A metabolomics approach investigating the functionality of the ESX-1 gene cluster in mycobacteria

CC Swanepoel

21617678

Dissertation submitted in partial fulfilment of the requirements for the degree *Magister Scientiae* in Biochemistry at the Potchefstroom Campus of the North-West University

Supervisor: Prof DT Loots

May 2015

"Science is a wonderful thing if one does not have to earn one's living at it."

Albert Einstein

CONTENTS

ACKNOWLEDGEMENTS

The contributions of the following institutions and individuals made towards the successful completion of this study are hereby acknowledged:

- First, I would like to thank Prof. Du Toit Loots for your ongoing support, guidance, encouragement and, most important, patience. I admire and respect your work ethic, dedication and fair criticism.
- The National Research Foundation (NRF) of South Africa, the Technology Innovation Agency (TIA) and North-West University (NWU) for the research grants provided.
- The DST/NRF Centre of Excellence for Biomedical TB Research (CBTBR), Stellenbosch University, South Africa, for providing the *M. smegmatis* samples used in this study.
- Fanie, Christa and Laneke, thank you for never hesitating to provide me with the necessary assistance regarding this study, either with “Anna-Bets”, general lab work, or data processing.
- To the rest of the staff and students at NWU Biochemistry and the Potchefstroom Laboratory of Inborn Errors of Metabolism (PLIEM), thank you for being there day in and day out, sharing in the good and the bad.
- To my parents, Awie and Eldorette, thank you for being my greatest fans with your words of encouragement, as well as your ongoing financial support.
- To my fiancé, Liandi, thanks for all the love, patience and support.
- Finally, I would like to thank our Heavenly Father, for the talent He generously bestowed upon me. “What, then, shall we say in response to these things? If God is for us, who can be against us? He who did not spare his own Son, but gave him up for us all – how will he not also, along with him, graciously give us all things?” (Rom 8: 31-32).

SUMMARY

Tuberculosis (TB) claims the lives of millions of individuals each year, and is consequently the world's second-most deadly infectious disease after acquired immune deficiency syndrome (AIDS), responsible for 1.4 million deaths in 2010 alone. Developing countries carry the heaviest burden, with the occurrence of multidrug-resistant (MDR) TB becoming more frequent, making more efficient vaccination and treatment strategies a necessity to combat this epidemic. The ESX-1 gene cluster (encoding the virulence-associated proteins ESAT-6 and CFP-10) and the Type VII secretion system are thought to be responsible for the transport of extracellular proteins across the hydrophobic, and highly impermeable, cell wall of *Mycobacterium*, and consequently are thought to play a role in the virulence of this organism. To date, our understanding of tuberculosis pathophysiology and virulence has been described primarily using proteomic and genomic approaches. Subsequently, using the relatively new research approach called metabolomics, and interpreting the data using systems biology, we aimed to identify new metabolite markers that better characterise virulence and the proteins involved, more specifically related to the ESX-1 gene cluster. Using a GCxGC-TOFMS metabolomics research approach, we compared the varying metabolomes of *M. smegmatis* ESX-1 knock-out (ΔESX-1_{ms}) to that of the wild-type parent strain and subsequently identified those metabolite markers differing between these strains. Multivariate and univariate statistical analyses of the analysed metabolome were used to identify those metabolites contributing most to the differences seen between the two sample groups. A general increase in various carbohydrates, amino acids and lipids, associated with cell wall structure and function, were detected in the ΔESX-1_{ms} strain relative to the wild-type parent strain. Additionally, metabolites associated with the antioxidant system, virulence protein formation and energy production in these mycobacteria, were also seen to differ between the two groups. This metabolomics investigation is the first to identify the metabolite markers confirming the role of the ESX-1 gene cluster with virulence and the underlying metabolic pathways, as well as its associated role with increased metabolic activity, growth/replication rates, increased cell wall synthesis and an altered antioxidant

mechanism, all of which are believed to contribute to this organism's increased pathogenicity and survival ability.

Key terms: metabolomics; *Mycobacterium*; tuberculosis; ESX-1; virulence; GCxGC-TOFMS.

OPSOMMING

Tuberkulose (TB) eis die lewens van miljoene mense elke jaar, en is gevolglik die wêreld se tweede grootste dodelikste aansteeklike siekte na die Verworwe Immuniteitsgebreksindroom (VIGS), en dra verantwoordelikheid vir 1,4 miljoen sterftes in 2010 alleen. Ontwikkelende lande dra die swaarste las, met 'n toenemende voorkoms van multimiddel weerstandige TB wat beteken dat beter inenting- en behandeling strategieë noodsaaklikheid is om hierdie epidemie te bestry. Wetenskaplikes glo dat die ESX-1 geen groep gebied (wat kodeer vir virulensie geassosieerde proteïene ESAT-6 en CFP-10), asook die tipe VII sekresie stelsel, dra nie net verantwoordelikheid vir die vervoer van ekstrasellulêre proteïene oor die hidrofobiese, en hoogs ondeurdringbare, selwand van *Mycobacterium* nie, maar lewer 'n verdere bydrae tot virulensie in hierdie organisme. Tot op hede is ons begrip van die patofisiologie en virulensie geassosieer met tuberkulose hoofsaaklik gekarakteriseer deur proteomiese en genomiese benaderings. Ons glo dat metabolomika, 'n relatiewe nuwe navorsings benadering, moontlik nuwe metaboliet merkers kan identifiseer met betrekking tot die ESX-1 geen groep gebied, wat virulensie en die geassosieerde proteïene beter kan karakteriseer, as die verkrygde data interpreteer word deur gebruik te maak van sisteme biologie. Met behulp van 'n GCxGC-TOFMS metabolomika navorsings benadering, het ons die metaboloom van die *M. smegmatis* ESX-1 uitklopmutant aan dié van die wilde-tipe ouer stam vergelyk en daarna metaboliet merkers ge-identifiseer wat die grootste verskil aantoon. Multi- en enkelveranderlike statistiese analises was uitgevoer op die data om die metabolome van die betrokke stamme te ontleed. Hierdie bevindinge was gebruik om die metaboliete wat die grootste verskil dra tussen die twee groepe, te identifiseer. 'n Algemene toename in verskillende koolhidrate, aminosure en lipiede, wat verband hou met die selwand struktuur en funksie, is waargeneem in die metaboloom van die ESX-1 uitklopmutant stam, relatief tot dié van die wilde-tipe. Daarbenewens, nadat die twee groepe met mekaar vergelyk is, was verdere metaboliete gevind wat verband hou met die anti-oksidant sisteem, virulensie proteïen vorming en energie produksie teenwoordig in mikobakterieë. Hierdie metabolomika studie is enig in sy soort deurdat dit metaboliet merkers identifiseer wat verband hou met die ESX-1 geen groep gebied. Verder word die funksie wat

hierdie geen groep gebied vervul ten opsigte van virulensie asook die onderliggende metaboliese weë teenwoordig in mikobakterieë, deeglik bespreek. Verder word die verhoogde metaboliese aktiwiteit, groei/replikasie tempo, verhoogde selwand sintese en die veranderde anti-oksidant meganisme bespreek. Vanuit die resultate gegenereer uit die studie is dit duidelik dat al hierdie verskynsels sal bydra tot die betrokke organisme se toenemende patogenisiteit asook vermoë om te oorleef.

Sleutel terme: metabolomika; *Mycobacterium*; tuberkulose; ESX-1; virulensie; GCxGC-TOFMS.

LIST OF TABLES

Chapter 1:

Chapter 5:

LIST OF FIGURES

Chapter 1:

Chapter 4:

Chapter 5:

LIST OF ABBREVIATIONS

AraLAM	- Arabinose lipoarabanomannan
bp	- Base pairs
BCG	- Bacille de Calmette et Guérin
C	- Cytosine
CV	- Coefficient of variation
CFP	- Culture filtrate protein
CLR	- C-type lectin receptor
Ecc	- ESX conserved components
EphA–F	- Epoxide hydrolase A–F
ESAT	- Early secreted antigenic target protein
Esp	- ESX-1 secretion-associated proteins
ESX	- ESAT gene cluster region
ΔESX-1_{ms}	- *M. smegmatis* ESX-1 knock-out
ΔESX-3_{ms}	- *M. smegmatis* ESX-3 knock-out
ext RD	- Extended region of difference
FDA	- US Food and Drug Administration
GABA	- γ-Aminobutyric acid
G	- Guanine
GC	- Gas chromatography

GCxGC-TOFMS	- Gas chromatography gas chromatography time-of-flight mass spectrometry
GC-MS	- Gas chromatography–mass spectrometry
HIV	- Human immunodeficiency virus
IdeR	- Iron-dependent regulator
IFNγ	- Interferon-γ
IL	- Interleukin
IRF	- Interferon regulatory factor
IS	- Insertion sequences
kb	- kilobase
LAM	- Lipoarabinomannan
LB	- Luria-Bertani
LC-MS	- Liquid chromatography–mass spectrometry
LM	- Lipomannan
mAGP	- Mycolyl arabinogalactan–peptidogalactan complex
ManLAM	- Mannose lipoarabinomannan
MDR	- Multidrug-resistant
meso-DAP	- L-alaninyl-D-isoglutaminyl-meso-diaminopimelate
mg	- Milligram
MPS	- Multi-purpose sampler
MPTR	- Major polymorphic tandem repeats
MSTFA	- N-methyl-N-(trimethylsilyl) trifluoroacetamide

NMR - Nuclear magnetic resonance

ORF - Open reading frame

PAMP - Pathogen-associated molecular patterns

PC - Principal component

PCA - Principal component analysis

PCR - Polymerase chain reaction

PE - Proline–glutamic acid

PE_PGRS - Polymorphic GC-rich repetitive sequences

PI - Phosphatidylinositol

PIM - Phophatidyl-*myo*-inositol mannoside

PLS-DA - Partial least squares–discriminant analysis

PPE - Proline-proline-glutamic acid

PRPP - 5-Phosphoribosyl-1-pyrophosphate

PRR - Pattern recognition receptor

QC - Quality control

QC-CV - Quality control–coefficient of variation

RD - Regions of difference

ROS - Reactive oxygen species

SAM - S-adenosyl methionine

S/N - Signal to noise

T7S - Type Vll secretion

TAG	- Triacylglycerol
TB	- Tuberculosis
TCA	- Tricarboxylic acid
TNF-α	- Tumour necrosis factor-α
TLR	- Toll-like receptor
TMCS	- Trimethylchlorosilane
VIP	- Variables important in projection
WHO	- World Health Organization
ZUR	- Zinc uptake regulons

CHAPTER 1:

PREFACE

1.1 BACKGROUND AND MOTIVATION

Tuberculosis (TB) and other infectious diseases are responsible for approximatley 14 million deaths annually (Morens *et al*., 2004). In 2010, the World Health Organization (WHO) reported 8.8 million new cases of TB, with a staggering 650 000 cases of multidrug-resistant TB (MDR-TB), contributing to the 1.4 million reported deaths globally for that year (3 800 deaths a day), 95% of which were reported in developing countries (WHO, 2010). Additionally, 350 000 of the recorded 1.4 million mortalities were associated with the human immunodeficiency virus (HIV) with Europe and Africa contributing most towards HIV and TB co-infection rates (WHO, 2011). Although HIV and TB co-infected patients' mortality rates have declined with the advent of antiretroviral therapy, MDR-TB still contributes heavily to the mortality rates in rural areas of southern Africa (Gandhi *et al*., 2006).

TB and HIV co-infection, coupled with MDR-TB, has caused serious health problems in especially the control, management and treatment of this epidemic in a variety of developing countries (Khan *et al*., 2001). Although TB death rates have fallen by 40% since 1990, the prevalence of TB is still a major health problem with 320 000 women reported having died in 2010 (WHO, 2010). Various factors are thought to contribute to the general annual increase in TB and include: rising HIV and MDR-TB infection, weakened efforts from public health departments to control the prevalence of tuberculosis, rising poverty, homelessness in conjunction with overcrowding, malnutrition and MDR-TB (Leinhardt, 2001). Additional factors contributing to the ever-increasing prevalence of TB in many developing countries is the lack of, or the ineffectiveness of, TB vaccines, time-consuming diagnostic methods and unsuccessful therapy due to patients not adhering to specific treatment regimens (McNerney & Zager, 2008). South Africa has an abnormally high prevalence of HIV infection and this epidemic accounted for over 17% of the global burden in 2007. Post-apartheid conditions such as overcrowding, rising poverty, malnutrition, MDR-TB and the diminishing efforts of public health organisations to control the TB epidemic, provides a favourable environment for *Mycobacterium tuberculosis* (*M. tuberculosis*), the causative agent of TB, to flourish (Karim *et al*., 2009). These statistics highlight the importance of further research, not only towards better

understanding and characterising *M. tuberculosis*, but also to achieve better vaccines and treatments to fight TB.

The genome of *M. tuberculosis* contains five copies of a cluster of genes known as the ESAT-6 loci (or ESX gene clusters). These different clusters (ESX-1–ESX-5) contain members of the CFP-10 and ESAT-6 gene families as well as other genes responsible for encoding proteins that have a secretory function (Gey van Pittius *et al*., 2001). The importance of secreted proteins in the host–pathogen interaction of *M. tuberculosis* cannot be emphasised enough. Studying these proteins and their corresponding secretion systems, therefore, may give information leading to a better understanding of TB disease mechanisms and ultimately improved anti-tuberculosis vaccines, drugs and treatment (Simeone *et al*., 2009).

In this study, I focus on elucidating the functionality of the ESX-1 gene cluster and the associated proteins, using a metabolomics research approach. Proteins generated by the ESX secretion system, better known as the Type VII secretion (T7S) system, have been a topic of various research studies over the last couple of years, as it is believed that these proteins play a major role in the virulence of mycobacteria (Abdallah *et al*., 2007; Das *et al*., 2011). By investigating these secretion systems using a metabolomics approach, we hope to broaden the general understanding of the functionality of this gene cluster, using a systems biology approach.

1.2 STRUCTURE OF DISSERTATION

This dissertation is a compilation of chapters written specifically to comply with the requirements of North-West University, Potchefstroom campus, for the completion of the degree of Master of Science (Biochemistry) in dissertation format.

Chapter 1 gives a brief background and clear motivation for the study conducted. This chapter discusses the structure of the dissertation as well as the outcomes of the study. Lastly, the contributions and roles of each co-worker and co-author towards the completion of this study, as well as the resultant papers which emanated, are reported.

As the literature on TB metabolomics, and especially on metabolomics studies pertaining to the ESX-1 gene cluster, is limited, Chapter 2 provides a literature overview of all the aspects relating to ESX-1 and *Mycobacterium*, from a genomics, proteomics and pathophysiological perspective. Additionally, a full description of metabolomics and its capacity for biomarker discovery, as well as the use of 2-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC–TOFMS) as an analytical tool for metabolomics investigations, is given in Chapter 2.

Chapter 3 expounds the problem statement, aims and objectives of this metabolomics study.

All the metabolomics methodology for this study – materials and methods used for sample collection, sample preparation, derivatisation, data processing, statistical analyses as well as the analytical conditions – are described in Chapter 4.

Chapter 5 describes all the results obtained using the metabolomics research methods described in Chapter 4, in an aim to identify metabolite markers best characterising the difference between *M. smegmatis* wild-type and ΔESX-1_{ms} strains. Before any attempt to answer the biological question, the methods used for the metabolome analysis were validated by determining the repeatability of the extraction procedure, the GCxGC-TOFMS used as well as the analyst's ability to successfully perform the extraction procedure, pertaining specifically to the measure of repeatability for method validation purposes. Thereafter, using various univariate and multivariate statistical approaches, metabolite markers were identified in order to explain the variation between the previously mentioned strains. The metabolite markers identified are discussed in three sections regarding the metabolism of *M. smegmatis*: 1) metabolites associated with altered cell wall synthesis in ΔESX-1_{ms}; 2) metabolites associated with altered energy production and redox state in ΔESX-1_{ms}; and 3) how this relates to the role of the ESX-1 gene cluster in *Mycobacterium* virulence?, followed by a conclusion.

Finally, Chapter 6 considers all the results by way of a conclusion, and suggests future areas of research with the aim of characterising the functionality of ESX-1 and the underlying T7S system.

1.3 OUTCOMES OF THE STUDY

The papers generated by this study are attached in Appendix A and Appendix B, respectively, should the reader be interested. The review article in Appendix A was published in *Disease Markers* (Impact Factor 2.174). The paper entitled “A metabolomics investigation of the function of the ESX-1 gene cluster in mycobacteria” in Appendix B is pending publication, on the condition that successful gene complementation be done on the samples used in the study.

1.3.1 Manuscripts

Appendix A

1) Conrad C. Swanepoel and Du Toit Loots. 2014. The use of metabolomics in conjunction with functional genomics and proteomics for *Mycobacterium tuberculosis* research. *Disease Markers*, http://dx.doi.org/10.1155/2014/124218.

Appendix B

2) C.C. Swanepoel, M. Newton-Foot, Gey van Pittius, Nicolaas C and D.T. Loots, A metabolomics investigation of the function of the ESX-1 gene cluster in mycobacteria. *To be submitted*

1.4 AUTHOR CONTRIBUTIONS

The principal author of this thesis is Conrad Cilliers Swanepoel. The contributions of the co-authors, co-workers and collaborators made towards this work are given in Table1.1.

The following is a statement from the co-authors confirming their individual roles in the study and giving their permission that the data generated and conclusions made may form part of this dissertation.

I declare that my role in this study, as indicated in Table1.1, is representative of my actual contribution and I hereby give my consent that this work may be published as part of the M.Sc. dissertation of Conrad Cilliers Swanepoel.

Prof. D.T. Loots

C.C Swanepoel

Table 1.1: Research team

Co-author	Co-worker	Collaborator	Contribution
C. C. Swanepoel (B.Sc. Hons. Biochemistry)			Responsible, together with the promoter, for the planning, execution, data analyses, and writing of the thesis, publication, and all other documentation associated with this study.
Prof. D.T. Loots (Ph.D. Biochemistry)			Promoter: Co-ordinated and supervised all aspects of the study including: study design, planning, execution, and the writing of the thesis, publication, and all other documentation associated with this study.
		DST/NRF Centre of Excellence for Biomedical TB Research (CBTBR), Stellenbosch University, South Africa	Provided all the *M. smegmatis* cultured samples used in this study. Also assisted in interpretation of the data.

1.5 REFERENCES

ABDALLAH, A.M., GEY VAN PITTIUS, N.C., DIGIUSEPPE CHAMPION, P.A.D., COX, J., LUIRINK, J., VANDENBROUCKE-GRAULS, C.M.J.E., APPELMELK, B.J., BITTER, W. 2007. Type VII secretion — mycobacteria show the way. *Nature*, 5, 883-891.

DAS, C., GHOSH, T.S., MANDE, S.S. 2011. Computational analysis of the ESX-1 Region of *Mycobacterium tuberculosis*. Insights into the Mechanism of Type VII Secretion System. *PLoS ONE*, 6(11), 1-12.

GANDHI, N.R., MOLL, A., STURM, A.W., PAWINSKI, R., GOVENDER, T., LALLOO, U., ZELLER, K., ANDREWS, A., FRIEDLAND, G. 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet*, 368, 1575-1580.

LEINHARDT, C. 2001. From exposure to disease: The role of environmental factors in suspectibility to and development of tuberculosis. *Epidemiologic Reviews*, 23(2), 288-301.

KHAN, M.Y., KINSARA, A.J., OSOBA A.O., WALI, S., SAMMAN, Y., MEMISH, Z. 2001. Increasing resistance of *M. tuberculosis* to anti-TB drugs in Saudi Arabia. *International Journal of Antimicrobial Agent*s, 17, 415-418.

GEY VAN PITTIUS, N.C., GAMIELDIEN, J., HIDE, W., BROWN, G.D., SIEZEN, R.J., BEYERS, A.D. 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C gram-positive bacteria. *Genome Biology*, 2(10), 1-18.

KARIM, S.S.A., CHURCHYARD, G.J., KARIM, Q.A., LAWN, S.D. 2009. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *Lancet*, 374, 921-933.

MCNERNEY, R., ZAGER, E.M. 2008. Multidrug-resistant tuberculosis. *BMC Infectious Diseases*, 8(10), 1-5.

MORENS, D.M., FOLKERS, G.K., FAUCI, A.S. 2004. The challenge of emerging and re-emerging infectious diseases. *Nature*, 430, 242-249.

SIMEONE, R., BOTTAI, D., BROSCH, R. 2009. ESX/type VII secretion systems and their role in host–pathogen interaction. *Current Opinion in Microbiology*, 12, 4-10.

WORLD HEALTH ORGANIZATION (WHO). 2010. Global tuberculosis control 2010. World Health Organization, Geneva, Switzerland.

WORLD HEALTH ORGANIZATION (WHO) 2011. World Health Organization 2011/2012 tuberculosis global facts. Geneva, Switzerland.

CHAPTER 2:
INTRODUCTION

2.1 Introduction

Since the WHO declared TB a global public health emergency approximately 20 years ago (WHO, 2011), mortality rates had gradually declined by approximately 50%. However, despite falling mortality rates, an estimated 8.6 million individuals were infected with TB in 2012, which resulted in the 1.3 million global deaths (320 000 of which were HIV co-infected cases) (WHO, 2013). Elucidating TB disease mechanisms is still considered the key for developing better vaccination strategies and treatment, with the aim of eventually curbing this epidemic.

In this chapter, I describe the pathogenesis of *M. tuberculosis*, as well as all the factors contributing towards mycobacterial virulence, with particular focus on the ESAT gene cluster region (ESX) 1 of the Type VII secretion (T7S) system. I also describe the research approach called metabolomics, which is used in this study to accomplish the aims of the research, and how this can be used in combination with other "omics" approaches (termed systems biology), for deciphering the functionality of the T7S system and its associated proteins.

2.2 Pathophysiology of *M. tuberculosis*

Mycobacterium tuberculosis infection occurs through airborne transmission from already infected individuals coughing, laughing, singing or sneezing. The bacteria end up in the upper part of the host's airways where the mucus-secreting goblet cells are located (Knechel, 2009). After the tubercle bacillus evades the initial host defence mechanism, these infected aerosols end up in the lungs of the host where the bacteria first encounter alveolar macrophages (Knechel, 2009; Kleinnijenhuis *et al*., 2011). These specialised cells exhibit germline-encoded receptors, namely pattern recognition receptors (PRRs), which recognise characteristic patterns of the unfamiliar structure (Kapsenberg, 2003), classifying the foreign substance as either a microbe or a harmless substance. These patterns recognised by the PRRs are termed the pathogen-associated molecular patterns (PAMPs) and serve as an initiation step before any immune response can commence (Kapsenberg, 2003; Kleinnijenhuis *et al*., 2011). Additional mechanisms and other macrophage receptors

also contribute to phagocytosis and the recognition of the mycobacterial microbes (Knechel, 2009; Russell, 2011), and include the complement system, which binds C3 proteins onto the cell wall of *M. tuberculosis,* optimising recognition by the host macrophages (Knechel, 2009). After recognition of the pathogen by various PRRs, Toll-like receptors (TLRs) and C-type lectin receptors (CLRs) recognise polysaccharide structures of pathogens and initiate the production of intracellular signals, eventually leading to activated macrophages, followed by phagocytosis and intracellular killing (Kleinnijenhuis *et al*., 2011).

Various pathogenic mycobacteria escape the initial intracellular destruction, which leads to multiplication and disruption of host macrophages (Kleinnijenhuis *et al*., 2011). Proteolytic enzymes, chemokines and pro-inflammatory cytokines such as tumour necrosis factor-α (TNF-α), interleukin (IL)-1β and IL-18, are readily secreted by macrophages in the inflammatory region (Beltan *et al*., 2000; Knechel, 2009; Kleinnijenhuis *et al*., 2011). These proteins act as signal inducers, luring monocytes, T-lymphocytes and neutrophils to the area of inflammation (Russell, 2011). Monocytes differentiate into tissue-specific macrophages where these immunological cells ingest the *Mycobacterium* pathogen (Knechel, 2009; Kleinnijenhuis *et al*., 2011). In time, the number of blood-derived macrophages increases, as mycobacteria replication increases logarithmically. T-cell immunity develops and antigen-specific T-lymphocytes arrive 14–21 days after the initial infection (Kleinnijenhuis *et al*., 2011), which results in the secretion of pro-inflammatory cytokines such as interferon-γ (IFNγ), activating macrophages for the purpose of eliminating the intracellular mycobacteria (Knechel, 2009). A direct consequence of this is a decline in the early logarithmic bacillary growth rate with the central solid necrosis present in the primary lesions, or the formation of a granuloma (responsible for the inhibition of mycobacterial growth) (Knechel, 2009; Kleinnijenhuis *et al*., 2011). This in turn may lead to a number of scenarios. In some patients, *M. tuberculosis* infection becomes dormant as a direct consequence of the granuloma suppressing extracellular growth (Kleinnijenhuis *et al*., 2011). Disease prognosis is usually determined by the quality of the host defences, as well as the equilibrium between the invading mycobacterial pathogen and host defences (Knechel, 2009). A few infected individuals, however, present with an active *Mycobacterium* infection leading to progressive TB in the lungs (Russell, 2011), while the minority of patients

exhibit hematogenous dissemination (Kleinnijenhuis *et al*., 2011), which can subsequently lead to an active TB infection in any organ of the host (Galimi, 2011).

Understanding the pathophysiology of *M. tuberculosis* not only improves our general understanding of the mechanisms causing infection by mycobacteria (Russell, 2011), but is also essential to developing novel strategies for improved vaccination and treatment of this disease (Abdallah *et al*., 2006).

2.3 *Mycobacterium* in general

Improved characterisation of mycobacteria may aid in the overall understanding of the mechanisms associated with TB, and ultimately lead to improved TB treatment, diagnosis and vaccination, in the quest eventually to rid the world of this epidemic. *Mycobacterium* species are characterised as rod-shaped, aerobic, non-motile, Gram-positive bacteria, with a high guanine (G) + cytosine (C) content (Cole *et al*., 1998; Abdallah *et al*., 2007). Although species within this genus are characterised as being Gram-positive bacteria, the cell wall of *M. tuberculosis* exhibits strong features of Gram-negative bacteria as well (Fu and Fu-Liu, 2002). The ESX gene clusters have been identified in the genome of *M. tuberculosis* H37Rv as well as in the genomes of eight other strains and species of mycobacteria (Gey van Pittius *et al*., 2001), including: *M. tuberculosis* CDC1551 and 210, *M. bovis* AF2122/97, *M. bovis* Bacille Calmette Guerin (BCG) Pasteur 1173P2, *M. leprae* TN, *M. avium* 104, *M. paratuberculosis* K10 as well as *M. smegmatis* MC^2155 (Gey van Pittius *et al*., 2001). Furthermore, pathogenicity varies among the different species, with *M. tuberculosis*, *M. africanum*, *M. canetti*, *M. bovis* and *M. avium* being virulent towards a human host, whereas the *M. smegmatis* and *M. phlei* strains are considered avirulent, and hence are classified as non-pathogenic (Beltan *et al*., 2000; Gey van Pittius *et al*., 2001; Abdallah *et al*., 2007).

Various studies have postulated that the variation between intra-species (pathogenic and non-pathogenic) doubling times, characterising different strains of mycobacteria, are due to the observed dissimilarity of their respective genome lengths, with longer doubling times associated with shorter and simpler genomes (Brosch *et al*., 2000). However, the morphological variations between various *Mycobacterium* strains give

expression to differences in their respective doubling times (Hoffmann *et al*., 2008). Interestingly, the *M. bovis* genome shares a 99.95% similarity with the genome of *M. tuberculosis*, as a number of deletions (region of difference) results in a much shorter *M. tuberculosis* genome length (Garnier *et al*., 2003), as will be explained later.

Morphological differences in the subspecies of *M. tuberculosis* are thought to be due to variation in their mycolic acids, glycolipids and various other membrane-bound molecules located within their respective cell walls (Hoffmann *et al*., 2008). Mycolic acids fulfil critical roles in biofilm formation, whereupon an extracellular polymeric substance forms colonies of *M. tuberculosis* and also contributes to resistance of this pathogen to TB antibiotics (Ojha *et al*., 2008; Verschoor *et al*., 2012). Furthermore, when comparing *M. tuberculosis* and *M. bovis* with BCG *M. bovis* and *M. canetti*, it is apparent that the latter two consist of more wrinkled colonies and this phenotypical characteristic is a direct result of larger amounts of phenolic glycolipids present in their respective cell walls (Reed *et al*., 2004). As most of the proteins associated with the T7S system are secreted over the mycobacterial cell wall, characterisation of the cell wall is essential to understanding this secretion system (Abdallah *et al*., 2007) and the various factors associated with it.

