues). Protein abundance expressed as emPAI values was calculated using the number of observable peptides and the number of observed parent ions per identified peptide. The number of observable (or expected) peptides for a protein was calculated through *in silico* trypsin digestion of the *M. tuberculosis* H37Rv database, and the resulting peptide fragments were compared with the scan range of the mass spectrometry. The emPAI values were calculated using a script developed at the Keio University (Japan) (available at http://empai.iab.keio.ac.jp/), using the following parameters: trypsin enzyme, Carbamidomethyl (C) fixed modification, mass range from 300 to 8,000 Da, no retention time filtering, Bold red peptides only (i.e., unique peptides in the Mascot result), peptides filtered by peptide Mascot score higher than 21. The Protein Abundance Index (PAI) was obtained by division of the observed parent ions with the number of theoretical observable

peptides; emPAI is obtained using the formula emPAI = $10^{PAI} - 1$. The protein concentration in the sample was calculated as mol% by dividing individual emPAI values by the sum of all values within a sample, and multiplied by 100. This step is not only important for measuring the mol% of a protein, but also to normalize any difference in emPAI values observed between samples due to differences in instrument efficiency during different. To determine differentially represented proteins between the hypo- and the hyper-virulent *M. tuberculosis* strains, we merged the peptide list identifications of two independent replicates per strain. Individual mol% values were compared, and proteins were divided in two categories as follows: i) for proteins identified in both samples, the difference in relative concentration between the strains had to be higher than 4 fold; b) for a protein identified in only one of the strains, we required that it had to be identified with a minimal of three parent ions, and we selected a mol% of 0.02 as a threshold. Such stringent criteria are required to guarantee that a protein identified in only one sample is most probably due to differences in abundance between the samples, and not because parent ions were not identified (but still present) in the MS analysis due to random fluctuation of the MS/MS data-dependant acquisition procedure.

3.7.2. MaxQuant

The acquired raw files (total proteome MS/MS files of susceptible, rifampicin (RIF)-mono-resistant, and MDR) were analysed by MaxQuant (version 1.2.2.5) (Cox and Mann, 2008). Andromeda, a probabilistic search engine incorporated into the MaxQuant framework was used to search the peak lists against the Tuberculist R21 database. Common contaminants like keratins, BSA and trypsin were automatically added to this database by the software.

The search included enzyme specificity: Trypsin/P; maximum missed cleavages: 2; carbamidomethyl (Cys) as fixed modification; and N-acetyl (protein), oxidation (Met), pyro-glu (Gln) and pyro-glu (Glu) as variable modifications. For statistical evaluation of the data

obtained, the posterior error probability and false discovery rate were used. The false discovery rate of 0.01 for proteins and peptides were required. The peptide identification was based on a search with an initial mass deviation of precursor ion of up to 20 ppm, with further data recalibrations and new search with a 6 ppm deviation. MS/MS mass deviation was set to 20 ppm for Q Exactive data, and 0.5 Da for any LTQ-Orbitrap data. To match identifications across different replicated and adjacent fractions the "match between runs" options in MaxQuant was enabled within a time window of 2 min.

3.7.3. Perseus statistical analysis tool

The bioinformatics analysis was done with Perseus tools available in MaxQuant environment (http://Perseusframeowrk.org). For the analysis the missing values were replaced using data imputation. The aim to perform imputation of missing values is that they should simulate signals of low abundant proteins instead of presenting absent proteins (not published).

For the comparative label-free analysis of the proteome of the hypo and hyper-virulent *M. tuberculosis* significant proteins in each strain was determined using a t-test with Permutation – based FDR correction to calculate the p-value with a threshold < 0.01 using the Perseus tool. The differentially abundant proteins in each strain were determined by the odds ratio. To determine the differentially abundant proteins in each strain relative to the other strain we used the significant p-value combined with the average iBAQ ratio [hyper-virulent strain$^{\text{average iBAQ intensities}}$/hypo-virulent strain$^{\text{average iBAQ intensities}}$].

3.8: Phosphoproteome

3.8.1. Filter Aided Sample Preparation (FASP)

Four milligrams of concentrated whole cell lysate protein was heated in 4% SDS buffer to ensure complete homogenisation and denaturation. The protein concentration was measured by using the RCDC assay (as described in above) and 250 µl of the lysate proteins was loaded onto a 15 ml Amicon filtration device (10kDAa MWCO) (Millipore) and centrifuged at 2000 x *g* for 40 min at 25^{o}C. The flow through was collected in a clean falcon tube. Subsequently, the rest of the lysate proteins were diluted in the filtration device and centrifuged twice more for 15 min and the elute was collected in a separate falcon tube. The flow through of the second and third centrifugation was collected in separate falcon tubes. The concentrated proteins were washed with 20 ml Tris buffer (UA buffer) containing 8M Urea in 0.1M Tris/HCl pH 8.5 to remove SDS. The concentrated proteins (final volume of 250 µl), were alkylated by mixing proteins with 1.5 ml 50 mM iodacetamide (IAA) and incubated in the dark for 20 min. The alkylated proteins were concentrated by centrifugation for 17 min at 20 000 x *g* at 25^{o}C to obtain a volume of 250 µl. The excess IAA was removed by multiple washes with 2 ml UA buffer and centrifuged at 20000g for 15 min. Thereafter the proteins were washed with 4 ml 50 mM ammonium bicarbonate (ABC) and concentrate by centrifugation at 20000 x *g* for 17 min until 250 µl. The proteins were equilibrated by washing twice with 4 ml 20 mM ABC until final volume of 250 µl with trypsin (Promega) in a protein to enzyme ratio of 100:1 at 37^{o}C for 16 hrs. After digestion the peptides were eluted with 50 µl of water to avoid desalting for further processing of the peptides.

3.8.2. Phosphopeptide enrichment

3.8.2.1. Strong Cation Exchange Chromatography (SCX)

Strong cation exchange chromatography is frequently used as a separation method for complexity reduction in proteomic studies. Trypsin digested peptides were diluted in SCX solvent A (5 mM KH2PO4/30% ACN, pH 2.7) to 7 ml, the pH were adjusted to 2.7 and then adjusted to 10 ml with 100% ACN (Beausoleil *et al*., 2004). The peptides were then separated by strong cationic exchange chromatography (SCX) as described below. Briefly, the peptide mixture was loaded onto cationic exchanger column equilibrated (washed) with 2 ml 30% ACN containing 5 mM KH_2PO_4 (pH 2.7). The flow through which predominantly contains multiply phosphorylated peptides was collected. The peptides bound to the column were eluted in an increasing salt gradient with buffer containing 5 mM KH_2PO_4 (pH2.7) and 150 mM KCl. The flow-through from the SCX column was also used as one fraction (salvage fraction).

3.8.2.2. Enrichment of phosphopeptides with Titanium dioxide (TiO_2)

The flow-through and seven SCX fractions were subjected to TiO_2 enrichment with 4 consecutive incubations for the flow through and1 incubation each for the remaining 7 fractions as described. The UV absorbance of the fractions was measured using the nano drop and they were incubated with TiO_2 (MS-Analysentechnik, Germany) with a peptide to bead ratio of 1:2 (Qing-Run Li, 2009). Before mixing the fractions 5 mg TiO_2 beads re-suspended in 10 ul 30 mg/ml dihydrobenzoic acid (DHB) solution to prevent non-specific binding of non-phosphorylated peptides. DHB solution is prepared by dissolving 30 mg of DHB powder in 75% acetonitrile containing 0.1% (Trifluoracetic acid) TFA. The fractions were rotated for 30 min and after incubation and briefly centrifuged. The supernatant was transferred to a newly labelled tube. The phosphopeptides bound beads were washed twice

with 200 µl 30% ACN and 3% TFA followed by two washes with 200 ul 75% ACN and 0.3% TFA. The enriched phosphorylated peptides were then eluted with 40 µl elution buffer (25% DHB and ACN with pH 10) under basic conditions. The eluted phospho-peptides were desalted using C18 Stage tips (as described above) (Rappsilber *et al*., 2003) .

3.8.3. LTQ-Orbitrap-Velos (Mass spectrometry)

The peptides were separated in a column packed in-house with C18 beads on a Proxeon Easy-nLC system using a binary gradient provided by buffer A (0.5% acetic acid and buffer B (0.5% acetic acid and 80% ACN). The peptides (4 µl) were eluted directly without any trapping column with buffer A at a flow rate if 500 nl/min. Elution was carried out at a flow rate of 250 nl/min, with a linear gradient from 10% to 35% buffer B in 95 min followed by 50% B for 15 min. At the end of the gradient the column was washed with 90% B and equilibrated with 5% B for 10 min. The LC system was directly coupled in-line with the LTQ-Orbitrap Velos instrument (Thermo Fischer Scientific, Bremen, Germany) via the Proxeon Biosystems nanoelectrospray source (*Proxeon Biosystems,* Denmark). The source was operated at 2.1-2.25 kV, with no sheath gas flow, with the ion transfer tube at 200^{o}C.

The mass spectrometer was programmed to acquire in a data dependant mode. For the high-high strategy, survey scans were acquired in the Orbitrap mass analyser with resolution 30,000 at m/z 400 with lock mass option enabled for the 445.120025 ion. However the target lock mass abundance was set to 0% instead of 5-10% in order to save the injection time for the lock mass. For the full scans 1E6 ions were accumulated within a maximum injection time of 250 ms in the C trap and detected in the Orbitrap analyser. The ten most intense ions with charge states ≥ 2 were sequentially isolated (signal threshold of 10,000) to a target value of 3E4 with a maximum injection time of 150 ms and fragmented by HCD in the collision cell (normalised collision energy of 40%) and detected in the Orbitrap analyser at 7,500

resolution. For the high-low strategy using CID, full scans were acquired in the Orbitrap analyser at 60,000 resolution because parallel acquisition in enabled in the high-low mode. Up to the 20 most intense peaks with charge state ≥ 2 were selected for sequencing (signal threshold of 1000) to target value if 5, 000 with a maximum injection time if 25 ms, isolation window of 2Th and fragmented in the top ion trap by collision induced dissociation with normalised collision energy of 35%, activation q = 0.25, and activation time of 10ms. For CID wideband activation and multi-stage activation options were enabled with the appropriate neutral loss mass list for singly, doubly and triply phosphorylated peptides. The fragmentation spectra were acquired in the ion trap at normal scan rate by lateral ejection and recorded by the dynode-multiplier system. For all sequencing events dynamic exclusion was enabled to minimize repeated sequencing. Peaks selected for fragmentation more than once within 30 s were excluded from selection (10 ppm window) for 60 s.

3.8.4. Data analysis –MaxQuant

The raw data acquired were processed using MaxQuant software version (1.2.2.5) and processed as per standard flow. Since the HCD spectra were acquired in profile mode, de-isotoping was performed similar to survey MS scans to obtain singly charged peak list. Peak lists generated from the 'quant' module were searched against the *M. tuberculosis* H37Rv protein database (version R11 tuberculist.epfl.ch). For the high-low strategy these values were 7 ppm and 0.5 Th, respectively. The search included cysteine carboaminomethylation as a fixed modification, and N-acteylation of protein, oxidation of methionine, and phosphorylation of Ser, Thr and Tyr were used as variable modification. Up to two miss cleavages were allowed for protease digestion and peptide had to be fully tryptic. The identify module in MaxQuant was used to filter identifications at 1% false discovery rate (FDR) at three levels namely, site, peptide and protein. As such there is no fixed cut-off

score threshold but instead spectra are accepted until the 1% FDR rate is reached. Only peptides with minimum length of 7 amino acids were considered for identification.

3.8.5. Statistical analysis

To determine the intensity of each phosphorylated protein the average of the intensities of the phosphorylated peptides for each protein per replicate were calculated. The *p-values* for each of the phosphorylated proteins identified were manually calculated in this study. A significant p-value threshold <0.05 were used to determine the number of phosphorylated proteins identified with high confidence. To determine the relative abundance of each phosphorylated protein in each strain relative to the other strain we used the significant p-value combined the average iBAQ ratio [hyper-virulent strain$^{\text{average intensities}}$/hypo-virulent strain$^{\text{average intensities}}$].

3.9: Transcriptomics

3.9.1. Culturing for RNA extractions

All subsequent experiments were set up as 2 biological (to assess biological measurements for two independent experiments done on different days) and 2 technical replicates (repeated measures of one biological sample on the same day) for each isolate. Each experimental culture was set up by aliquoting 800 µl of the secondary sub-culture in 40 ml supplemented 7H9 medium (1:100 dilution) and incubated at 37°C until mid-log phase (OD_{600}= 0.7-0.8). Once the cultures reached mid-log phase, 200 ml of Guanidine-thiocyanate (GITC) (Sigma-Aldrich, St.Louis, USA) solution (5M GTC, 0.5% sodium N- lauroyl sarcosine, 0.1M H-mercaptoethanol, 12.5 ml of 1M sodium citrate pH7.0 and 1% Tween 80 made up to 500 ml with RNase free water) was added to each 40 ml culture. Cells were harvested by

centrifugation (3000 x *g*, 15 min, 20°C) and the supernatant was discarded. The pelleted cells were then resuspended in 1 ml of TRIzol (Life Technologies, Gaithersburg, MD). The suspension was transferred to a 2 ml screw capped tube containing silica beads (IEPSA Medical diagnostics, South Africa) and ribolyzed (reciprocal shaker, Hybaid) at 6 W for 20 seconds. Thereafter tubes were cooled on ice for 1 min between pulses. The ribolysis was repeated 3 times. Samples were centrifuged at 13000 rpm for 1 min and the TRIZOL solution above the beads was transferred to a 2 ml phase lock gel tube (Eppendorf, Hamburg 22331, Germany) containing 300 µl chloroform/isoamyl alcohol (24:1) (Sigma-Aldrich, St. Louis, Germany). Tube were inverted several times and then centrifuged at 13000rpm for 10 min. The top aqueous layer was transferred to a new 1.5 ml tube and the crude RNA was precipitated with the addition of an equal volume of isopropanol (Merck, Darmstadt, Germany). Samples were incubated at -20°C overnight. The crude RNA were collected by centrifugation at 12000 x g, 30 min at 4°C and the pellet was washed with 70% ethanol (Merck, Darmstadt, Germany). RNA pellets was air-dried and dissolved in 70 µl RNase-free water. Contaminating chromosomal DNA was digested by typically adding 4 µl DNase and 4 µl DNase buffer (Whitehead Scientific) to 15 µl extracted RNA, followed by incubation at 37°C for 30 min. The DNase treated RNA was made up to a final volume of 200 µl with RNase-free water. An equal volume of Phenol:Chloroform (4:1) (Sigma-Aldrich, St. Louis, Germany) was added to the diluted RNA, gently mixed and left on ice for 10 min. Tubes were centrifuged at top speed for 10 min at room temperature. The top aqueous layer was transferred to a new tube, 0.1 volumes of RNase-free sodium acetate pH 5.2 and 2.5 volumes of 100% RNase-free Ethanol (Merck, Darmstadt, Germany) were added and incubated at -20°C overnight. The RNA was collected by centrifugation at 12000 x *g*, 30 min at 4°C , washed with RNase-free 70% ethanol (Merck, Darmstadt, Germany), recollected by centrifugation at 12000 x *g*, for 10 min at 4°C and the ethanol was aspirated off. Purified

RNA was air-dried and re-dissolved in 70 µl RNase-free water. The quantity and quality of the RNA extracted from each culture was determined by measuring the A260/A280 ratio spectrometrically and electronically observed on a virtual gel electrophoresis on the Experion Software version 2.01 (Bio-rad). The RNA preparations were only considered acceptable for subsequent analysis if the presence of the dominant 16S and 23S rRNA species appeared as fairly sharp bands, the A260/A280 ratio was between 1.8 and 2.1 and no high molecular weight DNA was detected.

3.9.2. cDNA synthesis

cDNA was synthesized from 0.5 µg highly purified RNA using a Reverse transcriptase kit according to the manufacturer's instructions (Southern Cross Biotechnologies). In short, 0.5 ug RNA was mixed with 2 µl genomic DNA Whipeout buffer and RNase-free water (total volume of 14 µl) and incubated for 2 min at 42^{o}C. After the incubation 14 µl Quantscript Reverse Transcriptase, 4 µl Quantscript RT buffer and 1 µl RT Primer mix was added to the RNA mix and incubated for 20 min at 42^{o}C. Therafter the mix was incubated for 3 min at 95^{o}C to inactivate the Quantscript Reverse Transcriptase.

3.9.3. Primer design for Quantitative Real time PCR of candidate genes

Oligo nucleotides (primers) for quantitative reverse transcriptase analysis of the candidate genes were designed against the whole genome sequence of the *M. tuberculosis* H37Rv reference strain (http://genolist.pasteur.fr/TubercuList) using the internet based Primer software 3 version 0.4.0 (Whitehead Scientific, South Africa).

The primers were designed to be 18 and 20 bp and a G+C content between 50-60% and to anneal to unique sequences within the target genome. For any primer pair, a maximum melting temperature (Tm) difference of 2^{o}C was allowed.

Table 3.1: Quantitative real time PCR of candidate genes

Rv number	Gene name	Primer sequence	Tm	Reference
Rv0015c	*PknA*	Forward: 5'gcaggtgaagatcaccgact 3'	60°C	Current
		Reverse: 5'tacccgaaaccgcttcatac3'		
Rv0014c	*PknB*	Forward: 5'ggacatgaacccggatattg3'	60°C	Current
		Reverse: 5'tcgtttgactaccagcaacg3'		
Rv0931c	*PknD*	Forward: 5'gtttatgaggccgaggacac3'	60°C	Current
		Reverse: 5'ggtcagcggaccatactgtt3'		
Rv1743	*PknE*	Forward: 5'gacgtcaaaccggagaacat3'	60°C	Current
		Reverse: 5'gagctggtctccctgatacg3'		
Rv1746	*PknF*	Forward: 5'atcccttctcagggatcgtt3'	60°C	Current
		Reverse: 5'catgattcgacgatcaggtg3'		
Rv0410c	*PknG*	Forward: 5'tcaactttgtcgagcacacc3'	60°C	Current
		Reverse: 5'atgatgttttccggcttcag3'		
Rv2176	*PknL*	Forward: 5'atgtcaagcccgagaacatc3'	60°C	Current
		Reverse: 5'ccgttagcagctcgtagacc3'		
Rv1747	-	Forward: 5'cagacgacgtcgatacgc3'	60°C	PhD thesis (Louw GE, 2009)
		Reverse: 5'catcgatggtccaggtcac3'		

3.9.4. Quantitative real-time PCR (qRT-PCR)

The standard curve method was used for relative quantification and the mycobacterial 16S ribosomal RNA for reference constitutive gene expression. The standard curve method for relative quantification is based on the construction of standard curve of cycle number at a threshold (C_T) against initial input amount of total RNA or copy number. This method assumes an approximately equal amplification efficiency of the PCR among the diluted samples of the same gene, thus producing a linear relationship of C_T against initial input amount of total RNA or copy number. In this study the standard curve method was used for

relative quantification of the reference gene and the candidate genes on both biological and technical duplicate samples.

A standard 20 µl PCR reaction was setup by adding 2 µl of light Cycler Fast start DNA Master plus SYBR Green I reaction mix (Roche Applied Science, Germany), 1 µl forward and reverse primer (10 µM of each candidate gene, 14 µl of RNase-free water and 2 µl of diluted cDNA (dilution series of hypo- and hyper-virulent samples). A negative control was included which contained the same mix without the sample. The PCR mix was transfered to the capillary tubes (Roche Applied Science) in a pre-cooled adapter block after which it was removed and centrifuged and centrifuged at 15 rpm for 3 seconds. The capillaries were placed in a the light cycler 2.0 instrument (Roche Applied Science) and cycled through a four step PCR parameter protocol: (i) Activation program (95^{o}C for 15 min); (ii) RT- PCR programme repeated for 45 cycles (95^{o}C for 15s, 60^{o}C for 30s, 72^{o}C for 30s); (iii) Melting curve programme (95^{o}C for 0s, 60^{o}C for 15s, 95^{o}C for 0s with a heating rate of 0.1^{o}C/s); (iv) Cooling down programme for 10 sec. The primer optimisation and the PCR efficiency of each gene were optimised using H37Rv.

3.9.5. Data analysis

16S rRNA was included as reference genes and used for normalisation of RNA levels, because in the experiments 16S rRNA expression levels are the most stable. The level of gene transcription of each individual gene was quantified using the delta-delta C_T calculation in which the relative abundance of the target gene was normalised relative to the levels of the reference RNA transcripts (16S rRNA). Significant fold changes were identified based in The Relative Expression Software Tool (REST-384) that assigns significance with a significance level of 5%. The data analysis were done according to the delta-delta CT equation R= PCR efficiency of the target gene $^{-(\Delta CT \text{ hyper-virulent } M.tb \text{ strain} - \Delta CT \text{ hypo-virulent } M.tb \text{ strain})}$.

3.10: Bioinformatics analysis tools

Table 3.2 summarises a list of free online available bioinformatics tools used in this thesis. The search engine MASCOT uses mass spectrometry peptide data to identify proteins from primary sequence databases (Perkins *et al*., 1999). The Tuberculist database and website intergrates genome-, proteome-, transciptome- and drug data, literature, structural views, mutatnt and operon annotation in a structured manner inorder to broaden the view of *M. tuberculosis* (Lew *et al*., 2011). The Pathosystems Resource Intergration Centre (PATRIC) is a freely available bacterial bioinformatics resource centre providing vital information on infectious diseases and pathogen information with data and analysis tools (Gillespie *et al*., 2011; Snyder *et al*., 2007). The Tuberculosis database (TBDB) is an open access source where published data on tuberculosis are freely available (Reddy *et al*., 2009). The TMHMM server v. 2.0 is a prediction server which can determine the number of transmembrane helices based on a hidden Markov model (Sonnhammer *et al*., 1998).

Tabel 3.2: List of online databases used for bioinformatic analysis of the proteins identified in this study:

Database	Online reference	Reference
MASCOT	http://www.matrixscience.com/	(Perkins *et al*., 1999)
Tuberculist	http://tuberculist.epfl.ch/	(Lew *et al*., 2011)
PATRIC	http://www.patricbrc.org/	(Snyder *et al*., 2007; Gillespie *et al*., 2011;)
TBDB	http://www.tbdb.org/	(Reddy *et al*., 2009)
TMHMM Server v. 2.0	http://www.cbs.dtu.dk/services/TMHMM/	(Sonnhammer *et al*., 1998)

3.11. List of buffers and solutions

3.11.1. Culturing:

Middlebrook 7H9 liquid medium	7H9 Middlebrook medium	4.7	g
	dH_2O	900	ml
	Glycerol	2	ml
	Polysorbate (Tween80)	1	ml
ADC supplement	BSA	25	g
	Glucose	10	g
	Catalase	0.75	ml
	dH2O make up to	500	ml
DC supplement	Glucose	10	g
	Catalase	0.75	ml
	dH_2O make up to	500	ml

3.11.2. Protein extraction:

Extraction buffer	10 mM Tris-HCL pH 7.4	500	µl
	Tween-80	3	µl
	1 Protease inhibitor tablet per	25	ml
	1 Phosphatase inhibitor		
	cocktail tablet per	10	ml

	dH_2O make up to	25	ml

3.11.3. RNA extraction:

5M GITC	Guanidine thiocyanate	295.4	g
	Sodium N-Lauroyl Sarcosine	25	g
	ß-mercaptoethanol	3.5	ml
	1 M Sodium citrate pH 7.0	12.5	ml
	Tween80	5	ml

CHAPTER 4

Label-free quantification of closely related hypo-and hyper-virulent *M. tuberculosis* strains using Estimated Protein Abundance Index (emPAI)

This chapter is structured for publication (Published format see appendix)

My contribution to this chapter: **Culturing of *M. tuberculosis* strains**

Protein extractions

Sample preparation for Mass Spectrometry analysis

Writing of Manuscript

4.1: Aim:

In this study, we used high throught-put mass spectrometry in combination with Estimated Protein Abundance Index (emPAI) calculations to describe proteomic differences between two closely related *M. tuberculosis* Beijing genotype strains. These strains have vastly different pathogenic characteristics in terms of their ability to transmit and cause disease in humans and to cause pulmonary damage in mice.

4.2: Results:

4.2.1. Experimental model of pulmonary tuberculosis in BALB/c mice infected with hypo- and hyper-virulent *M. tuberculosis* Beijing strains

Mice infected with the actively transmitting strain started to die three weeks after infection and by 5 weeks all the mice had died (figure. 4.1A). Culture of lung homogenates showed very high lung bacillary loads (figure. 4.1B), while histopathology showed extensive tissue damage (pneumonia) which started at day 14 post infection (figure. 4.1C).