2.4 The *Mycobacterium* cell envelope

Various studies have been conducted on the biosynthesis and composition of the *Mycobacterium* cell wall. A key characteristic of this complex envelope is the presence of very-long-chain, α-alkyl branched and β-hydroxylated fatty acids comprising 70 to 90 carbon atoms, called mycolic acids (Zhang *et al*., 2003; Kleinnijenhuis *et al*., 2011). These fatty acids are divided into two groups of α-mycolates: 1) with oxygenated functional groups and 2) those without (Chatterjee, 1997). The assembly of mycolic acids located in the cell envelope of *M. tuberculosis* are categorised into three main classes: 1) α-mycolic acids, 2) keto-mycolic acids and 3) methoxy-mycolic acids (Verschoor *et al*., 2012). Each respective class, located within the mycomembrane of this pathogen, differs with regard to its level of contribution to pathogenicity (Verschoor *et al*., 2012). *M. tuberculosis*, and other slow growing pathogenic mycobacteria, are able to modify these mycolic acids with

cyclopropanation, whereas non-pathogenic fast growing species like *M. smegmatis* do not (Chatterjee, 1997). The rigidity and hydrophobicity of the cell membrane is greatly influenced by the abundance of mycolic acids (Zhang *et al*., 2003). Owing to the role that these mycolic acids play in pathogenicity, various anti-TB medications have been developed which act by inhibiting the synthesis of mycolic acids, as well as other essential entities situated in the mycomembrane, in an attempt to disrupt the cell and halt bacterial growth (Yuan *et al*., 1997; Olivier and Loots, 2011).

The cell wall in totality consists predominantly of three distinguishable layers, namely: 1) the capsule, 2) the previously mentioned mycolyl arabinogalactan-peptidogalactan (mAGP) complex (also known as the mycomembrane or cell wall core) (Brennan, 2003), and 3) the inner membrane, rich in phospholipids (Abdallah *et al*., 2007). The exterior of the cell envelope is rich in various polysaccharides, including glucan, mannan and arabinomannan (Brennan, 2003). The mAGP layer, on the other hand, is far more complex and consists of the mycolic acids mentioned above, covalently bound to arabinogalactan by disaccharide–phosphate bonds, with D-galactan subsequently providing a strong covalent interaction between these components (Chatterjee, 1997; Zhang *et al*., 2003). Peptidoglycan and arabinogalactan are essential components, and together these molecules span the periplasmic space between the inner phospholipid bilayer and the mycomembrane (Abdallah *et al*., 2007). In so doing they form the backbone of the mAGP complex (Brennan, 2003; Zhang *et al*., 2003). Peptidoglycan comprises multiple interchanging units of *N*-acetylglucosamine and *N*-glucolylmuramic acid (Chatterjee, 1997), with tetrapeptide side chains linking the latter two molecules to muramic acid and arabinogalactan (Brennan, 2003). These side chains consist of L-alaninyl-D-isoglutaminyl-meso-diaminopimelyl-D-alanine (meso-DAP) with the diaminopimelic acids being amidated (Chatterjee, 1997). Various bacteria contain peptidoglycan in their cell walls (Chatterjee, 1997); however, mycobacterial peptidoglycan is distinct from these in that it provides extreme cell wall strength by cross-linking with muramic acid (Azuma *et al*., 1970), diaminopimelic acid and D-alanine (McNeil *et al*., 1990). Studies have shown that the *N*-glycolyl group attached to muramic acid, making up peptidoglycan, is exceptionally resistant to lysozyme attack (Chatterjee, 1997) and, although this phenomenon is not fully understood, it is believed to contribute towards protecting the organism from enzymatic degradation (Birken, 2008). Mycobacterial

arabinogalactan, in contrast to that of peptidoglycan, does not contain alternating units; however, it comprises a small number of defined structural motifs (Chatterjee, 1997). Furthermore, this polysaccharide consists exclusively of D-galactofuranoses and D-arabinofuranoses (Chatterjee, 1997). Free lipids, including phenolic glycolipids, phthiocerol dimycocerosates, dimycolyltrehalose, sulpholipids and phophatidyl-*myo*-inositol mannosides (PIMs), complete the outer part of the mycomembrane (Chatterjee, 1997; Abdallah *et al*., 2007), as these entities are interlinked with mycolic acids (Verschoor *et al*., 2012).

Lipoarabinomannan (LAM) is a lipoglycan located within the mycobacterial cell envelope, where it exerts a pathogenic function, contributing to mycobacterial virulence (Chatterjee & Khoo, 1998). A further key characteristic of LAM is the presence of a phosphatidylinositol (PI) core located at the reducing end of the macromolecule, making this glycolipid the multiglycosylated version of PIMs (Chatterjee, 1997), as the biosynthesis of PIM is defined as: PI → PIM_1 → PIM_2 → PIM_3 → PIM_4 → PIM_5 → PIM_6 (Chatterjee and Khoo, 1998). LAM is predominantly rooted in the inner phospholipid membrane and extends through the mycomembrane, with the arabinan termini exposed on the exterior of the capsule, thereby making LAM interspersed throughout the cell wall (Chatterjee and Khoo, 1998; Brennan, 2003). Related precursors, lipomannan (LM) and the PIMs, illustrate identical cellular locations (Briken *et al*., 2004). Subsequently, the exposure of LM and PIMs to the extracellular environment, due to its positioning in the cell wall, contributes to this physiological functioning (Chatterjee & Khoo, 1998). Gelber *et al*. (1990) used *M. tuberculosis* and *M. leprae* to demonstrate that LAM interacts with its immediate environment. These pathogenic mycobacteria possess capped LAM at the arabinan termini with a mannose molecule (ManLAM), whereas the saphrophytic, non-pathogenic *M. smegmatis* is mostly uncapped (Gilleron *et al*., 1997; Kleinnijenhuis *et al*., 2011). However a minority of organisms exhibit LAM capped with inositol-phosphate (AraLAM) at their arabinan termini (Chatterjee, 1997). The differences in these caps between mycobacterial species are believed to contribute to their degree of pathogenicity, and act by suppressing the immune response and facilitating the production of cytokines derived from macrophages (Chatterjee, 1997). Furthermore, LAM is a contributing factor to various clinical manifestations and studies have demonstrated that it plays a role in the inhibition of host T-cell and IFN-

γ activation, protein kinase C activity, the scavenging potential of oxygen free radicals and also suppressing macrophage activity (Chatterjee, 1997; Kleinnijenhuis *et al*., 2011). All of these factors are therefore thought to contribute to the virulence demonstrated by pathogenic *Mycobacterium* species.

Considering the structure of the mycobacterial cell envelope discussed above, focusing on its impenetrable nature, the need for various secreted proteins and transport systems becomes evident. These include the SecA1 general pathway (Feltcher *et al*., 2010), the non-essential SecA2 pathway (DiGiuseppe Champion & Cox, 2007), and finally the Type VII secretion pathway (Simeone *et al*., 2009), which is the focus of this study.

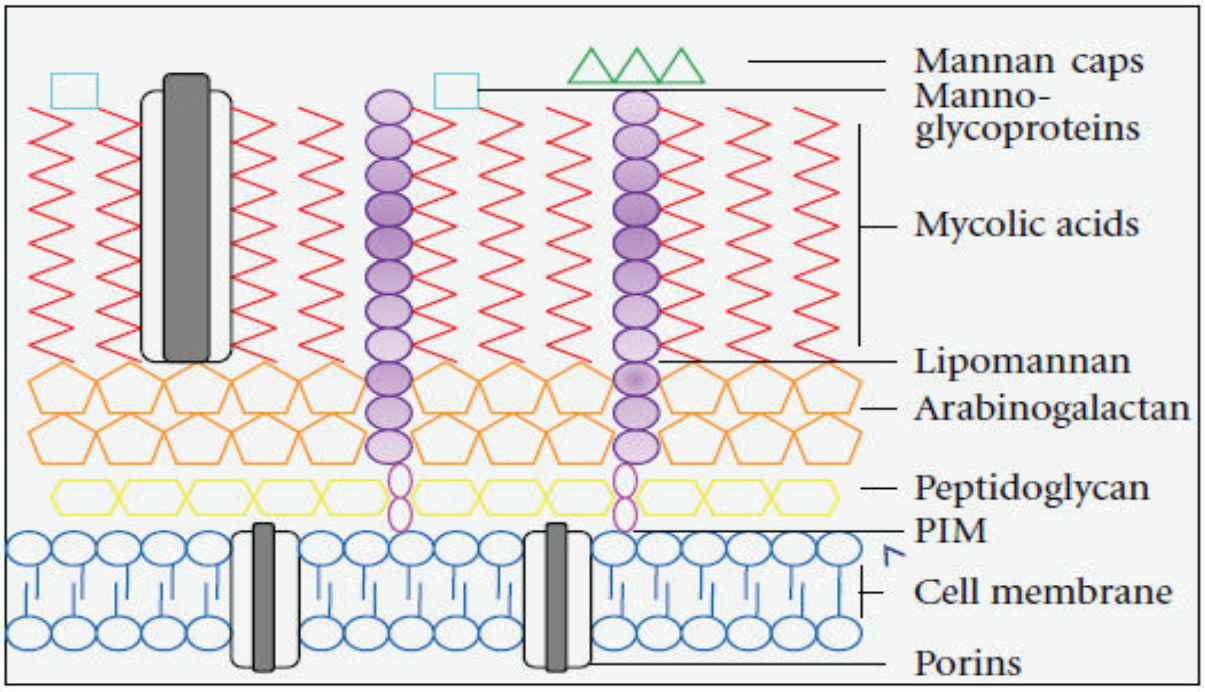


Figure 1.1: A schematic illustration of the cell envelope of *M. tuberculosis* (Kleinnijenhuis *et al*., 2011), illustrating two of the three layers identified in the cell envelope (the capsule being absent). More importantly, it indicates the mannan caps associated with LM and LAM, which contribute to mycobacterial virulence. Additionally, these caps are recognised by macrophages and initiate an immune response.

2.5 *Mycobacterium* genomics

2.5.1 Overview of the *Mycobacterium* genome

In 1882, Robert Koch isolated the causative agent of TB, *M. tuberculosis* (Koch, 1882). The genome sequence of *M. tuberculosis* H37Rv was completed and published in 1998 (Cole *et al*., 1998). Since the publication of the latter mentioned genomic data, our general understanding of the physiology, regulation and specific functions associated with the pathogen in particular, increased significantly (Domenech *et al*., 2001; Kleinnijenhuis *et al*., 2011). Following its annotation by Camus *et al*. (2000), it was confirmed that the genome of *M. tuberculosis* contains approximately 4.5 million nucleotide base pairs (bp), which makes it the second-largest bacterial genome so described (Cole *et al*., 1998). Furthermore, the genome consists of more than 3920 genes that encode for proteins larger than 80 amino acids in length (Tekaia *et al*., 1999) as well as an additional 50 genes that code for stable RNA (Cole *et al*., 1998). These genes account for no less than 91% of the total coding capacity of this pathogen (Cole *et al*., 1998). Interestingly, of all the proteins coded for by the genome of *M. tuberculosis*, only 40% have been associated with a function, with limited information available on the rest (Cole *et al*., 1998). Furthermore, 16% of the genome shows no similarity to proteins identified in other organisms and is thought to be *Mycobacterium* specific (Tekaia *et al*., 1999; Domenech *et al*., 2001). A large proportion of the genes in this genome (250 in all) codes for the enzymes involved in fatty acid metabolism, with *M. tuberculosis* producing over 100 enzymes involved in the β-oxidation of fatty acids (Domenech *et al*., 2001). Consequently, this organism exhibits a vast array of lipophilic molecules, varying in complexity, especially in the cell envelope (Tekaia *et al*., 1999; Domenech *et al*., 2001), as discussed above. Additionally, 11 genes encode for small proteins of approximately 100 amino acids in length, which share sequence homologies with ESAT-6, and are subsequently classified as part of the *esat-6* gene family (Gey van Pittius *et al*., 2001), which will be described in detail later.

Furthermore, 10% of the genome encodes for two new, unrelated proteins rich in glycine, termed Pro-Glu (PE) and Pro-Pro-Glu (PPE) (Cole *et al*., 1998; Gey van Pittius *et al*., 2006). Their names are derived from the proline and glutamic acid residues located in their respective N-terminal domains (Bottai & Brosch, 2009). More specifically, the PE family of proteins contains a proline-glutamic acid motif at

positions 8 and 9 in their N-terminal domains, stretching 110 amino acids in length, whereas proteins belonging to the PPE family exhibit a proline-proline-glutamic acid motif at both positions 7 and 9 (Gey van Pittius *et al*., 2006; Forrellad *et al*., 2013). Although information describing the functionality of these enzymes is very limited (Domenech *et al*., 2001), it is believed that the PE/PPE families play distinct roles in the antigenic variation of pathogenic *Mycobacterium* species (Voskuil *et al*., 2004).

From genetic information, *M. tuberculosis* is believed to synthesise vitamins, enzyme co-factors and all the essential amino acids required for ensuring viability (Tekaia *et al*., 1999; Domenech *et al*., 2001), in addition to all the enzymes required for the major metabolic pathways including lipid metabolism, glycolysis, the pentose phosphate pathway, the tricarboxylic acid (TCA) and the glyoxylate cycles (Cole *et al*., 1998), and, most recently, the citramalic acid cycle (Olivier and Loots, 2011).

Regulation is an essential factor in the viability of mycobacteria and the genome codes for at least 100 regulatory proteins, with 13 sigma factors responsible for initiating transcription (Cole *et al*., 1998). Studies have shown that insertion sequences (IS) are responsible for 3.4% of the genome's coding potential (Cole, 2002). Additionally, 16 copies of IS*6110* (a member of the IS3 family) was identified with this sequence, showing similarities to other IS elements. As IS*6110* is exclusively found in *M. tuberculosis* and *M. bovis*, it could potentially serve for the identification of *M. tuberculosis* complex (Thierry *et al*., 1989). It is also believed to play a role in genome flexibility (Cole, 2002), although the location on the genome of these IS sequences is believed to be near tRNA genes in a non-coding region (Cole *et al*., 1998).

The *M. tuberculosis* complex comprises a variety of tuberculosis-causing species, namely, *M. tuberculosis*, *M. canetti*, *M. africanum*, *M. microti*, *M. caprae*, *M. pinipedii*, and *M. bovis.* The variations in their genomes result in the phenotypic and morphological differences associated with these organisms (Brosch *et al*., 2000; Warren, *et al*., 2006; Forrellad *et al*., 2013). These genetic differences, however, are so small (0.1%) that differentiation between these groups is usually done on the basis of the observable phenotypic traits (Brosch *et al*., 2001). All *Mycobacterium* species can be divided broadly into two groups: 1) fast-growing strains (mostly non-pathogenic species) and 2) slow-growing strains (mostly pathogenic species)

(Brosch *et al*., 2001; Forrellad *et al*., 2013). As previously mentioned, morphological differences (due to variations in cell wall composition) play a distinct role in the doubling time of these species making up the *M. tuberculosis* complex (Hoffmann *et al*., 2008). Despite this, all exhibit an almost identical genetic composition, although their virulence varies to some degree (Brosch *et al*., 2002). It has been shown that the genome length (Hoffmann *et al*., 2008), in combination with the regions of difference (RD) (Lewis *et al*., 2003), contribute to the degree of virulence of difference *Mycobacterium* species (Garnier *et al*., 2003).

For the purpose of this study, the regions of difference, the ESAT-6 gene cluster (in particular ESX-1) and the PE and PPE proteins will be described in greater detail.

2.5.2 The regions of difference

It has been widely speculated that all species making up the *M. tuberculosis* complex share a common ancestor (Cole *et al*., 1998; Gey van Pittius *et al*., 2006), and through successive DNA deletions and/or insertions, every entity within this complex has evolved to exhibit the associated characteristic phylogenic and pathogenic differences (Forrellad *et al*., 2013). In the *M. tuberculosis* complex, genomic studies identified 14 regions of difference (RD), characterised as a small group of genes differing between species within the complex (Brosch *et al*., 2000). All 14 RD regions (RD1 to RD14) were successfully identified in the genome of *M. tuberculosis* H37Rv, which is normally used as a laboratory reference strain (Brosch *et al*., 2000; Forrellad *et al*., 2013). The first region, RD1, consists of 9 genes, Rv3871–Rv3879c, and is 9.5 kb in length (Pym *et al*., 2002; Forrellad *et al*., 2013). Included in these are the genes *esxA* and *esxB*, which encode for the secreted proteins ESAT-6 and CFP-10, respectively (Pym *et al*., 2002). These proteins are thought to contribute greatly to mycobacterial virulence upon successful secretion (Ganguly *et al*., 2008; Guo *et al*., 2012), which will be explained later. Furthermore, studies have shown the absence of these secreted antigenic proteins to be the main reason for attenuation of the *M. bovis* BCG strain (Domenech *et al*., 2001), as this vaccine strain does not contain RD1 (Lewis *et al*., 2003). These findings further confirm that RD1 contributes greatly to mycobacterial virulence, as loss of this region resulted in attenuation of virulent *M. tuberculosis* H37Rv (Andersen *et al*., 2005; Brodin *et al*., 2006). Other

studies have indicated that successful introduction of the *M. tuberculosis* RD1 region to the genome of *M. bovis* BCG restored secretion of ESAT 6 (Ganguly *et al*., 2008), which consequently increased virulence and immunogenicity (Lewis *et al*., 2003).

Furthermore, genes that encode for proteins belonging to the ESAT-6 family have also been identified in other RD regions, more specifically RD5 and RD8 (Brosch *et al*., 2000). Every mycobacterial RD region codes for different organism-specific entities fulfilling crucial roles with respect to the metabolism of these organisms (Brosch *et al*., 2000). It is believed that RD2 encodes for Mpb64, another immunogenic secretory protein, and a member of the regulatory LysR family of proteins (Brosch *et al*., 2000; Forrellad *et al*., 2013). Another example of this is RD8, which encodes for the majority of epoxide hydrolase (EphA–F) enzymes, which are responsible for detoxifying lipid peroxidation products originating from activated macrophages (Arand *et al*., 1994; Brosch *et al*., 2000). RD12 and RD13 (shown to be absent from both *M. bovis* strains) (Behr *et al*., 1999) encode for two cytochrome P450 enzymes, a thiosulfate sulfurtransferase, a methyltransferase as well as a molybdopterin converting factor (Brosch *et al*., 2002).

Apart from the above, RD1 comprises genes encoding for proteins secreted via a specialised secretion system termed the T7S system, which also implicates genes located outside the RD1 locus, thus forming part of an extended RD1 region (ext RD1) (Ganguly *et al*., 2008). These groups of genes, in combination with various other genes directly upstream of RD1, therefore collectively form the ESX-1 gene cluster region (Brosch *et al*., 2000: Ganguly *et al*., 2008), which is the focus of this study.

Although various hypotheses have been proposed, the exact functions of the secreted proteins, as well as the mechanism(s) the T7S system uses, is unknown (Abdallah *et al*., 2007; Ganguly *et al*., 2008). Future studies on the functionality of the RD1 region may give insight into the mechanisms associated with the virulence of these infectious organisms (Das *et al*., 2011).

2.5.3 The ESX gene cluster region

The 5 different ESX gene clusters present in *M. tuberculosis* H37Rv were identified as ESX-1 (Rv3866–Rv3883c), ESX-2 (Rv3884c–Rv3895c), ESX-3 (Rv0282–Rv0292), ESX-4 (Rv3444c–Rv3450c) and ESX-5 (Rv1782–Rv1798) (Das *et al*., 2011; Gey van Pittius *et al*., 2006; Forrellad *et al*., 2013). Collectively, these gene clusters are known as the ESX gene cluster region, as they show similarity both in gene content and gene order (Forrellad *et al*., 2013). Interestingly, also present in this array of genes are those conserved in four of the five regions, consequently termed the ESX conserved components (Ecc) (Abdallah *et al*., 2007; Bottai & Brosch. 2009). Furthermore, this region contains members of the gene family coding for ESX-1 secretion-associated proteins (Esp), namely, ESAT-6 (*EsxA*) and CFP-10 (*EsxB*) (Gey van Pittius *et al*., 2001; Das *et al*., 2011; Forrellad *et al*., 2013). More specifically, these genes (*EsxA* and *EsxB*) are located within RD1, a small portion of ESX-1 (Das *et al*., 2011; Feltcher *et al*., 2010), as stated previously. Additionally, four other gene pairs encoding for CFP-10/ESAT-6 were identified in regions ESX-2 to ESX-5 (one pair in each ESX region), as well as six other additional gene pairs outside any of the ESX loci, encoding CFP-10/ESAT-6-like proteins (Das *et al*., 2011).

The complex mechanism responsible for the transport of these and other synthesised proteins across the cell envelope of mycobacteria, as well as their secretion, is driven by the T7S system (Das *et al*., 2011). These proteins are encoded by a number of genes present in the ESX gene cluster region (Ganguly *et al*., 2008), as will be discussed later. Each ESX region exhibits a different, independent function (Gey van Pittius *et al*., 2001), but it is believed that the six genes shared between the different ESX regions encode for the necessary components of the T7S system (Abdallah *et al*., 2007; Das *et al*., 2011). ESX-5 is absent from the genome of the non-pathogenic *M. smegmatis* strain (Gey van Pittius *et al*., 2001), and further evidence indicates other ESX regions to be absent from various species within the *M. tuberculosis* complex (Gey van Pittius *et al*., 2006).

2.5.3.1 ESX-1

This gene cluster has received the most attention as it potentially provides an attractive target for anti-TB drug development. ESX-1 and the resultant secretion pathway are also associated with the virulence of pathogenic mycobacteria such as *M. tuberculosis*, *M. bovis* and *M. marinum* (Feltcher *et al*., 2010), as it is thought to function as the transport system for the RD-1-encoded proteins ESAT-6 and CFP-10 (Guo *et al*., 2012). Consequently, any genetic mutation within the ESX-1 locus leads to the attenuation of virulence (Pym *et al*., 2003). Lewis *et al*. (2003) observed that removal of ESX-1 from *M. tuberculosis* resulted in reduced virulence. Additionally, after supplementing avirulent BCG *M. bovis* with a functional ESX-1 locus, a dramatic increase in virulence was observed (Pym *et al*., 2002), further confirming the role of ESX-1 in pathogenicity (Forrellad *et al*., 2013). Furthermore, Sassetti & Rubin (2003) indicated that the ESX-1 gene region was also involved in cell viability, directly influencing the *in vivo* growth of *M. tuberculosis*. Various other proteins have also been shown to be secreted by the ESX-1 machinery, namely, Rv3872 (PE35), Rv3873 (PPE68) and Rv3881c (EspB) (McLaughlin *et al*., 2007; Raghavan *et al*., 2008; Das *et al*., 2011).

RD-1 encodes various other components complementing the functionality of the ESX-1 secretion system (Abdallah *et al*., 2007). Evidence of a number of other genes and proteins involved in ESAT-6 and CFP-10 secretion emphasises the complex nature of the *Mycobacterium* ESX-1-related secretion system and the associated virulence (Das *et al*., 2011). The ESX-1 gene cluster within *Mycobacterium* species exhibit different functions and in *M. marinum* it is involved in virulence as well as haemolysis (Tobin and Ramakrishnan, 2008), whereas in *M. smegmatis* it is believed to be involved in conjugation (Abdallah *et al*., 2007) and DNA transfer (Converse & Cox, 2005; Serafini *et al*., 2009).

Various other genes located outside the ESX-1 gene cluster contribute extensively to the secretion of ESAT-6, CFP-10 and other virulence proteins (Das *et al*., 2011), which will be discussed later.

2.5.3.2 ESX-2

Although little is known about the functionality of ESX-2 (Bitter *et al*., 2009), no genes for PE and PPE proteins have been located in the ESX-2 gene cluster (Gey van Pittius *et al*., 2001). Furthermore, neither ESX-2 nor ESX-4 is predicted to be essential for *in vitro* growth or virulence of mycobacteria (Feltcher *et al*., 2010). Interestingly, however, Abdallah *et al*. (2007) proved that during conditions of starvation, ESX-1 is down-regulated, whereas ESX-2 is up-regulated, illustrating genetic regulation within the ESAT-6 gene cluster region. The lack of ESX-2 complementation could be due to the divergent evolution of the secretion signals, as well as their differential regulation (Abdallah *et al*., 2007).

2.5.3.3 ESX-3

Previous studies have shown ESX-3 to play a major role in bacterial cell viability (Sassetti & Rubin, 2003) through divalent cation homeostasis (Rodriquez *et al*., 2002) of iron and zinc in particular (Serafini *et al*., 2009; Siegrist *et al*., 2009; Loots *et al.,* 2013). Serafini and co-workers (2009) discovered that ESX-3 is the only ESX gene cluster essential for the *in vitro* growth of *M. tuberculosis*, by comparing *M. tuberculosis* wild-type parent strains with the equivalent ESX-3 knock-out mutants. They further demonstrated that supplementing these knock-out strains with elevated iron or zinc supported their growth. Further evidence supporting this growth functionality was the discovery that ESX-3 is regulated by both an iron-dependent regulator (IdeR) (Simeone *et al*., 2009) and a zinc uptake regulator (Zur) in *M. tuberculosis* (Abdallah *et al*., 2007). Siderophores, produced by mycobacteria, fulfil critical roles in iron acquisition and uptake and are categorised into two groups, namely, 1) membrane-bound mycobactins, and 2) secreted exochelins (Ratledge & Dover, 2000; Loots *et al*., 2013). As a result, various studies have shown that ESX-3 secretes a unique factor, responsible for iron and zinc acquisition via their respective mycobactin and zinc-uptake systems (Loots *et al*., 2013; Siegrist *et al*., 2009). In a metabolomics investigation, Loots *et al*. (2013) confirmed these findings and indicated a reduction in metabolite products synthesised using zinc-dependent enzymes in *M. smegmatis* ESX-3 knock-out (ΔESX-3_{ms}) samples when compared

with an equivalent wild-type strain. Furthermore, Loots and co-workers (2013) illustrated the opposite for iron, and this manifestation in the ΔESX-3_{ms} strain was attributed to up-regulation of iron acquisition genes overcompensating for the loss of ESX-3. Moreover, ESX-3 is thought to be the first region to have been copied from ESX-4 (Gey van Pittius *et al*., 2001).

2.5.3.4 ESX-4

ESX-4 provides information on the evolution of mycobacteria, and is considered to be the evolutionary origin of this genus, as well as of all the ESX gene clusters (Gey van Pittius *et al*., 2001). It is also the only region for which an orthologues could be found in the genomes of organisms phylogenetically related to *Mycobacterium,* such as *Corynebacterium diphtheriae* and *Streptomyces coelicolor* (Gey van Pittius *et al*., 2001; Gey van Pittius *et al*., 2006). It has further been shown to be regulated by sigma factor SigM, and is considered to be a primitive T7S system, present in other high G+C-rich, Gram-positive bacteria (Bitter *et al*., 2009). Additionally, ESX-4 does not encode for any PE or PPE proteins (Abdallah *et al*., 2006), and is considered to be the smallest of all the ESX gene clusters, as it contains only 7 genes (Abdallah *et al*., 2007). As is the case with ESX-2, no ESX-4 ESAT-6 homologues have yet been detected in the extracellular content immediately adjacent to the *Mycobacterium* cell envelope; any disruption of the genes from these clusters has no direct outcome on the physiology of *M. tuberculosis* (Abdallah *et al*., 2007; Bitter *et al*., 2009).

2.5.3.5 ESX-5

In evolutionary terms, the ESX-5 gene cluster is considered the youngest of all the ESX gene clusters and has been associated with virulence as well as for encoding and secreting various PE and PPE proteins (Gey van Pittius *et al*., 2001; Simeone *et al*., 2009). In *M. marinum*, a group of PE proteins, polymorphic GC-rich repetitive sequences (PE_PGRS), are secreted via an ESX-5 secretion system (Bottai & Brosch, 2009; Forrellad *et al*., 2013). ESX-5 also plays an essential role in the propagation of genes encoding for ESAT-6/CFP-10 and PE/PPE (Gey van Pittius *et al*., 2006), and is primarily located in slow-growing *Mycobacterium* species (Bottai &

Brosch, 2009), functioning as a component for regulating cell growth (Abdallah *et al*., 2007). A further characteristic of ESX-5 is its susceptibility to duplication, which was indicated by the presence of copies of PE and PPE genes located within this cluster (Gey van Pittius *et al*., 2006).