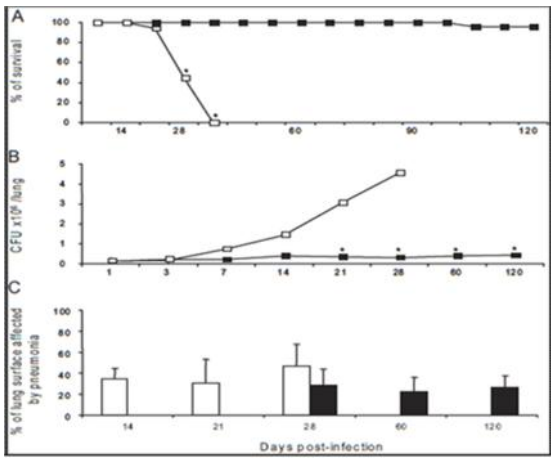


Figure 4.1: Virulence of *M.tuberculosis* clinical isolates in a murine infection model. BALB/c mice were infected by intratracheal injection (2.5x105 bacilli) with selected Beijing strains. The virulence of the strain representative of the largest cluster (white symbols) was compared to the virulence of the non-transmitting strain (black symbols). A) Kaplan–Meier survival curves. B) Bacillary loads. C) Morphometry (% of lung surface area affected by pneumonia). Asterisks represent statistical significance ($p<0.005$).

Based on these results we assigned a virulence status of hyper-virulent to this Beijing strain. Animals infected with the unique Beijing strain (non-transmitting) showed 80 to 90% survival after 4 months of infection (figure. 4.1A) and a tenfold lower number of colony forming units (CFU) in the lungs (figure. 4.1B). Histopathology showed extensive tissue damage (pneumonia) which started at day 14 post infection (figure. 4.1C). Based on these results we assigned a virulence status of hypo-virulent to this Beijing strain.

4.2.2. Proteome of the hypo-and hyper-virulent *M. tuberculosis* Beijing strains

To identify differences in abundance of proteins in the two Beijing genotype strains, whole cell lysate proteins were subjected to Gel-LC-MS/MS approach (figure 4.1) (Aebersold and Mann, 2003). The electrophoretic protein profile of both strains is almost identical (data not shown) (Lew *et al*., 2011).

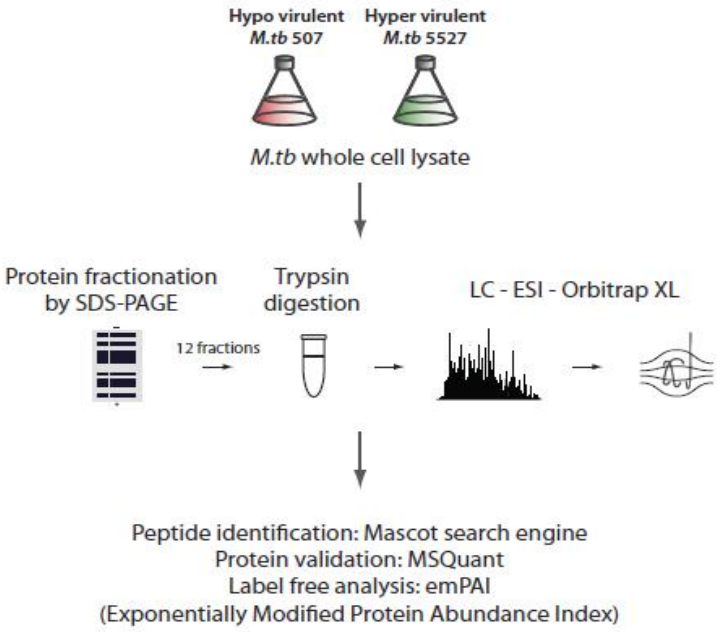


Figure 4.2: Workflow of the quantitative label-free proteomics approach using EmPAI

All mass spectrometry data obtained was identified using Mascot, and were submitted for identification against the *M. tuberculosis* H37Rv lab strain protein database (Tuberculist) (Cole *et al*., 1998; Lew *et al*., 2011). This reference database was used since previous studies have suggested that this gene annotation is the most reliable (Lew *et al*., 2011).

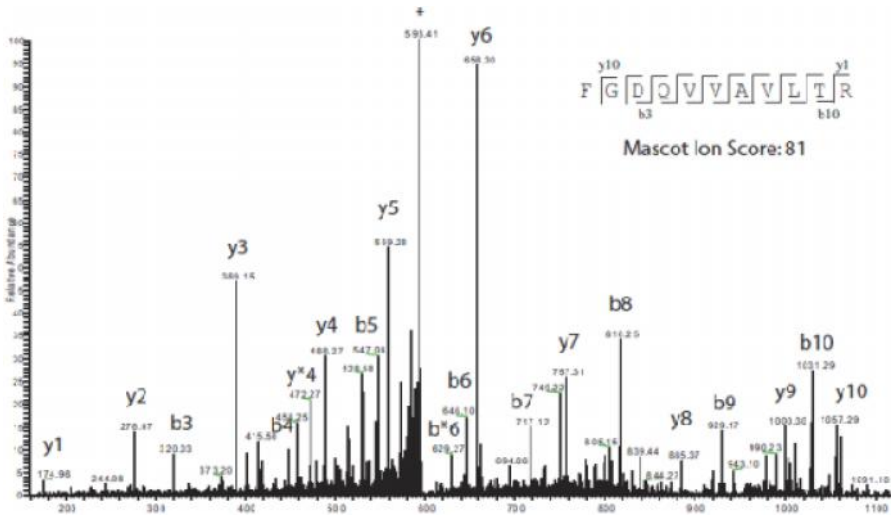


Figure. 4.3: MS/MS profile of ion m/z 601.41. Tandem mass spectrum of a prevalent ion on a particular time point in the LC gradient and ionized on the LTQ-Orbitrap. The peptide fragments randomly on each amide bond, resulting in carboxy-terminal y ions or amino-terminal b ions. After the fragment masses were submitted to Mascot, the peptide was identified as FGDQVVAVLTR (inset, with detected y and b ions represented) from protein Rv3220c – probable two component sensor kinase.

Figure 4.3 illustrates an example of a MS/MS spectrum of a peptide of m/z 602.84. The Mascot tool identified it as the peptide FGDQVVAVLTR, corresponding to the protein entry Rv3220c, annotated as "Probable two component sensor kinase". This identification had a mass accuracy of 4.4 ppm between observed and theoretical masses. Figure 4.3 also illustrates the fragmentation pattern and identification of y/b series of the sequence (sequence input). In this example Mascot was able to correlate the full y ion series, and 8 of 10 b ions, which resulted in a Mascot identification score of 81 (score 22 represents a $p<0.01$).

The whole proteome analysis of both replicates for each sample identified a total of 1443 proteins for the hypo-virulent strain, and 1524 proteins for the hyper-virulent *M. tuberculosis* strain. Merging both datasets, we identified the protein products of 1670 of the 4066 genes predicted by Tuberculist. From these, 145 proteins were identified only in the hypo-virulent cells and 226 were identified only in the hyper-virulent *M. tuberculosis* strain. The majority of these proteins observed in only one strain represent identifications based on one or two peptides, and in these cases those proteins were not used for quantitative comparison.

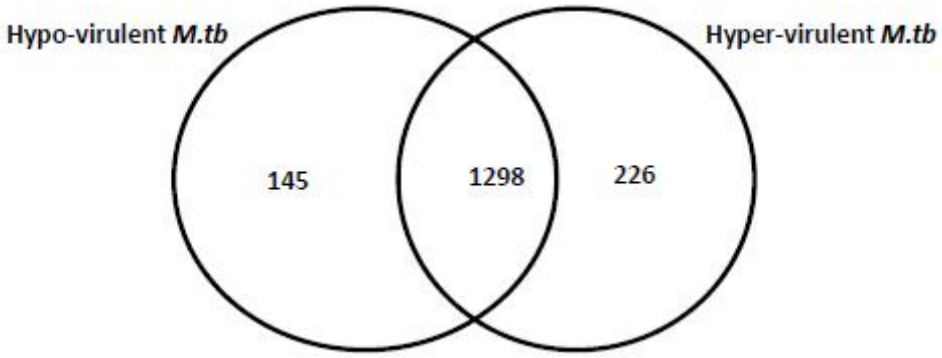


Figure 4.4: Venn diagram of proteins identified in hypo-and hyper-virulent *M. tuberculosis* strains

Appendix file 4.1 reports all peptides identified in whole lysate protein extract from the two Beijing genotype strains, their protein mass, number of peptides per identified protein (Pep/Prot), peptide length, observed charge, observed m/z ratio, measured peptide mass (Da), Mascot Score, the presence of modifications (as N-terminal acetylation or Met oxidation), and the error of the observed / theoretical mass in ppm.

4.2.3. emPAI analysis and data comparison

After normalization and mol% calculation, all mol% values from proteins present in both samples were plotted in a logarithmic linear distribution graph (figure. 4.2).

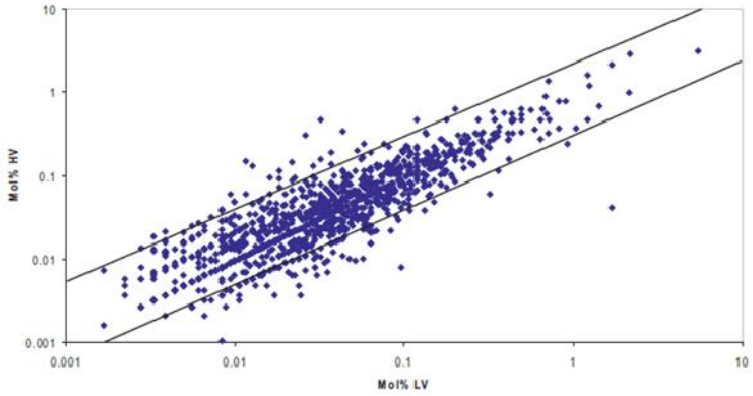


Figure 4.5: Protein abundance comparison of hypo- and hyper-virulent *M. tuberculosis* strain. Protein abundance was derived from emPAI values. Solid lines delimit differentially abundant proteins (external area) from similarly abundant (internal area).

This analysis showed that most of the 1298 proteins lie within a difference range of less than 3 fold (figure 4.5). This indicates not only that the normalization is reliable, but also that the relative quantitation provided by emPAI is relatively accurate, since the majority of proteins in this comparison are equally distributed between the two data sets. Appendix file 4.2 contains all emPAI values obtained for each strain and a sheet with the merged results for both strains.

4.2.4. Identification of differentially abundant proteins

Once mol% was calculated and samples compared, we set stringent thresholds to indicate proteins differentially represented in each Beijing strain, as shown in the Methods section. When these chosen criteria were applied to our dataset, only 53 entries from the hyper-virulent strain and 48 from the hypo-virulent *M. tuberculosis* strain were selected. Empai values for those entries can be seen in appendix file 4.2 in green and a short description of these entries is also reported in appendix file 4.3 (Correspond to supplementary file S3). Figure 4.6 shows these differential proteins grouped according to the functional category as given in Tuberculist. Interestingly, proteins in functional categories 3 (cell wall and cell processes) and 9 (regulatory proteins) are significantly over-represented in the hyper-virulent

strain while proteins from the functional category 1 (lipid metabolism) are more evident in the hypo-virulent strain.

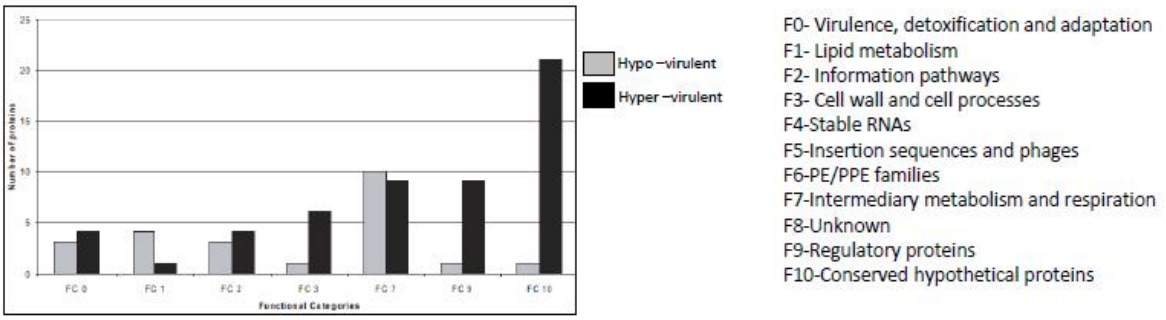


Figure 4.6: Functional classification of differentially abundant proteins. Proteins were classified according to Tuberculist's functional category groups. Grey bars represent proteins more abundant in the hypo-virulent strain, and black bars in the hyper-virulent *M. tuberculosis* strain. FC10 represents the group of conserved hypothetical proteins.

4.2.5. *In-vivo* RT-PCR measurements

Since ESAT-6 is an exported protein, intracellular accumulation of this protein demonstrated by proteomics could be a result of gene expression changes or changes in membrane transport machinery. Further validation of gene expression levels of ESAT-6 and CFP10 was performed in-vivo, during several time points of the infection (figure 4.7). Figure 4.7A shows the quantitative information of ESAT-6 from mice infected with one or the other strain was normalized against the number of copies of 16S, used as a control. Figure 4.7B demonstrates the absolute values of gene copies divided by 1 CFU. The results showed that ESAT-6 gene expression was reduced in the hyper-virulent strain (white bars). CFP10 showed a similar reduction, but to a lesser extent data (data not shown).

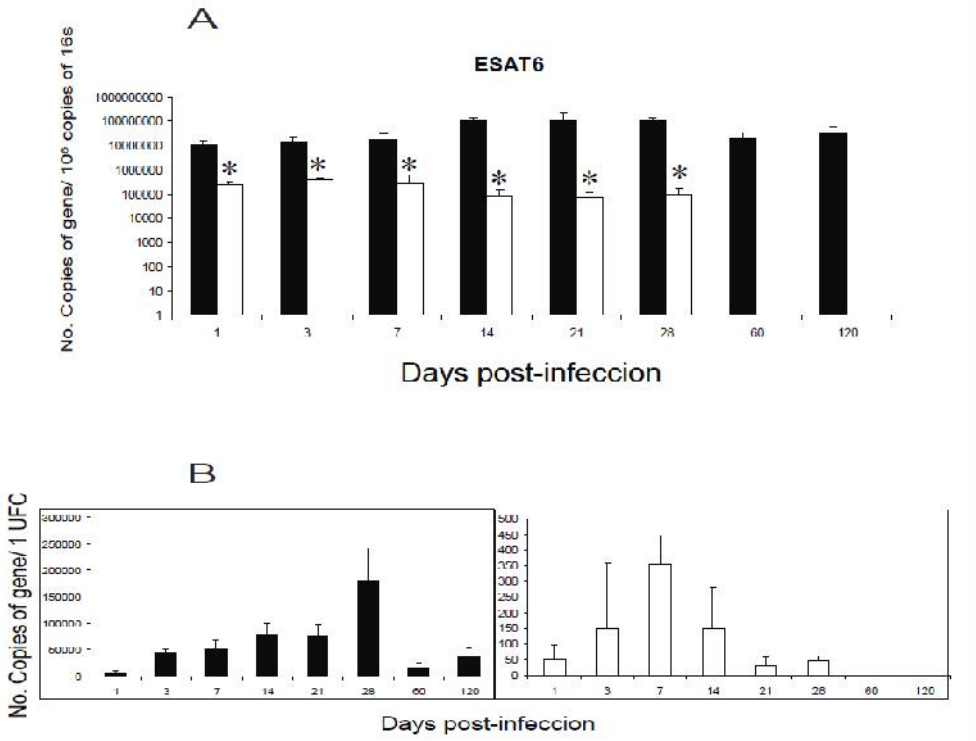


Figure 4.7: Validation of gene expression levels of ESAT-6 and CFP10 was performed in-vivo, during several time points of the infection. A) Quantitative information of ESAT-6 from mice infected with one or the other strain was normalized against the number of copies of 16S, used as a control. B) Absolute values of gene copies are divided by 1 CFU. Black bars represent hypo virulent strains and white bars represent hyper-virulent strain

4.3: Discussion

The identification of virulence factors from *M. tuberculosis* involved in the TB disease process is not only a crucial step to understand the pathogen's biology, but may also provide insights into regional TB epidemiology which in turn may improve treatment regimens in the long run. The release of the *M. tuberculosis* H37Rv lab strain genome (Cole *et al*., 1998) and further availability of other genomes of the *M. tuberculosis* complex such as *M. bovis*, the attenuated *M. bovis* BCG vaccine strain and the clinical isolate, *M. tuberculosis* CDC1551, has allowed many genomic comparisons in order to discover regions in the genome with gene mutations or gene deletions that result in a more or less virulent phenotype (Campuzano *et al*., 2007; de Souza *et al*., 2008). However, it can be expected that many factors participating in virulence of a strain cannot be directly identified at the genomic level, and information

regarding gene expression and protein abundance can contribute greatly to improve our understanding in this field.

To test this hypothesis, we employed a MS-based proteomic approach coupled with a label-free abundance estimate to identify proteins present in two clinical *M. tuberculosis* Beijing strains isolated from TB cases in South Africa. These isolates are of particular interest because they display striking differences in their level of virulence, as defined by their epidemiological and population characteristics, as well as virulence in a mouse infection model. The fact that the strain we defined as hyper-virulent using epidemiological studies was also found to be hyper-virulent using the definition of a rapid onset of disease and death in a mouse model is of itself an interesting, and in some ways counter-intuitive, result. A common and long-held view of pathogen evolution is that strains that rapidly kill their host may be negatively selected because the opportunities for transmission to new hosts are reduced. However, it is recognized that natural selection can still favor these strains as long as they possess a compensatory advantage, for example higher instantaneous transmissibility rates or increased resistance to host defenses (Gordon *et al*., 2009). Thus, it is possible that this strain is able to overcome the selective disadvantage of rapid killing of the host via mechanism(s) such as these. We also recognize that because of the significant differences in TB pathology between in-bred laboratory mice and humans (Ten Bokum *et al*., 2008) our animal model results are not necessarily translatable to our epidemiological ones.

From a total of 1670 identified proteins, we applied very stringent criteria to select a minimal observed difference threshold, and a total of 101 proteins were identified as being significantly differentially abundant in either sample. Our criteria are in accordance with the reported accuracy of emPAI, which has been shown to have an error range close to maximally a factor of 3 (Ishihama *et al*., 2005). Those 101 proteins were clustered into 3 functional categories, as defined in the Tuberculist resource database (established by(Cole *et*

al., 1998; Lew *et al*., 2011)). In addition, the remaining proteins not classified as differentially abundant were also used to check the presence of proteins previously described as virulence markers or factors. For example, one of the well characterized genomic regions associated with virulence is the phtiocerol dimycocerosates lipid (PDIM) locus. Mutants containing deletions of this locus were unable to multiply in mouse models (Lipsitch and Moxon, 1997). Some members of the PDIM family are sometimes linked to a phenoglycolipids **(**PGL) molecule, and such glycolipids are found to be very characteristic of highly virulent Beijing strains (Reed *et al*., 2004). The genes that participate in PGL synthesis and modification are located together within the PDIM locus. In recent years, this locus has been analyzed in detail and the functions of many of these genes determined (Apt and Kramnik, 2009). In this work we identified peptides representing proteins from 18 genes of a total of 35 genes present in that locus. From those, only 2 proteins (PpsA and FadD28) were not observed in a 1:1 ratio, but still the calculated differences were below our stringent threshold, being 2.5 fold and 2.8 fold over-represented respectively in the hypo-virulent strain, showing that in principle, PDIM and PGL synthesis are expected to be similar in this model.

In addition, our data show that proteins from functional category 1 (lipid metabolism) seem to be over-represented in the hypo-virulent strain. As discussed for PGL, lipids and lipoproteins (Camacho *et al*., 1999; Siméone *et al*., 2007) these proteins are believed to be key factors for acquisition of virulence in the strains. However, our data suggests that within a highly virulent strain family, the reducing part of the lipid repertoire may represent further steps in the evolution of virulence. A possible explanation for this process is that changing the repertoire of surface molecules may result in an enhanced ability of the pathogen to escape host recognition. This is corroborated by the fact that we observed that well described immunogenic proteins, ESAT-6, Esx-like proteins and MPT51 (FbpD) (Geiman *et al*., 2004),

were under-represented in the hyper-virulent *M. tuberculosis* strain. Interestingly, ESAT-6 and Esx-like proteins are highly-abundant and well defined virulence factors in *M. tuberculosis*, since the deletion of the RD1 region containing this gene result in loss of virulence (Sun *et al*., 2004) as observed for the attenuated *M. bovis* BCG strain (Mahairas *et al*., 1996). Previously in our group, we determined that the Esx-like proteins had a minimum 1.0 mol% in the avirulent *M. tuberculosis* H37Ra strain and up to 7.0 mol% in *M. tuberculosis* H37Rv (data not shown). However, the hyper-virulent Beijing strain showed only 0.23 mol%. Nonetheless, it is important to note that, while in this study we analyzed whole cell extracts, the decreased levels of ESAT-6 and Esx-like proteins in the hyper-virulent strain may be a result of faster export of these proteins to the extracellular environment.

Molecules involved in regulation of transcription (Functional category 9) could be main factors regulating virulence in our model. This could be particularly relevant because if a single transcriptional regulator effects the expression of numerous genes (either directly or indirectly) a large phenotypic effect could result from a minor genetic change. Interestingly, the relevance of these genes in pathogenesis is illustrated by examples where the inactivation of genes encoding sigma factors (Parish *et al*., 2003b; Malhotra *et al*., 2004) or 2CR systems (Sørensen *et al*., 1995; Kendall *et al*., 2007) causes attenuation of virulence *in vivo*. On the other hand, mutations or deletions in some of these 2CR systems also involve increased virulence (Parish *et al*., 2003a). Strikingly, this was shown for the regulon *devR* in mouse models, while alteration in the same gene has opposite effects in other animal infection models (Sørensen *et al*., 1995). In addition, most of the gene targets for these transcriptional regulators remain unknown up to the present date. This further increases the challenge in assessing the participation of regulons in virulence acquisition. This type of knowledge is essential, since some of these regulators might be involved in inducing or repressing

immunostimulatory molecules (de Souza *et al*., 2008). We identified 10 regulatory proteins, 9 of which were over-represented in the hyper-virulent strain. Of these, only Rv3574 has been investigated in-depth. It represses the expression of an estimated 74 genes in *M. tuberculosis* (Lewis *et al*., 2003), some of them involved in lipid metabolism. It was shown that deletion of Rv3574 leads to attenuation of virulence, which is in accordance with our data.

In summary, our data shows that many of the well characterized factors such as *devR*, *phoP* and others are equally abundant in our model and most probably do not participate in the attenuation of virulence in this case. However, our in depth proteomic analysis allowed us to question that, while lipid metabolism may be an important feature when comparing Beijing strains with other clinical *M. tuberculosis* isolates, such molecules may play a lesser role in virulence within the Beijing strain itself. In addition, we identified several transcriptional regulators that may play a role in virulence. Our results illustrate the full potential of a proteomic approach to select promising candidate molecules and genes for further characterization using the tools of molecular biology.

CHAPTER 5

Label-free quantification of the proteome of the closely related hypo- and hyper-virulent *M. tuberculosis* strains using MaxQuant analysis tool

My contribution to this project:

Culturing of *M. tuberculosis* strains

Protein extractions

Sample preparation for Mass Spectrometry analysis

Data analysis

Manuscript in preparation

5.1. Aim:

Here we used latest state of the art Q Exactive mass spectrometer in combination with the MaxQuant label-free proteomics approach to compare whole cell lysate protein abundance between two closely related *M. tuberculosis* Beijing genotype strains which demonstrate different levels of virulence. Furthermore, we highlight how the use of various label-free quantitation methodologies influences our interpretation of large scale proteome data.