The ESX clusters are incapable of complementing the functionality of each other (Abdallah *et al*., 2007), which was illustrated by the inability of any ESX gene cluster to compensate for the loss in virulence when ESX-1 was removed (Das *et al*., 2011). However, although the five ESX gene clusters show separate functionalities, there does seem to be a certain degree of interaction between the proteins encoded for by the different ESX gene clusters. This was confirmed by an elevated secretion of PPE41 encoded for by ESX-5, following ESX-1 deletion in *M. tuberculosis* (Gey van Pittius *et al*., 2001; Simeone *et al*., 2009). PPE41 is a protein associated with mycobacterial virulence and has a distinct advantage in that it lacks a Sec or Tat signal peptide sequence, hence, for the successful secretion thereof, no prior signalling is required (Abdallah *et al*., 2007; Feltcher *et al*., 2010). This implies that the ESX-1 and ESX-5 gene clusters have independent roles in the steps following infection of the host cells (Forrellad *et al*., 2013).

Finally, it was determined that six genes are shared between the five different ESX gene clusters, responsible for encoding the necessary components of mycobacterial T7S systems, as mentioned earlier. Moreover, the five ESX gene clusters also contain genes which are unique to each individual gene cluster (Abdallah *et al*., 2007).

2.6 The PE and PPE proteins

As previously stated, a common characteristic of these two unrelated proteins is that they contain Pro-Glu and Pro-Pro-Glu motifs in their N-terminals (Voskuil *et al*., 2004; Forrellad *et al*., 2013). These proteins are unique to mycobacteria, with 10% of the total coding capacity of the *M. tuberculosis* H37Rv genome devoted to encoding these entities (Cole *et al*., 1998), with sets of genes consequently termed *pe* and *ppe* (Sampson, 2011; Mukhopadhyay and Balaji, 2011). Although their precise function and subcellular location are unknown (Cole *et al*., 1998; Abdallah *et al*., 2007),

scientists believe these proteins contribute to lipolysis and lipogenesis of host macrophages (Cole *et al*., 1998). Furthermore, once secreted, these proteins are believed to induce other strong immune responses in the host (Forrellad *et al*., 2013). Only a minority of PE/PPE proteins have been shown to be relevant for the virulence of *M. tuberculosis* and these are largely encoded for by the ESX-1 and ESX-5 gene clusters (Gey van Pittius *et al*., 2006; Forrellad *et al*., 2013).

The PE protein family consists of 100 members, with relatively conserved N-terminals, of approximately 110 amino acids in length, and can be divided into three subgroups (Sampson, 2011). The first group consists of only 29 members and contains only the PE domain (Cole *et al*., 1998). The second group contains 8 members and consists of the PE domain followed by a unique sequence – it therefore demonstrates little homology with other members of the PE family (Voskuil *et al*., 2004; Mukhopadhyay and Balaji, 2011). The third and largest group is represented by 67 PE family members consisting of the PE domain, which contains Gly-Gly-Ala or Gly-Gly-Asn tandem repeats and consequently are termed the polymorphic GC-rich repetitive sequences (PE_PGRS) (Voskuil *et al*., 2004). Interestingly, PE_PGRS sequences illustrate considerable variety among different mycobacterial species (Mukhopadhyay and Balaji, 2011). This, together with the fact that the PE_PGRS sequences are key contributors to the polymorphisms in the *M. tuberculosis* complex, render these sequences essential for antigenic variation (Voskuil *et al*., 2004). Furthermore, antibodies against PE_PGRS33 were detected in mice infected with *M. tuberculosis*. Even though this PE protein was unable to elicit an antibody response, its secretion led to the activation of host interferon-γ (Delogu and Brennan, 2001). Various other PE_PGRS proteins demonstrated the same antigenic potential in other model organisms, confirming this finding (Mukhopadhyay and Balaji, 2011).

The PPE protein family, on the other hand, consists of 68 members and exhibits a conserved N-terminal domain that comprises 180 amino acids (Sampson, 2011). These amino acids possess C-terminal segments varying considerably in length and sequence (Cole *et al*., 1998). The PPE proteins are divided into three sub-families. The majority of these proteins (24 in all) belong to the PPE-SVP group and are characterised by a Gly-X-X-Ser-Val-Pro-X-X-Trp motif between amino acids 300 and

350 (Gey van Pittius *et al*., 2006). The second group, the major polymorphic tandem repeats (MPTR) (Cole *et al*., 1998), are characteristically comprised of multiple tandem copies of the Asn-X-Gly-X-Gly-Asn-X-Gly motif (Voskuil *et al*., 2004). The third group consists of unrelated proteins and the only similarity with the others is the presence of a common 180-residue PPE domain (Cole *et al*., 1998). More recent studies describe a fourth group that is distinct from the others in that it exhibits poor homology at the C-terminus (Gey van Pittius *et al*., 2006).

Although the function of these proteins is still unknown (Cole *et al*., 1998), they are deemed important as they make up 4% of the total proteins encoded for by pathogenic mycobacteria (Gey van Pittius *et al*., 2006). As previously mentioned, ESX-5 is directly dependent on the secretion of PPE41 (Abdallah *et al*., 2006; Abdallah *et al*., 2007), suggesting its role in ESX-5 virulence (Gey van Pittius *et al*., 2001). Evidence complementing this is the study done by Tundup *et al*. (2008), indicating secretion of PPE41 by pathogenic mycobacteria results in the up-regulation of $CD8^+$ and $CD4^+$ T cells, and elevated levels of TNF-α, IFN-γ and IL-2 cytokines (Mukhopadhyay and Balaji, 2011). Other studies indicated the involvement of another member of the PPE family, PPE46, in disease pathogenesis, with a transposon mutation resulting in the attenuation of growth in macrophages (Camacho *et al*., 1999; Gey van Pittius *et al*., 2006). Furthermore, Rv3872 (PE35) and Rv3873 (PPE 68), situated in the ESX-1 gene cluster region, were also found to play a role in the *in vivo* growth of *M. tuberculosis* (Gey van Pittius *et al*., 2006).

It is believed that there is a strong link between the PE/PPE proteins and the T7S system, as the secretion of multiple PE/PPE protein entities are ESX mediated (Tekaia *et al*., 1999; Gey van Pittius *et al*., 2006; Sampson, 2011). Considering this, a better understanding of the functionality of these proteins remains a priority today, as together with the T7S system they provide an attractive target for anti-TB drug development.

2.7 The *Mycobacterium* T7S system

ESX-1 and its secretion machinery secrete various proteins associated with mycobacterial virulence in pathogenic species (Abdallah *et al*., 2007; Das *et al*.,

2011). Although proteins secreted via non-pathogenic mycobacterial species exhibit extensive homology to those secreted by their pathogenic counterparts, the former fulfil organism-specific functions (Converse and Cox, 2005). In view of the strong association of the T7S system with virulence, a more detailed literature discussion will be given in the light of the corresponding functionality.

In contrast to other mycobacterial secretion systems, the T7S system secretes various proteins lacking any classical signal peptides (Das *et al*., 2011). Furthermore, every ESX locus contains genes coding for ESX machinery components, which are essential for the secretion of any protein. These components are termed EccABCDE and MycP1, without which ESAT-6/CFP-10 cannot be secreted (Feltcher *et al*., 2010). EccA1 and EccCb1 were identified as cytoplasmic AAA-ATPases (Pym *et al*., 2002) and therefore are believed to supply energy for the essential secretion process. Interestingly, a pairwise dependency, both in secretion and function (Fortune *et al*., 2005), is characteristic of most of the proteins secreted via the T7S system (Das *et al*., 2011). CFP-10 and ESAT-6, the first characterised substrates of the ESX-1 gene cluster associated with the T7S system, are secreted as a hetero-dimer complex forming a tight four-helix bundle (Das *et al*., 2011), illustrating strong mutual dependence (McLaughlin *et al*., 2007). Strong hydrophobic interactions between these two protein entities hold the hetero-dimer complex together (Abdallah *et al*., 2007), whereas the carboxy terminal end of CFP-10 does not participate in dimer formation (Fortune *et al*., 2005; Simeone *et al*., 2009). Subsequently, as a result of this interaction by the characteristic signal sequence exhibited in the C-terminus of CFP-10 with the FsK domain of $EccCb_1$ (Rv3871) (Abdallah *et al*., 2007), secretion initiation of the majority of proteins produced by this secretion system follows (Feltcher *et al*., 2010; Das *et al*., 2011). In addition to CFP-10/ESAT-6, ESX-1 secretion-associated proteins (Esp): EspA (Rv3616c) and EspC (Rv3615c), are encoded for by genes located outside the ESX-1 locus (Ohol *et al*., 2010), and exhibit a dependence on CFP-10/ESAT-6 for secretion and vice versa (mutual dependence of these four protein entities) (Das *et al*., 2011). It is believed that EspA and EspC also form a hetero-dimer complex (EspA/EspC) (Das *et al*., 2011; Ohol *et al*., 2010), so that a mutation in either CFP-10/ESAT-6 or EspA/EspC renders *M. tuberculosis* unable to secrete any of these entities (DiGiuseppe Champion & Cox, 2007). After interaction with CFP-10, $EccCb_1$ interacts with $EccCa_1$

(Rv3870), allowing for the recruitment of the CFP-10/ESAT-6 complex to the inner membrane pore $EccD_1$ (Rv3877) (Abdallah *et al*., 2007; Ohol *et al*., 2010). At this stage, CFP-10/ESAT-6 and EspA/EspC are translocated from within the intracellular environment, through the inner membrane of the mycobacterial cell envelope (MacGurn *et al*., 2005; McLaughlin *et al*., 2007; Das *et al*., 2011). Structural studies of ESAT-6 and CFP-10 revealed five additional components to be present in the ESX-1 gene cluster (Pym *et al*., 2003; Guinn *et al*., 2004). The mutual dependence of these entities for the expression and eventual secretion of the CFP-10/ESAT-6 complex has been established (Das *et al*., 2011). These additional components include: Rv3866, Rv3868, Rv3869, PE35 (Rv3872) and MycPI (Rv3866), although their exact functionality in this process is not well understood (Das *et al*., 2011). Little information exists on those components associated with the T7S system, responsible for translocating these protein entities across the mycomembrane (Ganguly *et al*., 2008; Das *et al*., 2011). Additionally, Rv3596c (AAA-ATPase chaperone), Rv0686 and its substrate Rv2151c are thought to contribute toward the secretion of various protein entities via the T7S system (Pym *et al*., 2002; Ganguly *et al*., 2008), although their exact functionality remains an enigma (DiGiuseppe Champion & Cox, 2007).

After the successful secretion of CFP-10/ESAT-6, these two entities disconnect and fulfil separate functions (Das *et al*., 2011). CFP-10 functions by de-phosphorylating macrophage proteins, and in so doing down-regulates their activities. This ultimately leads to a decrease in H_2O_2 and reactive oxygen species (ROS) and thereby a decreased capacity of the host to eliminate the threat. Additionally, cell-bound CFP-10 further compromises host defences, by accelerating the breakdown of the macrophage H_2O_2 (Ganguly *et al*., 2008). ESAT-6 has additionally been shown to bind to the surface of the macrophage through TLR-2, and in doing so inhibits the activation transcription factors, namely, interferon regulatory factors (IRFs) and NF-κB, thus preventing cell replication (Ganguly *et al*., 2008). The ESX-1 secretion system has also been shown to repress the fusion of phagosomes with lysosomes, and in doing so to aid in mycobacterial survival within host macrophages (Ganguly *et al*., 2008).

2.8 *Mycobacterium smegmatis* as a research model

Mycobacterium smegmatis, an avirulent saprophyte (Simeone *et al*., 2009), is a fast-growing mycobacterial species. It contains three fully functional ESX-gene clusters, namely, ESX-1, ESX-3 and ESX-4 whilst ESX-2 and ESX-5 are absent from its genome (Gey van Pittius *et al*., 2006). Comparative genomic studies revealed this avirulent strain to share 60–75% sequence homology with its more virulent relative, *M. tuberculosis* H37Rv (Gey van Pittius *et al*., 2001). Other studies complementing these findings demonstrated that ESX-3 is not an essential component required for viability of *M. smegmatis* (Abdallah *et al*., 2007), making this organism an ideal model for exploring ESX-3 gene cluster function. Because this strain does not possess an ESX-5 gene cluster, and therefore is unable to secrete PE25 and PPE41 unlike *M. bovis* and *M. marinum*, studies have shown that *M. smegmatis* nevertheless exhibits a cell wall similar to that of *M. tuberculosis* (Converse and Cox, 2005). Thus, ESX-5-secreted proteins, or other substrates formed in this gene cluster, do not play a role in the biosynthesis of the mycobacterial cell wall (Abdallah *et al*., 2006). Reconstruction of the entire ESX-5 gene cluster in this avirulent strain enabled *M. smegmatis* to partially secrete the homologous PPE41 (Abdallah *et al*., 2006), although it failed to secrete this protein efficiently (Abdallah *et al*., 2007). Two genes encoding for PE and PPE proteins, in ESX-1 and ESX-3, have been described in the genome of *M. smegmatis* (Abdallah *et al*., 2007). They are believed to have an ancestral origin, as PE and PPE proteins are present exclusively in *Mycobacterium* (Gey van Pittius *et al*., 2006).

Mycobacterium smegmatis contains 5 of the 11 *esx* loci present in *M. tuberculosis*. Present in these are the *esxA*/*esxB* gene pair as well as the surrounding genes complementing the T7S system (Converse and Cox, 2005; Simeone *et al*., 2009). It is important to note that *M. smegmatis* possesses a functional ESX-1 gene cluster and T7S system. Furthermore, although the proteins secreted via this system do not cause organismal virulence (most probably due to a lack of complementary mechanisms such as ESX-5 and their associated proteins), they contribute to an organism-specific function (MacGurn *et al*., 2005; McLaughlin *et al*., 2007), with CFP-10 and ESAT-6, both present, thought to modulate conjugal DNA transfer (Converse and Cox, 2005). Using *M. smegmatis* as a model to research mycobacteria has a number of advantages, the most attractive of these being its fast growth rates. *Mycobacteria smegmatis* has a doubling time of 3 hours (Kana *et al*.,

2004) and subsequently is ideal for growing knock-out construct strains. Considering that *M. tuberculosis* has a doubling time of 18–24 hours, *M. smegmatis* has proved to be a valuable substitute for studying the tubercle bacilli. Moreover, *M. smegmatis* is avirulent; hence *Mycobacterium* gene functionality can be investigated with little or no risk of infection. *Mycobacterium smegmatis* is therefore an attractive option for initial investigations of ESX-1 gene functionality.

2.9 Metabolomics

Metabolomics has been defined as "the systematic study of the unique chemical fingerprints that specific processes leave behind" (Nanda *et al*., 2011). A more accurate description is: "the unbiased identification and quantification of all metabolites present in a biological sample, by using highly sensitive and selective analytical techniques" (Dunn *et al*., 2005). These analyses are followed by data matrix interpretation and visualisation, via biostatistics (Schoeman *et al*., 2012). This research approach aims to analyse all of the metabolites present in a biological sample (metabolome), for the purpose of identifying biomolecules important to answering a specific biological question (van der Werf *et al*., 2007). Within a few years, metabolomics has become a popular research tool, for the detection of new biomarkers associated with various disease states (Weckwerth & Morgenthal, 2005). Metabolomics fills the gap in the current "omics" revolution (Nanda *et al*., 2011); complementing proteomics and genomics, all of which allow for a holistic or systems biology approach when these research techniques are used in combination (Aggerwal & Lee, 2003).

One of the limitations of metabolomics is that the human fluxome, together with a number of metabolic pathways are not well characterised (Nanda *et al*., 2011). Various studies have reported it to be difficult to detect all the metabolites present in a cell with optimum precision, resolution and sensitivity, when using only one analytical method (van der Werf *et al*., 2007). Thus, when using a metabolomics research approach, it is of the utmost importance to use several different analytical methods (van der Werf *et al*., 2007). However, when using only one analytical approach, it is important to select the method most suited for detecting and analysing as many metabolites as possible, related to the particular biological question and the

nature of the sample collected (Nanda *et al*., 2011). Considering this, metabolomics involves the application of a variety of analytical techniques, such as gas chromatography–mass spectrometry (GC-MS), liquid chromatography–mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR). Each technique has the potential of detecting hundreds of individual chemical structures and is carefully chosen, depending on the type of study, as every technique has its own advantages and disadvantages (Weckwerth & Morgenthal, 2005).

One of the main advantages of two-dimensional gas chromatography–time-of-flight mass spectrometry (GCxGC-TOFMS) over conventional GC-MS is 1) its increased dimensionality, hence its capacity to analyse highly complex samples (Welthagen *et al*., 2004; Weckwerth & Morgenthal, 2005); and 2) its sensitivity, which allows for one to detect and identify far more compounds in a relatively small sample (Weckwerth & Morgenthal, 2005). An advantage of this technique is its high scan rates, which provide accurate peak deconvolution of complex samples, so that, combined with mass spectral deconvolution, overlapping peaks can still be separated and identified (Welthagen *et al*., 2004).

CHAPTER 3:

AIMS AND OBJECTIVES

3.1 Problem statement

The functionality of the ESX-1 gene cluster and that of the associated T7S system are still poorly understood. Considering its proposed functions in mycobacteria, including its potential role in inducing virulence, a more complete characterisation of ESX-1 may not only lead to a better understanding of the *M. smegmatis* organism, but may ultimately lead to improved vaccination and treatment approaches. In this study we used a metabolomics approach to identify metabolite markers associated with ESX-1 expression, by comparing an *M. smegmatis* wild-type parent strain with an *M. smegmatis* ESX-1 knockout organism. Using a systems biology approach, we additionally interpret the results in light of the existing pool of proteomic and genomic information on ESX-1 in *M. smegmatis*. Although *M. smegmatis* is an avirulent strain of *Mycobacterium*, it codes the ESX-1 gene cluster and has a fully functional T7S system responsible for the secretion of the same homologue proteins (ESAT-6 and CFP-10) as *M. tuberculosis*. It has subsequently been used and described extensively in various genomics and proteomics investigations in its applications to understand mycobacteria better, including the various genes associated with the T7S system. To date, however, no metabolomics study on this organism has been conducted.

3.2. Aims

1. To validate the use of the GCxGC-TOFMS procedure and the corresponding extraction method, to generate data for the application of metabolomics to the current investigation. This was accomplished by determining the analytical repeatability of the data so generated.
2. Using the metabolomics method validated above, identify characteristic metabolite markers from GCxGC-TOFMS data, which best characterise the ESX-1 gene cluster function (using *M. smegmatis* as a model), from a metabolism perspective.

3.3 Objectives

- Related to Aim 1: Determine the repeatability of the metabolomics method used, by calculating the coefficient of variation (CV) for all compounds using data generated by either 1) a repeatedly injected quality control sample (validating the GCxGC-TOFMS procedure), or 2) from single injections of a repeatedly extracted quality control sample (validating the extraction method), with reference to the prescribed CV cut-off values.
- Related to Aim 2: Using the GCxGC-TOFMS data generated from the extracted experimental sample groups, characteristic metabolite markers were identified using both multivariate [principal component analysis (PCA)] and partial least squares–discriminate analysis (PLS-DA) and univariate (based on unpaired *t*-test and effect sizes) statistical methods. These metabolite markers were interpreted in the light of existing proteomic and genomic data. In doing so, we hoped to characterise ESX-1 more fully and the associated secretion system from a metabolomics perspective, in order to gain a holistic understanding of its functionality.

CHAPTER 4:

MATERIALS AND METHODS

4.1. Experimental design

The general outline of the experimental protocol followed in this study is summarised in Figure 4.1. This investigation had two experimental categories, where 1) the repeatability of both the analytical instrument (GCxGC-TOFMS) and the extraction procedure/the analyst's capacity to perform this in a repeatable manner were evaluated (illustrated on the left-hand side of the flow diagram), followed by 2) the application of the method to answer the biological question by generating metabolomic data using *M. smegmatis* as a model (illustrated on the right-hand side of the flow diagram), and then to interpret these results using a systems biology approach.

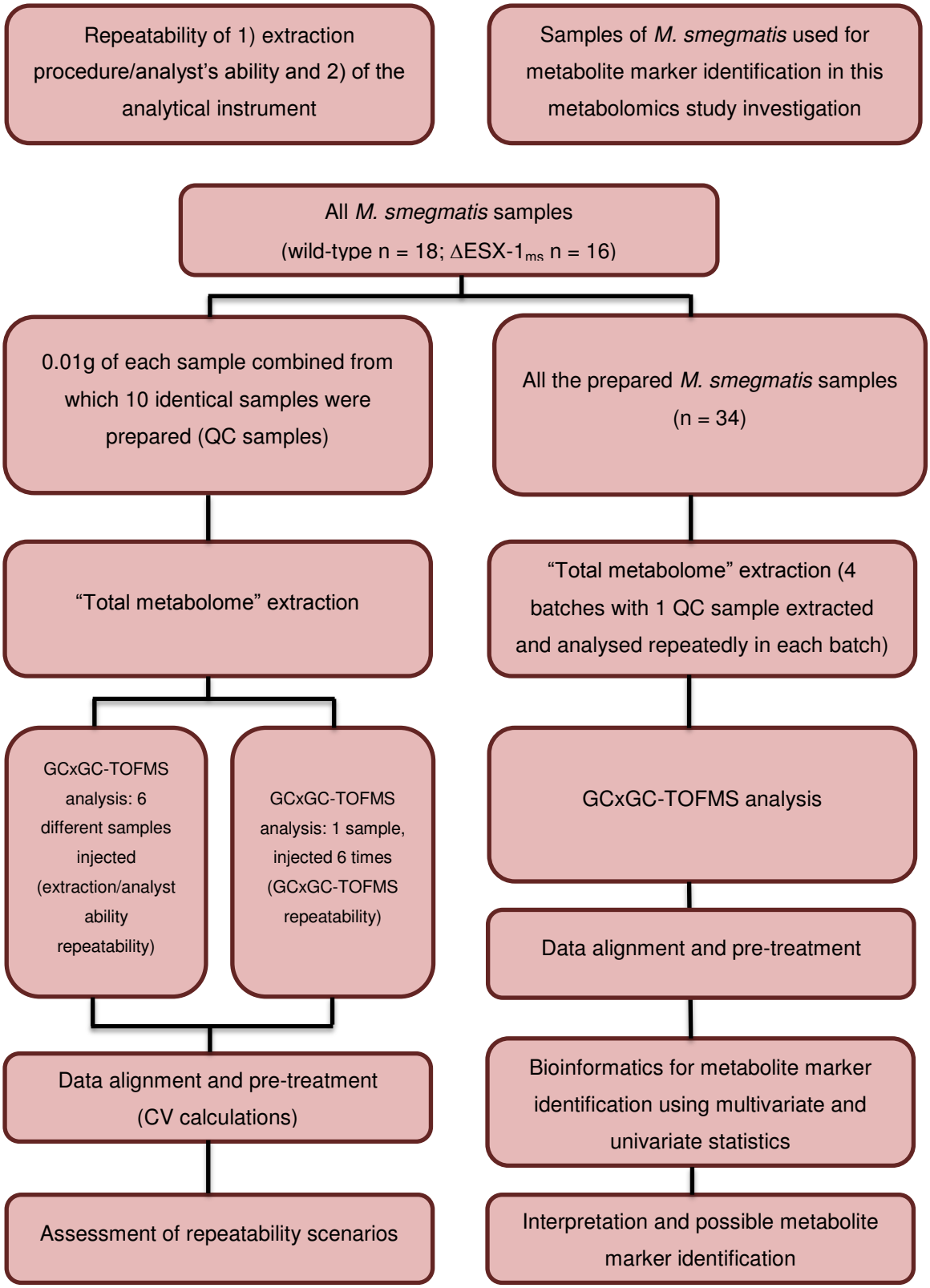


Figure 4.1: Schematic representation of the experimental design followed in this metabolomics investigation.

4.2 Reagents and chemicals

Ampicillin and X-gal were purchased from Roche (Indianapolis, IN, USA). The BBL Seven H11 Agar Base used as solid culture medium was purchased from Becton Dickinson & Company (Franklin Lakes, NJ, USA). All the additional supplementation for cultures was obtained from Roche (bovine serum albumin), Kimix (glucose), Sigma (kanamycin; Tween-80 and sucrose) or Merck (glycerol). The derivatisation reagents – methoxyamine hydrochloride, N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) and pyridine, as well as the internal standard (3-phenylbutyric acid) – were purchased from Sigma Aldrich (Darmstadt, Germany). The organic solvents used for the study (chloroform and methanol) were ultra-pure Burdick and Jackson brands (Honeywell International Inc., Muskegon, MI, USA).

4.3 *M. smegmatis* samples

4.3.1 Bacterial strains and culture conditions

The *M. smegmatis* cultures were prepared by the DST/NRF Centre of Excellence for Biomedical TB Research (CBTBR), Stellenbosch University, South Africa. Ethical approval for the use of these cultures was obtained from the University of Stellenbosch's and North-West University's respective Ethics Committees with corresponding numbers N10/04/126 and NWU-00127-11-A1.

Escherichia coli JM109 was used in the cloning procedures. *E. coli* was cultured overnight at 37°C in Luria-Bertani (LB) broth with shaking, and on LB agar plates. Solid and liquid media were supplemented with the antibiotics ampicillin (50 µg/mL) and kanamycin (50 µg/mL) and solid media with X-gal and 5% sucrose, as required. *M. smegmatis* mc2155 was used to construct the ΔESX-1_{ms} strain. For knock-out construction, *M. smegmatis* was grown in liquid LB medium with shaking, as well as on LB agar plates for 4–5 days at 37°C. LB was used as growth medium as Middlebrook 7H9 and 7H11 media contain sugars which are toxic to single crossover knock-out cultures containing a *sacB* marker gene. Media were supplemented with

Tween-80 (0.1%), kanamycin (25 µg/mL), X-gal and 5% sucrose, as required. For sample preparation for metabolite marker identification, *M. smegmatis* was grown on BBL Seven H11 Agar Base (Becton Dickinson & Company) supplemented with 0.2% glycerol, 0.5% bovine serum albumin and 0.2% glucose and incubated at 37ºC for 3 days.

4.3.2 Preparation of the *M. smegmatis* ESX-1 knock-out strain

The *M. smegmatis* ESX-1 knock-out strain (ΔESX-1_{ms}) was created by homologous recombination as described in Gordhan and Parish (2001). Briefly, 824 bp and 845 bp pieces of DNA, located directly upstream and downstream of the ESX-1 gene cluster, respectively, were amplified from *M. smegmatis* mc^2155 genomic DNA, using the following polymerase chain reaction (PCR) primers containing specific restriction sites: UpReg1smeg f (5'-GGTACCAGTCGAACCTGGGCAAGCG-3', *Kpn*I) and UpReg1smeg r (5'-AGATCTTCGGCGTATTCGCAGCACC-3', *Bg*/II) and DownReg1smeg f (5'-AGATCTCCCCAACAAGAATTGATCCG-3', *Bg*/II) and DownReg1smeg r (5'-AAGCTTTTGTTCAGCGAGCCCGAG-3', *Hind*/III), respectively. UpReg1smeg and DownReg1smeg were cloned into the appropriate restriction sites of the p2NIL suicide vector (Parish and Stoker, 2000), via three-way cloning, to create p2NIL_UpDownReg1Smeg. The *lacZ-sacB* marker gene cassette was digested out of pGOAL17 (Parish and Stoker, 2000), and inserted into the *Pac*I site of p2NIL_UpDownReg1Smeg to create the suicide vector p2NIL_ESX-1 KO.

p2NIL_ESX-1 KO was transformed into *M. smegmatis* mc^2155 and single recombinants were selected on kanamycin and X-gal, and passaged without selection to allow a second recombination event to occur. The resultant double recombinants were counter-selected on sucrose in the absence of kanamycin. PCR was used to distinguish between wild-type and ΔESX-1_{ms} using three combinations of primers. R1 F1 f (5'-GAGGTGCTCGACAGCGAGGC-3') and R1 F1 r (5'-TCGCCAGCGGAATCAGTGCC-3'), and R1 F2 f (5'-GAGGTGCTCGACAGCGAGGC-3') and R1 F2 r: (5'-GGTGTACCCGTATCCCTTTCCA-3') generate products in the wild-type strain, while R1 F1 f and R1 F2 r form a product in ΔESX-1_{ms}. The ΔESX-1_{ms} was confirmed by sequencing across the deleted region.

4.3.3 *S*ample preparation for metabolomic analyses

Separately cultured *M. smegmatis* wild-type (n = 18) and ΔESX-1_{ms} (n = 16) were prepared on BBL Seven H11 Agar Base solid culture medium, supplemented with 0.5% m/v bovine serum albumin, 0.2% m/v glucose and 0.2% v/v glycerol, in the absence of Tween-80, oleic acid and catalase (as described above). After 3 days of growth at 37°C, the cells were scraped off the plates and immediately snap frozen in liquid nitrogen. The samples were lyophilised and stored at –80°C prior to sample extraction for GCxGC-TOFMS metabolomic analysis.

4.3.4 Quality control samples

Quality control (QC) samples, extracted and analysed in conjunction with the experimental samples, ensured precise and repeatable results (Dunn *et al*., 2005; Fiehn *et al*., 2008) and enabled the analyst to evaluate the specific analytical technique used during experimentation, in addition to making the necessary correction should a batch effect be detected. Using the cultured *M. smegmatis* wild-type and ΔESX-1_{ms} samples, a small portion (0.01 g) of lyophilised cells from each of the separately prepared cultured groups were transferred to a microcentrifuge tube and homogenised using a carbide tungsten bead in a vibration mill for 3 minutes at 30 Hz. From this master mix, 10 samples were prepared by weighing off equal amounts of 0.01 g into 10 separate microcentrifuge tubes and subsequently used as QC samples for data correction and method validation. One QC sample was extracted, derivatised and analysed (as described in section 4.4) by repeatedly injecting these at regular intervals throughout each experimental sample batch. The data collected from all QC samples are illustrated in Chapter 5.

4.3.5 Repeatability and reliability

The remaining QC samples were used to evaluate repeatability of the analytical instrument (GCxGC-TOFMS) as well as the extraction procedure/the analyst's capacity to perform this in a repeatable manner. A single sample chosen at random was extracted and analysed repeatedly (n = 6) by GCxGC-TOFMS. The same was

done to evaluate the repeatability of the extraction procedure/the analyst's ability, but in this experiment 6 identical QC samples were extracted 6 times followed by GCxGC-TOF-MS analysis (as indicated in Figure 4.1).

4.4 Methods

4.4.1 Sample extraction and derivatisation method

A modified "total metabolome" extraction approach, as described in Meissner-Roloff *et al*. (2012), was used for the extraction of all samples. To the 10 mg of each sample weighed into a microcentrifuge tube (QC, wild-type and ΔESX-1_{ms}), 50 µL phenylbutyric acid (0.525 mg/mL H_2O), 25 µL tricarballylic acid (0.394 mg/mL ddH_2O) and 100 µL nonadecanoic acid (0.5 mg/mL chloroform) were added. The addition of a 1.25 mL mixture containing chloroform, methanol and water followed in a ratio of 2.5:7.5:2.5. Each sample was briefly vortexed (Retsch, Haan, Germany), followed by the addition of a carbide tungsten bead (Retsch), and then placed in a vibration mill for 10 minutes at 30 Hz. The samples were subsequently centrifuged for 5 minutes at 12 000 rpm, after which the supernatant were transferred to a GC vial and dried under a light stream of nitrogen gas. The dried, extracted samples were subsequently derivatised using 50 µL of methoxyamine hydrochloride (containing 15 mg/mL pyridine) at 50ºC for 90 minutes, after which 80 µL MSTFA with 1% TMCS was added, and the samples were further derivatised at 50ºC for 60 minutes (Kanani *et al*., 2007). Finally, the samples were transferred to a 0.1 mL insert and placed in a GC vial prior to GCxGC-TOFMS analysis.

4.4.2 GCxGC-TOFMS

4.4.2.1 GCxGC-TOFMS analysis

A Leco Pegasus 4D GCxGC-TOFMS instrument was used to analyse the prepared samples. GCxGC-TOFMS relies on the principle of two different separation mechanisms (volatility and polarity), to ensure optimum separation of components (Dallüge *et al*., 2002). Two-dimensional gas chromatography provides a greater

separation capability than that of a one-dimensional system (Dallüge *et al*., 2002; Kalinova *et al*., 2006). A thermal modulator situated between the primary and secondary columns ensures optimal compound separation by focusing the primary column eluent before exposing the compounds to the secondary column (Dallüge *et al*., 2002; Welthagen *et al*., 2004). With time-of-flight mass spectrometer (TOFMS) detection completing the technique, GCxGC-TOFMS is considered to be a powerful tool for total metabolome analysis. Furthermore, TOFMS also allows for a high acquisition rate (500 spectra per second) and comparatively high system accuracy due to the absence of spectral shifting in a continuous acquisition mode (Kalinova *et al*., 2006).

One microlitre of the prepared samples was injected in a splitless mode on an Agilent 7890A two-dimensional gas chromatograph (GCxGC) (Agilent, Atlanta, GA, USA), coupled to a time-of-flight mass spectrometer (TOFMS) (LECO Africa (Pty) Ltd.) equipped with a Gerstel Multi-Purpose Sampler (MPS) (Gerstel GmbH, Mülheim an der Ruhr, Germany). A 30-m Rxi-5Sil MS primary column with a film thickness of 0.25 µm and an internal diameter of 250 µm (Restch GmbH, Haan, Germany) enabled first-dimensional separation, and a 1.0 m Rxi-17 capillary column with a film thickness of 0.1 µm and internal diameter of 100 µm was fitted as the secondary column. The cryomodulator concentrated the effluent of the primary column before subjecting the compounds to the secondary column with a hot pulse of nitrogen gas for 0.7 seconds at 3-second intervals. The front inlet temperature was held at a constant 280°C for the entire run. The GC ramp for the primary oven was initially held at 70°C for 2 minutes, followed by a rise in temperature of 4ºC per minute until the oven reached a final temperature of 300°C, where it was held for 2 minutes. The secondary column's oven temperature was initially held at 75°C for 2 minutes, after which the temperature was raised by 4°C per minute until a final temperature of 305°C was achieved and then maintained for 2 minutes.

Every sample run started with a 560-second solvent delay in which the filament was switched off and consequently no mass spectrum was recorded during this period. These 560 seconds, however, were included in the primary column run time to ensure accurate retention times of the compounds of interest. The transfer line temperature was held at a constant 275°C, with the ion source temperature at a

constant 230°C. The detector voltage was adjusted to 1800 V with the filament bias at –70 eV. Spectra were collected over the range 50–600 m/z at an acquisition rate of 200 spectra per minute.

4.4.2.2 Batch composition and run order

All the samples were analysed by GCxGC-TOFMS in an untargeted manner. The samples utilised for metabolomic data acquisition, excluding those used for determining repeatability, were randomly grouped into four batches of ten samples and extracted accordingly. One QC sample per batch was similarly extracted and injected seven times throughout each batch in a sequence indicated below with "Q" denoting the QC sample and "S" the experimental sample (either *M. smegmatis* wild-type or ΔESX-1_{ms}).

$$[QQQQQ]S_1S_2S_3S_4S_5[Q]S_6S_7S_8S_9S_{10}[Q]$$

Before each batch analysis, leak and tune checks were done to ensure that every batch was subjected to the same analytical conditions and to minimise machine bias.

4.5 GCxGC-TOFMS data processing

4.5.1 Deconvolution and peak identification

Deconvolution and peak identification were carried out using Leco ChromaTOF software version 4.44 (Leco, St. Joseph, MI, USA). ChromaTOF is primarily used for peak finding, mass spectral deconvolution (spectral determination) and peak identification. Furthermore, the software provides in-line acquisition and analysis of chromatograms collected by the GCxGC-TOFMS instrument and differs from other software packages in that it is able to acquire mass spectra of high frequency (200 spectra per second), resulting in large file sizes (Luedemann *et al*., 2008). The process of identifying individual compounds in a complex co-elution is termed peak finding. Mass spectral deconvolution is a mathematical process used to separate co-eluting peaks after the sample was subjected to chromatography (Kusano *et al*., 2007). Deconvolution therefore enables the analyst to identify co-eluting compounds

with their own assigned mass spectra as well as respective retention times. This is a great advantage when studying the metabolome of *M. smegmatis* (Luedemann *et al*., 2008). All compounds were identified and named by comparing the mass fragmentation patterns and retention times generated by the TOFMS with those collected from commercial and in-house libraries constructed from previously injected standards. The mass spectra generated from the samples injected for this study were deconvoluted at a signal to noise (S/N) ratio of 200.

4.5.2 Peak alignment

With retention-time shifts complicating peak identification, an alignment program compensated for any difference and successfully aligned peaks with identical mass spectra and corresponding retention times across all samples (Schoeman *et al*., 2012). Thus statistical comparison, an optional function on the ChromaTOF software was used for peak alignment prior to data processing. Peak alignment uses the same variables as peak identification, namely, retention times and mass spectra to align peaks across all samples.

4.6 Statistical data analyses

4.6.1 Data pretreatment

Prior to any statistical data analyses, the peak areas were normalised relative to the internal standard (by calculating relative concentrations of all compounds using internal standard phenylbutyric acid). A zero value replacement and 50% filter followed. All metabolites with a concentration value of zero were replaced by a value calculated as half of the minimum value (i.e. the detection limit) detected for each batch, and all compounds not detected in at least 50% in any of the two experimental groups were excluded from the entire data set analysed. This was followed by a quality control–coefficient of variation (QC-CV) filter. The coefficient of variation (CV) is a parameter in terms of which repeatability is measured and can be defined as $CV = s/\tilde{x} \times 100$, where $\tilde{x}$ is the sample mean and *s* is the sample standard deviation. All CV calculations were done using the “R” (version 2.13.0) statistical

package. When the concentrations of all compounds across all the samples are known, determination of CV values serves as a powerful tool to quantify analytical measurement error. In the study conducted by Dunn *et al*. (2011), it is reported that the US Food and Drug Administration (FDA) allows for a 20% CV when the aim of the study is the identification of biomarkers. Furthermore, the FDA stressed that researchers must aim for the best achievable repeatability as it improves validity when identifying biomarkers. For our metabolomics study, however, as described by Schoeman *et al. (*2012) a CV value of 50% was chosen as the cut-off (Schoeman *et al*. 2012), and those metabolites not satisfying this condition in the QC samples were removed from the entire data matrix (wild-type and ΔESX-1_{ms}). This QC-CV filter allows for a "cleaner" data matrix, by removing those compounds which inherently demonstrate large analytical variation. Data clean-up simplifies metabolite marker identification in the subsequent multivariate and univariate statistical analyses.

When analysing samples in individual batches, the ideal is to subject each sample and batch of samples to the same analytical conditions. Despite all precautions, a batch drift may still occur, especially when analysing large sample batches over several days. As a result of this, the data are influenced in such a way that their validity may be questioned. It therefore becomes necessary to correct any drift by exposing the data to a QC correction step, in which QC samples belonging to a particular batch are used to correct for all non-biological systematic differences which may exist among the data generated by the GCxGC-TOFMS. These differences may be attributed to small analytical changes which may occur during the analysis. This data transformation method, also known as quantile equating, corrects for any linear and/or non-linear differences in distribution which may be present among batches of semi-quantitative data, obtained from the same analytical method (Draisma *et al*., 2010).

Both multivariate and univariate statistical approaches were used in this metabolomics study, as illustrated in Figure 4.2.

4.6.2 Multivariate statistical analyses

All statistical data analyses, both multivariate and univariate, were conducted using Statistica (version 10) and MetaboAnalyst. The latter is a web server designed exclusively to facilitate metabolic data processing as well as statistical analysis of data matrices (Xia *et al.*, 2012). Using MetaboAnalyst, each data set, prior to multivariate statistical analysis, was cube root transformed and/or log transformed for the purpose of univariate statistical analysis. All data were also centred using this web server as it converts all metabolite concentrations to fluctuate around zero instead of the mean of each concentration.

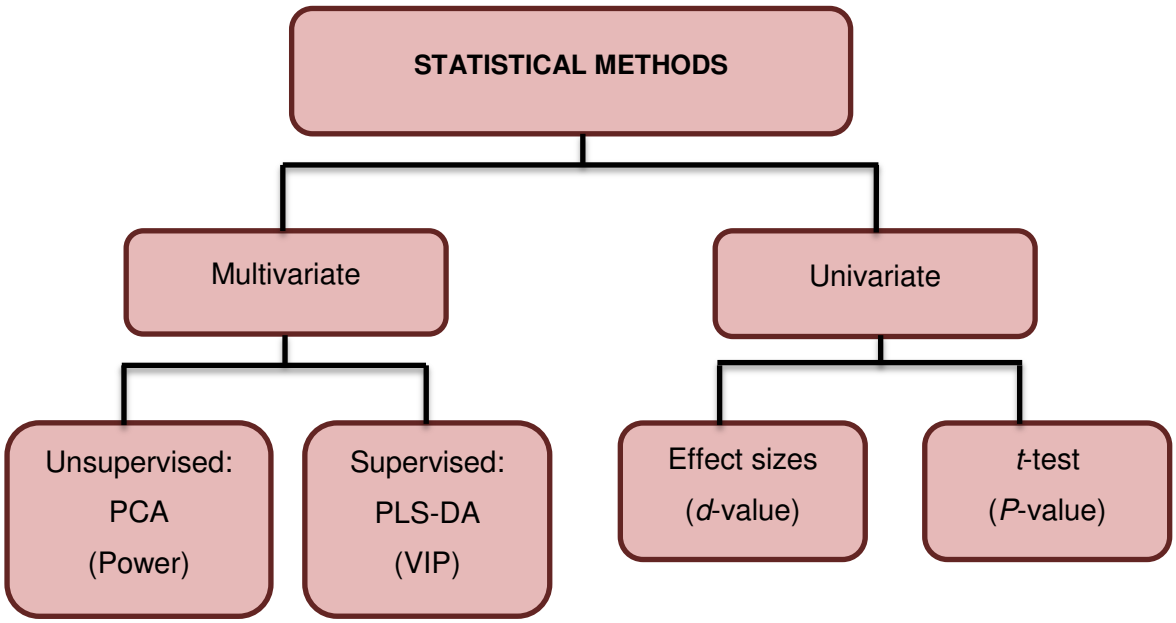


Figure 4.2: Illustration of the statistical methods used in this metabolomic investigation.

4.6.2.1 Principal component analysis

Principal component analysis (PCA) was used to determine if a natural grouping existed between the two groups in question (wild-type and ΔESX-1_{ms}) based on their different metabolite profiles. PCA is an example of an unsupervised statistical data analysis procedure; which attempts to elucidate the relationship between different groups, without any prior knowledge that these groups should differ. Furthermore, PCA transforms a large set of related variables (metabolites) into a set of new,

unrelated variables, smaller in size, termed principal components (PCs). The main focus of these PCs is to express the maximum variation in the data set by characterising each component in a multidimensional space. PC1 represents the biggest variation in the data set followed by PC2, PC3, PC4, etc., which in turn determines the next highest variation. Finally, the modelling power of each variable in the PCA was determined and used to select those variables of importance in the projection (Meissner-Roloff *et al*., 2012; Schoeman *et al*., 2012), by selecting only those which had a modelling power which exceeded 0.5 (Brereton, 2003).

4.6.2.2 Partial least squares–Discriminant analysis

Partial least squares–discriminant analysis (PLS-DA) is an example of a supervised data analysis method, and subsequently used as a means to identify compounds differing the most between two comparable groups. This is done via multivariate regression techniques to extract the information that predicts class or group membership (X), via a linear combination of the original variables (Y). This model is also used to determine variables important in projection (VIPs) and ranks these VIPs in order of discriminative contribution. VIPs are defined as a combined sum of squares of the collected PLS-DA weights, ranking the importance of a specific metabolite for differentiation between two assigned groups (Schoeman *et al*., 2012). VIPs were calculated and used to select those variables/metabolites of importance. Variables with VIP > 1 were identified and subsequently characterised as potential markers (Chong & Jun, 2005) using the PLS-DA statistical analysis model.

4.6.3 Univariate statistical analysis

The relevance of a particular statistical result was also determined using univariate statistical methods – effect sizes in combination with an unpaired *t*-test model. These models are different from the multivariate methods as both examine each variable separately, without taking into account any effect of multiple comparisons. As this is a metabolomics approach and the relationship of the markers to one another is important, univariate methods were not used for marker selection but only to indicate relevance from a univariate perspective.

4.6.3.1 Effect sizes

This univariate statistical analysis model is broadly defined as a qualitative referral to illustrate the significance a change must have to produce a substantial outcome. Effect sizes can alternatively be defined as the smallest change in a data set which can reliably be detected (Goodpaster *et al*., 2010) or alternatively as a measure of practical significance and is used as a univariate analysis to determine the importance of a variable. For parametric data, the effect size (d) can be defined by the formula $d = |X_1 - X_2|/S_{max}$, where X_1 and X_2 signify the group means and S_{max} is the standard deviation of the two groups analysed. Effect sizes of d > 0.5 and > 0.8 signifie medium and large practical significance, respectively (Ellis & Steyn, 2003).

4.6.3.2 *t*-Test

Using unpaired *t*-test calculations between two naturally grouped items (wild-type and ΔESX-1_{ms} samples), it is possible to use one measurement to test if there is a statistically significant difference between the means of metabolites identified in these two subgroups. This univariate statistical analysis model differs from paired *t*-test by comparing one measurement between two naturally differentiated groups, and not a comparison of two measurements in the same population. First, a *t*-value is determined and then a *P*-value can be calculated, which is subsequently used to establish if the differences within metabolite averages is statistically significant (Xia *et al*., 2009). A value of $P < 0.05$ was selected as the level of significance (Ellis & Steyn, 2003).

CHAPTER 5:
RESULTS AND DISCUSSIONS

5.1 Introduction

In the first part of this chapter, I discuss the results of the validation experiment (Aim 1) for 1) the GCxGC-TOFMS and 2) the extraction procedure used. Later in the chapter, I describe the results obtained from the metabolomics comparison of *M. smegmatis* wild-type and the ΔESX-1_{ms} using the validated approach (Aim 2).

5.2 Repeatability and reliability of the methodology and data

5.2.1 Repeatability of the GCxGC-TOFMS and extraction method

To determine the repeatability of the GCxGC-TOFMS experiments for generating metabolomics data (Aim 1), a single QC sample was extracted once (as described in section 4.3.4), and analysed repeatedly (n = 6) on the GCxGC-TOFMS (as described in section 4.4.2.1). The data generated were processed as described in section 4.5. The CV value for each compound was subsequently determined from the relative concentrations calculated (as described in section 4.6.1) for all compounds, using the resulting database of the six injection repeats. The repeatability was evaluated using these CV values.

Ten of the most prominent compounds, detected at regular intervals throughout the chromatographic run and representing different compound classes, were selected for the first evaluation. These 10 compounds and their respective CV values are given in Table 5.1.

Similarly, in order to determine the repeatability of the extraction procedure (or the analyst's capacity to perform this repeatedly), 6 identical QC samples were extracted and each separate extract injected onto the GCxGC-TOFMS instrument. The CV values of the same 10 compounds, detected at regular intervals throughout the chromatographic run and representing different compound classes, as selected for the first evaluation, were once again used in order to determine the repeatability of the extraction procedure (Table 5.1).

As seen in Table 5.1, all 10 compounds detected fell within the prescribed CV cut-off value of 20% (Dunn *et al*., 2011; FDA Guidance for Industry, 2001; *t*'Kindt *et al*., 2009) when considering the repeatability of both GCxGC-TOFMS and that of the extraction method, indicating that the analytical method was suitable for application to the metabolomics question, and the samples collected.

Table 5.1: CV values of 10 compounds with the highest concentrations detected at regular retention time intervals, representing compounds from various compound classes

				GCxGC-TOFMS repeatability	Extraction method/analyst's repeatability
		Retention time (s)			
Rank	Compound	1st dimension	2nd dimension	CV value (%)	CV value (%)
1	L-Valine	800	1.035	15.78	12.18
2	2-Amino-adipic acid	1817	1.2	11.39	18.55
3	Hexadecanoic acid	3110	1.085	12.75	19.37
4	L-Aspartic acid	1538	1.06	4.13	12.23
5	Propanedioic acid	767	0.95	8.58	11.12
6	Butanoic acid	2798	1.26	14.67	18.66
7	L-Threonine	1205	1.043	6.94	14.71
8	Glycine	1325	0.88	16.38	15.91
9	Benzoic acid	1610	1.235	8.54	9.33
10	Glucuronic acid	2510	0.915	13.81	15.46

An important consideration is that the CV value of a compound is also influenced by its concentration – those compounds occurring at higher concentrations being associated with less variation and hence lower CV values, and *vice versa* (*t*'Kindt *et al*., 2009; Schoeman *et al*., 2012). As per definition, in an untargeted, unbiased metabolomics approach, "all" compounds present in the *M. smegmatis* metabolome would be of importance, including those present at relatively low concentrations, potentially accompanied by higher CV values. For this reason, Schoeman *et al*. (2012) proposed that for metabolomics applications a 50% CV cut-off be used when processing data, prior to statistical analyses, in order to include compounds at low concentrations that could potentially be important markers (Schoeman *et al.,* 2012). In this regard, and indicated in Figure 5.1 , 92% of all the detected compounds from the 6 QC injections fell below this 50% CV value, consequently confirming the excellent repeatability of the GCxGC-TOFMS results (see Figure 5.1). Similarly, when evaluating the repeatability of the extraction procedure using this approach, 95% of all the compounds detected in this repeatability scenario fell below the 50% hypothetical cut-off point. This metabolomics method, therefore, generates data which fell well within acceptable repeatability criteria, suitable for evaluation of the research question. Furthermore, a large majority of the compounds analysed could thus be included for biomarker identification considering the 50% cut-off for metabolomics applications, as suggested by Schoeman *et al*. (2012). Examination of those compounds excluded on the grounds of their high CV values, showed them to be contaminants or at concentrations close to the detection limit of the GCxGC-TOFMS equipment.

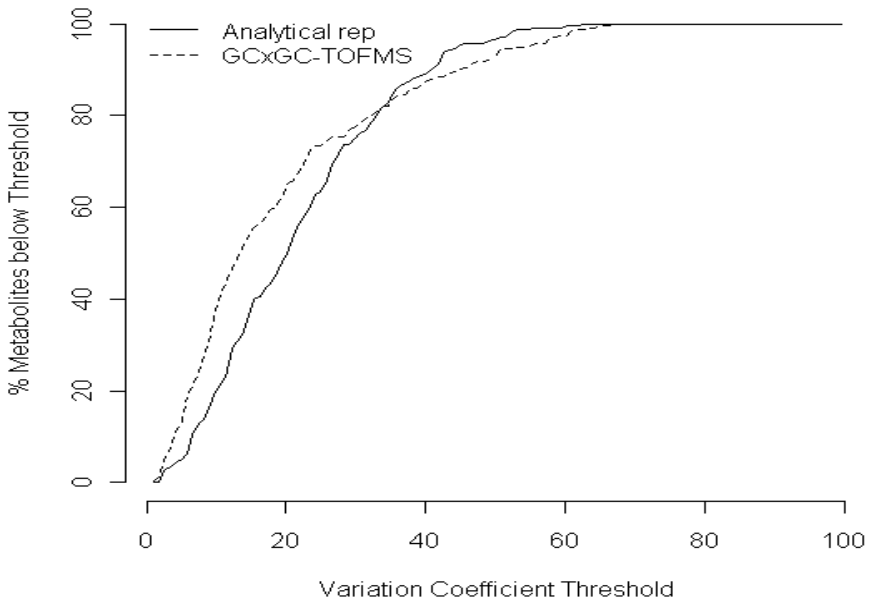


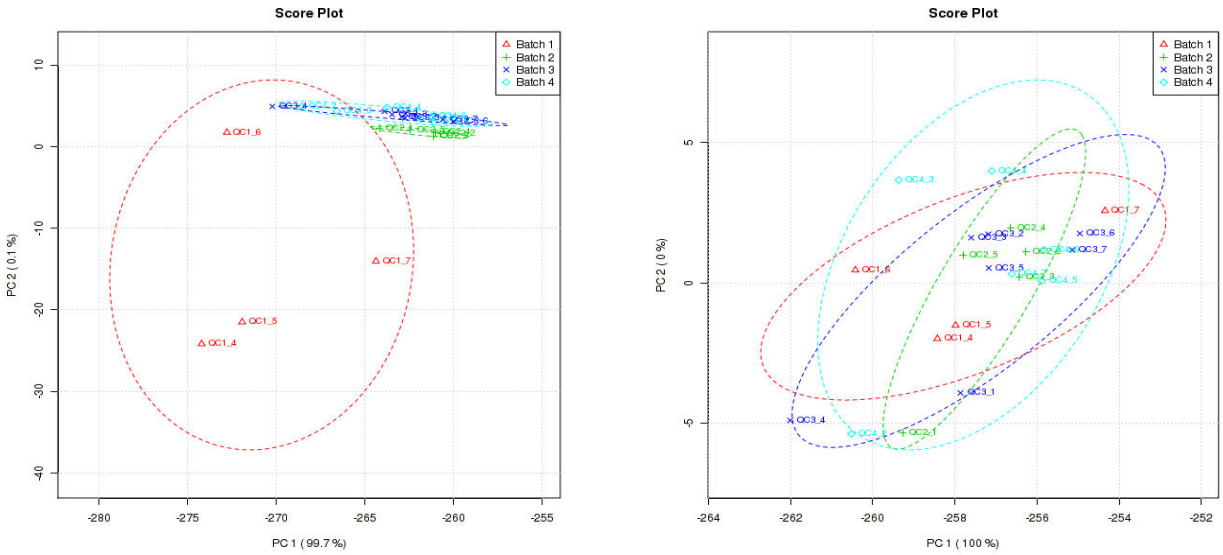
Figure 5.1: Distribution of coefficient of variation (CV) values, using the relative concentrations of all compounds determined for 1) the extraction method and 2) the GCxGC-TOFMS experiment.

5.2.2 Batch effect and QC correction of data

As previously discussed, batch effects may occur as a result of non-biological systematic errors, either from changes or drift in the extraction or analytical procedure. Despite one taking all possible precautions to prevent inconsistencies, batch effects are usually inevitable and the necessary monitoring and their correction are essential.

As mentioned previously, this data transformation method, also known as quantile equating, corrects for any linear and/or non-linear differences in distribution which may be present among batches of semi-quantitative data, obtained from the same analytical method (Draisma *et al.*, 2010). After calculating the CV values for all metabolites (results not shown), as well as PCA analyses before and after this correction (Figures 5.2a and 5.2b), results revealed the dramatic effects that this QC correction step had on the data, thus illustrating its contribution to database correction for better biomarker identification. In Figure 5.2a the batch effect is clear

(big red elips compared to smaller, overlapping elipses) when analysing the QC sample data set, and after applying the quantile equating correction step, better overlap between the batches occurs (Figure 2b), thus indicating correction of the batch effect.

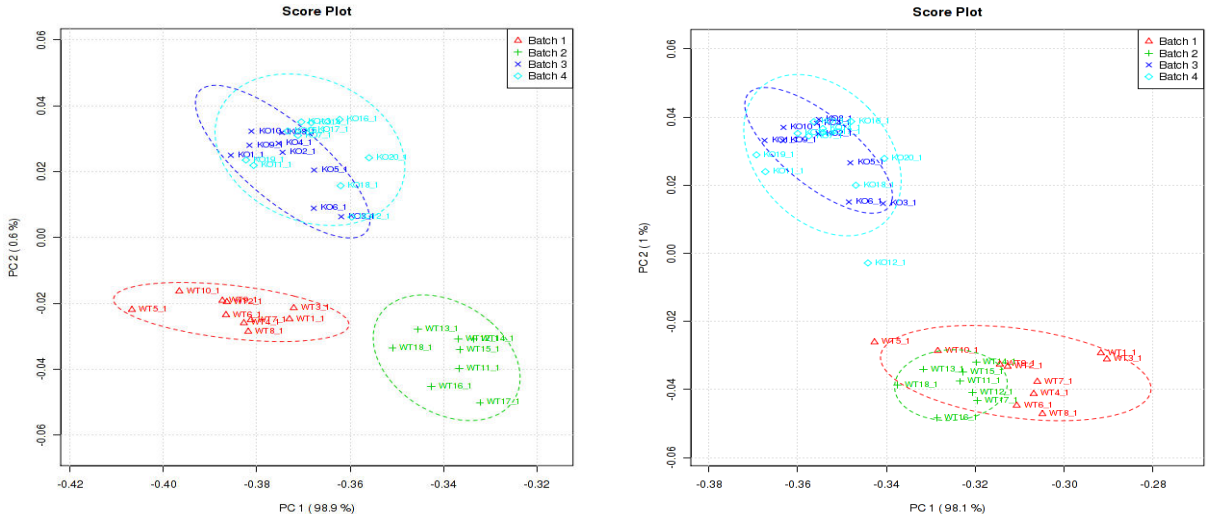


a) b)

Figure 5.2: a) Illustration of the PCA scores (PC1 plotted against PC2) of the QC samples, injected in four separate batches, prior to QC correction, illustrating a possible batch effect. b) Corresponding plot of PC1 against PC2 of the same samples after quantile equating QC correction, showing no batch effects (better overlapping of QC samples from different batches).