5.2. Results and discussion:

Label-free quantification of whole cell lysate proteins in the hypo- and hyper-virulent clinical *M. tuberculosis* Beijing strains

For an in depth proteomic characterization of whole cell lysate proteins in the hypo- and hyper-virulent *M. tuberculosis* Beijing strains, we adopted a strategy which included the analysis of the whole proteomes of both strains (figure 5.1)

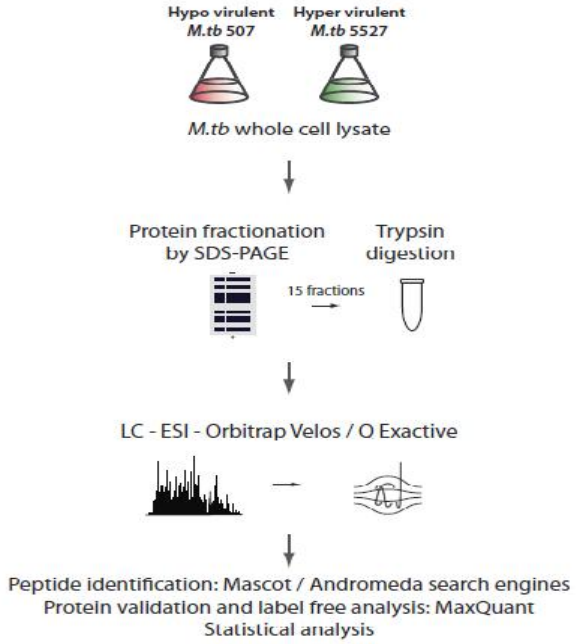


Figure 5.1: Workflow of the quantitative proteomics

For the in depth analysis of the total proteome, whole cell lysates of the hypo- and hyper-virulent *M. tuberculosis* Beijing strains were fractionated by SDS-PAGE (figure. 5.2). For our comparative analysis of these proteomes we used the mass spectrometry data obtained from the resulting 60 LC-MS/MS files (30 files for the hypo-virulent strain and 30 files for the hyper-virulent strain) in MaxQuant. The intensity based absolute quantification (iBAQ) algorithm was used to normalise the summed peptide intensities by the number of the theoretically expected peptides of the protein (Schwanhäusser *et al*., 2011). Using these analysis methods we identified a total of 2790 proteins which were present in both strains.

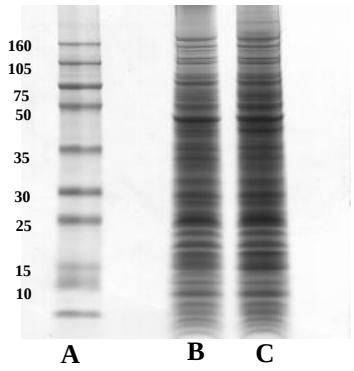


Figure 5.2: Fractionation of whole cell lysate protein using 4-12% SDS PAGE. A) Protein marker ~ 10 to 160 kDa. B) The whole cell lysate fraction of the hypo-virulent *M. tuberculosis* strain. C) The whole cell lysate fraction of the hyper-virulent *M. tuberculosis* strain.

In total we obtained 70% protein coverage of the expected *M. tuberculosis* proteome in one single experiment. To date this is the largest proteome reported for clinical *M. tuberculosis.*

The total iBAQ normalized intensities of the identified proteins of each strain were distributed across 6 orders of magnitude (figure. 5.3A), and when plotted against each other present a *Pearson* correlation of R=0.9755 (figure. 5.3B), showing the high similarity between the proteomes of each strain. All together, the Perseus software for statistical analysis identified 703 differentially abundant proteins based on t-test with a permutation-

based FDR threshold value of >0.01 based on the iBAQ normalised intensities (see Appendix). Figure 5.3C demonstrates a volcano plot of the average iBAQ ratio (hyper/hypo virulent) as a function of the t-test p-values.

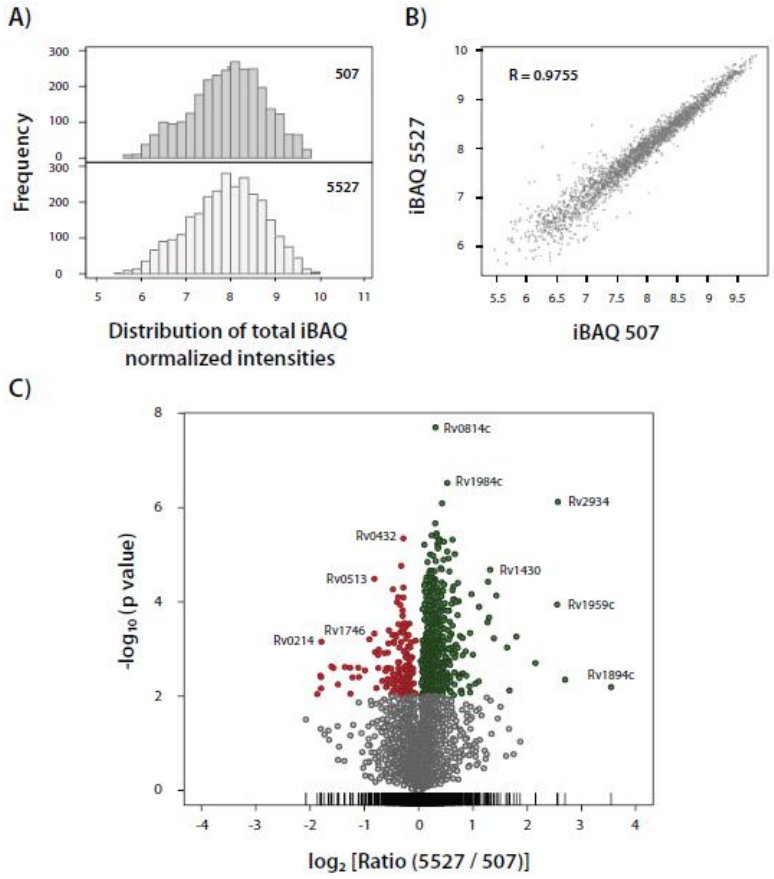


Figure 5.3: Quantitative analysis of the proteome of the hypo- and hyper-virulent clinical *M. tuberculosis* Beijing strains. A) Distribution of the total iBAQ normalized intensities. B) Plot correlation between hypo- (507) and hyper- (5527) virulent strains, inset the Pearson coefficient of correlation (*R*). C) Volcano plot of the average iBAQ ratio hyper/hypo virulent as a function of the t-test p-values. The Rv numbers demonstrate several of the proteins that were not identified using the emPAI label-free quantitation analysis tool.

Among the differential proteins, 565 were significantly abundant in the hyper-virulent *M. tuberculosis* Beijing genotype strain while 138 were significantly abundant in the hypo-virulent *M. tuberculosis* strain. Figure 5.3C also demonstrates some of the proteins that were not identified with the exponentially modified protein abundance index (emPAI) analysis (Chapter 4).

The over-abundant proteins were grouped according to their functional categories as given in Tuberculist (figure. 5.4) (http://tuberculist.epfl.ch/) (Cole *et al*., 1998; Lew *et al*., 2011). Proteins in the functional groups 0 (virulence, detoxification, adaptation), 1 (lipid metabolism), 2 (Information pathways), 7 (Intermediary metabolism and respiration), 9 (regulatory proteins) and 10 (Conserved hypothetical proteins) were significantly over-represented in the hyper-virulent *M. tuberculosis* strain (z-test; $p<0.05$).

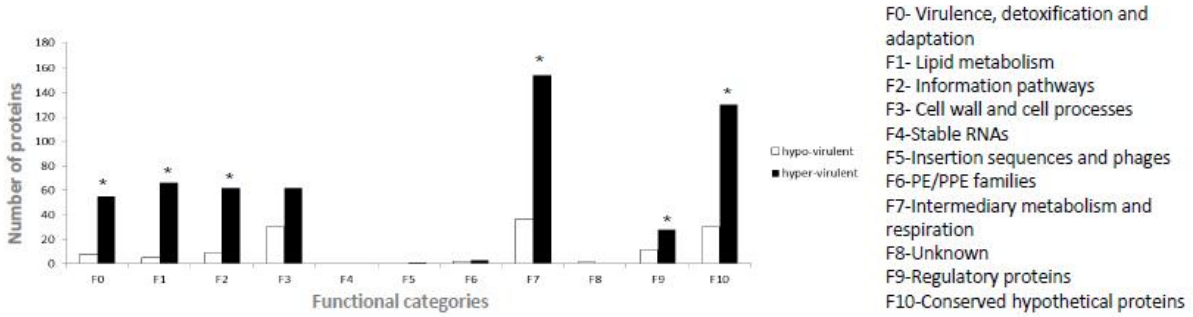


Figure 5.4: Functional categories of the differentially abundant proteins in the hypo- and hyper-virulent *M. tuberculosis* strains according to Tuberculist (http://genolist.pasteur.fr/tuberculisy/). Asterisks represent statistical significance using the z-test ($p<0.05$).

Proteins involved in lipid metabolism, information pathways and intermediary metabolism and respiration can inter-regulate the various biological processes. These proteins are unique since they form a biological network of processes, including lipid metabolism, glycolysis, gluconeogenesis, synthesis and modification of macromolecules as well as protein/peptide secretion (Mawuenyega *et al*., 2005).

We employed a stringent criteria threshold of 2 fold change based on the average iBAQ intensities to define the most abundant proteins in order to identify proteins that could contribute to our understanding of how altered gene expression and protein abundance could contribute to the variation in phenotype of these two closely related *M. tuberculosis* strains. Amongst the most differentially abundant proteins three hypothetical proteins were significantly (P-value <0.01 and greater than 10 fold) over-abundant in the hyper-virulent

strain. The conserved hypothetical protein Rv1894 demonstrated a 129.7 fold increase in the hyper-virulent *M. tuberculosis* strain and when searched against the bacterial bioinformatics resource centre, PATRIC (PathoSystems Resources Integration Centre) (Snyder *et al*., 2007; Gillespie *et al*., 2011) corresponded to locus tag VBIMycTub87468_2114 which encode for 2-nitropropane dioxygenase and is involved in nitrogen metabolism. The conserved hypothetical protein Rv2840c showed an 11.7 fold increase in the hyper-virulent strain and when search against PATRIC (Snyder *et al*., 2007; Gillespie *et al*., 2011) it corresponded to locus tag VBIMycTub87468_3167 which encodes for the a gene product that is a predicted nucleic-acid-binding protein which is implicated in transcription termination. The gene product of *Rv2570* has not been investigated for a probable function.

Sixty-six proteins involved in lipid metabolism were over represented in the hyper-virulent strain. The PDIM locus is a well characterised genome region associated with virulence and is involved in cell envelope trafficking suggesting that it is involved pathogenicity (Camacho *et al*., 2001; Astarie-Dequeker *et al*., 2009). The PDIM locus has been analysed for its involvement in the conditioning of the mycobacterial cell wall. Previous studies have shown that an avirulent H37Rv *M. tuberculosis* strain coated with a PDIM were able to persist longer than the uncoated H37Rv *M. tuberculosis* strain, *in vivo* (Kondo and Kanai, 1974; Camacho *et al*., 2001). Furthermore, it has been shown that mutants containing deletions of genes in this locus have a growth defect in the lungs of intravenously infected mice in comparison to the wild type parental strain (Camacho *et al*., 1999). In our study Rv2934 (PpsD) and Rv2937 (DrrB) which are part of the PDIM locus were significantly over-represented (p-value <0.01). PpsD which is a type-I modular polyketide synthase responsible for the synthesis of the lipid core common to pthtiocerol dimycocerosates (DIM) and PGL were 23.8 times more abundant in the hyper-virulent strain. The integral membrane protein

DrrB is an ABC transporter implicated in the efflux of daunorubicin DIM and was 4.6 times more abundant in the hyper-virulent strain (Schwab *et al*., 2009).

Two esat-6-like proteins Rv1793 (EsxN) and Rv1038 (EsxJ) were over-represented, 2.59 and 2.12 fold respectively, in the hyper-virulent *M. tuberculosis* strain. The ESX-5 secretion system which is the most recently evolved ESX system is responsible for the secretion of EsxN (Houben *et al*., 2012). In a recent study, mutants were generated in order to demonstrate the involvement of the ESX-5 secretion system in the export of PPE/PE transport and virulence. They demonstrated that the disruption of the ESX-5 system leads to loss of PPE secretion, the reduction of cell wall integrity and strong attenuation of virulence *M. tuberculosis*. Furthermore they hypothesise that the attenuation might not be due to the inactivation of the EsxN but rather due to the alteration in the integrity of the mycobacterial cell wall (Bottai *et al*., 2012).

Our comparative label-free quantitative analysis of the proteome of the two clinical *M. tuberculosis* Beijing strains also revealed a number of differentially abundant proteins involved in the main protein biosynthesis and glucose metabolism pathways, including the tricarboxylic acid cycle (TCA) and glycolysis/gluconeogenesis that were significantly (p-value <0.01) over represented in the hyper-virulent strain (Table 5.1).

Here we show proteins involved in glycolysis and the TCA cycle were over represented in the hyper-virulent *M. tuberculosis* strain suggesting that this strain could be a better scavenger of carbon. *M. tuberculosis* can utilise carbon from various sources i.e. (1) fatty acids produced during lipid metabolism, (2) amino acid biosynthesis and (3) glucose (Griffin *et al*., 2012).

Table 5.1: Proteins involved in central carbon metabolism according to PATRIC (http://www.patricbrc.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/pathway.html) which were over-represented in the hyper-virulent *M. tuberculosis* Beijing strain.

		Protein	
Rv number	**Protein name**	**P-value[$]**	**Fold change***
Rv0946c	Glucose-6-phosphate isomerase	$1.7E^{-4}$	1.22
Rv1436	Glyceraldehyde-3-phosphate dehydrogenase	0.005	1.15
Rv1475	Aconitate hydratase	$9.21E^{-5}$	1.28
Rv1098c	Fumarate hydratase	0.001	1.17
Rv2332	Malate dehydrogenase	0.004	2.20

*Calculated ratio using average iBAQ intensities
[$] Calculated using Perseus

It has previously been shown that strains with the *M. tuberculosis* Beijing genotype may have a duplication in the DosR regulon. The DosR regulon regulates several genes involved in lipid metabolism and respiration (Voskuil *et al*. 2004; Reed *et al*., 2007; Domenech *et al*. 2010). According to the proteomics data of the current study, several protein products of the DosR-dependant genes were over-represented (p-value <0.01) in the hyper-virulent *M. tuberculosis* strain (Table 5.2). A preliminary PCR analysis suggests that the 350 kb DosR duplication is absent in the hypo-virulent strain (data not shown).

Table 5.2: DosR-dependant genes products over-represented (p-value <0.01) in the hyper-virulent strain

Rv number	Gene	Predicted function	Fold change*
Rv1177	*fdxC*	Ferredoxin	1.5
Rv1738	-	Conserved hypothetical protein	2.2
Rv1955	*higB*	Regulatory role	100.6
Rv1996	-	Response to stress	1.8
Rv2005	-	ATP binding and response to stress	1.6
Rv2007	*fdxA*	Ferredoxin and transport	2.3
Rv2030	-	Nucleoside metabolic process	1.4
Rv2031	*hspX*	Alpha crystalin	1.8
Rv2032	*Acg*	Dinucleotide-utilizing enzyme	1.6
Rv2632	-	Conserved hypothetical protein	1.4
Rv3143	-	Response regulator	1.2

*Calculated using average iBAQ intensities

In a recent metabolomics study of the hypo and hyper-virulent *M. tuberculosis* strains, metabolite markers associated with virulence phenotype were identified, suggesting an increase in metabolic activity and subsequently leading to increase in growth rate which in turn all can contribute to the ability of the hyper-virulent *M. tuberculosis* Beijing strains' increased virulence and pathogenesis (Meissner-Roloff *et al*., 2012). We propose that an increased expression of proteins involved in the lipid and energy metabolism pathways together with other enzymes taking part in intermediate metabolism could suggest a higher intermediary metabolic activity leading to an increase in growth rate as was for the hyper-virulent strain *in vivo* which can be associated with increased virulence.

A direct correlation between the proteome and genome of the two strains is not obvious (Table 5.8). However, the presence of an nsSNP in the *whiB7* gene in the hypo-virulent strains suggests that the function of this global regulator may be impaired. This together with the absence of a copy of the DosR regulon suggests major alterations in the biology of

the hypo-virulent strain thereby possibly explaining its mild pathology in the mouse model (Reed *et al*., 2007).

Table 5.3: Presence and absence of SNP in hypo- and hyper-virulent *M. tuberculosis* strains

Strain	SNP/indel	Rv number (annotation)	Protein abundance
LV	4bp del	Rv0070c	n.s
LV	prob. IS insertion	Rv0095c	n.s
LV	SNP	Rv0278c	n.s
LV	2SNP	Rv0278c	n.s
LV	SNP	Rv0580c	1.012
LV	3SNPs	Rv1223	n.s
LV	SNP	Rv1384	1.018
LV	SNP	Rv1450c	n.s
LV	SNP	Rv1793	1.051
LV	Ins	Rv1894c	n.s
LV	SNP	Rv1917c	n.s
LV	SNP	Rv1948c	n.s
LV	SNP	Rv1963c;	n.s
LV	SNP	Rv2316	n.s
LV	SNP	Rv2543	n.s
LV	SNP	Rv3197A	n.s
LV	SNP	Rv3347c	n.s
LV	SNP	Rv3425	n.s
HV	SNP	Rv0213c	n.s
HV	SNP	Rv0214	0.883
HV	SNP	Rv0430	0.984
HV	SNP	Rv0753c	1.015
HV	SNP	Rv0777	n.s
HV	SNP	Rv0988	n.s
HV	SNP	Rv1226c	n.s
HV	SNP	Rv2134c	1.033
HV	SNP	Rv2664	n.s
HV	SNP	Rv2687c	n.s
HV	SNP	Rv3542c	n.s

LV- hypo-virulent *M. tuberculosis strain*
HV– Hyper-virulent *M. tuberculosis* strain
n.s – Not significant

This study also highlighted how the use of different methodologies for label-free proteomics analysis and mass spectrometry instruments could impact on the interpretation of results. Analysis of the total proteomes of the hypo- and hyper-virulent strains using the Q Exactive instrument, which has a higher sensitivity and specificity than the LTQ Orbitrap XL instrument (Michalski *et al*., 2011), resulted in the identification of 1146 additional proteins when compared to the emPAI analysis. The emPAI method is based on the rate of identification of peptide/protein in the mass spectrometer, assuming that the more abundant a

protein within a sample, the better the sequence coverage of the protein characterization. Therefore, emPAI estimates the protein abundance in a sample using a ratio derived from the number of observed peptides divided by the number of theoretical peptides of a specific protein. However, as mentioned its quantitation relies on peptide identification and specific chemical weight features, and thereby tends to overestimate the quantities of low molecular weight proteins.

The proteomics data that is presented in this chapter were obtained by using an Area Under Curve (AUC) approach, where chromatographic elution features of each peptide measured by the Q Exactive and peptide peak intensities are calculated by MaxQuant. Furthermore, MaxQuant uses the accurate mass and retention time between runs for protein quantification which means that it is less influenced by proteins size.

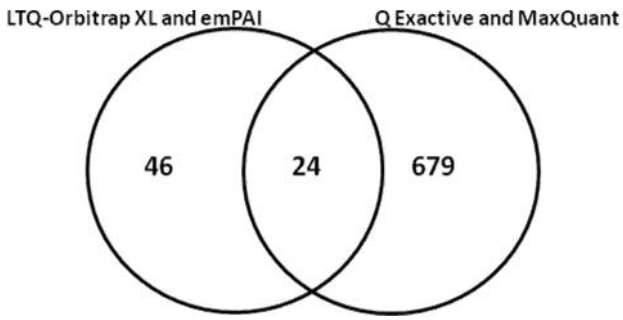


Figure. 5.5: Venn diagram of comparison of over-abundant proteins using LTQ-Orbitrap combined with emPAI versus Q Exactive combined with MaxQuant of hypo- and hyper-virulent *M. tuberculosis* strains.

Even though both label-free quantitative analysis software tools (emPAI and MaxQuant) provide useful information we demonstrate here that by using a more sensitive mass spectrometer, the Q Exactive in combination with MaxQuant analysis. The difference observed in results between the two approaches (Chapter 4 and Chapter 5) is probably because the latest proteomics technology, Q Exactive, is more sensitive and can perform more peptide sequences per second than the Orbitrap XL used in the previous analysis of the hypo- and hyper-virulent strains (Chapter 4) (de Souza *et al*., 2010). Furthermore, weaker

statistical value was added to the emPAI data analysis tool (Chapter 4) while the MaxQuant data analysis package can be used in combination with the Perseus tool to incorporated statistical value to the identification of differentially abundant proteins. Therefore, the differentially abundant proteins calculated in MaxQuant are much more reliable. When we compared the over-abundant protein identified using both methods (Chapter 4 and 5) we identified 679 more over-abundant proteins using MaxQuant (figure 5.5). In addition to the differences observed we cannot exclude that the protein extracts of whole cell lysate proteins from the hypo- and hyper-virulent strains analysed for the two approaches were done on two separate occasions.

In summary, here were present the largest quantitative proteome for *M. tuberculosis* to date. Furthermore we have identified numerous factors that might alter metabolic pathways of the hyper-virulent strain to adapt to survive over a longer period within the host and contribute to the attenuation the virulence phenotype of the transmitting *M. tuberculosis* strain. For example, we show an increase abundance of proteins involved in biosynthesis pathways of several carbon sources, i.e. lipid metabolism, central carbon metabolism (glycolysis/gluconeogenesis and TCA cycle), and amino acid biosynthesis that the mycobacterium bacilli can for energy production.

CHAPTER 6

The phosphoproteome of closely related hypo- and hyper-virulent *M. tuberculosis* strains

My contribution to this project:

- **Culturing of *M. tuberculosis* strains**
- **Protein extractions**
- **Sample preparation for Mass Spectrometry analysis**
- **Data analysis**
- **Manuscript in preparation**

6.1. Aim:

The aim of this study was to use a proteomics approach (LTQ-Orbitrap-Velos combined with MaxQuant label-free proteomics) to identify and characterise the phosphoproteome of the two closely related hypo- and hyper-virulent clinical *M. tuberculosis* Beijing strains with the view to determine the possible role of phosphorylation in pathogenesis. In order to understand the biology and pathogenic mechanisms of *M. tuberculosis* it is relevant to establish which proteins are phosphorylated at the whole proteome level which could identify metabolic pathways associated with virulence and pathogenesis.

6.2. Results and discussion:

Identification of phosphorylated proteins in the hypo- and hyper-virulent clinical *M.tuberculosis* Beijing strains

Whole cell lysate proteins were extracted from the hypo and hyper-virulent *M. tuberculosis* strains (described in chapter 4 and 5) cultured in Middlebrooks media supplemented with catalase and dextrose at midlog growth phase (OD_{600} of 0.7). Phosphorylation states were preserved with the addition of phosphatase inhibitors. Four milligram whole cell lysate proteins from each strain were processed using the FASP method and technical triplicates hypo-and hyper-virulent strains was generated for further analysis (Wiśniewski *et al*., 2009) demonstated in figure 6.1.

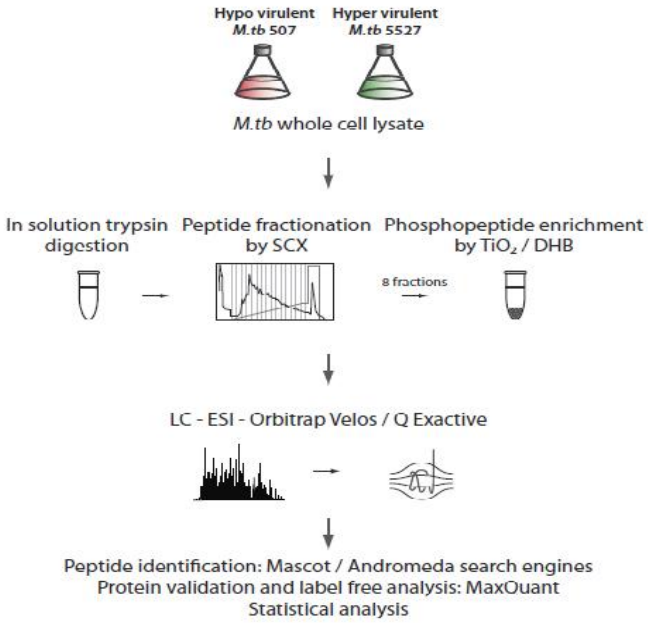


Figure 6.1: Workflow of the quantitative phosphoproteomics

LC-MS/MS enabled the identification of a total of 718 (S/T/Y) phosphorylation sites present on 620 peptides. These peptides mapped to 286 "phosphorylated" proteins (Appendix Table 6.1). This compared favourably with the most extensive data set on Ser/Thr phosphorylation available for *M. tuberculosis* where a slightly higher number of phosphorylated proteins were identified for H37Rv compared to the current study of clinical *M. tuberculosis* (301 *vs.* 266). Interestingly only 98 phosphorylated proteins were identified in both studies which may reflect both differences in the approaches used to increase the identification of phosphorylated proteins. Pooled data from six different growth conditions (Prisic *et al*., 2010) were used while the phosphoproteome presented in the current study represents only one growth condition, midlog growth phase, and we identified more phosphorylated peptides and phosphorylation-sites (Table 6.2). Most of the identified phosphorylated peptides (73%) showed the presence of a single phosphorylation site, while the remaining peptides (23%) showed phosphorylation at two sites. Phosphorylated peptides with >2 phosphorylation sites were not identified (figure 6.2B).