Further confirmation of this batch effect is evident when studying the PCA generated from the data of the extracted *M. smegmatis* wild-type and ΔESX-1_{ms} samples (Figure 5.3). The wild-type samples showed differentiation between the analysed batches, whereas the ΔESX-1_{ms} samples showed good overlap, with very little drift (Figure 5.3a). Quantile equating correction of the entire data set corrected for this (Figure 5.3b), resulting in overlap of the batches of each of the groups, and better differentiation between the two experimental groups. Any differentiation between *M.*

smegmatis wild-type and ΔESX-1_{ms} samples in the subsequent PCA analyses can therefore be ascribed to differences in the metabolite composition of the two groups.



a) b)

Figure 5.3: a) PCA scores plot (PC1 vs. PC2) of the wild-type and ΔESX-1_{ms} samples prior to quantile equating QC correction, indicating a possible batch effect due to the differentiation seen in the wild-type sample groups. b) The corresponding plot of PC1 against PC2 of the *M. smegmatis* wild-type and ΔESX-1_{ms} samples, after quantile equating QC correction was done, to correct the effect of drift. As shown in the figure, all the wild type samples were grouped in batches 1 (red) and 2 (green), wereas the ΔESX-1_{ms} samples were grouped in batches 3 (blue) and 4 (turquoise) respectively.

5.3. Metabolomic comparison of *M. smegmatis* wild-type and ΔESX-1_{ms} samples

As stated in section 3.2, the aim of this study was to use a metabolomics research approach to identify those metabolite markers best characterising ESX-1 gene cluster function, by comparing and interpreting the metabolite profiles of *M. smegmatis* wild-type and ΔESX-1_{ms} from a metabolism perspective. As discussed in

section 4.6.1, various data preprocessing steps were applied to the data, prior to any multi- or univariate statistical analyses.

5.3.1 PCA differentiation of the *M. smegmatis* wild-type and ΔESX-1$_{ms}$ samples

The data generated from the *M. smegmatis* wild-type and ΔESX-1$_{ms}$ samples analysed was subjected to PCA. Figure 5.4 illustrates clear differentiation between the experimental groups compared, which arose from the differences in the metabolite profiles of the respective strains. The total variance explained by the first three PCs (R^2 X cum) was 99.9%, of which PC1 contributed 99.8%, PC2 contributed 0.1%, and PC3 contributed 0%, as seen in Figure 5.4.

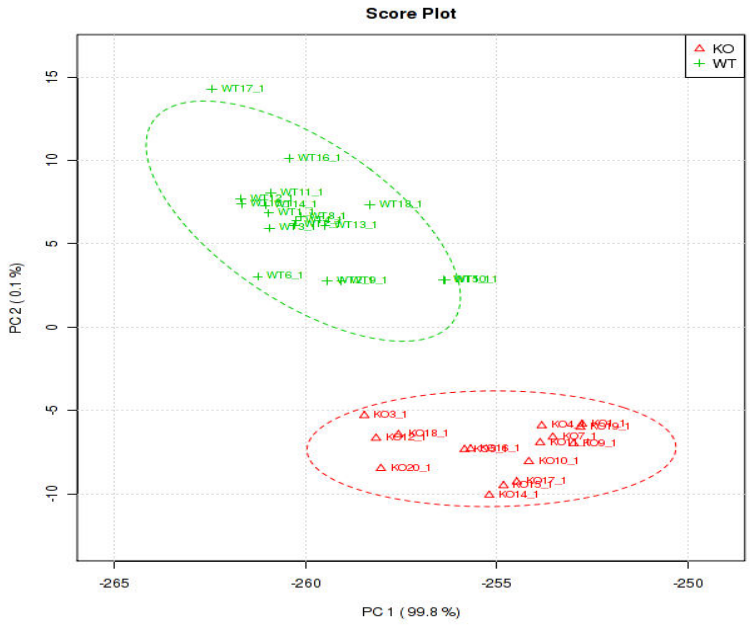


Figure 5.4: PCA scores plot (PC1 vs. PC2) of the GCxGC-TOFMS acquired data via a total metabolome extraction. The differentiation of the individual samples into distinct groups is based on their different metabolome characteristics, wild-type samples in green and ΔESX-1$_{ms}$ samples in red. As mentioned above and illustrated in this figure, the total variance illustrated by the first two PCs (R^2 X cum) was 99.9%, of which PC1 contributed 99.8%, PC2 contributed 0.1%

5.3.2 Metabolite marker identification

In metabolomics research, the differentiation observed between groups is a consequence of differences in their respective metabolite compositions. We can subsequently determine which metabolite markers contribute most to any differentiation of the experimental results observed. These markers, in turn, may be used to explain different phenotypic traits or, in the case reported here, explain the underlying mechanisms by which ESX-1 functions, from a metabolic perspective.

These metabolite markers were selected using a combination of univariate (*t*-test and effect size) and multivariate (PLS-DA and PCA) statistical approaches, as discussed in sections 4.6.2 and 4.6.3 As previously mentioned, those metabolites with a PLS-DA VIP value >1, and a PCA modelling power >0.5, were considered possible metabolite markers. The PLS-DA model used 3 components, with the modelling parameter R^2Y (cum) being 92.9%, indicative of the total accounted for variation of the response Y. Q^2 (cum), the cross-validated variation explained by the response Y, was 89.5%. These high values indicate the potential of this model with an excellent ability to explain the data. Figure 5.5 depicts the selection of these markers.

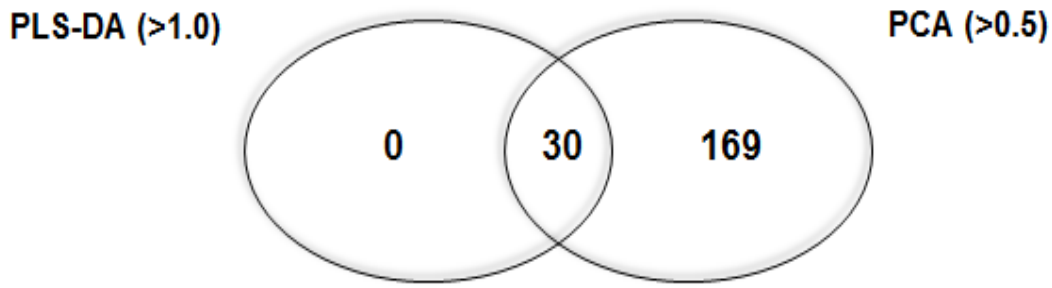


Figure 5.5: Venn diagram depicting selection of metabolite markers using the selected statistical methods.

Following statistical analyses (multivariate), 30 PLS-DA and 199 PCA metabolites adhered to their respective cut-offs, and all 30 PLS-DA markers selected also have PCA powers of >0.5. Of these 30 metabolite markers, 22 could be annotated by comparison of their mass fragmentation patterns and retention times to libraries compiled from previously injected standards. These metabolite markers, as well as their corresponding relative mean concentrations, retention times, PLS-DA VIP values, PCA powers, d-values and *P*-values, are given in Table 5.2. These results provide a basis for describing the functionality of the ESX-1 gene cluster, as discussed below.

Table 5.2: Metabolite markers identified as best describing the variation between the *M. smegmatis* wild-type and ΔESX-1_{ms} samples together with their respective relative mean concentrations, retention elution times, PLS-DA VIP values, PCA powers, effect size d-values and *t*-test *P*-values.

Compound (Chemspider)	Mean relative concentration (µg/g cells)		Retention time (s)	PCA power	PLS-DA VIP	Effect size	*t*-test
	***M. smegmatis* wild-type**	***M. smegmatis* ΔESX-1_{ms}**					
Glycerol (733)	2407.243 (1716.3)	4773.515 (2214.2)	330, 0.424	0.623	5.164	0.398	<0.001
Citric acid (305)	448.276 (287.4)	1061.352 (217.8)	2119, 5.070	0.97	5.051	2.655	<0.001
Mannose (1419823)	64.454 (45)	401.41 (232.4)	2234, 2.442	0.811	4.799	1.33	<0.001
Galactose (76670)	92.133 (42.8)	380.483 (73.9)	2041, 0.189	0.969	4.26	3.573	<0.001
Cadaverine (13866593)	9.023 (6.1)	54.407 (17.9)	665, 9.132	0.985	3.099	2.655	<0.001
L-Alanine (5735)	38.278 (20.1)	154.582 (58.4)	821, 5.906	0.952	2.854	2.033	<0.001
D-Gulonic acid (134243)	19.31 (10.8)	86.084 (19.4)	1592, 0.878	0.968	2.771	3.546	<0.001
γ-Aminobutyric acid (GABA) (116)	5.018 (4.4)	47.339 (8.6)	1649, 0.975	0.953	2.417	4.804	<0.001
L-Valine (6050)	57.945 (19.6)	184.038 (46.3)	797, 4.542	0.949	2.355	2.423	<0.001
L-Leucine (5880)	23.014 (9.1)	74.373 (18.1)	932,1.552	0.937	1.656	3.43	<0.001
D-Gulonolactone (588)	41.948 (9.3)	112.867 (32.3)	850, 5.478	0.921	1.979	2.494	<0.001

L-Lysine (5747)	32.496 (10.1)	79.84 (10.5)	2274, 0.953	0.985	1.859	5.305	<0.001
L-Glutamine (5746)	175.204 (52.3)	195.79 (60.4)	1802, 1.050	0.917	1.736	1.772	<0.001
Glucuronic acid (58552)	24.794 (5.4)	77.333 (20.8)	2282, 4.035	0.941	1.718	2.748	<0.001
L-Isoleucine (6067)	17.396 (6.7)	50.748 (10)	1205, 0.439	0.949	1.521	3.339	<0.001
Phosphoric acid (979)	34.372 (12.1)	95.677 (21.9)	2798, 1.091	0.833	1.43	2.045	<0.001
L-Norvaline (58608)	3.166 (1.1)	14.086 (5.8)	1049, 0.925	0.935	1.371	2.19	<0.001
L-Methionine (5907)	5.338 (1.7)	14.553 (5.1)	1630, 4.121	0.931	1.348	2.228	<0.001
Dodecanoic acid (3756)	10.244 (4.5)	33.709 (8.6)	2820, 0.585	0.861	1.281	2.008	<0.001
L-Phenylalanine (5910)	2.367 (0.9)	12.764 (3.2)	1814, 1.168	0.963	1.22	2.359	<0.001
D-Ribose-5-phosphate (3560875)	24.575 (11.8)	64.569 (17)	1558, 3.600	0.558	1.169	1.625	<0.001
3,4,5 Trihydroxy-pentanoic acid (9703151)	2.088 (0.6)	5.444 (1.2)	600, 7.795	0.97	1.168	3.535	<0.002

5.3.3 Discussion

The functions of ESX-1 in *M. smegmatis*, a non-pathogenic *Mycobacterium* species, are not well defined. Although this gene cluster has been shown to contribute towards conjugal DNA transfer in this organism (Flint *et al*., 2004; Coros *et al*., 2008), its potential involvement or role in metabolism has not been determined. As shown in Table 5.2, several metabolites including carbohydrates, fatty acids and amino acids were found to be comparatively elevated in ΔESX-1_{ms} relative to the *M. smegmatis* wild-type parent strain. Interestingly, 9 of the 22 metabolite markers identified in this study are identical to those described by Meissner-Roloff *et al*. (2012), who used a similar metabolomics approach to compare a hypo- and hypervirulent Beijing strain of *M. tuberculosis*, whereas the remaining 13 are intermediates in the same or related metabolic pathways. Considering the previously described role of ESX-1 in the secretion of various proteins involved in the virulence of *M. tuberculosis*, this could be expected. These findings not only support the work of Meissner-Roloff *et al*. (2012) but most of the previous studies describing the role of ESX-1 in mycobacterial virulence. Furthermore, metabolites identified as being elevated in ΔESX-1_{ms} may indicate alterations in cell wall synthesis, as well as additional effects on energy demand and redox stress correlating with pathways influencing virulence in *M. tuberculosis*. Furthermore, the majority of the metabolites identified were found to be elevated in the ΔESX-1_{ms} strain which indicates reduced utilisation of various essential processes in *M. smegmatis* due to the loss of the ESX-1 gene cluster. These observations will be discussed in detail below and are summarised in Figure 5.6.

5.3.3.1 Metabolites associated with altered cell wall synthesis in the ΔESX-1_{ms} strain

Several of the metabolites identified (Table 5.2) contribute to cell envelope structure, possibly highlighting a role for ESX-1 in cell envelope biogenesis in *M. smegmatis*. A component of ESX-1, $EccA_1$, has been shown to contribute to mycolic acid synthesis in *M. marinum,* thereby contributing to cell envelope biogenesis (Joshi *et al*., 2012). Deletion of $EccA_1$ in *M. marinum* resulted in the synthesis of 30–40% less

mycolic acids. $EccA_1$ was predicted to bind to and transport the various enzymes involved in mycolic acid synthesis, to the sites of mycolate synthesis in *M. marinum* (Joshi *et al., 2012*). The synthesis of mycolic acids relies on the incorporation of methyl, cyclopropane and oxygenated groups into fatty acid chains. These reactions are catalysed by S-adenosyl methionine (SAM)-dependent methyl-transferases, which transfer a methyl group from methionine to the unsaturated fatty acid (Asselineau *et al*., 2002). Considering this, the greatly increased methionine detected in ΔESX-1_{ms} relative to the wild-type strain suggests a reduced requirement for methionine for mycolic acid methylation and the role of the ESX-1 component $EccA_1$ in *M. smegmatis* cell wall synthesis. Furthermore, the mycolic acids of the mycomembrane are covalently linked to the arabinogalactan in the cell wall (Crick *et al*., 2001). Arabinogalactan and arabinomannan, which are both cell envelope components, are comprised of galactose and arabinose (1.0:2.7), and galactose and mannose (1.0:1.1), respectively (Azuma *et al*., 1970). The elevated levels of mannose and galactose in ΔESX-1_{ms} may be indicative of reduced arabinogalactan and arabinomannan production in this strain relative to the wild-type strain, resulting in accumulation of these substrates, further confirming the proposed role of ESX-1 in cell wall synthesis.

Various amino acids also contribute to the synthesis of structural components of the mycobacterial cell wall. Miyakawa *et al*. (1972) indicated large amounts of alanine and glutamic acid as being present in the bacterial cell wall; it is believed that these two amino acids, bound to muramic acid, are essential components of peptidoglycan and comprise roughly 22.8% of the total amino acid content of the cell wall (Acharya and Goldman, 1970). Subsequently, the increased concentrations of glutamine, γ-aminobutyric acid (GABA) (derived from glutamate), alanine and valine (required for alanine synthesis) are possibly a result of reduced peptidoglycan synthesis. Additionally, the increased concentrations of glutamine, lysine and phenylalanine in ΔESX-1_{ms} may also indicate that aspartic acid is being used preferentially to produce these amino acids, rather than for the synthesis of meso-diaminopimelate (meso-DAP), which is also required for peptidoglycan biosynthesis (Umbarger, 1978; Pavelka and Jacobs, 1996), further confirming these observations.

Cadaverine was also elevated in the ΔESX-1_{ms} strain relative to the wild-type strains. Cadaverine is a polyamine antioxidant, synthesised from lysine, which is involved in membrane permeability, acid tolerance and RNA synthesis, in various bacteria (Kamio *et al*., 1981; Neely and Olson, 1996; Samartzidou and Delcour, 1999). Cadaverine has been shown to induce porin closure and reduce membrane permeability (Samartzidou and Delcour, 1999) and is also found covalently linked to peptidoglycan (Kamio *et al*., 1981). Increased cadaverine levels within the cell may be due to elevated synthesis, or reduced utilisation. ΔESX-1_{ms} may increase cadaverine production for the purpose of reducing membrane permeability, to compensate for the under-developed cell envelope structure. Alternatively, defective cell envelope synthesis may prevent the efficient incorporation of cadaverine into the structure, resulting in its accumulation within the cell.

Lastly, the secretion, together with the proper structure and functioning of the ESX-1-associated protein, EspA, has previously been shown to be required for suitable cell wall integrity of *M. tuberculosis* (Garces *et al*., 2010). This observation, in addition to the above metabolomics results, provides strong evidence for the role of ESX-1 in cell wall synthesis in *M. smegmatis*, and that the defects due to the absence of ESX-1 may involve the mycolic acids in the mycomembrane, the peptidoglycan and arabinogalactan in the cell wall, as well as the LAMs which span the capsule. Type VII or ESX secretion depends on an intact cell wall structure; however, these results show that ESX-1 may mutually be required for the efficient synthesis of the cell envelope. The conserved localisation of various ESX-1 proteins in *M. smegmatis, M. marinum* and *M. tuberculosis* at the polar regions, where growth and therefore maximum cell envelope synthesis occurs, supports a role for ESX-1 in cell envelope biogenesis in *M. smegmatis*.

5.3.3.2 Metabolites associated with altered energy production and redox state in ΔESX-1_{ms} strain

Elevated levels of various antioxidants and their intermediates were detected in ΔESX-1_{ms}. For example, gulonic acid, glucuronic acid and gulonolactone, intermediates in the bacterial biosynthesis of ascorbic acid (Metzler, 2003), an antioxidant in mycobacterial species (Taneja *et al*., 2010), were elevated in ΔESX-

1_{ms}. Cadaverine can also function as an antioxidant to scavenge superoxide radicals and regulate extracellular pH by protecting the cells from acid stress (Kim *et al*., 2006). Inadequate formation of the cell envelope results in increased permeability of the cell in ΔESX-1_{ms} and hence elevated oxidative stress; consequently, there is then a need to generate more antioxidants.

Several metabolites representing carbon and nitrogen sources were also elevated in ΔESX-1_{ms}, suggestive of reduced energy utilisation. The raised concentrations of citric acid in the ΔESX-1_{ms} suggest partial or complete inhibition of the TCA cycle in this organism. Supporting this is the elevated norvaline detected in this strain, as its successful catabolism yields α-ketogluturate, an essential intermediate in the TCA cycle (Ramakrishnan *et al*., 1972). Glycerol, mannose and galactose, the primary energy substrates for mycobacteria (Merrill, 1931; Ramakrishnan *et al*., 1972; Patterson *et al*., 2003), were all elevated in ΔESX-1_{ms}, relative to *M. smegmatis* wild-type samples, indicating a reduction in the utilisation of these simple sugars for energy production. Ribose-5-phosphate is an essential component in the synthesis of nucleotides (Wood *et al*., 1973), of important factors such as ATP and NAD (which regulate growth) (Roos *et al*., 2004), the amino acids histidine and tryptophan, and is utilised in the pentose phosphate pathway (Pilz *et al*., 1984). The elevated concentration of ribose-5-phosphate in ΔESX-1_{ms} supports the hypothesis of reduced energy utilisation in this strain. Lipid catabolism becomes essential for mycobacterial viability after depletion of simple sugars as a carbon source (Micklinghoff *et al*., 2009), leading to the production of energy via beta-oxidation and the glyoxylate cycle (Upton and McKinney, 2007). Increased concentrations of glycerol and dodecanoic acid in the ΔESX-1_{ms} strain may indicate reduced utilisation of these fatty acids for energy in the absence of ESX-1. Furthermore, nitrogen is essential for mycobacterial growth and cell division (Amon *et al*., 2008); glutamic acid (provided in the culture medium and likely to be the main nitrogen source in this study) can be used to synthesise both glutamine and GABA (Jones, 2002), both of which are elevated in ΔESX-1_{ms}. Glutamine is a well-described nitrogen reservoir for various bacterial species and GABA is capable of supporting bacterial growth as either a carbon or a nitrogen source (Seth and Connell, 2000). The increased levels of these two metabolites in ΔESX-1_{ms} may result from a reduced nitrogen and carbon

requirement. The elevated levels of simple sugars, nitrogen metabolites and fatty acids in ΔESX-1_{ms}, relative to wild-type *M. smegmatis*, may indicate reduced utilisation due to a lower energy requirement, possibly associated with a reduction in the metabolic state, growth and cell envelope components in this gene-altered strain.

Interestingly, as previously mentioned, nine of the 22 metabolite markers identified in this study are identical to those described by Meissner-Roloff *et al*. (2012), who used a metabolomics approach to compare hypo- and hypervirulent Beijing strains of *M. tuberculosis*, whereas the other 13 metabolites are intermediates in the same, or related, metabolic pathways. Most strikingly, seven of the nine metabolites attributed to altered cell envelope biogenesis in this ESX-1 study were similarly elevated in the hypo-virulent strain, relative to the hypervirulent strain. These include mannose, galactose, alanine, glutamine, lysine, phenylalanine and cadaverine. Increased levels of antioxidants and energy metabolites (mannose, galactose, glycerol, fatty acids and glutamate) in the hypo-virulent *M. tuberculosis* strain also correspond with the observations in the absence of ESX-1 in *M. smegmatis*.

5.3.3.3 How do these results relate to the role of ESX-1 in *Mycobacterium* virulence?

Defective cell envelope biogenesis could affect virulence in several ways. The effect of EspA inactivation/deletion on virulence was shown to be independent of ESX-1-mediated secretion, suggesting that the effect was due to the resultant altered cell wall composition (Joshi *et al*., 2012). The altered cell wall structure may affect virulence by disrupting the bacterial interface, or by increasing its susceptibility to innate host defences (Garces *et al*., 2010). Various ESX-1-associated proteins have been identified in the capsular layer, including Esx, Esp and PPE proteins, which are produced by the secretion system (Sani *et al*., 2010). Disruption of this capsular layer may release some of these proteins, or prevent their insertion into it, thereby preventing efficient host–pathogen interactions with these virulence-mediating proteins, and in so doing alter bacterial virulence. ESX-1 may also be involved in transporting other proteins to the cell envelope in order for them to perform their functions, or to assist in cell envelope synthesis and maintenance. As ESX-1 codes for ESAT-6 and CFP-10 – which are also secreted over the cell envelope – the subsequent loss of the ESX-1 gene cluster encoding these proteins may serve as a

possible contributor to the elevation of leucine, isoleucine, valine, methionine, alanine and phenylalanine, because the side chains of these proteins consist predominantly of these amino acids (Renshaw *et al*., 2005). Furthermore, certain substrates and components themselves could be involved in cell wall biogenesis. This was proposed by Joshi *et al*. (2012) to be the case for $EccA_1$, which is predicted to transport enzymes responsible for mycolic acid synthesis to the site of mycolic acid biogenesis. Various cell surface lipids, proteins and carbohydrates are implicated in virulence via regulation of phagocytosis (Cywes *et al*., 1997; Stokes *et al*., 2004), immune modulation (Gagliardi *et al*., 2007; Geurtsen *et al*., 2009) and persistence (Sambou *et al*., 2008). Even small changes to the mycolic acid profile can affect virulence (Yuan *et al*., 1998; Dubnau *et al*., 2000; Glickman *et al*., 2000). An altered cell envelope structure, associated with ESX-1 deletion, is therefore likely to influence virulence, and hence the strong correlations in the metabolite profiles of the ESX-1 wild-type strain in the current metabolomics investigation, and the hypervirulent *M. tuberculosis* strain investigated by Meissner-Roloff *et al*. (2012) compared to their respective controls.

No significant change in ESX-1 protein expression was detected between the hypo- and hypervirulent *M. tuberculosis* strains investigated by de Souza *et al*. (2010), suggesting that all perturbations in these strains are not directly associated with ESX-1. The absence of anomalies in methionine levels in the differentially virulent *M. tuberculosis* strains further suggests that the altered cell wall structure described by Meissner-Roloff *et al*. (2012) results from mechanisms other than the defective mycolic acid synthesis described for $\Delta ESX\text{-}1_{ms}$ (Joshi *et al*., 2012). However, the same pathways influenced by ESX-1 in *M. smegmatis* appear to be affected by alternative mechanisms that influence virulence in *M. tuberculosis*. ESX-1 is, however, required for full virulence of *M. tuberculosis*, therefore the combined roles of ESX-1 and other mechanisms contributing to the variation between the hypo- and hypervirulent *M. tuberculosis* strains, in mycolic acid and cell wall biosynthesis, as well as other related pathways, promote mycobacterial virulence.

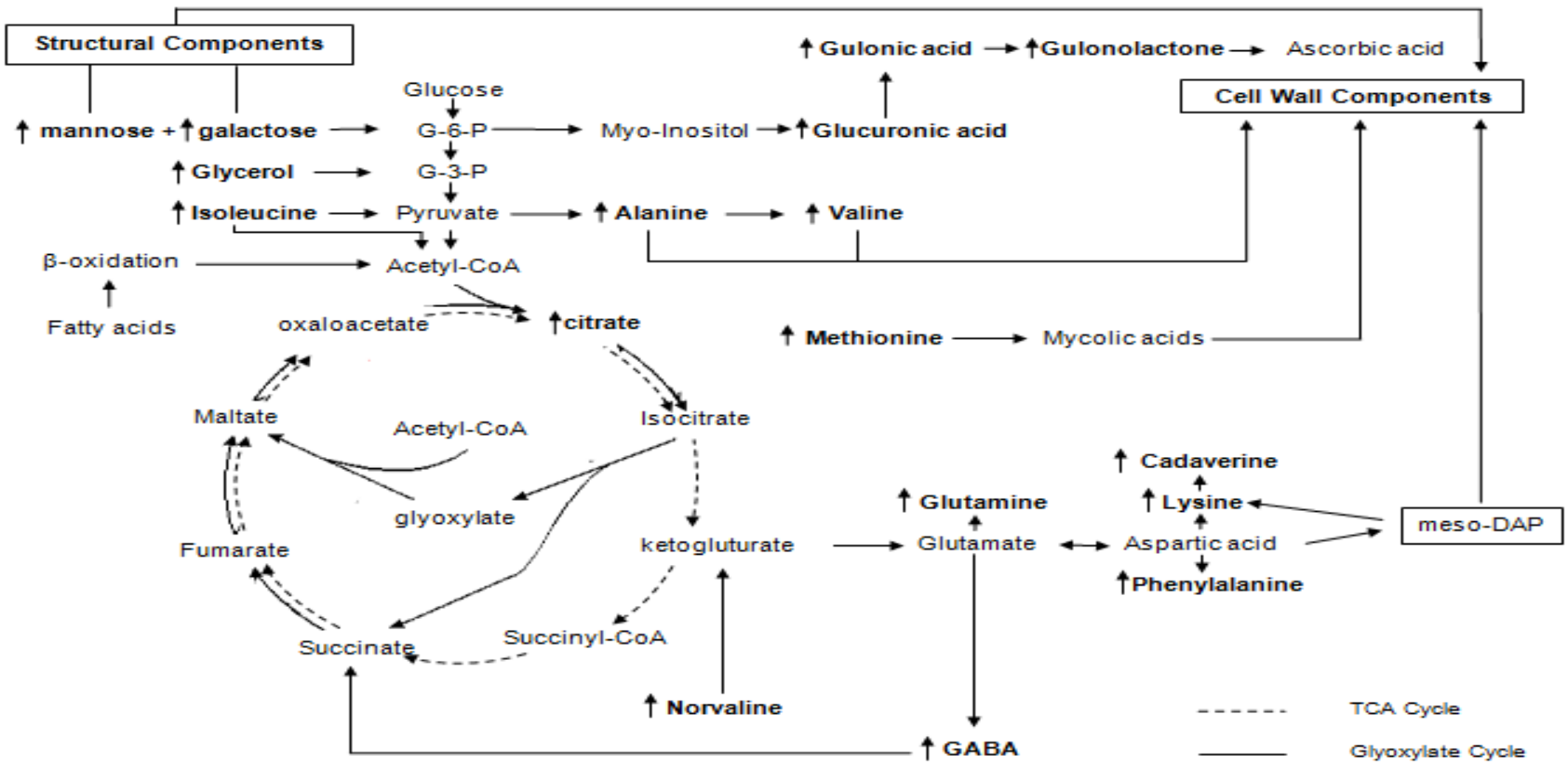


Figure 5.6: The metabolic pathways altered in ΔESX-1_{ms} compared to *M. smegmatis* wild-type strains, showing increases (indicated by ↑) in various metabolites associated with glycolysis, the TCA and glyoxylate cycles, structural components of the cell envelope, and oxidative stress/antioxidant effect.