Table 6.1: Comparative phosphoproteome profile of the previously described H37Rv *M.tuberculosis* and both the clinical *M. tuberculosis* Beijing from the current study.

Variable	Prisic *et al.*, 2010	Current study	Hypo-virulent *M. tb*	Hyper-virulent *M. tb*
No. of Phosphorylated proteins	301	286	95	191
No. of Unique Phosphorylated peptides	381	620	185	435
No. of Phosphorylation events	516	718	218	500
Phosphorylated serine	40%	43.0%	101	208
Phosphorylated threonine	60%	53.4%	106	278
Phosphorylated tyrosine	0%	3.5%	11	14

Of the 286 phosphorylated proteins 70 were uniquely identified in the hypo-virulent strain and 166 in the hyper-virulent with just 28 were that were common to both strains (figure 6.2A). We observed the site of phosphorylation on Ser/Thr/Tyr with the ratio being 101:106:11 for the hypo-virulent and 208:278:14 for the hyper-virulent strain respectively. Not surprisingly, the distribution of serine, threonine and tyrosine is in accordance with previously described phosphoproteome of other prokaryotes (Macek *et al*., 2007). We acknowledge that the over-representation of phosphorylation sites in the hyper-virulent strain might be due to technical limitations of the study and only represents a snapshot taken during midlog growth phase of two strains with varying levels of virulence. If these growth conditions might change numerous other phosphorylation sites may be detected and could possibly lead to an overlap of phosphorylation sites during various growth conditions.

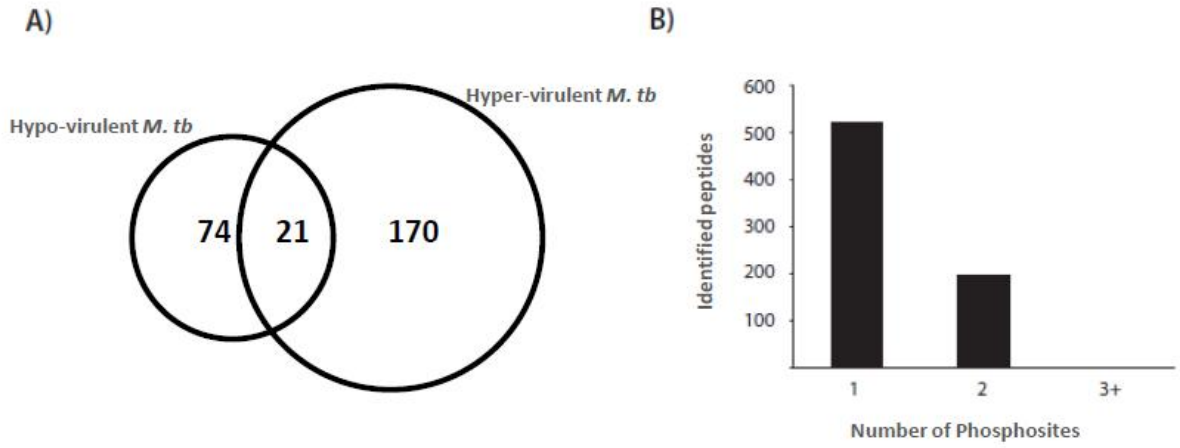


Figure 6.2: The phosphoproteome of the hypo- and hyper-virulent clinical *M.tuberculosis* Beijing strains. A) Phosphorylated protein identification. B) Distribution of phosphorylation across the identified peptides.

The MS/MS spectra of a representative phosphopeptide LGIPQIpSTGELFR from protein Rv0733 (*adk*) is demonstrated in figure 6.3 (where pS is phosphoserine). This adenylate kinase is a small ubiquitous enzyme essential in intracellular nucleotide metabolism. It has also been shown that this protein acts as both a nucleoside mono- and diphosphate kinase and could be implicated in RNA and DNA biosynthesis (Meena *et al*., 2003).

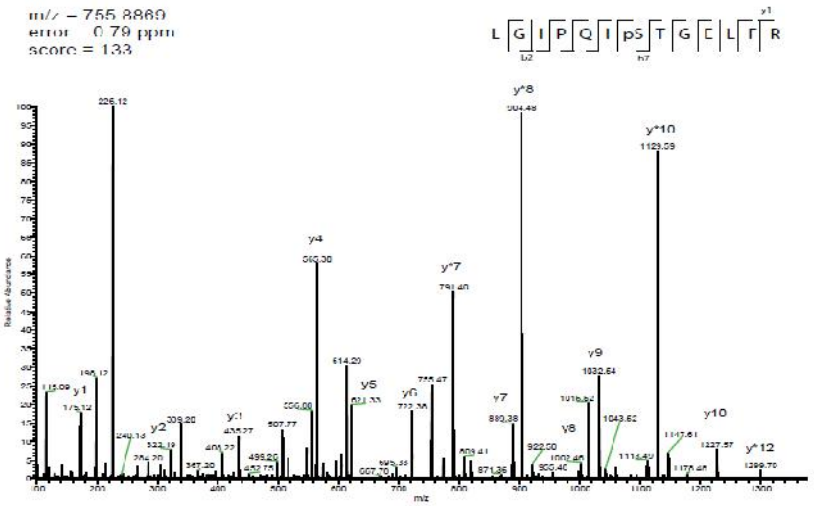


Figure 6.3: MS/MS spectra of the serine phosphorylated peptide LGIPQIpSTGELFR from protein Rv0733 (*adk*)

This is the first study to identify phosphorylated tyrosine in *M. tuberculosis*. The detection of tyrosine phosphorylation using mass spectrometry is normally limited due to its low abundance (Macek and Mijakovic, 2011). Most bacterial phosphorylation sites are on serine (70%) and threonine (20%) while tyrosine phosphorylation events account for less than 10% of the overall phosphoproteome (Macek and Mijakovic, 2011) while serine- and threonine phosphorylation account for 43% and 53.4% respectively. Tyrosine phosphorylated proteins have been previously shown to play important regulatory roles through their involvement in exopolysaccharide production, virulence, DNA metabolism, stress responses amongst other biological functions of the bacterial cell (Kobir *et al*., 2011). Furthermore, the level of tyrosine phosphorylation appears to have a positive correlation with the pathogenicity (Ge and Shan, 2011). In this study, 3.5% of the phosphorylated peptides were tyrosine phosphorylated peptides (Table 6.2). We identified 14 tyrosine phosphorylation sites in 10 phosphorylated proteins in the hyper-virulent strain relative to the 8 phosphorylation sites in 4 phosphorylated proteins in the hypo-virulent strain. Six of the tyrosine phosphorylated proteins identified in the hyper-virulent strain were essential genes for *in vitro* growth (Sassetti and Rubin, 2003; Griffin *et al*., 2011). One of these essential genes Rv0440 (GroEL2) were also amongst the tyrosine phosphorylated proteins. GroEL2 is involved in the prevention of misfolding and promotes refolding and the proper assembly of unfolded polypeptides that is generated under stress conditions. Another tyrosine phosphorylated protein identified in the hyper-virulent strain is *glnA1* which has been implicated in glutamine synthesis and is the only class I glutamine synthase gene that is essential for *M. tuberculosis* growth(Harth *et al*., 2005). It has been shown that the disruption if *glnA1* in *M. tuberculosis* result in glutamine auxotrophy where the mycobacteria cannot synthesize glutamine and could have implications on the amino acids biosynthesis. It has also been shown that this enzyme are released into the culture also plays a crucial role in pathogenesis (Harth *et al*.,

2005). Multiple tyrosine phosphorylated peptides were identified for proteins Rv0020c (fhaA) which is one of the 6 fha-domain containing proteins in *M. tuberculosis*. FhaA is a regulatory protein and have been implicated in cell wall biosynthesis (Fernandez *et al*., 2006) and has a strong association with PknA and PknB (Pallen *et al*., 2002; Roumestand *et al*., 2011).

Table 6.2: Tyrosine phosphorylated proteins identified

Rv number	Gene name	Functional category	Protein function	Strain	Tyrosine phosphorylated peptide
Rv0001	*dnaA*	F2	DNA replication initiation, regulation of DNA replication, ATP-binding	HV	_GDNQHSWPSpYFTERPHNTDSATAGVTSLNR_
Rv0020c	*fhaA*	F9	Conserved hypothetical protein with FHA domain	LV	_HEEGSpYVPSGPPGPPEQR_
				LV	_HPDQGDpYPEQIGYPDQGGYPEQR_
				LV	_HPDQGDYPEQIGpYPDQGGYPEQR_
				LV	_VPGpYAPQGGGYAEPAGR_
				HV	_GGYPPETGGYPPQPGpYPRPR_
				HV	_QDpYGGGADYTRYTESPR_
				HV	_(gl)QDYGGGADpYTR_
				HV	_GGQGQGRPDEpYYDDR_
				HV	_GGQGQGRPDEYpYDDR_
Rv0312	-	F10	Proline and threonine rich protein, ATP-binding	HV	_FEGDpSYNEGGPCWSM(ox)R_
Rv0421c	-	F10	Conserved hypothetical protein	HV	_GLAEGPLIAGGHpSpYGGR_
Rv2198c	*mmpS3*	F3	Integral membrane protein	HV	_AYpSAPESEHVTGGPpYVPADLR_
				HV	_ASGNHLPPVAGGGDKLPSDQTGEpTDApYSR_
Rv0632c	*echA3*	F1	Probable Enoyl-CoA hydratase	HV	_pSDPVSYTR_
Rv0440	*groEL2*	F0	60 kDa Chaperonin 2	HV	_(gl)QEIENSDSDpYDREK_
Rv0613c	-	F10	Conserved hypothetical protein	HV	_IVLAGpYDEELLER_
Rv1513	-	F10	Conserved hypothetical protein	HV	_HQDAFPPANpYVGAQR_
Rv2220	*glnA1*	F7	Glutamine synthase	HV	_DLpYELPPEEAASIPQTPTQLSDVIDR_
Rv3147	*nuoC*	F7	NADH-ubiquinone activity, transport	HV	_GM(ox)FGVpSGTGDTSGYGR_
Rv3398c	*idsA1*	F1	Probable multifunctional geranylgeranyl pyrophosphate synthase	HV	_pYGLPPQPDSDR_
Rv3628	*ppa*	F7	Inorganic diophosphatase activity, Magnesium ion binding	HV	_HFFVHpYK_

LV- hypo-virulent *M. tuberculosis strain*
HV– Hyper-virulent *M. tuberculosis* strain

All the 286 phosphorylated proteins identified in the hypo- and hyper-virulent strains were grouped into their functional categories accoding to Tuberculist (Cole *et al*., 1998; Lew *et al*., 2011) (figure. 6.4). The functional distribution of these phosphorylated proteins shows an over-representation in most of the functional groups in the hyper-virulent *M. tuberculosis* Beijing strain suggesting a relationship between post-translational regulation and virulence.

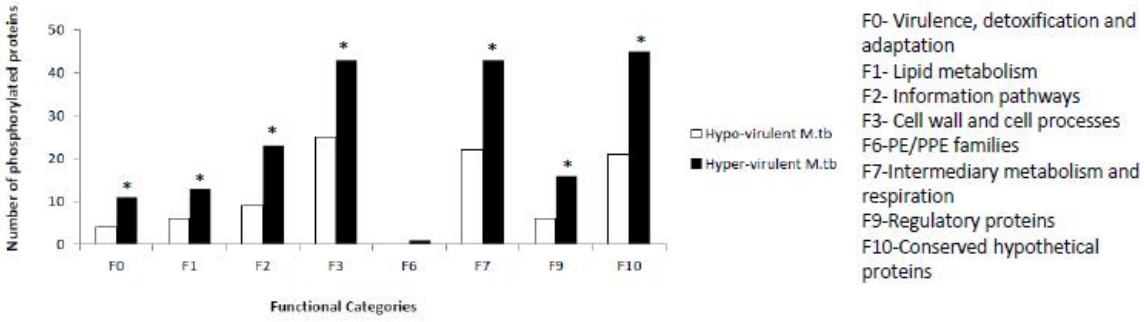


Figure 6.4: Functional categorisation of phosphorylated proteins according to Tuberculist. Asterisks represent statistical significance ($p<0.05$).

In order to verify whether the identified phosphorylation events were post-translationally regulated, we compared the RNA expression of 7 of the 11 STPK's encoded by the *M. tuberculosis* genome. RNA abundance of the seven autophosphorylated STPK's were analysed by qRT-PCR. Figure 6.5 shows a virtual electrophoretic fractionation of purified RNA isolated on separate days. The RNA extractions was considered acceptable further analysis given the presence of the dominant 16S and 23S rRNA species, the A_{260}/A_{280} ratio being between 1.8 and 2.1 and no high molecular weight DNA was visible.

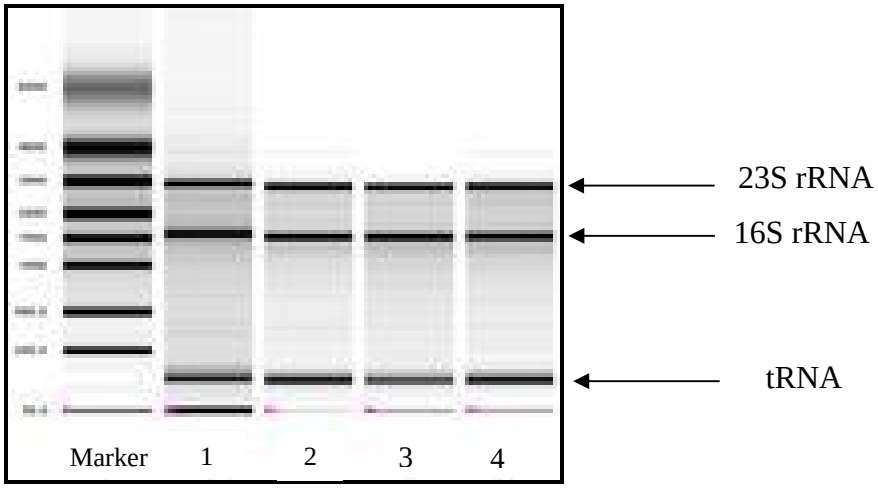


Figure 6.5: Assessment of RNA quality by a virtual gel (Experion software-BIORAD)

From Table 6.3 it is evident that there is no significant difference in the levels of RNA expression of the respective kinases based on the ratio of relative abundance of the protein abundance (chapter 5) and phosphorylated proteins identified (described in this chapter) in the hypo-and hyper-virulent strains. This contrasts with the over-abundance of the phosphorylated form of PknD (14.6 fold) and PknF (5.9 fold) in the hyper-virulent strain.

Table 6.3: Ratio of non-phosphorylated (average iBAQ ratio) and phosphorylated (average intensities) and relative expression of 7 STPKs

Rv number	Gene name	Strain	Proposed regulatory role	Average iBAQ ratio non-phosphorylated protein	Average intensity ratio phosphorylated protein	Relative level of transcription
Rv0015c	*PknA*	HV	Cell elongation and division	0.9	1	n.s
Rv0014c	*PknB*	HV	Cell elongation and division	0.9	2.3	n.s
Rv0931	*PknD*	HV	Phosphate transport	1.1	14.6	n.s
Rv1743	*PknE*	HV	Membrane transport	1.7	not in hypo-virulent *M.tb*	n.s
Rv1746	*PknF*	HV	Membrane transport	0.4	5.9	n.s
Rv0410c	*PknG*	LV, HV	Amino acid uptake, stationary phase metabolism	1.1	0.5	n.s
Rv2176	*PknL*	HV	Transcription	0.4	0.4	n.s

LV- Hypo-virulent *M.tuberculosis* strain –
HV- Hyper-virulent *M.tuberculosis* strain –
n.s –not significant

Seventeen proteins identified with the highest confidence of identification (based on p-value <0.05 and peptides identified in at least 2 technical replicates) were further investigated. Twelve of these phosphorylated proteins identified in the hyper-virulent strain were also identified in the H37Rv phosphorylome in a variety of conditions (Prisic *et al*., 2010). Amongst them, the Rv0503 and Rv2606c, involved in lipid metabolism and intermediary metabolism and respiration, were identified during stationary phase. Rv2606c encodes for pyridoxal biosynthase lyase (Pdx1/SnzP/PdxS), involved Vitamin B6 biosynthesis and is essential for the survival (Dick *et al*., 2010) and has been implicated in the virulence of *M. tuberculosis*. Bacteria that have the ability to utilise pyridoxine pathway can use carbon source providing a precursor for the tyrosine synthesis pathway ((http://umbbd.ethz.ch/pdi/pdi_map.html) Biocatalyst/Biodegeration database, University of Minnesota). The protein Rv0503 has not been investigated for its functional role. Furthermore, 7 phosphorylated proteins that were identified in the hyper-virulent strain were identified in H37Rv when it was grown in acetate as a carbon source (Prisic *et al*., 2010), suggesting that the hyper-virulent strain has the ability or has adapted to utilise carbon and other sources.

In our study we also identified the ATP transporter (Rv1747) and STPK (PknF) that have previously been shown to be involved in the transport of glucose across the membrane (Deol *et al*., 2005). PknF and Rv1747 are membrane proteins transcribed from in the same operon (Curry *et al*., 2005). Both of these proteins (in their phosphorylated state) were identified in the hyper-virulent *M. tuberculosis* strain with high confidence; PknF showed a 5.9 fold increase while Rv1747 showed a 2.3 fold increase based on the ratio of their average intensities. PknF interact with its substrate Rv1747 via a phospho-specific interaction (Spivey *et al*., 2011). Rv1747 has also been characterised as an efflux pump and has also

been implicated in rifampicin resistance (Louw *et al*., 2011). Rv1747 is also the substrate for 4 of the 11 *M. tuberculosis* STPK's, i.e. PknB, PknD, PknE and PknF (Grundner, *et al*., 2005). Importantly, Rv1747 contain 2 of the 7 Forkhead-associated (FHA) domains encoded by the 6 FHA domain containing proteins in the *M. tuberculosis* genome. FHA domains are phosphopeptide recognition motifs spanning 80-100 amino acid residues that is folded in an 11-stranded beta sandwich which recognises and mainly binds phospho-threonine residues in specific peptides or proteins (Nott *et al*., 2009). Most FHA domains recognise phospho-threonine with specificity provided by residues C-terminal to the phospho-threonine residue, particularly the +3 position. PknF interacts with the phosphorylation sites at Thr-150 and Thr-208 of Rv1747 in order to exhibit ATPase activity to perform in the uptake of glucose (Molle *et al*., 2004). It has previously been shown that Rv1747 is required for virulent infection by *M. tuberculosis* in mice (Curry *et al*., 2005). We propose that an increase in uptake of glucose could lead to an increased metabolic activity of the cell which may be associated with hyper-virulence.

The autophosphorylated PknD was also identified in the hyper-virulent strain. It has been previously shown that increasing PknD activity in *M. tuberculosis* resulted in the specific phosphorylation of a single anti-anti-sigma factor homolog, Rv0516c (Greenstein, MacGurn, *et al*., 2007). Even though the phosphorylated form of Rv0516c was not identified, the non-phosphorylated protein was over represented in the hyper-virulent strain (chapter 5).

Only four proteins (Rv0458, Rv0178, Rv0813 and Rv3799) containing phosphorylated peptides were uniquely identified with high confidence (based on P-values <0.05) in the hyper-virulent *M. tuberculosis* strain. Three of these proteins have been implicated in lipid metabolism while Rv0178 (mce-associated membrane protein) in the mce operon 1 (*mce1*) and has directly been implicated in virulence (Casali *et al*., 2006).

Whole genome sequencing of the two closely related strains, the hypo- and hyper-virulent *M. tuberculosis* strains revealed 51 non-synonymous single nucleotide polymorphisms (nsSNP) were identified in the two strains (data not published). The SNPs identified with high confidence in either the hypo- and hyper-virulent strain with the highest confidence is given in table 5.3. Amino acid changes in genes that are involved in regulatory function, or that are part of the upstream signal transduction pathways were significantly over represented among the SNP's that were intragenic and fixed in the typical Beijing lineage. Numerous biological processes in bacteria are regulated by transcription and these changes in the regulatory networks are thought to significantly contribute to phenotypic diversity (Perez *et al*., 2009). It has previously been shown that strains of the W/Beijing show up-regulation of the DosR regulon which is also responsible for the regulation of various genes involved in lipid metabolism (Reed, 2007). Proteins involved in DNA transcription was observed in the hyper-virulent strain when compared to the hypo-virulent strain (described in chapter 4) (de Souza *et al*., 2010). A double mutation identified in the PknA sensor domain which has shown to be crucial in the response to signals (Thakur *et al*., 2008) and also contributes to the specificity of the kinase. Strikingly PknA was also found to be phosphorylated and was identified with high confidence (peptides identified by MS/MS in two technical replicates) in the hyper-virulent strain. Furthermore, an amino acid mutation was also identified in the response regulator KdpE which is also crucial in responding to signals. Interestingly, the sensor kinase of this two-component system KpdD was found to be phosphorylated in the hyper-virulent strain (Vlisidou *et al*., 2010). These mutations of the mycobacterial regulatory network might have contributed to the alteration of cellular responses of these two strains to environmental signals and lead to varying levels of virulence.

In summary, we propose that the increased identification of phosphorylation events in the hyper-virulent strain contributes to altered gene expression due to post-translational regulation and contributes significantly to the phenotypic differences, i.e. varying level of virulence, between these two closely related *M. tuberculosis* strains. Most importantly, we believe that our dataset extend and complement the previous knowledge regarding the phosphorylated peptides and phosphorylation sites. This depth phosphoproteomics analysis of the hypo and hyper-virulent *M.tuberculosis* Beijing strains provide a novel tool to identify and characterise proteins involved in various biological processes. The identification of an altered metabolic pathway in the hyper-virulent *M.tuberculosis* strain provides novel insight into understanding virulence mechanisms.

CHAPTER 7

Proteome analysis of a rifampicin mono-resistant *M. tuberculosis* strain and the wild type progenitor strain

My contribution to this project:

Culturing of *M. tuberculosis* strains

Protein extractions

Sample preparation for Mass Spectrometry analysis

Data analysis

Manuscript in preparation

7.1: Introduction

Drug resistance in *Mycobacterium tuberculosis* develops spontaneously and the resulting drug resistant mutants are selected during periods of poor adherence, mono-therapy and the administration of inappropriate treatment. *In vitro* studies analysing spontaneously generated drug resistant *M. tuberculosis* mutants have shown a direct correlation between mutations in a specific target gene and resistance to the anti-tuberculosis (TB) drug used to select the resistant mutants (Bobadilla-del-Valle *et al*., 2001). Consequently, the relationship between a defined mutation and resistance to a specific anti-TB drug has become a dogma and now forms the basis for the development of molecular-based drug-resistance diagnostic assays (Georghiou *et al*., 2012). However, the above studies have failed to exclude the possibility that additional modulatory mutations may occur concurrently in association with the evolution of drug-resistance. Furthermore, it is not known whether mutation in certain target genes could influence the transcriptome/proteome. Non-synonymous single nucleotide polymorphisms (nsSNPs) conferring RIF resistance occur exclusively in the *rpoB* gene, with 95% occurring within the 81bp region known as the RIF resistance-determining region (RRDR) (Telenti *et al*., 1993). Generally, these snSNPs result in the replacement of aromatic amino acids with non-aromatic amino acids causing drug resistance by altering the forces that bind RIF to the RNA polymerase (Campbell *et al*., 2001). Furthermore, amino acid substitutions in the essential ß-subunit of RNA polymerase are thought cause alteration in the expression of genes in RIF resistant strains leading to subsequent changes in various metabolic pathways (Du Preez and Loots, 2012).

7.2. Aim:

In this study we aimed to determine whether the *in vivo* emergence of RIF resistance altered protein abundance in a clinical isolate of *M. tuberculosis*, with the view to determine the impact of RIF resistance on the physiology of the bacilli. We used a proteomic approach which combined the latest state of the art mass spectrometer (Q Exactive) with the label-free data analysis tool MaxQuant to perform a quantitative proteomics analysis of a RIF-mono-resistant mutant and wild type parent *M. tuberculosis* strain.

7.3. Results and discussion:

7.3.1. Selection of RIF mono-resistant mutant and wild type *M. tu*berculosis progenitor strains

DNA sequencing of the *rpoB* gene of a RIF-mono-resistant clinical isolate demonstrated both a wild type (TCG) and a mutant (TTG) codon at position 531, suggesting hetero-resistance. Single colony forming units were selected in the presence and absence of RIF and the respective *rpoB* genes were sequenced to identify resistant and susceptible strains. Hetero-resistance was confirmed by spoligotyping which demonstrated that both the RIF-mono-resistant mutant and wild type progenitor *M. tuberculosis* strain were members of the Beijing genotype (Kamerbeek *et al*., 1997).

7.3.2. Label free comparative quantification of proteomes of the *M. tuberculosis* Beijing *rpoB* mutant and the wild type parent strain

Whole cell lysate proteomes extracted from the *M. tuberculosis* Beijing RIF-mono-resistant mutant and the wild type progenitor strain were fractionated by gel-LC-MS/MS. Label-free

quantification of proteins was done using the MaxQuant software package; comparing the levels of protein abundance between the RIF-mono-resistant mutant and the wild type progenitor strain as described in Chapter 3 (figure 7.1). A total of 90 LC-MS/MS files were analyzed from 3 replicates of the RIF-mono-resistant mutant and wild type strain, respectively.

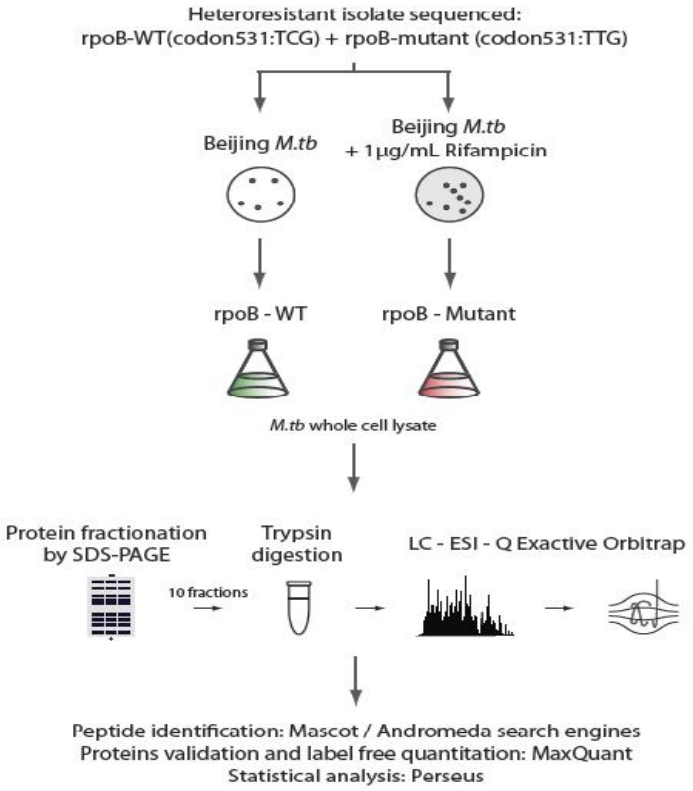


Figure 7.1: Workflow of the proteome analysis of the RIF-mono-resistant mutant and the wild type progenitor *M. tuberculosis* strain.

Figure 7.2 demonstrates a volcano plot of the average iBAQ ratio (described in Chapter 5) of proteins extracted from the two strains as a function of their t-test p-values. Using this algorithm we identified 2703 proteins present in both the RIF-mono-resistant mutant and the

wild type progenitor strain. Of these, 1885 (69%) proteins were present in equal abundance, while 818 (31%) showed varying degrees of abundance.

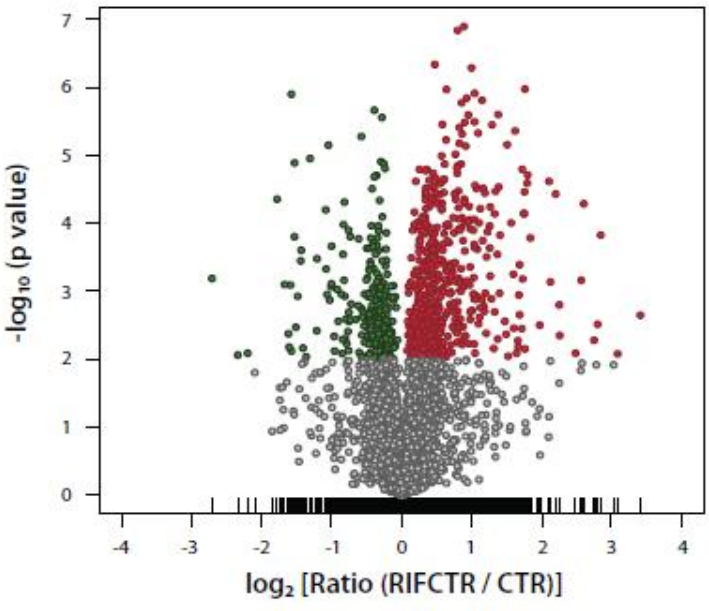


Figure 7.2: Volcano plot of the average iBAQ for the RIF-mono-resistant mutant and the wild type progenitor strain as a function of the t-test p-values. The spots in green indicate decrease in abundance, while spots in red indicate proteins with an increased abundance after acquisition of the *rpoB* mutation.

From this data set we showed that the abundance of 189 proteins was significantly decreased, while the abundance of 434 proteins was significantly increased as a consequence of the evolution of RIF resistance [Perseus software for statistical analysis (unpublished) tool in combination with t-test (FDR threshold set to a p-value of ≤0.01 – calculated from the normalised iBAQ intensities (Chapter 5)] Proteins with an increased abundance could be grouped into 7 functional categories according to Tuberculist (Cole *et al*., 1998; Lew *et al*., 2011).

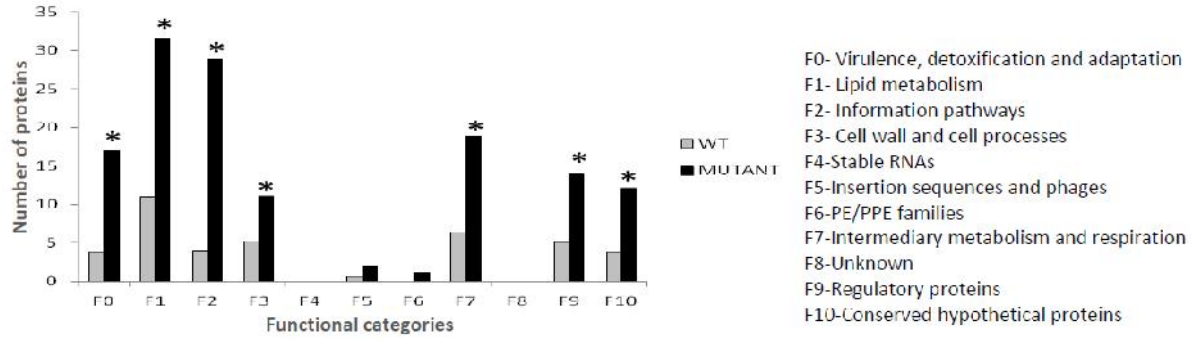


Figure 7.3: Differentially abundant proteins of RIF-mono-resistant mutant compared to the wild type progenitor strain according to the different functional categories. * indicates significant difference (z-test P<0.05) between the two strains.

Proteins involved in virulence, detoxification and adaptation (functional category 0), lipid metabolism (functional category 1), information pathways (functional category 2), cell wall and cell processes (functional category 3), intermediary metabolism and respiration (functional category 7), regulatory proteins (functional category 9) and conserved hypothetical proteins (functional category 10) were significantly over-abundant in the RIF-mono-resistant mutant. This suggests that the S531L *rpoB* mutation influences the regulation of several biological processes including lipid metabolism, energy metabolism and amino acid metabolism which all influences the mycobacterial cell wall composition.

Our comparative analysis showed that 8 significantly abundant proteins selected with the Perseus-tool to be significant (p-value < 0.01) and had a 10 fold increase in the RIF-mono-resistant mutant strain (Table 7.1). Probable protein functions have been assigned to only 5 proteins (Rv1738, RsfB, MoaE2, VapC32 and AhpC); however these proteins have not been investigated for their role in drug resistance. The hypothetical protein Rv1738 is an essential protein (Sassetti and Rubin, 2003) and its expression is regulated by the DosR regulon (Bartek *et al.*, 2009).

RsfB is one of 6 molecules with putative anti-sigma factor antagonist or stress signalling proteins and is over-expressed during cold and oxidative stress (Dhandayuthapani, 2007). In

our dataset we also identified the probable molybdenym cofactor biosynthesis E2 (MoaE2) protein which showed a 17.2 fold increase in abundance in the RIF-mono-resistant mutant. In mycobacteria molybdenum enzymes are involved in the global carbon, sulphur and nitrogen cycles which catalyse important redox reactions which have implications for several biological functions including dormancy regulation (Shi and Xie, 2011). The abundance of the VapC32 protein was increased by 18.1 fold in the RIF-mono-resistant mutant and forms part of the toxin-antitoxin systems which has been implicated in bacterial programmed cell death (PCD) and persistence in stress conditions (Arcus *et al*., 2005). It has been hypothesised TA systems may be important for changing growth rate, inducing long term dormancy and in generating bacilli that can survive multidrug chemotherapy (Dahl *et al*., 2003; Beste *et al*., 2009; Han *et al*., 2010; Frampton *et al*., 2012). The abundance of the alkyl hydroperoxide reductase C protein (AhpC) was increased significantly by 10.9 fold in the RIF-mono-resistant mutant and has been shown to compensate for the catalase-peroxidase deficiency caused by oxidative and nitrosative stress (Guimarães *et al*., 2005).

Table 7.1: Eight significantly abundant proteins with (p-value < 0.01 and 10 fold increase) in the RIF-mono-resistant mutant strain

Protein IDs	Gene name	Essential/ Non-essential	Protein description	Functional category	Ratio*	p-value[$]
Rv1489	-	Nonessential	Conserved hypothetical protein	F10	27.4	< 0.01
Rv1738	-	Essential	Conserved hypothetical protein	F10	11.2	< 0.01
Rv1352	-	Non-essential	Conserved hypothetical prototein	F10	10.0	< 0.01
Rv1117	-	Non-essential	Conserved hypothetical protein	F10	10.0	< 0.01
Rv3687c	*rsfB*	Non-essential	Anti-anti-sigma factor	F2	17.8	< 0.01
Rv0866	*moaE2*	Non-essential	Probable molybdenym cofactor biosynthesis E2	F7	17.2	< 0.01
Rv1114	*vapC32*	Non-essential	Conserved hypothetical protein with PIN domain	F0	18.1	< 0.01
Rv2428	*ahpC*	Non-essential	Alkyl hydroperoxide reductase C protein	F0	10.9	< 0.01

[*]Calculated using average iBAQ intensities

[$]Calculated using Perseus

Together this suggests that the perturbation in the structure of the RNA polymerase up-regulates certain stress-responses which may directly impact on the growth fitness of the mutant (Gagneux *et al*., 2006).

Within the theme of a stress response, we observed that of 32 proteins (of the 48) genes regulated by the DosR regulon (Kendall *et al*., 2004; Voskuil *et al*., 2004b) were significantly over-abundant (p-value < 0.01) in the RIF-mono-resistant mutant (Table 7.2). Previously it has been shown that the DosR regulon is up-regulated under various stress conditions, including hypoxia (Kendall, 2003) and its expression allows for long term survival under aerobic conditions in a non-replicating state (Park, Guinn, *et al*., 2003). Bartek and colleagues showed that the DosR regulon in itself did not play an active role in *M. tuberculosis* survival but suggested that the induction of a metabolically dormant-like state could allow antibiotic tolerance (Bartek *et al*., 2009). Together this suggests that the perturbation in the structure of the RNA polymerase up-regulates certain stress-responses which may directly impact on the growth state of the RIF-mono-resistant mutant which may make it more resilient to antibiotic treatment (Louw *et al*., 2011). Furthermore, the observed change in gene expression may explain the fitness cost associated with the evolution of RIF resistance (Gagneux *et al*., 2006).

Table 7.2: Significantly abundant proteins regulated by DosR regulon in RIF-mono-resistant mutant

Rv number	Gene name	Function	Functional category	P-value*	Fold change[$]
Rv0080	-	Unknown	F10	0.03	2.9
Rv0081	-	Involved in transcriptional mechanism	F9	0.001	3.89
Rv0569	-	Unknown	F10	$8.25E^{-5}$	3.06
Rv0570	*NrdZ*	Involved in DNA replication pathway	F2	$3.05E^{-5}$	4.43
Rv0571	-	Phosphoribosyl transferase domain protein	F10	$3.88E^{-4}$	2.55
Rv0574	*CapA*	Capsule biosynthesis protein	F10	0.08	4.79
Rv1736	*NarX*	Involved in nitrate reduction and persistence in the host	F7	$7.10E^{-5}$	6.54
Rv1737	*narK2*	Involved in excretion of nitrate	F3	0.011	3.02
Rv1738	-	Unknown	F10	$1.05E^{-6}$	11.2
Rv1813	-	Possible membrane protein	F10	0.007	20.61
Rv2004	-	Unknown	F10	$1.54E^{-6}$	4.50
Rv2005	-	Universal stress protein	F0	$2.55E^{-6}$	4.09
Rv2006	*otsB1*	Trehalose biosynthesis	F0	$3.05E^{-5}$	3.45
Rv2007	*FdxA*	Involved in electron transfer	F7	$6.93E^{-6}$	6.58
Rv2028c	-	Universal stress protein	F0	$4.37E^{-6}$	8.18

Rv2029	*PfkB*	Involved in glycolysis	F7	$1.49E^{-4}$	6.19
Rv2030	-	Erythromycin esterase homolog	F10	$2.52E^{-6}$	6.72
Rv2031	*acr, hspX*	Universal stress protein	F0	$4.04E^{-4}$	2.23
Rv2032	*Acg*	Dinucleotide-utilizing enzyme	F10	$1.44E^{-7}$	3.13
Rv2623	*TB31.7*	Universal stress protein	F0	$1.41E^{-5}$	3.50
Rv2624	-	Universal stress protein	F0	$2.75E^{-5}$	4.08
Rv2625	-	Probable conserved transmembrane alanine and leucine rich protein	F3	$2.86E^{-4}$	5.34
Rv2626	*hrp1*	CBS domain protein	F10	$1.93E^{-4}$	2.64
Rv2627	*DosT*	Unknown	F10	$1.7E^{-6}$	3.21
Rv2629	*PfkB*	Unknown	F10	$1.41E^{-4}$	2.33
Rv2630	-	Archease	F10	0.017	3.57
Rv3127	-	Involved in lipid metabolism	F10	$3.88E^{-6}$	3.33
Rv3130c	*tgs1*	Involved in synthesis of triacycglycerol	F1	$5.18E^{-7}$	3.85
Rv3131	-	Unknown	F10	$4.72E^{-6}$	4.57
Rv3132	*DosS*	Regulatory protein	F9	0.011	2.25
Rv3133	*DosR*	Regulatory protein	F9	$9.87E^{-5}$	3.84
Rv3134	-	Universal stress protein	F0	0.011	3.78

[*]Calculated using average iBAQ intensities

[$]Calculated using Perseus

In a recent metabolomics study of two *rpoB* mutants, it was shown that the S531L mutant had an advantage for using alternative energy sources in the form of fatty acids and cell wall lipids. Furthermore they also demonstrated that the *rpoB* mutation contributed to an altered fatty acid metabolism (Du Preez and Loots, 2012). We identified 39 of the protein products of the genes involved in fatty acid biosynthesis listed by Takayama *et al*. (Takayama *et al*., 2005). However, only twelve of these proteins were significantly (p-value <0.05) over-abundance in the RIF-mono-resistant strain (Table 7.3). Interestingly, the IrtB protein was 58 times more abundant in RIF-mono-resistant strain. This protein together with IrtA are also required for growth in iron deficient conditions (Rodriguez and Smith, 2006). It is predicted that these proteins form an ABC transporter involved in iron sequestration (Rodriguez and Smith, 2006). No function has been associated with drug resistance.

Table 7.3: Over-abundant (p-value < 0.05) in the RIF-mono-resistant strain proteins involved in the anabolic pathway of fatty acid metabolism of *M. tuberculosis*

Rv number	Gene name	Protein function	P-value[$]	Fold Change*
Rv2243	*fadD*	Catlaysis malonyl-CoA-ACP transacylase activity	0.048	1.3
Rv2244	*acpM*	Fatty acid biosynthesis	0.006	1.2
Rv0533	*fabH*	Fatty acid biosynthesis	0.049	1.3
Rv1142c	*echA10*	Oxidise fatty acids	0.001	2.1
Rv1141c	*echA11*	Oxidise fatty acids	0.001	1.7
Rv1484	*inhA*	Mycolic acid biosynthesis	0.004	1.3
Rv0644c	*mmaA2*	Mycolic acid modification	0.007	1.2
Rv0643	*mmaA3*	Mycolic acids modification	0.007	1.3
Rv3400	-	Probably involved in cellular metabolism	0.006	1.4
Rv2006	*otsB1*	Involved in trehalose biosynthesis	$3.05E^{-5}$	3.5
Rv1349	*irtB*	Involved in iron homeostasis	0.043	58
Rv1747	-	Involved in active transport of lipooligosaccharide across membrane	$2.44E^{-4}$	1.7

*Calculated according to average intensities (iBAQ)
[$] Calculated using Perseus

In summary, this is the first comprehensive analysis of the effect of the *rpoB* mutation on the proteome of a RIF-mono-resistant mutant strain containing a S531L mutation. Based on the over-representation of 32 proteins in the DosR regulon, the proteome analysis presented in this study suggests that the *rpoB* mutation allows for the bacteria to enter into a metabolically altered state. Furthermore, we identified over-represented proteins, vapC32, moaE2, rsfB and aphC in the RIF-mono-resistant mutant that has previously been shown to be over-expressed under various stress conditions. These findings support a previous studies that the DosR regulon may be important for bacteria to tolerate exposure to antibiotics (Bartek *et al*., 2009).

CHAPTER 8

The label-free quantification of the proteome of a multi-drug resistant *M. tuberculosis* strain before and after exposure to rifampicin

8.1. Aim:

In this study we aimed to determine the influence of RIF on the proteome of a multi-drug resistant (MDR) *M. tuberculosis* strain using a quantitative proteomics approach. This study aimed to build on the previous findings which showed that RIF induced an up-regulation of certain efflux and transporter genes in RIF resistant *M. tuberculosis* isolates. We hypothesized that RIF induced a stress response in these strains which in turn lead to the activation of efflux and transporter genes which would regulate the intracellular concentration of RIF. For this study we used an LC/MS/MS (LTQ-Orbitrap-Velos) approach and analyzed the MS data with MaxQuant label-free proteomics software to quantify whole cell lysate proteins extracted from an MDR *M. tuberculosis* isolate (Louw *et al*., 2011) with and without exposure to RIF for 24 hours.

8.2. Results and discussion:

8.2.1. Label-free quantification of the MDR *M. tuberculosis* proteome with and without 24 hours exposure to 2 µg/ml RIF

To identify the difference in abundance of proteins between the RIF exposed and unexposed MDR *M. tuberculosis* cultures, the whole cell lysate proteins were analysed using the gel-LC-MS/MS approach as described in figure 8.1. Label-free quantification of the proteins was done using the MaxQuant software package; we compared the levels of protein abundance between the RIF exposed and unexposed MDR *M. tuberculosis* isolate as described in Chapter 3. A total of 90 LC/MS/MS files were analysed from 3 replicates of each growth condition.

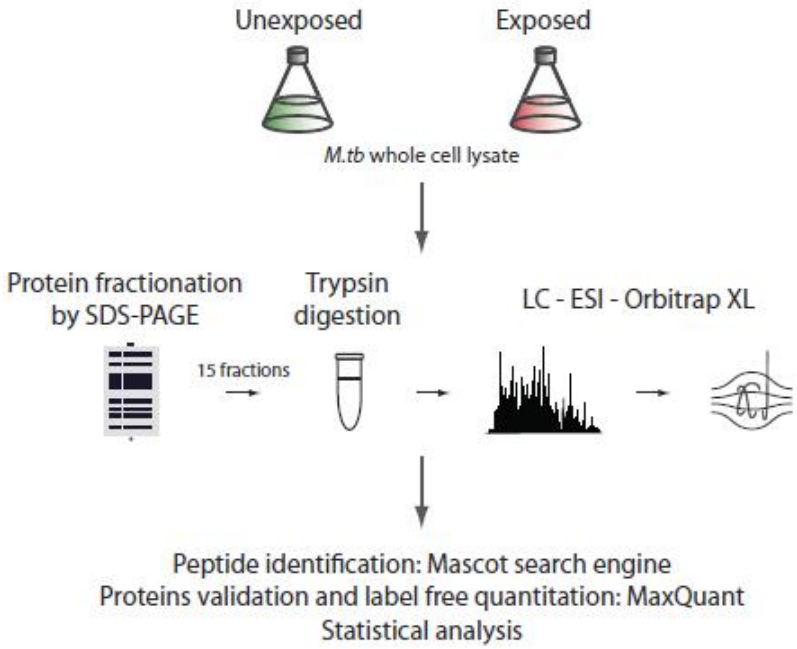


Figure 8.1: Workflow for the analysis of the MDR *M. tuberculosis* isolate with and without exposure to the critical concentration of RIF (2 µg/ml)

In our comparative analysis we used the intensity based absolute quantification (iBAQ) algorithm, which normalizes the summed peptide intensities by the number of the theoretical peptides of the protein (described in Chapter 5). Using this approach we identified a total of 2044 proteins that were present under both growth conditions. Of these, 1553 (76%) proteins were present in equal abundance. Figure 8.2 demonstrates a volcano plot of the average iBAQ ratio of proteins extracted from the exposed and unexposed MDR *M. tuberculosis* isolate as a function of the t-test p-values.

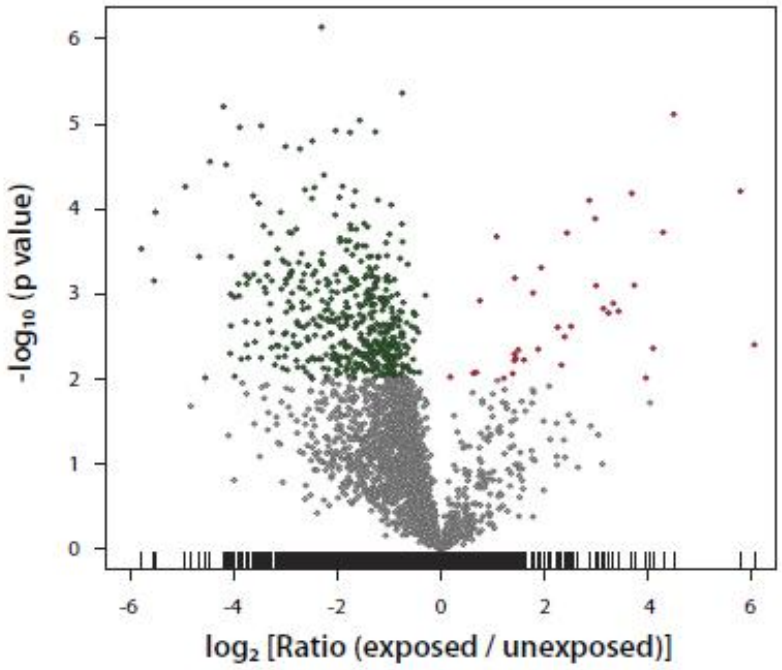


Figure 8.2: Volcano plot of the RIF exposed/unexposed *M. tuberculosis*. Average iBAQ ratio as a function of the t-test p-values. Green spots represent the differentially abundant proteins of unexposed MDR *M. tuberculosis* isolate; red spots represent the differentially abundant of RIF exposed MDR *M. tuberculosis* isolate. Grey spots represent proteins in equal abundance

Over-represented proteins in exposed and unexposed MDR *M. tuberculosis* isolates

Perseus software (unpublished) in combination with the t-test (FDR threshold set to a *P-value* of 0.01 - calculated from the normalized iBAQ intensities) were used to determine the proteins that were significantly over-represented in either exposed and unexposed MDR *M. tuberculosis* isolate (as described in Chapter 5). We showed that the abundance of 502 proteins was significantly decreased on exposure to RIF, while the abundance of 44 proteins was significantly increased under these conditions. Proteins with a decreased abundance could be grouped into 7 functional categories according to Tuberculist (figure 8.3).