CHAPTER 6:
CONCLUSION

Here I summarise the major findings reported in Chapter 5 in the context of the aims of this study as listed in Chapter 3. Finally, I discuss future research prospects based on the outcomes of this metabolomics study.

TB is considered to be the world's second-most deadly infectious disease, after HIV. The WHO recently reported that one third of the global population is infected with *M. tuberculosis* (Mathema *et al*., 2006), and in 2010 documented 8.8 million new cases, 650 000 of which were infected with multidrug-resistant strains, contributing to the 1.4 million deaths from TB reported globally (equivalent to 3 800 deaths a day) for that year. Developing countries are by far the worst affected by this epidemic and account for 95% of the reported global deaths due to this disease (WHO, 2010). Additionally, 25% of the individuals who succumb as a result of this epidemic are reported to be co-infected with HIV, with Europe and Africa exhibiting the highest HIV co-infection rates (WHO, 2011). Although the mortality rates of these co-infected individuals have declined since the advent of antiretroviral therapy, MDR-TB still contributes greatly to the high mortality rates in rural areas, especially in southern Africa, as a direct consequence of these patients failing to comply with the necessary treatment regimens (Gandhi *et al*., 2006; Caminero *et al*., 2010). These findings are alarming, as TB is considered a curable disease and those affected are capable of a full recovery (Knechel, 2009; Russell, 2011) after the successful completion of the correct course of treatment (Olivier and Loots, 2011).

Many research groups have studied the role of the ESX-1 gene cluster region of *M. tuberculosis,* and the associated proteins, due to their accompanying involvement in the virulence of this pathogen. Proteins encoded by the ESX-1 gene cluster and secreted via the associated T7S system, including ESAT-6 and CFP-10, are known to disrupt host defences and consequently support *M. tuberculosis* infection. Better characterisation of ESX-1 functionality and the mechanism(s) associated with the secretion of various entities encoded for by this gene cluster is expected to contribute substantially towards fully understanding *M. tuberculosis* virulence. Subsequently, this may lead to better vaccination strategies and approaches to treatment, in the fight to curb this epidemic.

All prior knowledge on ESX-1 functionality was acquired from either a genomics, transcriptomics or proteomics perspective. In contrast, I used a GCxGC-TOFMS

metabolomics research approach to explore the function of the ESX-1 gene cluster in *M. smegmatis*. This was done by extracting and analysing "all" metabolites present in the respective metabolomes of *M. smegmatis* wild-type and ΔESX-1_{ms} samples. As *M. smegmatis* is non-pathogenic, my findings do not directly correlate with the virulence of the bacteria. However the changes in virulence due to the absence or presence of the ESX-1 gene cluster helps to understand its role in *Mycobacterium* as a whole – including the virulent strains – and how this may contribute to virulence, including its dependence on other gene clusters for virulence in related pathogenic *Mycobacterium* species.

6.1 Conclusions pertaining to method validation

It is clear that the experimental method selected showed acceptable CV values when considering both the repeatability of the GCxGC-TOFMS and that of the extraction procedure selected, for generating metabolomics data, considering the 20% cut-offs for major metabolites as proposed by Dunn *et al*. (2005). Furthermore, when considering the 50% metabolomics CV cut-offs proposed by Schoeman *et al*. (2012) in Figure 5.1, 92% and 95% of all the detected compounds from the QC samples fell within range, consequently confirming excellent GCxGC-TOFMS repeatability and the extraction procedure/analyst's ability repeatability scenarios, proving these methods suitable for use for this metabolomics investigation.

Furthermore, PCA analyses confirmed the necessity for using QC samples when analysing multiple batches and batch correction of the metabolomic data generated (Figures 5.2a and b).

6.2 Conclusions pertaining to *M. smegmatis* wild-type and ΔESX-1_{ms} metabolomic comparison

This is the first study to use a metabolomics approach to describe the metabolome associated with the ESX-1 gene cluster in mycobacteria, using *M. smegmatis* as a model. The metabolite profiles of an ΔESX-1_{ms} and its *M. smegmatis* wild-type parent strain were compared and various metabolites – predominantly

carbohydrates, amino acids and fatty acids – were found to be overabundant in the absence of ESX-1. A total of 22 metabolites were identified as showing the most variation when comparing the two sample groups, on the basis of both multivariate and univariate statistical analyses. All 22 metabolites correlate with the findings of Meissner-Roloff *et al*. (2012), who used a similar metabolomics approach to compare a hypo- and hypervirulent Beijing strain of *M. tuberculosis* (9 metabolites are identical and 13 are of the same or related metabolic pathways), implying that ESX-1 influences the same pathways associated with virulence in *M. tuberculosis*. This influence of ESX-1 supports its previously proposed role in virulence and the metabolism involved in its optimisation. Nevertheless, no substantial conclusions regarding virulence can be drawn from this study, because *M. smegmatis* is non-pathogenic. Despite this, however, the alterations seen due to the absence or presence of ESX-1, does better assist in an understanding of the role of this gene cluster in *Mycobacterium* as a whole, including the virulent strains, and how this may contribute to virulence, as well as its dependence on other gene clusters for virulence in the related pathogenic species.

The altered metabolite profiles of ΔESX-1$_{ms}$ were attributed to alterations in 1) the cell envelope structure, and associated effects on 2) redox stress and 3) energy metabolism. These results correlate the function of the ESX-1 gene cluster in *M. smegmatis* with the roles in mycolic acid synthesis and cell wall biogenesis previously ascribed to ESX-1 in the pathogenic species *M. marinum* and *M. tuberculosis*. ESX-1-associated secretion is dependent on an intact cell wall structure; however, it appears also to be involved in generating such a structure. This role for ESX-1 in cell envelope biogenesis may contribute to virulence in *M. tuberculosis* by ensuring the incorporation and maintenance of various immune-modulating components at the interface between the host and pathogen. It should also be mentioned that these results have provided new insights regarding the functionality of ESX-1. As metabolomics is particularly focused on generating hypotheses, future studies investigating these findings using a more direct approach should be considered.

With the uncertainty surrounding the T7S system and the proteins contributing to virulence of *Mycobacterium*, a better understanding of the mechanisms associated

with these and the role ESX-1 and the factors related to it, are of great value. Understanding these mechanisms will enable modern medicinal approaches to target the T7S system and eventually lead to improved vaccination strategies and more effective treatment options.

Despite the difficulties associated with interpreting metabolomic data due to the limited literature available on metabolic pathways related to *Mycobacterium* and especially ΔESX-1_{ms} strains, I succeeded in identifying a number of metabolite markers, implicating the ESX-1 gene cluster in playing a role in cell growth and replication, cell wall synthesis, antioxidant capacity and the synthesis of compounds involved in *M. tuberculosis* virulence. Furthermore, this study highlights the potential use of metabolomics in generating metabolite biomarkers to describe the mechanisms associated with an altered phenotype, subsequently bridging the gap between previously generated genomics/proteomics data, and the resultant altered phenotypical characteristics.

6.3 Recommendations for future research

Considering the findings above suggesting the role of the ESX-1 gene cluster in cell wall structure, opens up a variety of research questions. As discussed in Chapter 2, the composition of the mycobacterial cell wall enables secretion of virulence-associated proteins in *M. tuberculosis*. Mycolic acids form an essential part of the backbone of this cell envelope and these very long-chain fatty acids have been shown to contribute to virulence in mycobacteria. As mentioned in the discussion, an ESX-1 component, EccA1, has previously been shown to be involved in mycolic acid biogenesis in *M. marinum*. Very little literature on the involvement of ESX-1 on mycolic acid synthesis in other mycbacteria, including *M. tuberculosis*, is available. Exploring the contribution of *M. tuberculosis* ESX-1 to mycolic acid synthesis may therefore lead to the identification of key mycolic acids or associated metabolic pathways linked to virulence in this organism. These entities belong exclusively to *Mycobacterium*, making them an attractive target for drug research in the fight to rid the world of TB.

Furthermore, ESAT-6 and CFP-10, together with other entities implicated in the T7S system, are encoded for by RD-1, another component of the ESX-1 gene cluster. This 9.5-kb region is absent from all avirulent BCG strains, but present in all *M. tuberculosis* strains. Lewis *et al*. (2003) successfully deleted RD-1 from the genome of *M. tuberculosis* H37Rv and subsequently constructed the H37Rv:ΔRD1 strain. This study group subsequently assessed the virulence of this strain using various animal and host cell models. Comparing the infection and replication severity results of the H37Rv:ΔRD1 strain to a wild-type *M. tuberculosis* H37Rv, they found that the former strain mimics avirulent characteristics of the BCG strains. A metabolomics research approach that compares metabolite profiles of the same H37Rv:ΔRD1 strain with the wild-type *M. tuberculosis* H37Rv strain, may potentially lead to the identification of metabolite markers that better explain the contribution of RD-1, in the ESX-1 gene cluster, to *M. tuberculosis* virulence.

Finally, TB is still considered a fatal infectious disease if left untreated – more effective treatment and vaccination strategies will become a reality only through better characterising and understanding the associated disease mechanisms. Therefore all opportunities need to be considered in devising the next phase of medicinal approaches to curb this epidemic, especially the use of metabolomics in future research.

CHAPTER 7:
REFERENCES

ABDALLAH, A.M., GEY VAN PITTIUS, N.C., DIGIUSEPPE CHAMPION, P.A.D., COX, J., LUIRINK, J., VANDENBROUCKE-GRAULS, C.M.J.E., APPELMELK, B.J., BITTER, W. 2007. Type VII secretion — mycobacteria show the way. *Nature*, 5, 883-891.

ABDALLAH, A.M., VERBOOM, T., HANNES, F., SAFI, M., STRONG, M., EISENBERG, D., MUSTERS, R.J.P., VANDENBROUCKE-GRAULS, C.M.J.E., APPELMELK, B.J., LUIRINK, J., BITTER, W. 2006. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Molecular Microbiology*, 62(3), 667-679.

ACHARYA, P.V.N., GOLDMAN, D.S. 1970. Chemical composition of the cell wall of the H37Ra strain of *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 102(3), 733-739.

AMON, J., BRU, T., GRIMRATH, A., HÄNBLER, E., HASSELT, K., HÖLLER, M., JEBBERGER, N., OTT, L., SZÖKÖL, J., TITGEMEYER, F., BURKOVSKI, A. 2008. Nitrogen control in *Mycobacterium smegmatis*: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-Type Regulator GlnR, *Journal of Bacteriology*, 190(21), 7108-7116.

ANDERSEN, P., DOHERTY, T.M. 2005. The success and failure of BCG — implications for a novel tuberculosis vaccine. *Nature Reviews Microbiology*, 3, 656-662.

ARAND, M., GRANT, D.F., BEETHAM, J.K., FRIEDBERG, T., OESCH, F., HAMMOCK, B.D. 1994. Sequence similarity of mammalian epoxide hydrolases to the bacterial haloalkane dehalogenase and other related proteins. *FEBS Letters*, 338, 251-256.

AZUMA, I., AJISAKA, M., YAMAMURA, Y. 1970. Polysaccharides of *Mycobacterium bovis Ushi* 10, *Mycobacterium smegmatis*, *Mycobacterium phlei*, and Atypical *Mycobacterium* P1. *Infection and Immunity*, 2(3), 347-349.

BEHR, M.A., SMALL, P.M. 1999. A historical and molecular phylogeny of BCG strains. *Vaccine*, 17, 915-922.

BELTAN, E., HORGEN, L., RASTOGI, N. 2000. Secretion of cytokines by human macrophages upon infection by pathogenic and non-pathogenic mycobacteria. *Microbial Pathogenesis*, 28, 313-318.

BIRKEN, V. 2008. Molecular mechanisms of host-pathogen interactions and their potential of the discovery of new drug targets. *Current Drug Targets*, 9(2), 150-157.

BITTER, W., HOUBEN, E.N.G., BOTTAI, D., BRODIN, P., BROWN, E.J., COX, J.S., DERBYSHIRE, K., FORTUNE, S.M., GAO, LY., LIU, J., GEY VAN PITTIUS, N.C., PYM. A.S., RUBIN, E.J., SHERMAN.

D.R., COLE. S.T., BROSCH, R. 2009. Systematic genetic nomenclature for type VII secretion systems. *PLoS Pathogens*, 5(10), 1-6.

BOTTAI, D., BROSCH, R. 2009. Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families. *Molecular Microbiology*, 73(3), 325-328.

BRENNAN, P.J. 2003. Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis*, 83, 91-97.

BRERETON, R. 2003. Data Analysis for the laboratory and chemical plant. John Wiley & Sons Ltd., Chichester.

BRODIN, P., MAJLESSI, L., MARSOLLIER, L., DE JONGE, M.I., BOTTAI, D., DEMANGEL, C., HINDS, J., NEYROLLES, O., BUTCHER, P.D., LECLERC, C., COLE, S.T., BROSCH, R. 2006. Dissection of ESAT-6 System 1 of *Mycobacterium tuberculosis* and impact on iImmunogenicity and virulence. *Infection and Immunity*, 74(1), 88-98.

BROSCH, R., GORDON, S.V., PYM, A., EIGLMEIER, K., GARNIER, T., COLE, S.T. 2000. Comparative genomics of the mycobacteria. *International Journal of Medical Microbiology*, 290, 143-152.

BROSCH, R., PYM, A.S., GORDON, S.V., COLE, S.T. 2001. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends in Microbiology*, 9(9), 452-458.

BROSCH, R., GORDON, S.V., MARMLESSE, M., BRODIN, P., BUCHRLESER, C., ELGLMEIER, K., GARNIER, T., GUTIERREZ, C., HEWINSON, G., KREMER, K., PARSONS, L.M., PYM, A.S., SAMPER, S., VAN SOOLINGEN, D., COLE, S.T. 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *PNAS*, 99(6), 2684-3689.

CAMINERO, J.A., SOTGIU, G., ZUMLA, A., MIGLIORI, G.B. 2010. Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *Lancet Infectious Diseases*, 10(9), 621-629.

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. *Molecular Microbiology*, 34(2), 257-267.

CHATTERJEE, D. 1997. The mycobacterial cell wall: structure, biosynthesis and sites of drug action. *Current Opinion in Chemical Biology*, 1, 579-588.

CHATTERJEE D., KHOO, K.H. 1998. Mycobacterial lipoarabinomannan: an extraordinary lipoheteroglycan with profound physiological effects. *Glycobiology*, 8(2), 113-120.

CHONG, I.G., JUN, C.H. 2005. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78, 103–112.

COLE, S.T., BROSCH, R., PARKHILL, J., GARNIER, T., CHURCHER, C., HARRIS, D., GORDON, S.V., EIGLMEIER, K., GAS, S., BARRY III, C.E., 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., J. SKELTON, 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.

COLE, S.T. 2002. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology*, 148, 2919-2928.

CONVERSE, S.E., COX, J.S. 2005. A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *Journal of Bacteriology*, 187(4), 1238-1245.

COROS, A., CALLAHAN, B., BATTAGLIOLI, E., DERBYSHIRE, K.M. 2008. The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Molecular Microbiology*, 69(4), 794-808.

DALLÜGE, J., VREULS, R.J.J., BEENS, J., BRINKMAN, U.A.T. 2002. Optimization and characterization of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection (GCxGC-TOF MS). *Journal of Separation Science*, 25, 212-216.

DAS, C., GHOSH, T.S., MANDE, S.S. 2011. Computational analysis of the ESX-1 region of *Mycobacterium tuberculosis*. insights into the mechanism of type VII secretion system. *PLoS ONE*, 6(11), 1-12.

DELOGU, G., BRENNAN, M.J. 2001. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. *Infection and Immunity*, 69(9), 5606-5611.

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. *Molecular and Cellular Proteomics*, 9, 2414-2423.

DIGIUSEPPE CHAMPION, P.A., COX, J.S. 2007. Protein secretion systems in mycobacteria. *Cellular Microbiology*, 9(6), 1376-1384.

DOMENECH, P., BARRY III, C.E., COLE, S.T. 2001. *Mycobacterium tuberculosis* in the post-genomic age. *Current Opinion in Microbiology*, 4, 28-34.

DRAISMA, H.H.M, REIJMERS, T.H., VAN DER KLOET, F., BOBELDIJK-PASTOROVA, I., SPIES-FABER, E., VOGELS, J.T.W.E., MEULMAN, J.J., BOOMSMA, D.I., VAN DER GREEF, J., HANKEMEIER, T. 2010. Equating, or correction of between-block effects with application to body fluid LC-MS and NMR metabolomics data sets, *American Chemical Soceity*, 82, 1039-1046.

DUNN, W.B., BAILEY, N.J.C., JOHNSON, H.E. 2005. Measuring the metabolome; current analytical technologies. *Analyst*, 130, 606-625.

DUNN, W.B., BROADHURST, D., BEGLEY, P., ZELENA, E., FRANCIS-MCINTYRE, S., ANDERSON, N., BROWN, M., KNOWLES, J.D., HALSALL, A., HASELDEN, J.N., NICHOLLS, A.W., WILSON, I.D., KELL, D.B., GOODACRE, R., The Human Serum Metabolome (HUSERMET) Consortium. 2011. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 6(7), 1060-1083.

ELLIS, S.M., STEYN, H.S. 2003. Practical significance (effect sizes) versus or in combination with statistical significance (p-values). *Management Dynamics: Journal of the Southern African Institute for Management Scientists*, 12(4), 51-53.3

FELTCHER, M.E., SULLIVAN, J.T., BRAUNSTEIN, M. 2010. Protein export systems of *Mycobacterium tuberculosis*: novel targets for drug development?. *Future Microbiology*, 5, 1581-1597.

FIEHN, O., WOHLGEMUTH, G., SCHOLZ, M., KIND, T., LEE, D.Y., LU. Y., MOON, S., NIKOLAU, B. 2008. Quality control for plant metabolomics: reporting MSI-compliant studies. *The Plant Journal*, 53, 691-704.

FLINT, J.L., KOWALSKI, J.C., KARNATI, P.K., DERBYSHIRE, K.M. 2004. The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. *PNAS*, 101(34), 12598-12603.

FOOD AND DRUG ADMINISTRATION. 2001. *Guidance for Industry: Bioanalytical Method Validation*. Rockville, MD: US Department of Helath and Human Services, FDA, Center for Drug Evaluation and Research.

FORRELLAD, M.A., KLEPP, L.I., GIOFFRÉ, A., GARCIÁ, J.S.Y., MORBIDONI, H.R., SANTANGELO, M. CATALDI, A.A., BIGI, F. 2013. Virulence factors of the *Mycobacterium tuberculosis* complex. *Landes Bioscience*, 4(1), 1-64.

FORTUNE, S. M., JAEGER, A., SARRACINO, D.A., CHASE, M.R., SASSETTI, C.M., SHERMAN, D.R., BLOOM, B.R., RUBIN, E.J. 2005. Mutually dependent secretion of proteins required for mycobacterial virulence. *PNAS*, 102(30), 10676-10681.

FU, L.M., FU-LIU, C.S. 2002. Is *Mycobacterium tuberculosis* a closer relative to Gram-positive or Gram–negative bacterial pathogens?. *Tuberculosis*, 82(2-3), 85-90.

GALIMI, R. 2011. Extrapulmonary tuberculosis: tuberculous meningitis new developments. *European Review for Medical Pharmacological Sciences*, 15, 365-386.

GANDHI, N.R., MOLL, A., STURM, A.W., PAWINSKI, R., GOVENDER, T., LALLOO, U., ZELLER, K., ANDREWS, A., FRIEDLAND, G. 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet*, 368, 1575-1580.

GANGULY, N., SIDDIQUI, I., SHARMA, P. 2008. Role of *M. tuberculosis* RD-1 region encoded secretory proteins in protective response and virulence. *Tuberculosis*, 88, 510-517.

GARNIER, T., EIGLMEIER, K., CAMUS, JC., MEDINA, N., MANSOOR, H., PRYOR, M., DUTHOY, S., GRONDIN, S., LACROIX, C., MONSEMPE, C., SIMON, S., HARRIS, B., ATKIN, R., DOGGETT, J., MAYES, R., KEATING, L., WHEELER, P.R., PARKHILL, J., BARRELL, B.G., COLE, S.T., GORDON, S.V., HEWINSON, R.G. 2003. The complete genome sequence of *Mycobacterium bovis*. *PNAS*, 100(13), 7877-7882.

GELBER, R.H., BRENNAN, P.J., HUNTER, S.W., MUNN, M.W., MONSON, J.M., MURRAY, L.P., SIU, P., TSANG, M., ENGLEMEN, E.G., MOHAGHEGPOUR, N. 1990. Effective vaccination of mice against leprosy bacilli with subunits of *Mycobacterium leprae*. *Infection and Immunity*, 58(3), 711-718.

GILLERON, M., HIMOUDI, N., ADAM, O., CONSTANT, P., VENISSE, A., RIVIÈRE, M., PUZO, G.1997. *Mycobacterium smegmatis* phosphoinositols-glyceroarabinomannans. Structure and localization of alkali-labile and alkali-stable phosphoinositides. *Journal of Biological Chemistry*, 272(1), 117-124.

GEY VAN PITTIUS, N.C., GAMIELDIEN, J., HIDE, W., BROWN, G.D., SIEZEN, R.J., BEYERS, A.D. 2001. The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C gram-positive bacteria. *Genome Biology*, 2(10), 1-18.

GEY VAN PITTIUS, N.C., SAMPSON, S.L., LEE, H., KIM, T., VAN HELDEN, P.D., WARREN, R.M. 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. *BMC Evolutionary Biology*, 6(95), 1-31.

GOODPASTER, A.M., ROMICK-ROSENDALE, L.E., KENNEDY, M.A. 2010. Statistical significance analysis of nuclear magnetic resonance-based metabonomics data. *Analytical Biochemistry*, 401, 134–143.

GORDHAN, B.G. AND PARISH, T. 2001. Gene replacement using pretreated DNA. In *Mycobacterium tuberculosis* protocols. *Methods in Molecular Medicine*, 54, 77-92.

GUINN, K.M., HICKEY, M.J., MATHUR, S.K., ZAKEL, K.L., GROTZKE, J.E., LEWINSOHN, D.M., SMITH, S., SHERMAN, D.R. 2004. Individual RD-1 region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology*, 51(2), 359-370.

GUO, S., XUE, R., LI, Y., WANGA, S.M., REN, L., XU, J.J. 2012. The CFP10/ESAT6 complex of *Mycobacterium tuberculosis* may function as a regulator of macrophage cell death at different stages of tuberculosis infection. *Medical Hypotheses*, 78, 389-392.

HOFFMANN, C., LEIS, A., NIEDERWEIS, M., PLITZKO, J.M. AND ENGELHARDT, H. 2008. Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proceedings of the National Academy of Sciences USA*, 105, 3963-3967.

JONES, E.A. 2002. Ammonia, the GABA neurotransmitter system, and hepatic encephalopathy. *Metabolic Brain Disease*, 17(4), 275-281.

KALINOVA, B., JIROS, P., ZDAREK, J., WEN, X., HOSKOVEC, M. 2006. GC×GC-TOF MS technique—A new tool in identification of insect pheromones: Analysis of the persimmon bark borer sex pheromone gland. *Talanta*, 69, 542-547.

KAMIO, Y., ITOH, Y., TERAWAKI, Y., KUSANO, T. 1981. Cadaverine is covalently linked to peptidoglycan in *Selenomonas ruminantium*. *Journal of Bacteriology*, 145(1), 122–128.

KANANI, H., CHRYSANTHOPOULOS, P.K., AND KLAPA, M.I. (2008). Standardizing GC-MS metabolomics. *Journal of Chromotography*. 871, 191-201.

KAPSENBERG, M.L. 2003. Dendritic-cell control of pathogen-driven T-cell polarization. *Nature Reviews Immunology*, 3, 984-993.

KARIM, S.S.A., CHURCHYARD, G.J., KARIM, Q.A., LAWN, S.D. 2009. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *Lancet*, 374, 921-933.

KHAN, M.Y., KINSARA, A.J., OSOBA A.O., WALI, S., SAMMAN, Y., MEMISH, Z. 2001. Increasing resistance of *M. tuberculosis* to anti-TB drugs in Saudi Arabia. *International Journal of Antimicrobial Agent*s, 17, 415-418.

KIM, JS., CHOI, S.H., LEE, J.K. 2006. Lysine decarboxylase expression by *Vibrio vulnificus* is induced by SoxR in response to superoxide stress. *Journal of Bacteriology*, 188(24), 8586-8592.

KLEINNIJENHUIS, J., OOSTING, M., JOOSTEN, L.A.B., NETEA, M.G., VAN CREVEL, R. 2011. Innate immune recognition of *Mycobacterium tuberculosis*. *Clinical and Developmental Immunology*, Vol. 2011, 405310.

KNECHEL, N.A. 2009. Tuberculosis: Pathophysiology, clinical features, and diagnosis. *Critical Care Nurse*, 29, 34-43.

KOCH, R. 1882. Die aetiogie der Tuberculose. *Berl Klin Wochenschr*, 19, 221-230.

KUSANO, M., FUKUSHIMA. A., KOBAYASHI, M., HAYASHI, N., JONSSON, P., MORITZ, T., EBANA, K., SAITO, K. 2007. Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. *Journal of Chromatography*, 855, 71-79.

LEINHARDT, C. 2001. From exposure to disease: The role of environmental factors in suspectibility to and development of tuberculosis. *Epidemiologic Reviews*, 23(2), 288-301.

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. *The Journal of Infectious Diseases*, 187, 117-123.

LOOTS, DT., MEISSNER-ROLOFF, R.J., NEWTON-FOOT, M., GEY VAN PITTIUS, N.C. 2013. A metabolomics approach exploring the function of the type VII secretion system of *M. smegmatis*. *Metabolomics*, 9, 631-641.

LUEDEMANN, A., STRASSBURG, K. ERBAN, A., KOPKA, J. 2008. TagFinder for the quantitative analysis of gas chromatography—mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics*, 24(5), 732-737.

MACGURN, J.A., RAGHAVAN, S., Stanley, S.A., Cox, J.S. 2005. A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*. *Molecular Microbiology*, 57 (6), 1653-1663.

MATHEMA, B., KUREPINA, N.E., BIFANI, P.J., KREISWIRTH, B.N. 2006. Molecular epidemiology of tuberculosis: Current Insights. *Clinical Microbiology Reviews*, 19(4), 658-685.

MCLAUGHLIN, B., CHON, J.S., MACGURN, J.A., CARLSSON, F., CHENG, T.L., COX, J.S., BROWN, E.J. 2007. A *Mycobacterium* ESX-1–secreted virulence factor with unique requirements for export. *PLoS Pathogens*, 3(8), 1051-1061.

MCNERNEY, R., ZAGER, E.M. 2008. Multidrug-resistant tuberculosis. *BMC Infectious Diseases*, 8(10), 1-5.

MCNEIL, M., DAFFE, M., BRENNAN, P.J. 1990. Evidence of the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. *Journal of Biological Chemistry*, 265(30), 18200-18206.

MERRILL, M.H. 1931. Studies on carbon metabolism of organisms of the genus *Mycobacterium*: utilization of organic compounds in a synthetic medium. *Journal of Bacteriology*, 21(5), 361-374

MEISSNER-ROLOFF., R.J., KOEKEMOER, G., WARREN, R.M., LOOTS, DT. 2012. A metabolomics investigation of a hyper- and hypo-virulent phenotype of Beijing lineage *M. tuberculosis*. *Metabolomics*, 8, 1194-1203.

METZLER, D.E. 2003. *Biochemistry: the chemical reactions of living cells*. Sand Diego, USA: Elsevier Science.

MICKLINGHOFF, J.C., BREITINGER, K.J., SCHMIDT, M., GEFFERS, R., EIKMANNS, B.J., BANGE, FC. 2009. Role of the transcriptional regulator RamB (Rv0465c) in the control of the glyoxylate cycle in *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 191(23), 7260-7269.

MIYAKAWA, T., MATSUZAWA, H., MATSUHASHI, M., SUGINO, Y. 1972. Cell wall peptidoglycan mutants of *Escherichia coli* K-12: Existence of two clusters of genes, *mra* and *mrb*, for cell wall peptidoglycan biosynthesis. *Journal of Bacteriology*, 112(2), 950-958.

MORENS, D.M., FOLKERS, G.K., FAUCI, A.S. 2004. The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996), 242-249.

MUKHOPADHYAY, S., BALAJI, K.N. 2011. The PE and PPE proteins of *Mycobacterium tuberculosis*. *Tuberculosis*, 91, 441-447.

NANDA, T., DAS, M., TRIPATHY, K., TEJA Y, R. 2011. Metabolomics: The Future of systems biology. *Journal of Computer Science Systems Biology*, 13, 1-6.