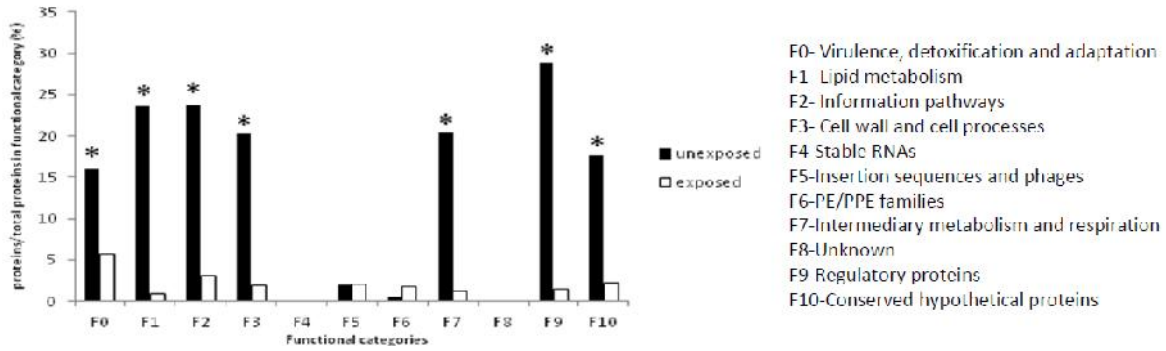


Figure 8.3: Functional categorisation of differentially abundant proteins in MDR *M. tuberculosis* isolate with and without exposure to RIF. Abundant proteins represented as % of the identified proteins divided by the total number of proteins characterised according to Tuberculist. Asterisks represent statistical significance using the z-test ($p<0.05$).

This result suggested that RIF induces a rapid decrease in the abundance of a subset of the proteins involved virulence, detoxification and adaptation (functional category 0), information pathways (functional category 2), cell wall and cell wall processes (functional category 3), intermediary metabolism and respiration (functional category 7), regulatory proteins (functional category 9) and conserved hypothetical proteins (functional category 10). Conversely, RIF exposure did not influence the abundance of proteins involved in stable RNAs (functional categories 4), insertion sequences and phages (functional categories 5) and PE/PPE (functional categories 6).

Results obtained in this study indicate that the presence of RIF continues to cause an environmental stress on the RIF resistant MDR *M. tuberculosis* isolate, leading to the reduction in several biological processes. This differs from the dogma which suggests that RIF resistance is simply a reflection of the change in the binding constant between RIF and the RNA polymerase due to mutation in the *rpoB* gene (Campbell *et al*., 2001).

More recently it has been suggested that prolonged exposure may contribute to the evolution and adaptation of *M. tuberculosis* whereby it could acquire putative compensatory mechanisms to survive (Gagneux *et al*., 2006). The putative compensatory mechanisms

could include the production of enzymes that could alter or inactivate the antibiotic, decrease the cell wall permeability or alternative mechanisms such as efflux pumps (Singh *et al*., 2011). Previous studies have demonstrated the association between efflux pump activation and RIF resistance where a number of these efflux pumps (*Rv1258c*, *Rv1410c*, *Rv1819c*, and *Rv2136c*) were up-regulated in the presence of RIF (Wilson *et al*., 1999; Siddiqi *et al*., 2004; Geiman *et al*., 2006; Jiang *et al*., 2008; Louw *et al*., 2009). To test this hypothesis we interrogated the proteomic data to determine if these efflux pumps were over-represented in the RIF exposed MDR *M. tuberculosis* isolate (Table 8.1).

Table 8.1: Efflux pumps identified in the current study which were previously shown to be up-regulated in the presence of RIF.

Rv number	Gene name	Fold change (RIF exposed vs. unexposed)*	P-value[$]
Rv2989	-	1.5	0.155
Rv0342	*iniA*	0.2	0.005
Rv1819	*bacA*	0.3	0.341
Rv1463	-	0.7	0.224
Rv1747	-	0.13	0.006
Rv2994	-	0.02	0.006
Rv3806	*ubiA*	0.1	0.005
Rv3679	-	0.4	0.043
Rv2942	*mmpL7*	0.3	0.027
Rv2688	-	0.3	0.172
Rv2936	*drrA*	0.2	0.001
Rv2937	*drrB*	1.41	0.579

* Calculated ratio based on average iBAQ intensities
[$] Calculated in Perseus

Only 12 of the previously described efflux pumps were identified in this study (Cole *et al*., 1998; Gupta *et al*., 2010; Louw *et al*., 2011). However significant increase in their abundance in response to RIF could not be demonstrated. Similarly, our analysis could not demonstrate an increase in the abundance in the ATP synthase subunits in response to RIF. This differs from a Quantitative Real-Time PCR (qRT-PCR) study which demonstrated that the ATP synthase operon was up-regulated in response to RIF (unpublished data, M de Vos, PhD, 2012). This may be explained by the fact that a direct comparison cannot always be made between the transcriptome and proteome due to methodological limitations. Protein sequence coverage of membrane proteins may be limited due to their inherent physical properties which include the insolubility as well as the absence of charged residues in the membrane spanning domains. This limitation can be partially overcome by the inclusion of detergents (such as Triton X-114) in the extraction protocol to increase the solubility of membrane proteins (Målen *et al*., 2008). However, in the current study we did not employ this strategy which leads to the under-representation of cell wall and lipid membrane proteins (48%) *vs*. proteins involved in information pathways (84%). Furthermore, PTMs could also pose a challenge to the detection of these membrane associated proteins as both the efflux pumps and ATP Synthase may exist in a phosphorylated state (Barrera and Robinson, 2011).

Table 8.2: Differentially abundant proteins with a fold change greater than 10 in the RIF exposed *MDR- M. tuberculosis* isolate

Rv number	Gene name	Corresponding PATRIC annotation	Protein function	Functional category	P-value*	Fold change[$]
Rv3555c	-	VBIMycTub87468_2314	DNA binding	F10	0.000	1078.27
Rv0531	-	VBIMycTub87468_0589	Possible transmembrane protein	F3	0.009	274.68
Rv3919c	*gid*	VBIMycTub87468_4369	Probable glucose-inhibited division protein B	F3	0.000	175.04
Rv0298	*ydcD*	VBIMycTub87468_0330	Programmed cell death toxin	F10	0.004	64.51
Rv2325c	-	VBIMycTub87468_2595	Transmembrane component of energizing module of ECF transporters in Mycobacteria	F10	0.000	51.15
Rv0598c	*vapC27*	VBIMycTub87468_0664	Toxin 1, PIN domain	F0	0.000	29.46
Rv3576	*lppH*	VBIMycTub87468_3993	Non-specific serine/threonine protein kinase	F3	0.001	23.09
Rv1959c	*parE1*	VBIMycTub87468_2187	Plasmid stabilization protein	F0	0.000	19.55
Rv3817	*aph*	VBIMycTub87468_4250	Phosphotransferase	F7	0.002	17.38
Rv3203	*lipV*	VBIMycTub87468_3582	Possible lipase	F7	0.005	15.08
Rv2303c	-	VBIMycTub87468_2566	probable antibiotic resistance protein	F0	0.006	14.39
Rv0826	-	VBIMycTub87468_0921	Hypothetical protein	F10	0.001	13.63
Rv0201c	-	VBIMycTub87468_0228	Hypothetical protein	F10	0.000	11.40

* Calculated in Perseus
[$] Calculated ratio based on average iBAQ intensities

Our comparative analysis showed that 13 proteins were over-abundant (p-value < 0.01 and average iBAQ ratio >10 fold) in the RIF exposed MDR *M. tuberculosis* isolate (Table 8.2.). Four of these proteins (Rv3555c, Rv2325c, Rv0826 and Rv0201c) have been assigned as conserved hypothetical category with no known function. Rv3817 and Rv3203 were involved in intermediary metabolism and respiration. The phosphotransferase Rv3817 is involved in the lipopolysaccharide production which is a key component of the outer membrane (Brennan, 2003). However, its role in drug resistance has not been investigated, although changes in the lipopolysaccharide structure of the cell envelope have been shown to cause changes in resistance to a variety of antibiotics (McKay *et al*., 2003). *Rv3203* (LipV) is one of the 21 genes in the *M. tuberculosis* genome that encode for lipases/esterases (Deb *et al*., 2006). The authors hypothesised that lipases release fatty acids from triacylglycerol (TG) stores when the bacilli are under nutrient stress. Lipases/esterases hydrolyse TG which accumulates following various stress conditions in *M. tuberculosis* (Daniel J, *et al*. 2004; Sirakova TD *et al*. 2006; Natalie J. Garton et. al., 2008). Hydrolysis of TG by lipases/esterases provides fatty acids as a carbon source which in turn contributes to an alteration of the mycobacterial cell envelope (Garton et. al. 2008).

Three cell wall proteins (Rv0531 essential, Rv3576, and Rv3919c (Gid)) were significantly over-abundant in the RIF exposed MDR *M. tuberculosis* isolate (Table 8.2). However, their functions have not been described. Analysis of the protein sequences using the online transmembrane prediction tool that uses the Hidden Markov Model (TMHMM Server v. 2.0) predicted that Rv0531 has two transmembrane domains while Rv3576 only had one transmembrane domain. The glucose-inhibited division protein (Gid), also known as GidB, was not predicted to have a transmembrane domain but has been associated with low-level streptomycin

resistance in *M. tuberculosis* (Spies *et al*., 2011). This is supported by a recent study that demonstrated that streptomycin resistant isolates contained a point mutation in the E92D of the *gidB* gene (Via *et al*., 2010).

Three proteins (Rv2303, Rv0598 and Rv1959) involved in virulence, detoxification and adaptation were significantly over-abundant in the RIF exposed MDR *M. tuberculosis* isolate. Rv2303 is thought to possibly be involved in antibiotic resistance (Gillespie *et al*., 2011). Rv0598 (VapC27) and Rv1959 (ParE1) were previously identified in the cell envelope of *M. tuberculosis* H37Rv after membrane enrichment withTritonX-114 (Målen *et al*., 2008). Both of these proteins are thought to be toxins and part of the toxin-antitoxin system (Jiang *et al*., 2002; Ramage *et al*., 2009). Toxin-antitoxin (TA) systems are the most common mechanisms involved in bacterial programmed cell death (PCD). Similarly, the hypothetical protein, Rv0298, when searched against the bacterial bioinformatics resource centre, PATRIC (PathoSystems Resources Integration Centre) (Snyder *et al*., 2007; Gillespie *et al*., 2011) the corresponding locus tag (VBIMycTub87468_0330) lists this gene to encode for an antitoxin (YdcD) which is proposed to be involved in PCD. The mass spectrometer failed to detect the toxins Rv1246c (RelE) and Rv1495 (MazF4), in the unexposed MDR *M. tuberculosis* isolate. These were however detected in the exposed MDR isolate. PCD in prokaryotes allows bacteria to survive under environmental stresses such as nutrient deprivation, antibiotics exposure, hypoxia in the macrophages, etc. (Dahl *et al*., 2003). It has been hypothesised TA systems may be important for changing growth rate, inducing long term dormancy and in generating bacilli that can survive multidrug chemotherapy (Dahl *et al*., 2003; Beste *et al*., 2009; Han *et al*., 2010; Frampton *et al*., 2012). The TA system are organised in operons composed of two genes in which the first open reading frame (ORF) encodes

a labile antitoxin that combines with and neutralizes a stable regulatory toxin encoded by the second ORF (Hayes, 2003; Gerdes *et al*., 2005; Singh *et al*., 2010). The *M. tuberculosis* H37Rv and CDC1551 genome encodes for 88 putative TA systems (Arcus *et al*., 2005; Pandey and Gerdes, 2005; Ramage *et al*., 2009; Frampton *et al*., 2012) clearly suggesting the importance of these systems in the physiology of *M. tuberculosis*. TA systems are associated with translational maintenance, DNA replication and bacterial persistence (Hayes, 2003). These TA system pairs also forms part of the dormancy regulon (DosR) (Florczyk *et al*., 2001; Shi *et al*., 2003) and have been shown to be induced after exposure to vancomycin (Provvedi *et al*., 2009). This suggests that they may play an important role in the regulation if mycobacterial growth in response to diverse stress conditions and also could act as cell cycle arrest factors that induce the mycobacterial cell to enter a state of dormancy to protect it against environmental stress conditions. Furthermore, RelE and MazF are regulators of translation by cleavage of mRNA to block translation (Christensen *et al*., 2003). Both RelE and MazF cleave mRNA: RelE requires the ribosome for RNA cleavage while MazF does not (Christensen *et al*., 2003). Accordingly, it is suggested that toxins, RelE and MazF function in the quality control of gene expression. The toxins, RelE and MazF, are the best studied TA toxins in *E. coli* (Korch *et al*., 2009). Both RelE and MazF induced reversible dormancy in *E. coli* and *M. tuberculosis* (Korch and Hill, 2006; Korch *et al*., 2009). However, toxin-induced growth arrest was reversed upon the expression of each toxin's cognate antitoxin (Korch *et al*., 2009). Interestingly, over-expression of RelE in *M. tuberculosis* increased bacterial survival in the presence of RIF (Singh *et al*., 2010). Similar studies have not been done for MazF in *M. tuberculosis*, however, exposure to RIF induced the expression of *mazF* in *E. coli* (Kolodkin-Gal *et al*., 2008).

In summary, this is the first comprehensive analysis of the effect of RIF on the proteome of a RIF resistant MDR *M. tuberculosis* isolate. The analysis highlights two physiological events; 1) remodeling of the cell envelope and 2) initiation of a dormant/persistent growth phase which is associated with a significant decrease in the abundance of 25% of the proteome. Together this suggests that RIF continues to influence the biology of *M. tuberculosis* despite the presence of an *rpoB* mutation. These findings support a recent study which showed that RIF exposure induced the MDR *M. tuberculosis* isolate to become less refractive to other anti-TB drugs (i.e. ofloxacin) (Louw *et al*., 2011).

CHAPTER 9

Conclusion and Prospective studies

Since the advent of whole genome sequences of laboratory strain (H37Rv) and the first *M. tuberculosis* clinical isolate (CDC1551) a new era has dawned in the field of TB research. Subsequently, an ever increasing number of genomes from clinical isolates are sequenced with the view to correlate mutation to phenotype. Despite these advances, our understanding of the mechanisms of *M. tuberculosis* pathogenesis remains limted. This can in part be ascribed to the fact that the field of proteomics has lagged behind the genomics era. Furthermore, most proteome studies are focused on identifying individual novel drug targets. However, with the advancement in mass spectrometry technology it is clearly evident that the field of proteomics will enhance our understanding of the physiology of *M. tuberculosis*. Ideally, genomics, proteomics, transcriptomics, lipidomics and metabolomics platforms need to be combined to provide a systems biology approach to unravel the biology of *M. tuberculosis*.

The studies described in this thesis were designed to use the latest state of the art high throughput mass spectrometry to identify proteomic differences to develop hypotheses to explain the mechanisms underlying varying levels of virulence and drug resistance. This is the first study to characterise the proteome of members of the *M. tuberculosis* Beijing lineage which have been linked to outbreaks and drug resistance.

Using a label-free proteomics approach we were the first to characterise the proteomes of two closely related hypo- and hyper-virulent clinical *M.tuberculosis* strains with different levels of virulence (Chapter 4) (de Souza *et al*., 2010). The resolution of this analysis was further enhance when using a more advanced mass spectrometer with an increased sensitivity. Through this analysis it was possible to identify an additional 1146 proteins in whole cell lysate (Chapter 5). This represents the highest number of *M. tuberculosis* proteins identified to date within a single

experiment. This comparative analysis highlighted numerous factors that could alter metabolic pathways which allowed the hyper-virulent strain to rapidly replicate within the host and cause pathology as well as factors which contributed to the attenuation of virulence in the hypo-virulent and non- transmitting *M. tuberculosis* strain. For example, we showed an increased abundance of proteins involved in biosynthesis pathways of several carbon sources, i.e. lipid metabolism, central carbon metabolism (glycolysis/gluconeogenesis and TCA cycle), and amino acid biosynthesis that can be utilised by the mycobacterium bacilli as a source of energy for long-term survival

In chapter 6 we identified novel phosphorylation sites as well as confirming previously described phosphorylation sites. Furthermore, for the first time we were able to demonstrate tyrosine phosphorylation in mycobacteria. This work was complemented with a review (Chapter 2) describing the state of our knowledge in the field of phosphoproteomics in *M. tuberculosis*.

In chapter 7 we present the first comprehensive analysis of the effect of *in vivo* evolution of RIF resistance on the proteome of a RIF-mono-resistant mutant strain containing a S531L mutation. We identified the presence of over-abundant proteins which could provide novel insight into potential compensatory mechanisms that the bacillus uses to reduce susceptibility to anti-TB drugs. Our findings show that proteins involved in a stress response may relate to altered physiology enabling the pathogen to tolerate and persist when exposed to anti-TB drugs. These findings support a previous studies that showed that the DosR regulon may be important for bacteria to tolerate exposure to antibiotics (Bartek *et al*., 2009).

Furthermore we also present the first comprehensive analysis of the effect of RIF on the proteome of a RIF resistant MDR *M. tuberculosis* isolate (Chapter 8). Our

analysis suggests that RIF continues to influence the biology of *M. tuberculosis* despite the presence of an *rpoB* mutation. We observed a significant decrease in the abundance of 25% of the proteome suggesting exposure to RIF causes an alteration in the cell envelope composition and further leads to the bacterium to survive in a metabolically dormant/ persistent growth phase. Our findings also suggest that the presence of RIF may further induce a reversible persistent growth state which is refractive to RIF.

In conclusion, in this study we were able to use a novel proteomic approach to characterise the proteomes of *M. tuberculosis* strains which demonstrate different physiological phenotypes, including varying levels of virulence and drug resistance. Our results illustrate the full potential of a proteomic approach to select promising candidate molecules and genes for further characterization using the tools of molecular biology.

Prospective studies

The results presented in this thesis provides a resource of clinical importance for further investigation into the significance of the candidate proteins involved in fundamental biological processes that broaden our knowledge of *M. tuberculosis* pathogenesis.

1. Identified candidate phosphorylated proteins can be further analysed by specialised bioinformatics tools to identify novel substrates that could be implicated in virulence.

2. The whole phosphorylome could be analysed to determine the involvement of a group of phosphorylated proteins in a cascade of events in a network leading to a physiological change, i.e. levels of virulence and drug resistance.

3. Furthermore phosphorylated proteins can be assessed by determining the level of expression of a potential candidate's influence on other proteins and on the phenotype of a particular strain of *M. tuberculosis*.

4. Candidate proteins identified in proteome studies of drug resistant isolates can be futher analysed to determine and provide insight into other potential compensatory mechanisms the *M. tuberculosis* bacilli in response to the S531L *rpoB* mutation as well as other mutations.

5. Since we have identified one of the potential compensatory mechanism, such as the toxin-antitoxin system, that is involved in the physiological and biological alteration *M. tuberculosis* in response to RIF exposure we can assess how these bacilli respond within the host.

6. The ultimate validation would be to challenge these hypotheses created in this study by the proteome analysis of the *M. tuberculosis* strains in macrophages. This would potentially give more insight on how these candidate proteins identified in this current study is influenced in the host environment and in turn provide more clinical significance of these phenotypes.

REFERENCE LIST

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422, 198–207.

Apt, A., Kramnik, I., 2009. Man and mouse TB: contradictions and solutions. Tuberculosis (Edinb) 89, 195–198.

Arcus, V.L., Rainey, P.B., Turner, S.J., 2005. The PIN-domain toxin-antitoxin array in mycobacteria. Trends Microbiol. 13, 360–365.

Arora, G., Sajid, A., Gupta, M., Bhaduri, A., Kumar, P., Basu-Modak, S., Singh, Y., 2010. Understanding the Role of PknJ in *Mycobacterium tuberculosis* : Biochemical Characterization and Identification of Novel Substrate Pyruvate Kinase A. PLoS ONE 5, e10772.

Astarie-Dequeker, C., Le Guyader, L., Malaga, W., Seaphanh, F.-K., Chalut, C., Lopez, A., Guilhot, C., 2009. Phthiocerol Dimycocerosates of *M. tuberculosis* Participate in Macrophage Invasion by Inducing Changes in the Organization of Plasma Membrane Lipids. PLoS Pathog 5, e1000289.

Av-Gay, Y., Everett, M., 2000. The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis* . Trends Microbiol. 8, 238–244.

Bach, H., Wong, D., Av-Gay, Y., 2009. *Mycobacterium tuberculosis* PtkA is a novel protein tyrosine kinase whose substrate is PtpA. Biochem. J. 420, 155–160.

Bagchi, G., Chauhan, S., Sharma, D., Tyagi, J.S., 2005. Transcription and autoregulation of the Rv3134c-devR-devS operon of *Mycobacterium tuberculosis* . Microbiology 151, 4045–4053.

Barford, D., Das, A.K., Egloff, M.P., 1998. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. Annu Rev Biophys Biomol Struct 27, 133–164.

Barrera, N.P., Robinson, C.V., 2011. Advances in the Mass Spectrometry of Membrane Proteins: From Individual Proteins to Intact Complexes. Annual Review of Biochemistry 80, 247–271.

Bartek, I.L., Rutherford, R., Gruppo, V., Morton, R.A., Morris, R.P., Klein, M.R., Visconti, K.C., Ryan, G.J., Schoolnik, G.K., Lenaerts, A., Voskuil, M.I., 2009. The DosR regulon of *M. tuberculosis* and antibacterial tolerance. Tuberculosis (Edinb) 89, 310–316.

Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villén, J., Li, J., Cohn, M.A., Cantley, L.C., Gygi, S.P., 2004. Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc. Natl. Acad. Sci. U.S.A. 101, 12130–12135.

Bellinzoni, M., Wehenkel, A., Shepard, W., Alzari, P.M., 2007. Insights into the catalytic mechanism of PPM Ser/Thr phosphatases from the atomic resolution structures of a mycobacterial enzyme. Structure 15, 863–872.

Bendt, A.K., Burkovski, A., Schaffer, S., Bott, M., Farwick, M., Hermann, T., 2003. Towards a phosphoproteome map of *Corynebacterium glutamicum*. Proteomics 3, 1637–1646.

Beresford, N., Patel, S., Armstrong, J., Szöor, B., Fordham-Skelton, A.P., Tabernero, L., 2007. MptpB, a virulence factor from *Mycobacterium tuberculosis* , exhibits triple-specificity phosphatase activity. Biochem. J. 406, 13–18.

Beresford, N.J., Saville, C., Bennett, H.J., Roberts, I.S., Tabernero, L., 2010. A new family of phosphoinositide phosphatases in microorganisms: identification and biochemical analysis. BMC Genomics 11, 457.

Beste, D.J.V., Espasa, M., Bonde, B., Kierzek, A.M., Stewart, G.R., McFadden, J., 2009. The genetic requirements for fast and slow growth in mycobacteria. PLoS ONE 4, e5349.

Betts, J.C., Dodson, P., Quan, S., Lewis, A.P., Thomas, P.J., Duncan, K., McAdam, R.A., 2000. Comparison of the proteome of *Mycobacterium tuberculosis* strain H37Rv with clinical isolate CDC 1551. Microbiology 146, 3205–3216.

Bobadilla-del-Valle, M., Ponce-de-Leon, A., Arenas-Huertero, C., Vargas-Alarcon, G., Kato-Maeda, M., Small, P.M., Couary, P., Ruiz-Palacios, G.M., Sifuentes-Osornio, J., 2001. rpoB Gene mutations in rifampin-resistant *Mycobacterium tuberculosis* identified by polymerase chain reaction single-stranded conformational polymorphism. Emerg Infect Dis 7, 1010–1013.

Boitel, B., Ortiz-Lombardía, M., Durán, R., Pompeo, F., Cole, S.T., Cerveñansky, C., Alzari, P.M., 2003. PknB kinase activity is regulated by phosphorylation in two Thr residues and dephosphorylation by PstP, the cognate phospho-Ser/Thr phosphatase, in *Mycobacterium tuberculosis* . Mol. Microbiol. 49, 1493–1508.

Bott, M., Brocker, M., 2012. Two-component signal transduction in *Corynebacterium glutamicum* and other *corynebacteria*: on the way towards stimuli and targets. Appl. Microbiol. Biotechnol. 94, 1131–1150.

Bottai, D., Di Luca, M., Majlessi, L., Frigui, W., Simeone, R., Sayes, F., Bitter, W., Brennan, M.J., Leclerc, C., Batoni, G., Campa, M., Brosch, R., Esin, S., 2012. Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Molecular Microbiology 83, 1195–1209.

Brennan, P.J., 2003. Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis* . Tuberculosis (Edinb) 83, 91–97.

Camacho, L.R., Constant, P., Raynaud, C., Lanéelle, M.-A., Triccas, J.A., Gicquel, B., Daffé, M., Guilhot, C., 2001. Analysis of the Phthiocerol Dimycocerosate Locus of*Mycobacterium tuberculosis* Evidence that this lipid is involved in the cell wall permeability barrier. J. Biol. Chem. 276, 19845–19854.

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

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., Darst, S.A., 2001. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. Cell 104, 901–912.

Campuzano, J., Aguilar, D., Arriaga, K., León, J.C., Salas-Rangel, L.P., González-y-Merchand, J., Hernández-Pando, R., Espitia, C., 2007. The PGRS domain of *Mycobacterium tuberculosis* PE_PGRS Rv1759c antigen is an efficient subunit vaccine to prevent reactivation in a murine model of chronic tuberculosis. Vaccine 25, 3722–3729.

Canova, M.J., Veyron-Churlet, R., Zanella-Cleon, I., Cohen-Gonsaud, M., Cozzone, A.J., Becchi, M., Kremer, L., Molle, V., 2008. The *Mycobacterium tuberculosis* serine/threonine kinase PknL phosphorylates Rv2175c: mass spectrometric profiling of the activation loop phosphorylation sites and their role in the recruitment of Rv2175c. Proteomics 8, 521–533.

Casali, N., White, A.M., Riley, L.W., 2006. Regulation of the *Mycobacterium tuberculosis* mce1 operon. J. Bacteriol. 188, 441–449.

Cavazos, A., Prigozhin, D.M., Alber, T., 2012. Structure of the Sensor Domain of *Mycobacterium tuberculosis* PknH Receptor Kinase Reveals a Conserved Binding Cleft. J. Mol. Biol. 422, 488–494.

Chaba, R., Raje, M., Chakraborti, P.K., 2002. Evidence that a eukaryotic-type serine/threonine protein kinase from *Mycobacterium tuberculosis* regulates morphological changes associated with cell division. Eur. J. Biochem. 269, 1078–1085.

Chiaradia, L.D., Mascarello, A., Purificação, M., Vernal, J., Cordeiro, M.N.S., Zenteno, M.E., Villarino, A., Nunes, R.J., Yunes, R.A., Terenzi, H., 2008. Synthetic chalcones as efficient inhibitors of *Mycobacterium tuberculosis* protein tyrosine phosphatase PtpA. Bioorg. Med. Chem. Lett. 18, 6227–6230.

Chopra, P., Singh, B., Singh, R., Vohra, R., Koul, A., Meena, L.S., Koduri, H., Ghildiyal, M., Deol, P., Das, T.K., Tyagi, A.K., Singh, Y., 2003. Phosphoprotein phosphatase of *Mycobacterium tuberculosis* dephosphorylates serine–threonine kinases PknA and PknB. Biochemical and Biophysical Research Communications 311, 112–120.

Chow, K., Ng, D., Stokes, R., Johnson, P., 1994. Protein tyrosine phosphorylation in *Mycobacterium tuberculosis* . FEMS Microbiol. Lett. 124, 203–207.

Christensen, S.K., Pedersen, K., Hansen, F.G., Gerdes, K., 2003. Toxin–antitoxin Loci as Stress-response-elements: ChpAK/MazF and ChpBK Cleave Translated RNAs and are Counteracted by tmRNA. Journal of Molecular Biology 332, 809–819.

Cohen, P., 2002. The origins of protein phosphorylation. Nat. Cell Biol. 4, E127–130.

Cohen, P.T.W., 1997. Novel protein serine/threonine phosphatases: Variety is the spice of life. Trends in Biochemical Sciences 22, 245–251.

Cohen, T., Sommers, B., Murray, M., 2003. The effect of drug resistance on the fitness of *Mycobacterium tuberculosis* . Lancet Infect Dis 3, 13–21.

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

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., Galagan, J., Niemann, S., Gagneux, S., 2011. Whole-genome sequencing of rifampicin-resistant *M. tuberculosis* strains identifies compensatory mutations in RNA polymerase. Nat Genet 44, 106–110.

Cowley, S., Ko, M., Pick, N., Chow, R., Downing, K.J., Gordhan, B.G., Betts, J.C., Mizrahi, V., Smith, D.A., Stokes, R.W., Av-Gay, Y., 2004. The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. Molecular Microbiology 52, 1691–1702.

Cowley, S.C., Babakaiff, R., Av-Gay, Y., 2002. Expression and localization of the *Mycobacterium tuberculosis* protein tyrosine phosphatase PtpA. Res. Microbiol. 153, 233–241.

Cox, H., Kubica, T., Doshetov, D., Kebede, Y., Rüsch-Gerdess, S., Niemann, S., 2005. The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia. Respiratory Research 6, 134.

Cox, J., Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372.

Cozzone, A.J., 1998. Post-translational modification of proteins by reversible phosphorylation in prokaryotes. Biochimie 80, 43–48.

Cozzone, A.J., Grangeasse, C., Doublet, P., Duclos, B., 2004. Protein phosphorylation on tyrosine in bacteria. Arch. Microbiol. 181, 171–181.

Curry, J.M., Whalan, R., Hunt, D.M., Gohil, K., Strom, M., Rickman, L., Colston, M.J., Smerdon, S.J., Buxton, R.S., 2005. An ABC transporter containing a forkhead-associated domain interacts with a serine-threonine protein kinase and is required for growth of *Mycobacterium tuberculosis* in mice. Infect. Immun. 73, 4471–4477.

Dahl, J.L., Kraus, C.N., Boshoff, H.I.M., Doan, B., Foley, K., Avarbock, D., Kaplan, G., Mizrahi, V., Rubin, H., Barry, C.E., 3rd, 2003. The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of *Mycobacterium tuberculosis* in mice. Proc. Natl. Acad. Sci. U.S.A. 100, 10026–10031.

Dasgupta, N., Kapur, V., Singh, K.K., Das, T.K., Sachdeva, S., Jyothisri, K., Tyagi, J.S., 2000. Characterization of a two-component system, devR-devS, of *Mycobacterium tuberculosis* . Tuber. Lung Dis. 80, 141–159.

de Souza, G.A., Fortuin, S., Aguilar, D., Pando, R.H., McEvoy, C.R.E., van Helden, P.D., Koehler, C.J., Thiede, B., Warren, R.M., Wiker, H.G., 2010. Using a label-free proteomics method to identify differentially abundant proteins in closely related hypo- and hypervirulent clinical *Mycobacterium tuberculosis* Beijing isolates. Mol. Cell Proteomics 9, 2414–2423.

de Souza, G.A., Målen, H., Søfteland, T., Saelensminde, G., Prasad, S., Jonassen, I., Wiker, H.G., 2008. High accuracy mass spectrometry analysis as a tool to verify and improve gene annotation using *Mycobacterium tuberculosis* as an example. BMC Genomics 9, 316.

Deb, C., Daniel, J., Sirakova, T.D., Abomoelak, B., Dubey, V.S., Kolattukudy, P.E., 2006. A Novel Lipase Belonging to the Hormone-sensitive Lipase Family Induced under Starvation to Utilize Stored Triacylglycerol in *Mycobacterium tuberculosis* . J. Biol. Chem. 281, 3866–3875.

Deol, P., Vohra, R., Saini, A.K., Singh, A., Chandra, H., Chopra, P., Das, T.K., Tyagi, A.K., Singh, Y., 2005. Role of *Mycobacterium tuberculosis* Ser/Thr kinase PknF: implications in glucose transport and cell division. J. Bacteriol. 187, 3415–3420.

Dhandayuthapani, S., 2007. Stress response of genes encoding putative stress signaling molecules of *Mycobacterium tuberculosis* . Frontiers in Bioscience 12, 4676.

Dick, T., Manjunatha, U., Kappes, B., Gengenbacher, M., 2010. Vitamin B6 biosynthesis is essential for survival and virulence of *Mycobacterium tuberculosis* . Mol. Microbiol. 78, 980–988.

Domenech, P., Kolly, G.S., Leon-Solis, L., Fallow, A., Reed, M.B., 2010. Massive gene duplication event among clinical isolates of the *Mycobacterium tuberculosis* W/Beijing family. J. Bacteriol. 192, 4562–4570.

du Preez, I., Loots, D.T., 2012. Altered Fatty Acid Metabolism Due to Rifampicin-Resistance Conferring Mutations in the rpoB Gene of *Mycobacterium tuberculosis* : Mapping the Potential of Pharmaco-metabolomics for Global Health and Personalized Medicine. OMICS: A Journal of Integrative Biology 120911064527007.

Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon, S.J., Yaffe, M.B., 2000. The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. Mol. Cell 6, 1169–1182.

Ewann, F., Jackson, M., Pethe, K., Cooper, A., Mielcarek, N., Ensergueix, D., Gicquel, B., Locht, C., Supply, P., 2002. Transient requirement of the PrrA-PrrB two-component system for early intracellular multiplication of *Mycobacterium tuberculosis* . Infect. Immun. 70, 2256–2263.

Ewann, F., Locht, C., Supply, P., 2004. Intracellular autoregulation of the *Mycobacterium tuberculosis* PrrA response regulator. Microbiology (Reading, Engl.) 150, 241–246.

Fernandez, P., Saint-Joanis, B., Barilone, N., Jackson, M., Gicquel, B., Cole, S.T., Alzari, P.M., 2006. The Ser/Thr protein kinase PknB is essential for sustaining mycobacterial growth. J. Bacteriol. 188, 7778–7784.

Filliol, I., Motiwala, A.S., Cavatore, M., Qi, W., Hazbón, M.H., Bobadilla del Valle, M., Fyfe, J., García-García, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M.I., León, C.I., Crabtree, J., Angiuoli, S., Eisenach, K.D., Durmaz, R., Joloba, M.L., Rendón, A., Sifuentes-Osornio, J., Ponce de León, A., Cave, M.D., Fleischmann, R., Whittam, T.S., Alland, D., 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J. Bacteriol. 188, 759–772.

Fiuza, M., Canova, M.J., Zanella-Cléon, I., Becchi, M., Cozzone, A.J., Mateos, L.M., Kremer, L., Gil, J.A., Molle, V., 2008. From the characterization of the four serine/threonine protein kinases (PknA/B/G/L) of *Corynebacterium glutamicum* toward the role of PknA and PknB in cell division. J. Biol. Chem. 283, 18099–18112.

Florczyk, M.A., McCue, L.A., Stack, R.F., Hauer, C.R., McDonough, K.A., 2001. Identification and characterization of mycobacterial proteins differentially expressed understanding and shaking culture conditions, including Rv2623 from a novel class of putative ATP-binding proteins. Infect. Immun. 69, 5777–5785.

Frampton, R., Aggio, R.B.M., Villas-Bôas, S.G., Arcus, V.L., Cook, G.M., 2012. Toxin-antitoxin systems of Mycobacterium smegmatis are essential for cell survival. J. Biol. Chem. 287, 5340–5356.

Gagneux, S., Long, C.D., Small, P.M., Van, T., Schoolnik, G.K., Bohannan, B.J.M., 2006. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis* . Science 312, 1944–1946.

Garnak, M., Reeves, H.C., 1979a. Phosphorylation of Isocitrate dehydrogenase of *Escherichia coli*. Science 203, 1111–1112.

Garnak, M., Reeves, H.C., 1979b. Purification and properties of phosphorylated isocitrate dehydrogenase of *Escherichia coli*. J. Biol. Chem. 254, 7915–7920.

Ge, R., Shan, W., 2011. Bacterial Phosphoproteomic Analysis Reveals the Correlation Between Protein Phosphorylation and Bacterial Pathogenicity. Genomics, Proteomics & Bioinformatics 9, 119–127.

Gee, C.L., Papavinasasundaram, K.G., Blair, S.R., Baer, C.E., Falick, A.M., King, D.S., Griffin, J.E., Venghatakrishnan, H., Zukauskas, A., Wei, J.-R., Dhiman, R.K., Crick, D.C., Rubin, E.J., Sassetti, C.M., Alber, T., 2012. A Phosphorylated Pseudokinase Complex Controls Cell Wall Synthesis in Mycobacteria. Sci. Signal. 5, ra7.

Geiman, D.E., Kaushal, D., Ko, C., Tyagi, S., Manabe, Y.C., Schroeder, B.G., Fleischmann, R.D., Morrison, N.E., Converse, P.J., Chen, P., Bishai, W.R., 2004. Attenuation of late-stage disease in mice infected by the *Mycobacterium tuberculosis* mutant lacking the SigF alternate sigma factor and identification of SigF-dependent genes by microarray analysis. Infect. Immun. 72, 1733–1745.

Geiman, D.E., Raghunand, T.R., Agarwal, N., Bishai, W.R., 2006. Differential gene expression in response to exposure to antimycobacterial agents and other stress conditions among seven *Mycobacterium tuberculosis* whiB-like genes. Antimicrob. Agents Chemother. 50, 2836–2841.

Georghiou, S.B., Magana, M., Garfein, R.S., Catanzaro, D.G., Catanzaro, A., Rodwell, T.C., 2012. Evaluation of Genetic Mutations Associated with *Mycobacterium tuberculosis* Resistance to Amikacin, Kanamycin and Capreomycin: A Systematic Review. PLoS ONE 7, e33275.

Gerdes, K., Christensen, S.K., Løbner-Olesen, A., 2005. Prokaryotic toxin-antitoxin stress response loci. Nat. Rev. Microbiol. 3, 371–382.

Gillespie, J.J., Wattam, A.R., Cammer, S.A., Gabbard, J.L., Shukla, M.P., Dalay, O., Driscoll, T., Hix, D., Mane, S.P., Mao, C., Nordberg, E.K., Scott, M., Schulman, J.R., Snyder, E.E., Sullivan, D.E., Wang, C., Warren, A., Williams, K.P., Xue, T., Yoo, H.S., Zhang, C., Zhang, Y., Will, R., Kenyon, R.W., Sobral, B.W., 2011. PATRIC: the Comprehensive Bacterial Bioinformatics Resource with a Focus on Human Pathogenic Species. Infect. Immun. 79, 4286–4298.

Gopalaswamy, R., Narayanan, S., Chen, B., Jacobs, W.R., Av-Gay, Y., 2009. The serine/threonine protein kinase PknI controls the growth of *Mycobacterium tuberculosis* upon infection. FEMS Microbiology Letters 295, 23–29.

Gordon, S.V., Bottai, D., Simeone, R., Stinear, T.P., Brosch, R., 2009. Pathogenicity in the tubercle bacillus: molecular and evolutionary determinants. Bioessays 31, 378–388.

Graves, J.D., Krebs, E.G., 1999. Protein Phosphorylation and Signal Transduction. Pharmacology & Therapeutics 82, 111–121.

Greenstein, A.E., Echols, N., Lombana, T.N., King, D.S., Alber, T., 2007. Allosteric activation by dimerization of the PknD receptor Ser/Thr protein kinase from *Mycobacterium tuberculosis* . J. Biol. Chem. 282, 11427–11435.

Greenstein, A.E., Grundner, C., Echols, N., Gay, L.M., Lombana, T.N., Miecskowski, C.A., Pullen, K.E., Sung, P.-Y., Alber, T., 2005. Structure/function studies of Ser/Thr and Tyr protein phosphorylation in *Mycobacterium tuberculosis* . J. Mol. Microbiol. Biotechnol. 9, 167–181.

Greenstein, A.E., MacGurn, J.A., Baer, C.E., Falick, A.M., Cox, J.S., Alber, T., 2007. M. tuberculosis Ser/Thr Protein Kinase D Phosphorylates an Anti-Anti–Sigma Factor Homolog. PLoS Pathog 3, e49.

Griffin, J.E., Gawronski, J.D., Dejesus, M.A., Ioerger, T.R., Akerley, B.J., Sassetti, C.M., 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog. 7, e1002251.

Griffin, J.E., Pandey, A.K., Gilmore, S.A., Mizrahi, V., Mckinney, J.D., Bertozzi, C.R., Sassetti, C.M., 2012. Cholesterol Catabolism by *Mycobacterium tuberculosis* Requires Transcriptional and Metabolic Adaptations. Chemistry & Biology 19, 218–227.

Grundner, C., Gay, L.M., Alber, T., 2005. *Mycobacterium tuberculosis* serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains. Protein Science 14, 1918–1921.

Grundner, C., Ng, H.-L., Alber, T., 2005. *Mycobacterium tuberculosis* protein tyrosine phosphatase PtpB structure reveals a diverged fold and a buried active site. Structure 13, 1625–1634.

Guimarães, B.G., Souchon, H., Honoré, N., Saint-Joanis, B., Brosch, R., Shepard, W., Cole, S.T., Alzari, P.M., 2005. Structure and Mechanism of the Alkyl Hydroperoxidase AhpC, a Key Element of the *Mycobacterium tuberculosis* Defense System against Oxidative Stress. J. Biol. Chem. 280, 25735–25742.

Gupta, A.K., Katoch, V.M., Chauhan, D.S., Sharma, R., Singh, M., Venkatesan, K., Sharma, V.D., 2010. 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.

Han, J.-S., Lee, J.J., Anandan, T., Zeng, M., Sripathi, S., Jahng, W.J., Lee, S.H., Suh, J.-W., Kang, C.-M., 2010. Characterization of a chromosomal toxin-antitoxin, Rv1102c-Rv1103c system in *Mycobacterium tuberculosis*. Biochem. Biophys. Res. Commun. 400, 293–298.

Hanekom, M., van der Spuy, G.D., Streicher, E., Ndabambi, S.L., McEvoy, C.R.E., Kidd, M., Beyers, N., Victor, T.C., van Helden, P.D., Warren, R.M., 2007. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. J. Clin. Microbiol. 45, 1483–1490.

Harth, G., Maslesa-Galić, S., Tullius, M.V., Horwitz, M.A., 2005. All four *Mycobacterium tuberculosis glnA* genes encode glutamine synthetase activities but only *GlnA1* is abundantly expressed and essential for bacterial homeostasis. Mol. Microbiol. 58, 1157–1172.

Haydel, S.E., Dunlap, N.E., Benjamin, W.H., Jr, 1999. In vitro evidence of two-component system phosphorylation between the *Mycobacterium tuberculosis* TrcR/TrcS proteins. Microb. Pathog. 26, 195–206.

Haydel, S.E., Malhotra, V., Cornelison, G.L., Clark-Curtiss, J.E., 2012. The prrAB two-component system is essential for *Mycobacterium tuberculosis* viability and is induced under nitrogen-limiting conditions. J. Bacteriol. 194, 354–361.

Hayes, F., 2003. Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. Science 301, 1496–1499.

He, H., Zahrt, T.C., 2005. Identification and characterization of a regulatory sequence recognized by *Mycobacterium tuberculosis* persistence regulator MprA. J. Bacteriol. 187, 202–212.

Hernández-Pando, R., Orozcoe, H., Sampieri, A., Pavón, L., Velasquillo, C., Larriva-Sahd, J., Alcocer, J.M., Madrid, M.V., 1996. Correlation between the kinetics of Th1, Th2 cells and pathology in a murine model of experimental pulmonary tuberculosis. Immunology 89, 26–33.

Hernandez-Pando, R., Pavön, L., Arriaga, K., Orozco, H., Madrid-Marina, V., Rook, G., 1997. Pathogenesis of tuberculosis in mice exposed to low and high doses of an environmental mycobacterial saprophyte before infection. Infect. Immun. 65, 3317–3327.

Himpens, S., Locht, C., Supply, P., 2000. Molecular characterization of the mycobacterial SenX3-RegX3 two-component system: evidence for autoregulation. Microbiology (Reading, Engl.) 146 Pt 12, 3091–3098.

Houben, E.N.G., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S.R., Jiménez, C.R., Ottenhoff, T.H.M., Luirink, J., Bitter, W., 2012. Composition of the type VII secretion system membrane complex. Molecular Microbiology 86, 472–484.

Hunter, T., 1995. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. Cell 80, 225–236.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., Mann, M., 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell Proteomics 4, 1265–1272.

Jers, C., Soufi, B., Grangeasse, C., Deutscher, J., Mijakovic, I., 2008. Phosphoproteomics in bacteria: towards a systemic understanding of bacterial phosphorylation networks. Expert Review of Proteomics 5, 619–627.

Jiang, X., Zhang, W., Zhang, Y., Gao, F., Lu, C., Zhang, X., Wang, H., 2008. Assessment of efflux pump gene expression in a clinical isolate *Mycobacterium tuberculosis* by real-time reverse transcription PCR. Microb. Drug Resist. 14, 7–11.

Jiang, Y., Pogliano, J., Helinski, D.R., Konieczny, I., 2002. ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase. Mol. Microbiol. 44, 971–979.

Johnson, R., Streicher, E.M., Louw, G.E., Warren, R.M., van Helden, P.D., Victor, T.C., 2006. Drug resistance in *Mycobacterium tuberculosis*. Curr Issues Mol Biol 8, 97–111.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., van Embden, J., 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. 35, 907–914.

Kendall, S.L., Movahedzadeh, F., Rison, S.C.G., Wernisch, L., Parish, T., Duncan, K., Betts, J.C., Stoker, N.G., 2004. The *Mycobacterium tuberculosis* dosRS two-component system is induced by multiple stresses. Tuberculosis (Edinb) 84, 247–255.

Kendall, S.L., Withers, M., Soffair, C.N., Moreland, N.J., Gurcha, S., Sidders, B., Frita, R., Ten Bokum, A., Besra, G.S., Lott, J.S., Stoker, N.G., 2007. A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* . Mol. Microbiol. 65, 684–699.

Klumpp, S., Hermesmeier, J., Selke, D., Baumeister, R., Kellner, R., Krieglstein, J., 2002. Protein Histidine Phosphatase: A Novel Enzyme With Potency for Neuronal Signaling. Journal of Cerebral Blood Flow & Metabolism 22, 1420–1424.

Kobir, A., Shi, L., Boskovic, A., Grangeasse, C., Franjevic, D., Mijakovic, I., 2011. Protein phosphorylation in bacterial signal transduction. Biochim. Biophys. Acta 1810, 989–994.

Kolodkin-Gal, I., Sat, B., Keshet, A., Kulka, H.E.-, 2008. The Communication Factor EDF and the Toxin–Antitoxin Module mazEF Determine the Mode of Action of Antibiotics. PLoS Biol 6, e319.

Kondo, E., Kanai, K., 1974. Further studies on the increase in cholesterol ester content of the lungs of tuberculous mice. Jpn. J. Med. Sci. Biol. 27, 59–65.

Korch, S.B., Contreras, H., Clark-Curtiss, J.E., 2009. Three *Mycobacterium tuberculosis* Rel Toxin-Antitoxin Modules Inhibit Mycobacterial Growth and Are Expressed in Infected Human Macrophages. J. Bacteriol. 191, 1618–1630.

Korch, S.B., Hill, T.M., 2006. Ectopic Overexpression of Wild-Type and Mutant hipA Genes in *Escherichia coli*: Effects on Macromolecular Synthesis and Persister Formation. J Bacteriol 188, 3826–3836.

Koul, A., Choidas, A., Treder, M., Tyagi, A.K., Drlica, K., Singh, Y., Ullrich, A., 2000. Cloning and characterization of secretory tyrosine phosphatases of *Mycobacterium tuberculosis* . J. Bacteriol. 182, 5425–5432.

Koul, A., Choidas, A., Tyagi, A.K., Drlica, K., Singh, Y., Ullrich, A., 2001. Serine/threonine protein kinases PknF and PknG of *Mycobacterium tuberculosis* : characterization and localization. Microbiology (Reading, Engl.) 147, 2307–2314.

Kumar, D., Narayanan, S., 2012. pknE, a serine/threonine kinase of *Mycobacterium tuberculosis* modulates multiple apoptotic paradigms. Infect. Genet. Evol. 12, 737–747.

Kumar, P., Kumar, D., Parikh, A., Rananaware, D., Gupta, M., Singh, Y., Nandicoori, V.K., 2009. The *Mycobacterium tuberculosis* protein kinase K modulates activation of transcription from the promoter of mycobacterial monooxygenase operon through phosphorylation of the transcriptional regulator VirS. J. Biol. Chem. 284, 11090–11099.

Lan, N.T.N., Lien, H.T.K., Tung, L.B., Borgdorff, M.W., Kremer, K., van Soolingen, D., 2003. *Mycobacterium tuberculosis* Beijing Genotype and Risk for Treatment Failure and Relapse, Vietnam. Emerging Infectious Diseases 9, 1633–1635.

Lew, J.M., Kapopoulou, A., Jones, L.M., Cole, S.T., 2011. TubercuList--10 years after. Tuberculosis (Edinb) 91, 1–7.

Lewis, K.N., Liao, R., Guinn, K.M., Hickey, M.J., Smith, S., Behr, M.A., Sherman, D.R., 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation. J. Infect. Dis. 187, 117–123.