NEELY, M.N., OLSON, E. R. 1996. Kinetics of expression of the *Escherichia coli* cad operon as a function of pH and lysine. *Journal of Bacteriology*, 178(18), 5522-5528.

OHOL, Y.M., GOETZ, D.H., CHAN, K., SHILOH, M.U., CRAIK, C.S., COX, J.S. 2010. *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host & Microbe*, 7, 210-220.

OJHA, A.K., BAUGHN, A.D., SAMBANDAN, D., HSU, T., TRIVELLI, X, GUERARDEL, Y., ALAHARI, A., KREMER, L., JACOBS JR., W.R., HATFULL, G.F. 2008. Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Molecular Microbiology*, 69(1), 164-174.

OLIVIER, I., LOOTS, DT. 2011. An overview of tuberculosis treatments and diagnostics. What role could metabolomics play?. *Journal of Cell and Tissue Research*, 11(1), 2655-2671.

PARISH, T., STOKER, N.G. 2000. Use of a flexible cassette method to generate a double unmarked *Mycobacterium tuberculosis tlyA plcABC* mutant by gene replacement. *Microbiology*, 146, 1969-1975.

PATTERSON, J.H., WALLER, R.F., JEEVARAJAH, D., BILLMAN-JACOBE, H. MCCONVILLE, M.J. 2003. Mannose metabolism is required for mycobacterial growth. *Journal of Biochemistry*, 372, 77-86.

PAVELKA JR, M.S., JACOBS JR, W.R. 1996. Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of *Mycobacterium smegmatis*. *Journal of Bacteriology*, 178(22), 6496-6507.

PILZ, R.B., WILLIS, R.C., BOSS, G.R. 1984. The influence of ribose 5-phosphate availability on purine synthesis of cultured human lymphoblasts and mitogen-stimulated lymphocytes. *The Journal of Biological Chemistry*, 269(5), 2927-2935.

PYM, A.S., BRODIN, P., BROSCH, R., HUERRE, M., COLE, S.T. 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Molecular Microbiology*, 46(3), 709-717.

PYM, A.S., BRODIN, P., MAJLESSI, L., BROSCH, R., DEMANGEL, C., WILLIAMS, A., GRIFFITHS, K.E., MARCHAL, G., LECLERC, C., COLE, S.T. 2003. Recombinant BCG exporting ESAT-6 confers enhances protection against tuberculosis. *Nature Medicine*, 9(5), 533-539.

RAGHAVAN, S., MANZANILLO, P., CHAN, K., DOVEY, C., COX, J.S. 2008. Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. *Nature*, 454, 717-722.

RAMAKRISHNAN, T., SURYANARAYANA MURTHY, P., GOPINATHAN, K.P. 1972. Intermediary metabolism of mycobacteria. *Bacteriological Reviews*, 36(1), 65-108.

RATLEDGE, C., DOVER, L.G. 2000. Iron metabolism in pathogenic bacteria. *Annual Review of Microbiology*, 54(1), 881-941.

REED, M.B., DOMENECH, P., MANCA, C., SU, H., BARCZAK, A.K., KREISWIRTH, B.N., KAPLAN, G., BARRY III, C.E. 2004. A glyco-lipid of hyper-virulent tuberculosis strains that inhibits the innate immune response. *Nature*, 431, 84-87.

RENSHAW, P.S., LIGHTBODY, K.L., VEVERKA, V., MUSKETT, F.W., KELLY, G., FRENKIEL, T.A., GORDON, S.V., HEWINSON, R.G., BURKE, B., NORMAN, J., WILLIAMSON, R.A., CARR, M.D. 2005. Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *The EMBO Journal*, 24, 2491-2498.

ROOS, A.K., ANDERSSON, C.E., BERGFORS, T., JACOBSSON, M., KARLEN, A., UNGE, T., JONES, T.A., MOWBRAY, S.L. 2004. *Mycobacterium tuberculosis* ribose-5-phosphate isomerase has a known fold, but a novel active site. *Journal of Molecular Biology*, 335, 799-809.

RUSSELL, D.G. 2011. *Mycobacterium tuberculosis* and the intimate discourse of a chronic infection. *Immunological Reviews*, 240, 253-268.

SAMARTZIDOU, H., DELCOUR, A.H. 1999. Excretion of endogenous cadaverine leads to a decrease in porin-mediated outer membrane permeability. *Journal of Bacteriology*, 181(3), 791-798.

SAMPSON, S.L. 2011. Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clinical and Developmental Immunology*, 2011, 497203.

SASSETTI, C.M. & RUBIN, E.J. 2003. Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences USA*, 100, 12989-12994.

SCHOEMAN, J.C., DU PREEZ, I., LOOTS, DT. 2012. A comparison of four sputum pre-extraction preparation methods for identifying and characterising *Mycobacterium tuberculosis* using GCxGC-TOFMS metabolomics. *Journal of Microbiological Methods*, 91(2), 301-311.

SERAFINI, A., BOLDRIN, F., PALÙ, G., MANGANELLI, R. 2009. Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc. *Journal of Bacteriology*, 191(20), 6340-6344.

SETH, A., CONNELL, N.C. 2000. Amino acid transport and metabolism in mycobacteria: Cloning, interruption, and characterization of an L-arginine/γ-aminobutyric acid permease in *Mycobacterium bovis* BCG. *Journal of Bacteriology*, 182(4), 919-927.

SIEGRIST, M.S., UNNIKRISHNAN, M., MCCONNELL, M.J., BOROWSKY, M., CHENG, T.Y., SIDDIQI, N., FORTUNE, S.M., MOODY, D.B, RUBIN, E.J. 2009. Mycobacterial ESX-3 is required for mycobactin-mediated iron acquisition. *PNAS*, 106(44), 18792-18797.

SIMEONE, R., BOTTAI, D., BROSCH, R. 2009. ESX/type VII secretion systems and their role in host–pathogen interaction. *Current Opinion in Microbiology*, 12, 4-10.

TANEJA, N.K., DHINGRA, S., MITTAL, A., NARESH, M., TYAGI, J.S. 2010. *Mycobacterium tuberculosis* transcriptional adaptation, growth arrest and dormancy phenotype development is triggered by Vitamin C. *PLoS Pathogens*, 5(5), 1-13.

TEKAIA, F., GORDON, S.V., GARNIER, T., BROSCH, R., BARRELL., B.G., COLE, S.T. 1999. Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tubercle and Lung Disease*, 79(6), 329-342.

THIERRY, D., CAVE,M.D., EISENACH, K.D., CRAWFORD, J.T., BATES, J.H., GICQUEL, B., GUESDON, J.L. 1989. *IS*6110, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Research*, 18(1), 188.

T'KINDT, R., MORREEL, K., DEFORCE, D., BOERJAN, W., VAN BOCXLAER, J. 2009. Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: Repeatability and sample pre-treatment. *Journal of Chromatography B*, 877, 3572-3580.

TOBIN, D.M, RAMAKRISHNAN, L. 2008. Comparative pathogenesis of *Mycobacterium marinum* and *Mycobacterium tuberculosis*. *Cellular Microbiology*, 10(5), 1027-1039.

TUNDUP, S., PATHAK, N., RAMANADHAM, M., MUKHOPADHYAY, S., MURTHY, K.J.R., EHTESHAM, N.Z., HASNAIN, S.E. 2008. The co-operonic PE25/PPE41 protein complex of *Mycobacterium tuberculosis* elicits increased humoral and cell mediated immune response. *PLoS ONE*, 3(10), e3586.

UMBARGER, H. E. 1978. Amino acid biosynthesis and its regulation. *Annual Review Biochemistry*, 47, 532-606.

UPTON, A.M., MCKINNEY, J.D. 2007. Role of the methylcitrate cycle in propionate metabolism and detoxification in *Mycobacterium smegmatis*. *Microbiology*, 153, 3973-3982.

VAN DER WERF. M.J., OVERKAMP, K.M., MUILWIJK, B., COULIER, L., HANKEMEIER, T. 2007. Microbial metabolomics: Toward a platform with full metabolome coverage. *Analytical Biochemistry*, 370, 17-25.

VERSCHOOR, J.A., BAIRD, M.S., GROOTEN, J. 2012. Towards understanding the functional diversity of cell wall mycolic acids of *Mycobacterium tuberculosis*. *Progress in Lipid Research*, 51, 325-339.

VOSKUIL, M.I., SCHNAPPINGER, D., RUTHERFORD, R., LIU, Y., SCHOOLNIK, G.K. 2004. Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis*, 84, 256-262.

WARREN, R., GEY VAN PITTIUS, N.C., BARNARD, M., HESSELING, M., ENGELKE, E., DE KOCK, M., GUTIERREZ, M.C., CHEGE, G.K., VICTOR, T.C., HOAL, E.G., VAN HELDEN, P.D. 2006. Differentiation of *Mycobacterium tuberculosis* complex by PCR amplification of genomic regions of difference. *The International Journal of Tuberculosis and Lung Disease*, 10(7), 818-822.

WECKWERTH, W., MORGENTHAL, K. 2005. Metabolomics: from pattern recognition to biological interpretation. *Drug Discovery Today:Targets*, 10(22), 1551-1558.

WELTHAGEN, W., SHELLIE, R.A., SPRANGER, J., RISTOW, M., ZIMMERMANN, R., FIEHNG, O. 2004. Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GCxGC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice. *Metabolomics*, 1(1), 65-73.

WOOD, A.W., BECKER, M.A., SEEGMILLER, J.E. 1973. Purine nucleotide synthesis in lymphoblasts cultured from normal subjects and a patient with lesch-nyhan syndrome. *Biochemical Genetics*, 9(3), 261-274.

WORLD HEALTH ORGANIZATION (WHO). 2010. Global tuberculosis control 2010. World Health Organization, Geneva, Switzerland.

WORLD HEALTH ORGANIZATION (WHO) 2011. World Health Organization 2011/2012 tuberculosis global facts. Geneva, Switzerland.

WORLD HEALTH ORGANIZATION (WHO). 2013. Global tuberculosis control 2010. World Health Organization, Geneva, Switzerland.

XIA, J., MANDAL, R., SINELNIKOV, I.V., BROADHURST, D. AND WISHART, D.S. 2012. MetaboAnalyst 2.0—a comprehensive serverfor metabolomic data analysis. *Nucleic Acids Research*, 40, 10-22.

YUAN, Y., CRANE, D.C., MUSSER, J.M., SREEVATSAN, S., BARRY III, C.E. 1997. MMAS-1, the branch point between cis- and trans-cyclopropane-containing oxygenated mycolates in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, 272(15), 10041-10049.

ZHANG, N., TORRELLES, J., MCNEIL, M., ESCUYER, V. E., KHOO, K. H., BRENNAN, P. J. & CHATTERJEE, D. 2003. The Emb proteins of mycobacteria direct arabinosylation of lipoarabinomannan and arabinogalactan via an N-terminal recognition region and a C terminal synthetic region. *Molecular Microbiology*, 50, 69-76.

APPENDIX A

The use of functional genomics in conjunction with metabolomics for *Mycobacterium tuberculosis* research

Conrad C. Swanepoel, Du Toit Loots

Hindawi Publishing Corporation
Disease Markers
Volume 2014, Article ID 124218, 12 pages
http://dx.doi.org/10.1155/2014/124218

Review Article

The Use of Functional Genomics in Conjunction with Metabolomics for *Mycobacterium tuberculosis* Research

Conrad C. Swanepoel and Du Toit Loots

School for Physical and Chemical Sciences, Centre for Human Metabonomics, North-West University, Private Bag x6001, Box 269, Potchefstroom 2531, South Africa

Correspondence should be addressed to Du Toit Loots; dutoit.loots@nwu.ac.za

Received 20 August 2013; Revised 3 December 2013; Accepted 14 February 2014; Published 18 March 2014

Academic Editor: Andreas Pich

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is a fatal infectious disease, resulting in 1.4 million deaths globally per annum. Over the past three decades, genomic studies have been conducted in an attempt to elucidate the functionality of the genome of the pathogen. However, many aspects of this complex genome remain largely unexplored, as approaches like genomics, proteomics, and transcriptomics have failed to characterize them successfully. In turn, metabolomics, which is relatively new to the "omics" revolution, has shown great potential for investigating biological systems or their modifications. Furthermore, when these data are interpreted in combination with previously acquired genomics, proteomics and transcriptomics data, using what is termed a systems biology approach, a more holistic understanding of these systems can be achieved. In this review we discuss how metabolomics has contributed so far to characterizing TB, with emphasis on the resulting improved elucidation of *M. tuberculosis* in terms of (1) metabolism, (2) growth and replication, (3) pathogenicity, and (4) drug resistance, from the perspective of systems biology.

1. Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, is considered to be the world's second most deadly infectious disease, after that for which the human immunodeficiency virus (HIV) is responsible. The pathogen was discovered in 1882 by the German physician Robert Koch [1, 2]. The World Health Organization (WHO) recently reported that one-third of the global population is infected with *M. tuberculosis* [2] and in 2010 documented 8.8 million new cases of TB, 650 000 of which were infected with multidrug-resistant (MDR) strains, contributing to the 1.4 million deaths reported globally (equivalent to 3 800 deaths a day) for that year. Developing countries are by far the worst affected by this epidemic and account for 95% of the reported global deaths due to this disease [3]. Additionally, 25% of the individuals who succumb as a result of this epidemic are reported to be coinfected with HIV, with Europe and Africa exhibiting the highest HIV coinfection rate [4]. Although the mortality rates of these coinfected individuals have declined since the advent of antiretroviral therapy, MDR-TB still contributes greatly to the high mortality rates in the rural areas of especially southern Africa, as a direct consequence of these patients failing to comply with the treatment regimens [5, 6]. These findings are alarming, as TB is considered a curable disease and those affected are capable of a full recovery [7, 8] after the successful completion of the correct treatment regimens [9].

By applying different research methodologies (such as genomics, proteomics, transcriptomics, and lipidomics), a large body of knowledge has been generated and applied to new diagnostic and treatment protocols for infectious diseases, including TB [10, 11]. The latest addition to these "omics" methodologies, metabolomics, can be defined as the unbiased identification and quantification of all metabolites present in a biological sample (the metabolome) at a given time, using highly sensitive and selective analytical techniques [12], followed by the interpretation and visualization of the data generated via biostatistics [13]. Metabolic profiling, defined as the "detailed metabolome

analysis requiring highly specialized analytical techniques and accurate concentration determination for sample classification" [14], and metabonomics, "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification" [15], are key terms used to expand the definition of metabolomics [12]. Alterations in the metabolic profile of an organism can be directly linked to the corresponding genes in its genome, as Raamsdonk et al. [16] illustrated with the Functional Analysis by Coresponses in Yeast (FANCY) approach. Their method proves the principle that any genetic modification of an organism's functional gene will lead to alterations in its metabolite profiles [17]. Moreover, the FANCY approach provides additional information towards elucidating gene function by comparing the metabolic profiles of different organism strains with known or unknown gene modifications and processing the metabolite data generated using multivariate statistical analyses, in particular principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA), in order to identify those metabolite markers best accounting for the differences between strains [17, 18]. When used in conjunction with enzymology and proteomics, this approach can assist in deciphering enzyme/protein functionality by comparing the metabolic profiles of a wild-type strain to those of an identical strain with a deletion in the genome corresponding to a gene coding for the specific enzyme/protein involved [19]. The advantages of using metabolomics for such comparative investigations are that (1) it provides an excellent representation of the cellular metabolite state of an organism at the time of sampling, and of the influence of any perturbation induced by the environment, altered genes, or disease [20, 21]; (2) metabolites can be identified with a high degree of certainty and there are fewer metabolite types/classes that exist than there are genes or proteins [22]; and (3) thanks to the excellent analytical techniques currently available, accurate characterization and quantification of the metabolome can be achieved, using only small sample amounts, with minimum sample preparation beforehand [23, 24]. The most commonly used analytical techniques for metabolome/metabolic profile analyses are nuclear magnetic resonance (NMR) and mass spectrometry (MS) [23]. NMR provides a unique advantage in that it is nondestructive of the sample being analyzed and is the preferred technique when characterizing unknown compounds [25]. NMR spectra, however, are considered rather complex; when investigating metabolic profiles containing a great diversity of metabolites with many compound classes and concentrations [26], these criteria are judged to be less useful. For these applications, MS is the preferred approach, as it is less time consuming and simplifies detection. A further advantage of using MS for such applications is that it can be used to elucidate and monitor the metabolic pathway of a substrate and its intermediates active in metabolism, by labeling them with stable ^{13}C isotopes.

The addition of chromatographic separation of the compounds in a complex sample, prior to MS detection [27], greatly improves the sensitivity and accuracy of such analyses. Various "hyphenated" MS approaches exist for such applications and include those combined with gas and liquid chromatography, enabling the separation of the compounds in a mixture on the basis of such chemical properties as their solubility, volatility, and mass [23, 24]. Deconvolution of overlapping spectral peaks, in combination with accurate peak identification, further improves metabolome analyses and subsequently contributes to the better elucidation of the metabolic profiles [24].

Interpretation of the metabolomics data, in combination with those generated using other "omics" approaches, collectively defined as a systems biology research approach, provides a holistic view of the organism or cell under investigation [20, 28]. All the information generated from such research disciplines is used in combination to explain the phenomenon or perturbation of interest. In this review we discuss the previously reported functional genomics research conducted, in conjunction with recently generated metabolomics data, in an attempt to improve the characterization of *M. tuberculosis*. A detailed overview of the gene encoding entities contributing to (1) metabolism, (2) growth and replication, (3) pathogenicity, and (4) drug resistance of this microbe will be considered relevant. We then highlight those areas not yet fully understood and the potential role metabolomics can play in elucidating these.

2. Interpreting Functional Genomics in Conjunction with Metabolomics for Improved Characterization of *Mycobacterium tuberculosis*

The *M. tuberculosis* complex can be separated into a variety of subspecies, namely, *M. tuberculosis, M. canetti, M. africanum*, *M. microti, M. caprae, M pinnipedii, M. bovis,* and *M. bovis* Bacille Calmette Guerin (BCG) [29], all of which can potentially cause a TB-like disease state in mammals. Interestingly, these subspecies represent 0.01% genetic diversity in terms of their respective genomes [30]. Factors contributing to the minor genetic variation observed between these mycobacterial species include (1) differences in the copy number and location of an insertion sequence (IS) specific for *M. tuberculosis* and *M. bovis* (IS6*110*); (2) variations in a subgroup of the Pro-Glu (PE) family of proteins, namely, polymorphic GC-rich sequences (PGRS); and (3) the variation observed in short DNA sequences, characterized as the region of difference (RD) [31, 32]. The genome of *M. tuberculosis* exhibits more than 4 million nucleotide base pairs (bp), with guanine and cytosine the major components (65.6%), in addition to over 4 000 genes [33], 82 of which were added following the reannotation of the genome by Camus [34].

Considering these differences, one of the first TB metabolomics investigations conducted by the Laboratory of Infectious Disease Metabolomics, Centre for Human Metabonomics, North-West University, South Africa, was an initial test of the functionality of this approach for detecting minor variations in the metabolomes of different *Mycobacterium* species, including *M. tuberculosis*, *M. kansasii*, *M. avium*, and *M. bovis* BCG, on the basis of

the small genetic differences previously described between these species. Olivier and Loots [35] subsequently extracted the lipidome of these species and after successfully processing the data generated by gas chromatography-mass spectrometry (GC-MS), determined that these species could be differentiated from one another on the basis of their characteristic metabolite signatures, using a combination of 12 selected metabolite markers. More importantly, these markers were subsequently used to construct a diagnostic model able to identify these *Mycobacterium* species in sputum collected from patients, further highlighting the potential of metabolomics for diagnosis. These results not only showed that the minor genetic variations observed result in a significantly altered metabolite profile but also that metabolomics is sensitive enough to detect these changes. Later studies using this approach included metabolomics investigations of the functionality changes associated with virulence (ESX-1) and growth (ESX-3) [36], factors contributing to *M. tuberculosis* drug resistance [37], hyper- versus hypo-virulence [38], and adaptations of the host to *M. tuberculosis* and vice versa [39]. These topics are discussed in greater detail below in addition to some of the significant contributions to TB metabolomics by other research groups.

2.1. Systems Biology Related to General *M. tuberculosis* Metabolism

2.1.1. The Tricarboxylic Acid, Glyoxylate, and Citramalic Acid Cycles. Although some metabolic pathways in *M. tuberculosis* differ from those of other bacterial species, the microbe's genome has all the elements and pathways necessary to synthesize essential biomolecules (amino acids, vitamins, and enzyme cofactors) necessary for replication and growth [30, 33]. Evidence from various genomic studies showed that the H37Rv strain of *M. tuberculosis* is capable of metabolizing carbohydrates, hydrocarbons, alcohols, ketones, and carboxylic acids [33] via the glycolysis pathway, the pentose phosphate pathway, the tricarboxylic acid (TCA) cycle, and the glyoxylate cycle, respectively [33, 40]. Subsequently, Tian et al. [41] identified a new TCA cycle in *M. tuberculosis* using a biomedical metabolomics approach [42]. In their study they indicated that *M. tuberculosis* shows no α-ketogluturate dehydrogenase (KDH) activity and subsequently proved the existence of a split TCA cycle, with or without interconnecting pathways. This split originates after the decarboxylation of α-ketogluturate to the conventional TCA intermediate, succinate, via (1) succinic semialdehyde (SSA) or (2) glutamate's conversion to 4-aminobutyrate (GABA) and then SSA [41].

As mentioned above, ^{13}C-labeled metabolites/substrates permit the metabolic pathway to be elucidated and the metabolism of a substrate and its intermediates to be monitored [19, 43]. When used in combination with different genomic techniques, this approach may serve to clarify better the functionality of uncharacterized genes. Such a multifaceted approach was used successfully to determine the functionality of Rv1248c (*sucA*), which was originally believed to be the E1 component of the KDH complex. Prior to this, various genomic studies reported a lack of the E2 component of KDH, namely, dihydrolipoamide succinyltransferase. This compound was thought to render this enzyme complex inactive in *M. tuberculosis* and was subsequently proposed as an alternative working model for Rv1248c, which was believed to be responsible for the nonoxidative decarboxylation of α-ketogluturate to SSA. The enzymatic conversion of substrate to product by the recombinant Rv1248c was monitored using ^{1}H-NMR combined with activity-based metabolite profiling [43, 44]. Subsequently, Rv1248c was shown to be a carboligase enzyme, responsible for catalyzing the synthesis of 2-hydroxy-3-oxoadipate (HOA), by condensing the activated aldehyde group of α-ketoglutarate with glyoxylate [43, 45]. Of even greater importance was the realization that this approach could be used to elucidate the functionalities of other enzymes and their associated genes. Subsequently, the functionality of phosphoenol-pyruvate carboxykinase (PEPCK), encoded by *pckA* [46, 47], which was previously believed to be responsible for catalyzing the interconversion of oxaloacetate (OAA) and phosphoenol-pyruvate (PEP), was determined. Examining *M. bovis* BCG, using metabolite profiling of ^{13}C-labeled acetate carbon flux in wild-type and the *pckA*-deleted strains, revealed that this enzyme is responsible for only the unidirectional conversion of OAA to PEP [47]. This proved that although the TCA cycle produced pyruvate, it was not converted to PEP [43] as was previously thought. These findings further illustrated a lack of pyruvate phosphate dikinase (*ppdK*) activity in these organisms [47] and led to the conclusion that PEPCK was involved in gluconeogenesis in *M. tuberculosis* [48]. Further application of this "omics" approach revealed that during *in vitro* growth of *M. tuberculosis*, carbon flow between known intermediates of the TCA cycle, α-ketoglutarate, and succinate was intermittent, further confirming the lack of a functional KDH enzyme and that this organism operates via a bifurcated TCA cycle [43, 49]. Olszewski et al. [50] subsequently illustrated the same phenomenon in *Plasmodium falciparum*, the causative agent of malaria [51], using a similar metabolomics approach. They showed that this protozoan parasite also has a bifurcated TCA cycle at α-ketoglutarate, with the resultant two pathways simplifying the degradation of glutamate and/or glutamine into two carbon units [43].

Whereas the early genomics methods identified various genes related to the TCA cycle, their exact functionality was only truly determined after metabolic examinations, confirming how crucial metabolomics and the interpretation of these data using a systems biology approach can be.

2.1.2. Lipid Metabolism. Mycobacterium tuberculosis possesses a great number of lipophilic molecules, ranging from common fatty acids to glycolipids [52] and very-long-chain molecules (such as mycolic acids) [33, 53]. Although many enzymes have been identified as being directly responsible for the biosynthesis of these lipids, they are greatly outnumbered by those responsible for fatty acid oxidation in these organisms. Previous genomics studies revealed that the genome of *M. tuberculosis* encodes more than 250 enzymes related to fatty acid oxidation, which is remarkable considering

that the corresponding number for *E. coli* is only 50 [33]. In *M. tuberculosis*, lipids are generally metabolized to yield energy essential for mycobacterial growth and replication. Many, however, are also used in the synthesis of various components directly involved with the mycobacterial cell wall, which in turn plays a particular role in mycobacterial virulence, as will be explained in greater detail in what follows. The majority of complexes and enzymes required for lipid degradation, are encoded for by a vast amount of these genes and include (1) 36 acyl-CoA synthases; (2) 36 acyl-CoA dehydrogenases (capable of catalyzing the initial step in fatty acid degradation); (3) 21 enzymes associated with the enoyl-CoA hydratase/isomerase superfamily; (4) six 3-hydroxyacyl-CoA dehydrogenases [54] (responsible for the conversion of 3-hydroxy fatty acids to 3-keto-fatty acids) [33, 55]; and (5) six acetyl-CoA C-acetyltransferases, completing β-oxidation, eventually to yield two acetyl-CoA molecules [54].

Complementary to these findings is the more recently acquired metabolite information associated with these genes. More specifically, various predictions regarding the associated carboxylase systems have been made from this perspective but once again with considerable uncertainty [34]. Pyruvate carboxylase (PYC), encoded for by *pycA*, is thought to be necessary for the interconversion of oxaloacetate and pyruvate. The latter in turn can be metabolized to yield acetyl-CoA, the same end product following the catabolism of fatty acids via β-oxidation [45, 56]. Acyl-CoA carboxylase (ACC), on the other hand, has been shown to be responsible for the carboxylation of more specific short-chain acyl-CoA substrates to their respective lipid products, subsequently serving as substrates in the fatty acid synthase (FAS) systems and polyketide synthesis, ultimately producing mycolic acids [57]. Interestingly, both PYC and ACC require biotin as cofactor, with this vitamin donating a carboxyl anion to the very specific organic acid substrate [45, 57]. There is limited information on PYC in the metabolomics literature, but ACC has been the subject of more metabolic research due to it potentially serving as an antibiotic drug target, which Rabinowitz and coworkers [58] illustrated using isotope tracer metabolites and a metabolomics research approach.

Additional metabolomics-related approaches by de Carvalho et al. [19] pertaining to these genes associated with fatty acids revealed compartmentalized cocatabolism of carbon substrates in *M. tuberculosis* and a description of how various carbohydrates and fatty acids can be channeled simultaneously to their respective metabolic fates. They subsequently provided an entirely new understanding of this previously uncharacterized metabolic network. Using ^{13}C labeling of the metabolites, these authors confirmed that *M. tuberculosis* is not capable of diauxic growth and that the pathogen has the capacity to catabolize multiple carbon sources simultaneously, ensuring optimal monophasic growth. de Carvalho et al. [19] further concluded that each individual isotopically labeled carbon source, added to the growth media of the *M. tuberculosis* H37Rv cells, was successfully and differentially catabolized through the glycolytic, pentose phosphate or TCA pathways.

Further genomic studies illustrated two related enzyme systems as responsible for synthesizing essential fatty acids in mycobacteria, namely, (1) fatty acid synthase I (FASI) [59] and (2) fatty acid synthase II (FASII) [33, 60]. In contrast to the mycobacterial FASII system, FASI exhibits a multidomain enzyme necessary for fatty acid biosynthesis [61]. Interestingly, mycobacteria are the first prokaryote organisms in which both of these systems were found to be fully functional [61]. Genes encoding acyl carrier proteins (ACPs) were also identified in the genome and these proteins were shown to function by transporting all pathway intermediates between the active site centers of enzymes involved in the FASII biosynthetic system [59, 61]. The two systems function in unison: FASI synthesizes precursors necessary for mycolic acids and other complex lipids present in the cell wall [33, 62]; and FASII elongates the FASI products to form meromycolate precursors, which are chemically altered eventually to form mycolic acids [61–63]. As part of this elongation process, β-ketoacyl-ACP synthase (*KasA*) catalyzes the condensation of the two carbon units from malonyl-ACP to a rapidly growing chain characterized as acyl-ACPs [64]. Interestingly, in contrast to the metabolism of mycobacteria, the host FASII differs from FASI in that the latter uses a single multifunctional enzyme-ACP complex to initiate fatty acid biosynthesis [65]. Moreover, the characteristic differences between the structural arrangement of bacterial and mammalian synthases provide an attractive and unique opportunity for the development of antibacterial agents specifically targeting the bacterial FASII system. In addition, as these lipids make up approximately 60% of the mycobacterial cell wall [55], they are thought to play an important role in the viability and virulence of the organism, making them attractive targets in TB drug discovery [66].