Lin, M.-H., Hsu, T.-L., Lin, S.-Y., Pan, Y.-J., Jan, J.-T., Wang, J.-T., Khoo, K.-H., Wu, S.-H., 2009. Phosphoproteomics of *Klebsiella pneumoniae* NTUH-K2044 Reveals a Tight Link between Tyrosine Phosphorylation and Virulence. Mol Cell Proteomics 8, 2613–2623.

Lipsitch, M., Moxon, E.R., 1997. Virulence and transmissibility of pathogens: what is the relationship? Trends Microbiol. 5, 31–37.

Louw, G.E., Warren, R.M., Gey van Pittius, N.C., McEvoy, C.R.E., Van Helden, P.D., Victor, T.C., 2009. A balancing act: efflux/influx in mycobacterial drug resistance. Antimicrob. Agents Chemother. 53, 3181–3189.

Louw, G.E., Warren, R.M., Pittius, N.C.G. van, Leon, R., Jimenez, A., Hernandez-Pando, R., McEvoy, C.R.E., Grobbelaar, M., Murray, M., Helden, P.D. van,

Victor, T.C., 2011. Rifampicin Reduces Susceptibility to Ofloxacin in Rifampicin-resistant *Mycobacterium tuberculosis* through Efflux. Am. J. Respir. Crit. Care Med. 184, 269–276.

Macek, B., Mijakovic, I., 2011. Site-specific analysis of bacterial phosphoproteomes. Proteomics 11, 3002–3011.

Macek, B., Mijakovic, I., Olsen, J.V., Gnad, F., Kumar, C., Jensen, P.R., Mann, M., 2007. The Serine/Threonine/Tyrosine Phosphoproteome of the Model Bacterium *Bacillus subtilis*. Mol Cell Proteomics 6, 697–707.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., Stover, C.K., 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent M. bovis. J. Bacteriol. 178, 1274–1282.

Målen, H., Berven, F.S., Søfteland, T., Arntzen, M.Ø., D'Santos, C.S., De Souza, G.A., Wiker, H.G., 2008. 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 8, 1859–1870.

Malhotra, V., Sharma, D., Ramanathan, V.D., Shakila, H., Saini, D.K., Chakravorty, S., Das, T.K., Li, Q., Silver, R.F., Narayanan, P.R., Tyagi, J.S., 2004. Disruption of response regulator gene, *devR*, leads to attenuation in virulence of *Mycobacterium tuberculosis* . FEMS Microbiol. Lett. 231, 237–245.

Manai, M., Cozzone, A.J., 1979. Analysis of the protein-kinase activity of Escherichia coli cells. Biochemical and Biophysical Research Communications 91, 819–826.

Mawuenyega, K.G., Forst, C.V., Dobos, K.M., Belisle, J.T., Chen, J., Bradbury, E.M., Bradbury, A.R.M., Chen, X., 2005. *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. Mol. Biol. Cell 16, 396–404.

McKay, G.A., Woods, D.E., MacDonald, K.L., Poole, K., 2003. Role of Phosphoglucomutase of Stenotrophomonas maltophilia in Lipopolysaccharide Biosynthesis, Virulence, and Antibiotic Resistance. Infect. Immun. 71, 3068–3075.

Meena, L.S., Chopra, P., Bedwal, R.S., Singh, Y., 2003. Nucleoside diphosphate kinase-like activity in adenylate kinase of *Mycobacterium tuberculosis* . Biotechnol. Appl. Biochem. 38, 169–174.

Meier, A., Sander, P., Schaper, K.J., Scholz, M., Böttger, E.C., 1996. Correlation of molecular resistance mechanisms and phenotypic resistance levels in streptomycin-resistant *Mycobacterium tuberculosis* . Antimicrob. Agents Chemother. 40, 2452–2454.

Meissner-Roloff, R.J., Koekemoer, G., Warren, R.M., Loots, D.T., (2012). A metabolomics investigation of a hyper-and hypo-virulent phenotype of Beijing lineage *M. tuberculosis*. Metabolomics 1–10.

Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., Horning, S., 2011. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. Mol Cell Proteomics 10.

Mijakovic, I., Macek, B., 2012. Impact of phosphoproteomics on studies of bacterial physiology. FEMS Microbiol. Rev. 36, 877–892.

Möker, N., Brocker, M., Schaffer, S., Krämer, R., Morbach, S., Bott, M., 2004. Deletion of the genes encoding the MtrA–MtrB two-component system of

Corynebacterium glutamicum has a strong influence on cell morphology, antibiotics susceptibility and expression of genes involved in osmoprotection. Molecular Microbiology 54, 420–438.

Molle, V., Reynolds, R.C., Alderwick, L.J., Besra, G.S., Cozzone, A.J., Fütterer, K., Kremer, L., 2008. EmbR2, a structural homologue of EmbR, inhibits the *Mycobacterium tuberculosis* kinase/substrate pair PknH/EmbR. Biochem. J. 410, 309–317.

Molle, V., Soulat, D., Jault, J.-M., Grangeasse, C., Cozzone, A.J., Prost, J.-F., 2004. Two FHA domains on an ABC transporter, Rv1747, mediate its phosphorylation by PknF, a Ser/Thr protein kinase from *Mycobacterium tuberculosis* . FEMS Microbiol. Lett. 234, 215–223.

Morth, J.P., Gosmann, S., Nowak, E., Tucker, P.A., 2005. A novel two-component system found in *Mycobacterium tuberculosis* . FEBS Lett. 579, 4145–4148.

Nagaraj, N., D'Souza, R.C.J., Cox, J., Olsen, J.V., Mann, M., 2012. Correction to Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation. J. Proteome Res. 11, 3506–3508.

Narayan, A., Sachdeva, P., Sharma, K., Saini, A.K., Tyagi, A.K., Singh, Y., 2007. Serine threonine protein kinases of mycobacterial genus: phylogeny to function. Physiol. Genomics 29, 66–75.

Nott, T.J., Kelly, G., Stach, L., Li, J., Westcott, S., Patel, D., Hunt, D.M., Howell, S., Buxton, R.S., O'Hare, H.M., Smerdon, S.J., 2009. An Intramolecular Switch Regulates Phosphoindependent FHA Domain Interactions in *Mycobacterium tuberculosis* . Sci. Signal. 2, ra12.

O'Toole, R., Smeulders, M.J., Blokpoel, M.C., Kay, E.J., Lougheed, K., Williams, H.D., 2003. A two-component regulator of universal stress protein expression and adaptation to oxygen starvation in *Mycobacterium smegmatis*. J. Bacteriol. 185, 1543–1554.

Oda, Y., Nagasu, T., Chait, B.T., 2001. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat. Biotechnol. 19, 379–382.

Olsen, J.V., de Godoy, L.M.F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., Mann, M., 2005. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol. Cell Proteomics 4, 2010–2021.

Pallen, M., Chaudhuri, R., Khan, A., 2002. Bacterial FHA domains: neglected players in the phospho-threonine signalling game? Trends Microbiol. 10, 556–563.

Pandey, D.P., Gerdes, K., 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res. 33, 966–976.

Parish, T., Smith, D.A., Kendall, S., Casali, N., Bancroft, G.J., Stoker, N.G., 2003. Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. Infect. Immun. 71, 1134–1140.

Parish, T., Smith, D.A., Roberts, G., Betts, J., Stoker, N.G., 2003. The senX3–regX3 two-component regulatory system of *Mycobacterium tuberculosis* is required for virulence. Microbiology 149, 1423–1435.

Park, E.M., Lee, I.J., Kim, S.H., Song, G.Y., Park, Y.M., 2003. Inhibitory effect of a naphthazarin derivative, S64, on heat shock factor (Hsf) activation and glutathione status following hypoxia. Cell Biol. Toxicol. 19, 273–284.

Park, H.-D., Guinn, K.M., Harrell, M.I., Liao, R., Voskuil, M.I., Tompa, M., Schoolnik, G.K., Sherman, D.R., 2003. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis* . Mol. Microbiol. 48, 833–843.

Parwati, I., Alisjahbana, B., Apriani, L., Soetikno, R.D., Ottenhoff, T.H., van der Zanden, A.G.M., van der Meer, J., van Soolingen, D., van Crevel, R., 2010. *Mycobacterium tuberculosis* Beijing genotype is an independent risk factor for tuberculosis treatment failure in Indonesia. J. Infect. Dis. 201, 553–557.

Peirs, P., De Wit, L., Braibant, M., Huygen, K., Content, J., 1997. A serine/threonine protein kinase from *Mycobacterium tuberculosis*. Eur. J. Biochem. 244, 604–612.

Pérez, E., Samper, S., Bordas, Y., Guilhot, C., Gicquel, B., Martín, C., 2001. An essential role for phoP in *Mycobacterium tuberculosis* virulence. Mol. Microbiol. 41, 179–187.

Perez, J.C., Shin, D., Zwir, I., Latifi, T., Hadley, T.J., Groisman, E.A., 2009. Evolution of a bacterial regulon controlling virulence and Mg(2+) homeostasis. PLoS Genet. 5, e1000428.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoeresis 20, 3551–3567.

Preu, J., Panjikar, S., Morth, P., Jaiswal, R., Karunakar, P., Tucker, P.A., 2012. The sensor region of the ubiquitous cytosolic sensor kinase, PdtaS, contains PAS and GAF domain sensing modules. J. Struct. Biol. 177, 498–505.

Prisic, S., Dankwa, S., Schwartz, D., Chou, M.F., Locasale, J.W., Kang, C.-M., Bemis, G., Church, G.M., Steen, H., Husson, R.N., 2010. Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. Proc. Natl. Acad. Sci. U.S.A. 107, 7521–7526.

Provvedi, R., Boldrin, F., Falciani, F., Palù, G., Manganelli, R., 2009. Global transcriptional response to vancomycin in *Mycobacterium tuberculosis*. Microbiology (Reading, Engl.) 155, 1093–1102.

Qing-Run Li, Z.-B.N., 2009. Effect of Peptide-to-TiO2 beads Ratio on Phosphopeptide Enrichment Selectivity. Journal of proteome research.

Ramage, H.R., Connolly, L.E., Cox, J.S., 2009. Comprehensive Functional Analysis of *Mycobacterium tuberculosis* Toxin-Antitoxin Systems: Implications for Pathogenesis, Stress Responses, and Evolution. PLoS Genet 5, e1000767.

Ramaswamy, S., Musser, J.M., 1998. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tuber. Lung Dis. 79, 3–29.

Rappsilber, J., Ishihama, Y., Mann, M., 2003. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. 75, 663–670.

Ravichandran, A., Sugiyama, N., Tomita, M., Swarup, S., Ishihama, Y., 2009. Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. Proteome 9, 2764–2775.

Reddy, T.B.K., Riley, R., Wymore, F., Montgomery, P., DeCaprio, D., Engels, R., Gellesch, M., Hubble, J., Jen, D., Jin, H., Koehrsen, M., Larson, L., Mao, M., Nitzberg, M., Sisk, P., Stolte, C., Weiner, B., White, J., Zachariah, Z.K., Sherlock, G., Galagan, J.E., Ball, C.A., Schoolnik, G.K., 2009. TB database:

an integrated platform for tuberculosis research. Nucleic Acids Research 37, D499–D508.

Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N., Kaplan, G., Barry, C.E., 3rd, 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature 431, 84–87.

Reed, M.B., Gagneux, S., Deriemer, K., Small, P.M., Barry, C.E., 3rd, 2007. The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. J. Bacteriol. 189, 2583–2589.

Rigden, D.J., 2011. Identification and modelling of a PPM protein phosphatase fold in the Legionella pneumophila deAMPylase SidD. FEBS Letters 585, 2749–2754.

Roberts, G., Vadrevu, I.S., Madiraju, M.V., Parish, T., 2011. Control of CydB and GltA1 Expression by the SenX3 RegX3 Two Component Regulatory System of *Mycobacterium tuberculosis*. PLoS ONE 6, e21090.

Rodriguez, G.M., Smith, I., 2006. Identification of an ABC Transporter Required for Iron Acquisition and Virulence in *Mycobacterium tuberculosis* . J. Bacteriol. 188, 424–430.

Roumestand, C., Leiba, J., Galophe, N., Margeat, E., Padilla, A., Bessin, Y., Barthe, P., Molle, V., Cohen-Gonsaud, M., 2011. Structural Insight into the *Mycobacterium tuberculosis* Rv0020c Protein and Its Interaction with the PknB Kinase. Structure 19, 1525–1534.

Saini, D.K., Malhotra, V., Dey, D., Pant, N., Das, T.K., Tyagi, J.S., 2004. DevR-DevS is a bona fide two-component system of *Mycobacterium tuberculosis* that is hypoxia-responsive in the absence of the DNA-binding domain of DevR. Microbiology (Reading, Engl.) 150, 865–875.

Sassetti, C.M., Boyd, D.H., Rubin, E.J., 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. 48, 77–84.

Sassetti, C.M., Rubin, E.J., 2003. Genetic requirements for mycobacterial survival during infection. Proc. Natl. Acad. Sci. U.S.A. 100, 12989–12994.

Scherr, N., Müller, P., Perisa, D., Combaluzier, B., Jenö, P., Pieters, J., 2009. Survival of pathogenic mycobacteria in macrophages is mediated through autophosphorylation of protein kinase G. J. Bacteriol. 191, 4546–4554.

Schmidl, S.R., Gronau, K., Pietack, N., Hecker, M., Becher, D., Stülke, J., 2010. The Phosphoproteome of the Minimal Bacterium Mycoplasma pneumoniae analysis of the complete known ser/thr kinine suggests the existence of novel kinases. Mol Cell Proteomics 9, 1228–1242.

Schwab, U., Rohde, K.H., Wang, Z., Chess, P.R., Notter, R.H., Russell, D.G., 2009. Transcriptional responses of *Mycobacterium tuberculosis* to lung surfactant. Microb. Pathog. 46, 185–193.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M., 2011. Global quantification of mammalian gene expression control. Nature 473, 337–342.

Schwartz, D., Gygi, S.P., 2005. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat. Biotechnol. 23, 1391–1398.

Sharma, K., Gupta, M., Pathak, M., Gupta, N., Koul, A., Sarangi, S., Baweja, R., Singh, Y., 2006. Transcriptional control of the mycobacterial embCAB operon

by PknH through a regulatory protein, EmbR, *in vivo*. J. Bacteriol. 188, 2936–2944.

Shi, L., Jung, Y.-J., Tyagi, S., Gennaro, M.L., North, R.J., 2003. Expression of Th1-mediated immunity in mouse lungs induces a *Mycobacterium tuberculosis* transcription pattern characteristic of nonreplicating persistence. Proc. Natl. Acad. Sci. U.S.A. 100, 241–246.

Shi, L., Potts, M., Kennelly, P.J., 1998. The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait. FEMS Microbiol. Rev. 22, 229–253.

Shi, T., Xie, J., 2011. Molybdenum enzymes and molybdenum cofactor in mycobacteria. J. Cell. Biochem. 112, 2721–2728.

Shrivastava, R., Ghosh, A.K., Das, A.K., 2007. Probing the nucleotide binding and phosphorylation by the histidine kinase of a novel three-protein two-component system from *Mycobacterium tuberculosis*. FEBS Lett. 581, 1903–1909.

Siddiqi, N., Das, R., Pathak, N., Banerjee, S., Ahmed, N., Katoch, V.M., Hasnain, S.E., 2004. *Mycobacterium tuberculosis* isolate with a distinct genomic identity overexpresses a tap-like efflux pump. Infection 32, 109–111.

Silva, A.P.G., Tabernero, L., 2010. New strategies in fighting TB: targeting *Mycobacterium tuberculosis* -secreted phosphatases MptpA & MptpB. Future Med Chem 2, 1325–1337.

Siméone, R., Constant, P., Guilhot, C., Daffé, M., Chalut, C., 2007. Identification of the missing trans-acting enoyl reductase required for phthiocerol dimycocerosate and phenolglycolipid biosynthesis in *Mycobacterium tuberculosis*. J. Bacteriol. 189, 4597–4602.

Singh, M., Jadaun, G.P.S., Ramdas, Srivastava, K., Chauhan, V., Mishra, R., Gupta, K., Nair, S., Chauhan, D.S., Sharma, V.D., Venkatesan, K., Katoch, V.M., 2011. Effect of efflux pump inhibitors on drug susceptibility of ofloxacin resistant *Mycobacterium tuberculosis* isolates. Indian J Med Res 133, 535–540.

Singh, R., Barry, C.E., 3rd, Boshoff, H.I.M., 2010. The three RelE homologs of *Mycobacterium tuberculosis* have individual, drug-specific effects on bacterial antibiotic tolerance. J. Bacteriol. 192, 1279–1291.

Snyder, E.E., Kampanya, N., Lu, J., Nordberg, E.K., Karur, H.R., Shukla, M., Soneja, J., Tian, Y., Xue, T., Yoo, H., Zhang, F., Dharmanolla, C., Dongre, N.V., Gillespie, J.J., Hamelius, J., Hance, M., Huntington, K.I., Jukneliene, D., Koziski, J., Mackasmiel, L., Mane, S.P., Nguyen, V., Purkayastha, A., Shallom, J., Yu, G., Guo, Y., Gabbard, J., Hix, D., Azad, A.F., Baker, S.C., Boyle, S.M., Khudyakov, Y., Meng, X.J., Rupprecht, C., Vinje, J., Crasta, O.R., Czar, M.J., Dickerman, A., Eckart, J.D., Kenyon, R., Will, R., Setubal, J.C., Sobral, B.W.S., 2007. PATRIC: The VBI PathoSystems Resource Integration Center. Nucleic Acids Res 35, D401–D406.

Sonnhammer, E.L., von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6, 175–182.

Sørensen, A.L., Nagai, S., Houen, G., Andersen, P., Andersen, A.B., 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. Infect. Immun. 63, 1710–1717.

Spies, F.S., Ribeiro, A.W., Ramos, D.F., Ribeiro, M.O., Martin, A., Palomino, J.C., Rossetti, M.L.R., da Silva, P.E.A., Zaha, A., 2011. Streptomycin Resistance and Lineage-Specific Polymorphisms in *Mycobacterium tuberculosis gidB* Gene ▿. J Clin Microbiol 49, 2625–2630.

Spivey, V.L., Molle, V., Whalan, R.H., Rodgers, A., Leiba, J., Stach, L., Walker, K.B., Smerdon, S.J., Buxton, R.S., 2011. Forkhead-associated (FHA) domain containing ABC transporter Rv1747 is positively regulated by Ser/Thr phosphorylation in *Mycobacterium tuberculosis*. J. Biol. Chem. 286, 26198–26209.

Starck, J., Källenius, G., Marklund, B.-I., Andersson, D.I., Akerlund, T., 2004. Comparative proteome analysis of *Mycobacterium tuberculosis* grown under aerobic and anaerobic conditions. Microbiology (Reading, Engl.) 150, 3821–3829.

Stock, A.M., Robinson, V.L., Goudreau, P.N., 2000. Two-component signal transduction. Annu. Rev. Biochem. 69, 183–215.

Sun, R., Converse, P.J., Ko, C., Tyagi, S., Morrison, N.E., Bishai, W.R., 2004. *Mycobacterium tuberculosis* ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. Mol. Microbiol. 52, 25–38.

Sun, X., Ge, F., Xiao, C.-L., Yin, X.-F., Ge, R., Zhang, L.-H., He, Q.-Y., 2010. Phosphoproteomic Analysis Reveals the Multiple Roles of Phosphorylation in Pathogenic Bacterium *Streptococcus pneumoniae*. J. Proteome Res. 9, 275–282.

Takayama, K., Wang, C., Besra, G.S., 2005. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. Clin. Microbiol. Rev. 18, 81–101.

Tan, C.S.H., Bodenmiller, B., Pasculescu, A., Jovanovic, M., Hengartner, M.O., Jorgensen, C., Bader, G.D., Aebersold, R., Pawson, T., Linding, R., 2009. Comparative Analysis Reveals Conserved Protein Phosphorylation Networks Implicated in Multiple Diseases. Sci. Signal. 2, ra39.

Telenti, A., Imboden, P., Marchesi, F., Lowrie, D., Cole, S., Colston, M.J., Matter, L., Schopfer, K., Bodmer, T., 1993. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. Lancet 341, 647–650.

ten Bokum, A.M.C., Movahedzadeh, F., Frita, R., Bancroft, G.J., Stoker, N.G., 2008. The case for hypervirulence through gene deletion in *Mycobacterium tuberculosis*. Trends Microbiol. 16, 436–441.

Thakur, M., Chaba, R., Mondal, A.K., Chakraborti, P.K., 2008. Interdomain interaction reconstitutes the functionality of PknA, a eukaryotic type Ser/Thr kinase from *Mycobacterium tuberculosis*. J. Biol. Chem. 283, 8023–8033.

Tyagi, J.S., Sharma, D., 2004. Signal transduction systems of mycobacteria with special reference to M. tuberculosis. Current Science 86, 93–102.

van der Spuy, G.D., Kremer, K., Ndabambi, S.L., Beyers, N., Dunbar, R., Marais, B.J., van Helden, P.D., Warren, R.M., 2009. Changing *Mycobacterium tuberculosis* population highlights clade-specific pathogenic characteristics. Tuberculosis (Edinb) 89, 120–125.

van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., Hermans, P., Martin, C., McAdam, R., Shinnick, T.M., 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting:

recommendations for a standardized methodology. J. Clin. Microbiol. 31, 406–409.

Via, L.E., Cho, S.-N., Hwang, S., Bang, H., Park, S.K., Kang, H.S., Jeon, D., Min, S.Y., Oh, T., Kim, Y., Kim, Y.M., Rajan, V., Wong, S.Y., Shamputa, I.C., Carroll, M., Goldfeder, L., Lee, S.A., Holland, S.M., Eum, S., Lee, H., Barry, C.E., 3rd, 2010. Polymorphisms associated with resistance and cross-resistance to aminoglycosides and capreomycin in *Mycobacterium tuberculosis* isolates from South Korean Patients with drug-resistant tuberculosis. J. Clin. Microbiol. 48, 402–411.

Vlisidou, I., Eleftherianos, I., Dorus, S., Yang, G., ffrench-Constant, R.H., Reynolds, S.E., Waterfield, N.R., 2010. The KdpD/KdpE two-component system of Photorhabdus asymbiotica promotes bacterial survival within *M. sexta hemocytes*. J. Invertebr. Pathol. 105, 352–362.

Voisin, S., Watson, D.C., Tessier, L., Ding, W., Foote, S., Bhatia, S., Kelly, J.F., Young, N.M., 2007. The cytoplasmic phosphoproteome of the Gram-negative bacterium *Campylobacter jejuni*: Evidence for modification by unidentified protein kinases. Proteomics 7, 4338–4348.

Voskuil, M.I., Visconti, K.C., Schoolnik, G.K., 2004a. *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis (Edinb) 84, 218–227.

Voskuil, M.I., Visconti, K.C., Schoolnik, G.K., 2004b. *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis 84, 218–227.

Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., Klebl, B., Thompson, C., Bacher, G., Pieters, J., 2004. Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. Science 304, 1800–1804.

Wehenkel, A., Bellinzoni, M., Graña, M., Duran, R., Villarino, A., Fernandez, P., Andre-Leroux, G., England, P., Takiff, H., Cerveñansky, C., Cole, S.T., Alzari, P.M., 2008. Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and therapeutic potential. Biochim. Biophys. Acta 1784, 193–202.

WHO | Global tuberculosis control 2011 [WWW Document], 2012. . WHO. URL http://www.who.int/tb/publications/global_report/en/index.html

Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P., Rane, S., Brown, P.O., Schoolnik, G.K., 1999. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc. Natl. Acad. Sci. U.S.A. 96, 12833–12838.

Wiśniewski, J.R., Nagaraj, N., Zougman, A., Gnad, F., Mann, M., 2010. Brain phosphoproteome obtained by a FASP-based method reveals plasma membrane protein topology. J. Proteome Res. 9, 3280–3289.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., Mann, M., 2009. Universal sample preparation method for proteome analysis. Nat. Methods 6, 359–362.

Xing, T., Ouellet, T., Miki, B.L., 2002. Towards genomic and proteomic studies of protein phosphorylation in plant-pathogen interactions. Trends Plant Sci. 7, 224–230.

Zahrt, T.C., Deretic, V., 2000. An essential two-component signal transduction system in *Mycobacterium tuberculosis* . J. Bacteriol. 182, 3832–3838.

Zahrt, T.C., Wozniak, C., Jones, D., Trevett, A., 2003. Functional analysis of the *Mycobacterium tuberculosis* MprAB two-component signal transduction system. Infect. Immun. 71, 6962–6970.