Recently, du Preez and Loots [37] used a metabolomics approach to compare the lipid metabolomes of two genetically different *rpoB* mutant *M. tuberculosis* strains (S552L and S531L) to that of an isogenic *M. tuberculosis* wild-type strain, in an effort to characterize rifampicin resistance more fully. These authors subsequently detected reduced concentrations of various 10-methyl branched-chain fatty acids and the corresponding cell wall lipids in the metabolome of the mutant strains [37]. Their results suggested an inhibition of FASII and S-adenosylmethionine (SAM), which was confirmed by detection of elevated substrate concentrations in these enzyme pathways. Additionally, a decrease in various straight-chain fatty acids implied their increased utilization as an alternative energy source during the reduced growth rates associated with these mutants. The findings of this metabolomics investigation confirmed other studies indicating that the resistance of *M. tuberculosis* to rifampicin, due to a mutation in the β-subunit of the *rpoB* gene of RNA polymerase, comes at an overall fitness cost. These mutant organisms show reduced growth, replication, and viability, accompanied by various metabolic modifications in order to survive this state. This study was the first of its kind to indicate that the *M. tuberculosis rpoB* mutant strain is accompanied by an altered fatty acid metabolism and further illustrated the potential of metabolomics as a valuable genomics tool, able to

elucidate gene functionality. These findings will be discussed further under systems biology related to drug resistance.

Lastly, genes responsible for encoding 9 polyketide synthase enzymes (*pks*) were also identified in the *M. tuberculosis* genome [60]. Polyketides are multidomain proteins essential for the construction of complex lipids and various metabolites associated with the mycobacterial cell envelope [67, 68]. Interestingly, the *pks* gene complexis located directly upstream of *mas*, and both their respective products—phthiocerol and mycocerosic acid—are used to produce another necessary cell wall component, phthiocerol dimycocerosate (PDIM), which is also strongly associated with mycobacterial virulence [33, 52]. Bisson and associates [69] illustrated a significant upregulation of the polyketide synthase genes, *ppsA-ppsE,* and *drrA* (Rv2936), responsible for the transport of PDIM over the mycobacterial cell envelope in Beijing and Haarlem *rpoB* mutant *M. tuberculosis* strains, using a combined metabolomics and proteomics approach. Although intact PDIM was not detected in either of the *rpoB* mutant strains used by Bisson and colleagues [69], elevated concentrations of the various precursors to this were recorded [67]. It was further hypothesized by Cole et al. [33] that these enzymes elongate FASI-CoA primers to produce the associated fatty acids. However, this could not be confirmed as the related metabolite data are inadequate for the purpose [33].

2.2. Systems Biology Related to Growth, Replication, and Survival. Five copies of the ESX gene cluster (ESX-1–ESX-5), collectively termed the ESAT-6 gene cluster region, have been identified in the genome of *M. tuberculosis* [70]. It is believed that six genes located within the cluster encoding the essential components of Type VII secretion (T7S) system are responsible for the transport of virulence and growth associated proteins across the highly hydrophobic cell wall of the organism [68, 71]. These five copies function independently from one another. Whereas ESX-1 and ESX-5 have been linked to virulence of *M. tuberculosis*, ESX-3 has been implicated in the uptake of iron and zinc by the pathogen [36, 72, 73]. The nine genes associated with the latter, Rv0282–Rv0292 [36], are under the transcriptional regulation of IdeR (iron-dependent repressor) and Zur (zinc-uptake regulator). Their subsequent levels of regulation are determined by the immediate concentrations of zinc and iron, respectively [74]. The exact role of ESX-3 in the T7S system, as well as its role in divalent cation homeostasis and growth, however, remains unclear. With regard to the different ESX gene clusters, ESX-3 is considered the only region essential for the *in vitro* growth of *M. tuberculosis* [75]. In addition, intracellular iron acts as a cofactor for over 30 enzymes in this pathogen and together with zinc fulfills critical roles in the activation of enzymes involved in amino acid biosynthesis in particular [76]. Subsequent to these findings, Loots and coworkers [36] used a metabolomics approach to explore the functionality of the ESX-3 gene cluster, using the nonpathogenic *M. smegmatis* strain as a model. This was achieved by comparing the metabolite profiles of an ESX-3 knock-out with those of a wild-type parent strain. Loots and colleagues [36] anticipated both iron and zinc starvation in the ESX-3 knock-out strain, due to their previously described role in the homeostasis of these essential divalent ions. This was the case for zinc, with the metabolite products synthesized using zinc-dependent enzymes detected in greatly reduced concentrations in the ESX-3 knock-out strain. Interestingly, iron illustrated the completely opposite behavior. This was ascribed to upregulation of other iron acquisition genes of *M. smegmatis,* overcompensating for the loss of ESX-3. This study of Loots et al. [36] successfully combined metabolomics and functional genomics and led to a better understanding of the role of iron and zinc in the growth of *M. smegmatis* and the functionality of their associated genes.

Various environmental and metabolic challenges threaten the viability of *M. tuberculosis*. This organism has accordingly developed an extensive array of regulatory factors to adapt and survive [33]. Following the discovery of 3 new genes in 2002 [34], the genome is now known to exhibit 11 pairs of genes related to the sensor histidine kinases and their associated response regulators [33]. The regulatory proteins present in the genome of the H37Rv strain are now known to exceed 100, with 13 sigma factors being responsible for supervising gene expression at the level of transcription initiation [33]. These transmembrane proteins have been linked to virulence in some fungal strains identified in immunocompromised individuals [77]. Also identified in the genome of *M. tuberculosis* are genes named *pknA–pknL*, thought to be responsible for coding 11 serine/threonine kinases (STPKs) [78, 79], known to act as sensors to environmental signals that thereby regulate developmental changes and host-pathogen interactions [80]. Further discovery revealed *pknA* and *pknB* to be upregulated during mycobacterial growth, and both form part of an operon, encoding genes necessary for the control of cell shape and ultimately cell wall synthesis [78]. Interestingly, within the *Mycobacterium* genus, the number of these kinase enzymes varies between 4 and 24, depending on the species [80], with *M. smegmatis* exhibiting no homologue proteins encoded for by the *pknF* and *pknG* genes [81]. Furthermore, these enzymes and the regulatory proteins mentioned above are involved in regulating different stress responses and control mycobacterial growth and pathogenicity, as will be discussed below.

Although not directly working on these *pkn* genes specifically, du Preez and Loots [37] also investigated bacteria-host interactions and adaptions using a metabolomics research approach. They reported for the first time the existence of a citramalic acid cycle in *M. tuberculosis*. They speculated, on the basis of this cycle's function in *Rhodospirillum rubrum*,that *M. tuberculosis* uses this as an additional means for generating carbon intermediates for energy production in the TCA cycle, as a means of surviving attack by the host. du Preez and Loots [37] also detected elevated concentrations of GABA and subsequently the importance of the glutamine - GABA shunt as a major source of carbon intermediates in the TCA cycle via succinic acid. Additionally, their metabolomics investigation detected adaptions of the host to the pathogen, which included a more than 10-fold elevation of various neurotransmitters, including norepinephrine, and postulated this as an additional mechanism for hydrogen peroxide

synthesis via glucose autooxidation, as a means of eliminating the bacteria [37].

2.3. Systems Biology Related to Mycobacterial Pathogenicity. Also identified in the genome of *M. tuberculosis* are many genes coding for proteins strongly associated with virulence and pathogenicity. The identification of these genes stimulated further research regarding their respective products and functioning and considered their influence on both metabolism and their role in pathogenicity. Cole et al. [33] deduced that approximately 7.1% of the total coding capacity of the *M. tuberculosis* genome is devoted to the 169 PE and Pro-Pro-Glu (PPE) proteins, which are unique to mycobacteria. Their functions and location in the cell, however, are unknown [82]. Their names were derived from the Pro-Glu and Pro-Pro-Glu motifs identified near the N-terminus of these macromolecules. Due to their variation, the PE family of proteins are assigned to three groups, with most of the proteins belonging to the subclass PGRS [33, 83], whereas the majority of the PPE proteins exhibit major polymorphic tandem repeats (MPTR) [33, 82, 83]. Although not yet confirmed, it has been widely postulated that members of this extensive family are associated with immune evasion, antigenic variation, and virulence [84]. Additionally, some are thought to function as immunodominant proteins [82] or have been linked with the mycobacterial cell wall, as necessary entities for lipolysis and lipogenesis of host macrophages [33] following infection, subsequently promoting replication of the tubercle bacilli. Furthermore, various members of the Rv2123 family have been identified as being upregulated during conditions of depleted iron [82], whereas others were confirmed to play a role in disease pathogenesis by the upregulation of PPE41 [85]. Furthermore, other PE (Rv3872/PE35) and PPE (Rv1807/PPE31 and Rv3873/PPE68) proteins are thought to be necessary for the *in vivo* growth of *M. tuberculosis,* following infection of the host [82]. A further 3 PE (Rv0285/PE5, Rv0335c/PE6, and Rv1169c/PE11) and 5 PPE (Rv0286/PPE4, Rv0755c/PPE12, Rv1753c/PPE2, Rv3135/PPE50, and Rv3343c/PPE54) proteins were identified to be essential for the *in vitro* growth of the organism [75]. Interestingly, Rv3872 (PE35) and Rv3873 (PPE68) are encoded by the ESX-1 gene cluster region, strongly linked to the microbe's virulence and pathogenicity [82, 83]. As most of the PE and PPE proteins are encoded for by ESX-5, and these proteins are also related to pathogenicity, the result was not unexpected because no ESX-5 region was found to be present in the genome of nonpathogenic *Mycobacterium* species (such as *M. smegmatis* and *M. vanbaalenii*) [82]. Despite this, however, the fact that the virulent *M. bovis* strain exhibits a fully functional ESX-1 and ESX-5 gene cluster, gene alterations within the ESX-1 gene cluster region of this organism led to the construction of the *M. bovis* BCG vaccine strain, confirming that both these two gene clusters are required for pathogenicity and virulence [86]. Further research regarding the functionality, location, and eventual secretion of all PE and PPE proteins is therefore highly desirable, as these proteins are undoubtedly important to pathogenicity. Elucidating their exact function is considered essential for developing better vaccination and treatment strategies in the fight against the growing TB epidemic.

In the light of this, Meissner-Roloff and associates [38] compared the metabolomes of a hypo- and hypervirulent Beijing *M. tuberculosis* strain and subsequently identified a reduction in various metabolite markers in the relatively hypervirulent strain. Glutamine, alanine, and glycine, all precursors of glutamic acid, are characterized as the fundamental building blocks of the PGRS subfamily. The reduced concentration of these amino acids detected in the hypervirulent strain was ascribed to an increase in PGRS formation and hence to the associated virulence. Further confirmation for this was the reduced concentrations of alanine and glycine in the same strain, due to their role as a dipeptide (glycylalanine) in PGRS protein formation. These results not only confirmed the postulated role of PE and PPE proteins in virulence and the altered metabolism associated with them, but the researchers also identified a further 33 metabolite markers and thereby explained how an altered metabolic rate, varying metabolism related to growth and replication, cell wall synthesis, and antioxidant capacity, relates to the increased virulence associated with these strains [38]. Briefly, the reductions in various primary energy substrates—such as glucose, galactose, mannose, mannonic acid, myo-inositol, glutamine and glycerol—in the hypervirulent strain suggests an increased utilization of these compounds due to an elevated metabolic and growth rate in this organism in addition to the increased use of these and other structural metabolites—such as glycerol, galactofuranoside, arabinofuranose, cadaverine, alanine-glutamic acid, aspartic acid, and lysine—for cell wall synthesis [87, 88]. Reduced concentrations of commonly occurring fatty acids in the hypervirulent strain were ascribed to the previously reported upregulation of transcription factor Rv3574 and the dormancy survival regulon (DosR) [89], both of which are known to be involved in lipid metabolism and respiration. Lastly, reductions of phenylalanine and tyrosine, intermediates in the shikimate pathway and responsible for the synthesis of a number of compounds associated with virulence, in addition to alterations in the precursors of the antioxidant mycothiol (myo-inositol, glucoseamine, and the cysteine precursor, serine), indicated increased utilization of these in the synthesis of various factors associated with virulence [90].

2.4. Systems Biology Related to Drug Resistance. The first-line anti-TB drugs, isoniazid, ethambutol, pyrazinamide, streptomycin, and rifampicin [91, 92], target different processes in all mycobacteria that cause TB. These metabolic processes targeted by the above mentioned drugs are considered essential in mycobacteria for viability, pathogenesis, cell wall synthesis, protein synthesis, DNA replication, and extracellular transport [93]. The biosynthesis of fatty acids is another pathway essential for bacterial survival and includes the previously described FASII pathway, which is also considered an important target for anti-TB drug discovery [65]. Isoniazid and ethambutol have been shown to inhibit the cell wall synthesis of *M. tuberculosis*—isoniazid inhibits the synthesis

of mycolic acids [9, 92] and ethambutol inhibits the synthesis of aribinogalactan in the cell wall [92], both of which are essential components of the mycomembrane [55, 94] and necessary for secreting various virulence-associated proteins via the T7S system [71]. Pyrazinamide, a structural analogue of nicotinamide, targets energy metabolism, also via disruption of the extracellular membrane of *M. tuberculosis* [9, 92]. Lastly, streptomycin and rifampicin both inhibit the synthesis of proteins and nucleic acids [9]; streptomycin interferes more specifically with protein biosynthesis, inhibiting the initiation of mRNA translation [92], whereas rifampicin interferes with bacterial RNA synthesis, effectively inhibiting the transcription of bacterial DNA to RNA by binding to DNA-dependent RNA polymerase β-subunit encoded for by the *rpoB* gene [9, 95].

Mutations in the *M. tuberculosis* genome, or any other organism capable of causing TB-like disease states in mammals, may potentially lead to resistance to these first-line antibiotics. Mutations in the gene encoding enoyl-acyl carrier protein reductase (*inhA*), an essential enzyme in FASII, are thought to lead to the subsequent resistance of *M. tuberculosis* to isoniazid [61]. Furthermore, Zhang et al. [96] indicated that a Ser315Thr mutation in the primary mycobacterial catalase-peroxidase gene, *katG,* also results in resistance, which was confirmed by drastically lowered catalase and peroxidase activities in the strains exhibiting this mutation [97]. Resistance to ethambutol, on the other hand, is thought to be due to a mutation at amino acid residue 306 of the *EmbB* gene (Met306Leu, Met306Val, and Met306Ile), responsible for the synthesis of arabinogalactan [98]. The majority (72–97%) of the pyrazinamide-resistant strains are characterized by a mutation in the gene coding for pyrazinamidase (*pncA*), responsible for the conversion of pyrazinamide to its active form, pyrazinoic acid (POA) [97, 99]. Other defective mechanisms responsible for metabolizing pyrazinamide and regulating the *pncA* gene and/or the efflux or POA have also contributed to the high level of pyrazinamide resistance observed in some clinical isolates, confirming that pyrazinamide resistance in *M. tuberculosis* is not due only to a single factor [100]. Streptomycin resistance, however, has been shown to result from one or more single point mutations: 491 (C → T), 512 (C → T) and 904 (A → G) in the 16S rRNA gene (*rrs*) coding for the ribosomal protein S12 (*rpsL*) [9] in addition to a Lys88Gln mutation that illustrates the same phenotypic effect. Infection with *M. tuberculosis* resistant to rifampicin is usually accompanied by higher fatality rates compared to other first-line medications [97], as resistance to rifampicin is usually accompanied by resistance to one or more of the other first-line antibiotics as well [95]. Mutations in the β-subunit of the RNA polymerase encoding gene *rpoB*, an enzyme crucial for the transcription of bacterial DNA to RNA, have been shown to cause rifampicin resistance [37, 95]. Furthermore, a variety of additional gene mutations in the 81-bp region of the *rpoB* gene, known as the rifampicin-resistance determining region (RRDR), have also been identified as resulting in the same phenotype [101].

Only two studies have investigated drug resistance in organisms responsible for TB from a metabolomics perspective. du Preez and Loots [37] researched the altered metabolic profiles of two different *rpoB* mutant *M. tuberculosis* strains, by comparing these to the wild-type parent strain, using the total metabolome extraction method and GC-MS analyses. One of the major findings of their study was the absence of the two saturated mid-chain methyl-branched fatty acids required for mycolic acid synthesis [49]—10Me-C15:0 and 10Me-C16:0—in both of these rifampicin-resistant mutants. This result was accompanied by elevated levels of the well-known *Mycobacterium* branched-chain fatty acid, tuberculosteric acid (TBSA). Considering the mechanism by which these branched-chain fatty acids are synthesized, du Preez and Loots [37] suggested that the *rpoB* mutation results in a disturbance in the equilibrium of mRNA and the ribonucleoside 5′-triphosphates (ATP, GTP, CTP, and UTP), leading to reduced flavin adenine dinucleotide (FAD) required for the synthesis of these essential fatty acids [37, 102]. They furthermore indicated that as a result of the diminishing nucleoside triphosphates (NTPs), ATP in particular, less SAM is synthesized, further contributing to a reduction in these 10-methyl branched-chain fatty acids observed in the mutant strains [37]. These findings were further confirmed by the accumulation of their upstream fatty acid substrates, oleic acid (C18:1 ω9C), and palmitoleic acid (C16:1 ω7C). The consequential reduction in mycolic acids would be expected to result in a decreased use of lipoarabinomannan (LAM) for cell wall synthesis, resulting in accumulating LAM levels in circulation [103], which explains why the major fatty acid components of LAM, namely, TBSA and C16:0, were detected in elevated concentrations [37]. However, although this hypothesis is a possibility, the accumulation of LAM substrates may also indicate inhibition of LAM synthesis. Lastly, heptadecanoic acid (C17:0) was detected in reduced amounts in the two mutant strains when compared with the *M. tuberculosis* wild-type parent strain. du Preez and Loots [37] ascribed this to the ability of the mutant strains to use acetate and fatty acids as primary carbon sources via the glyoxylate cycle during conditions of stress [104, 105]. The catabolism that was a consequence, resulted in elevated concentrations of propionyl-CoA and acetyl-CoA [106], and in this metabolomics study benzene propionic acid (a derivative of propionic acid) was detected in raised concentrations in the metabolite profile of one of the rifampicin-resistant mutants. This observation confirmed their hypothesis and illustrated the potential of metabolomics and systems biology and its application to understand drug resistance better.

Loots [107] subsequently used the same metabolomics research approach to identify potentially new metabolic pathways and metabolite markers to explain many of the phenotypical characteristics associated with a *katG* mutation and the resulting isoniazid-resistance in *M. tuberculosis*. The isoniazid-resistant strains demonstrated increased susceptibility to oxidative stress and consequently adapted to this by upregulating the synthesis of a number of compounds. These involved the increased uptake and use of alkanes and fatty acids as sources of carbon and energy as well as the synthesis of various compounds and processes directly associated

with reducing oxidative stress, including an ascorbic acid degradation pathway. Such a pathway in these organisms has not been proposed before.

3. Further Applications of Metabolomics for Improved Characterization of *M. tuberculosis*

The genome of *M. tuberculosis* is constituted of genes that code for the proteins involved in both aerobic and anaerobic respiration [32]. As these tubercle bacilli are predominantly located in the lungs of the host, their survival depends on successfully competing with the host for oxygen [33]. This may be limited during the latent phase of infection, when the microbe is contained within the granuloma of the host's lung. Aerobic respiration by the bacterium is achieved by conventional oxidative phosphorylation eventually to yield ATP via the electron transport chain, involving a ubiquinone cytochrome *b* reductase complex and cytochrome *c* oxidase [33, 105]. Additionally, the necessary components linked to the anaerobic growth of *M. tuberculosis* have also been identified in its genome; when oxygen is depleted, nitrate serves as the terminal electron acceptor. During anaerobic conditions, upregulation in a number of genes has been shown, which include (1) *narX,* which encodes for nitrate reductase; (2) *nark2*, necessary for the transport of nitrite/nitrate; and (3) *fdxA,* which encodes feredoxin A, a protein crucial for alternative electron transport [108, 109]. Furthermore, anaerobic conditions also result in increased activity of ribonucleotide reductase class II (*nrdZ*) [108] as well as nitrite reductase (*nirBD*) [110]. The last uses a rearranged form of the *narGHJI* operon, which is also the operon for nitrate reductase [111]. Although not yet explored in any *Mycobacterium* species, Hirai et al. [112] used a combined postgenetic approach to investigate the gene-to-metabolite networks regulating sulfur and nitrogen as well as the secondary metabolism of the flowering plant *Arabidopsis thaliana* grown in nutritionally stressed conditions. This study illustrated the downregulation of the nitrate reductase genes, following the reduction of anaerobic photosynthetic activity, thereby illustrating the role of these genes during anaerobic conditions [112]. Various plants respond to changing nitrate or nitrite reductase activity, elevated CO_2 and decreased aerobic photosynthesis rates, by increasing intracellular concentrations of specific metabolites [113], which can be detected by qualitatively and quantitatively comparing the corresponding metabolomes. This study approach could potentially be applied also to investigate these mechanisms in mycobacterial respiration in *M. tuberculosis*.

4. Concluding Remarks

Improved understanding of the mechanisms involved in *M. tuberculosis* disease, using a combined systems biology research approach, should reveal new insights into the improved diagnosis and treatment of this devastating epidemic. Different research approaches—genomics, proteomics, and transcriptomics—have successfully contributed to expanding our knowledge and appreciation of these disease mechanisms. Despite the progress made, however, most of the functions of the genes involved are still unconfirmed, partly due to a lack of proteomics and metabolomics data. Over the last decade, systems biology has contributed greatly towards elucidating the underlying mechanisms associated with various infectious diseases [114, 115]. The eventual aim of such a multifaceted approach with regard to the largely uncontrolled TB epidemic is a better understanding of TB disease pathogenesis through the identification of specific genes, related proteins (or enzymes), and metabolite markers. This understanding could potentially be used in the development of improved diagnostic and treatment strategies. Metabolomics combined with functional genomics and proteomics would serve well as a new approach for revealing the mysteries of *M. tuberculosis* and attaining themillennium development goals, as set by the WHO, expanding on the research contributions made so far and summarized in this review.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] R. Koch, "Classics in infectious diseases. The etiology of tuberculosis: Robert Koch. Berlin, Germany 1882," *Reviews of Infectious Diseases*, vol. 4, no. 6, pp. 1270–1274, 1982.

[2] B. Mathema, N. E. Kurepina, P. J. Bifani, and B. N. Kreiswirth, "Molecular epidemiology of tuberculosis: current insights," *Clinical Microbiology Reviews*, vol. 19, no. 4, pp. 658–685, 2006.

[3] World Health Organization, *Global Tuberculosis Control: WHO Report, Global Tuberculosis Programme*, World Health Organization, 2010.

[4] World Health Organization, *Global Tuberculosis Control 2011: WHO Report*, WHO, Geneva, Switzerland, 2011.

[5] N. R. Gandhi, A. Moll, A. W. Sturm et al., "Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa," *The Lancet*, vol. 368, no. 9547, pp. 1575–1580, 2006.

[6] J. A. Caminero, G. Sotgiu, A. Zumla, and G. B. Migliori, "Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis," *The Lancet Infectious Diseases*, vol. 10, no. 9, pp. 621–629, 2010.

[7] N. A. Knechel, "Tuberculosis: pathophysiology, clinical features, and diagnosis," *Critical Care Nurse*, vol. 29, no. 2, pp. 34–43, 2009.

[8] D. G. Russell, "*Mycobacterium tuberculosis* and the intimate discourse of a chronic infection," *Immunological Reviews*, vol. 240, no. 1, pp. 252–268, 2011.

[9] I. Olivier and D. Loots, "An overview of tuberculosis treatments and diagnostics. What role could metabolomics play," *Journal of Cell and Tissue Research*, vol. 11, pp. 2655–2671, 2011.

[10] A. Aderem, J. N. Adkins, C. Ansong et al., "A systems biology approach to infectious disease research: innovating the pathogen-host research paradigm," *mBio*, vol. 2, no. 1, pp. 1–4, 2011.

[11] T. Sakata and E. A. Winzeler, "Genomics, systems biology and drug development for infectious diseases," *Molecular BioSystems*, vol. 3, no. 12, pp. 841–848, 2007.

[12] W. B. Dunn and D. I. Ellis, "Metabolomics: current analytical platforms and methodologies," *Trends in Analytical Chemistry*, vol. 24, no. 4, pp. 285–294, 2005.

[13] J. C. Schoeman, I. du Preez, and D. T. Loots, "A comparison of four sputum pre-extraction preparation methods for identifying and characterising *M. tuberculosis* using GCxGC-TOFMS metabolomics," *Journal of Microbiological Methods*, vol. 91, pp. 301–311, 2012.

[14] J. M. Halket, D. Waterman, A. M. Przyborowska, R. K. P. Patel, P. D. Fraser, and P. M. Bramley, "Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS," *Journal of Experimental Botany*, vol. 56, no. 410, pp. 219–243, 2005.

[15] J. K. Nicholson, J. C. Lindon, and E. Holmes, "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.

[16] L. M. Raamsdonk, B. Teusink, D. Broadhurst et al., "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations," *Nature Biotechnology*, vol. 19, no. 1, pp. 45–50, 2001.

[17] L. Wu, W. A. Van Winden, W. M. Van Gulik, and J. J. Heijnen, "Application of metabolome data in functional genomics: a conceptual strategy," *Metabolic Engineering*, vol. 7, no. 4, pp. 302–310, 2005.

[18] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz, "Systematic functional analysis of the yeast genome," *Trends in Biotechnology*, vol. 16, no. 9, pp. 373–378, 1998.

[19] L. P. S. de Carvalho, S. M. Fischer, J. Marrero, C. Nathan, S. Ehrt, and K. Y. Rhee, "Metabolomics of *Mycobacterium tuberculosis* reveals compartmentalized co-catabolism of carbon substrates," *Chemistry and Biology*, vol. 17, no. 10, pp. 1122–1131, 2010.

[20] W. Weckwerth, "Metabolomics in systems biology," *Annual Review of Plant Biology*, vol. 54, pp. 669–689, 2003.

[21] T. Nanda, M. Das, K. Tripathy et al., "Metabolomics: the future of systems biology," *Journal Computer Science Systems Biology*, vol. 1, pp. 2–6, 2011.

[22] Q.-Z. Wang, C.-Y. Wu, T. Chen, X. Chen, and X.-M. Zhao, "Integrating metabolomics into systems biology framework to exploit metabolic complexity: strategies and applications in microorganisms," *Applied Microbiology and Biotechnology*, vol. 70, no. 2, pp. 151–161, 2006.

[23] S. G. Villas-Bôas, S. Mas, M. Åkesson, J. Smedsgaard, and J. Nielsen, "Mass spectrometry in metabolome analysis," *Mass Spectrometry Reviews*, vol. 24, no. 5, pp. 613–646, 2005.

[24] T. J. Phelps, A. V. Palumbo, and A. S. Beliaev, "Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints," *Current Opinion in Biotechnology*, vol. 13, no. 1, pp. 20–24, 2002.

[25] S. Halouska, O. Chacon, R. J. Fenton, D. K. Zinniel, R. G. Barletta, and R. Powers, "Use of NMR metabolomics to analyze the targets of D-cycloserine in mycobacteria: role of D-alanine racemase," *Journal of Proteome Research*, vol. 6, no. 12, pp. 4608–4614, 2007.

[26] H. K. Kim, Y. H. Choi, and R. Verpoorte, "NMR-based plant metabolomics: where do we stand, where do we go?" *Trends in Biotechnology*, vol. 29, no. 6, pp. 267–275, 2011.

[27] N. Glassbrook, C. Beecher, and J. Ryals, "Metabolic profiling on the right path," *Nature Biotechnology*, vol. 18, no. 11, pp. 1142–1143, 2000.

[28] H. Kitano, "Systems biology: a brief overview," *Science*, vol. 295, no. 5560, pp. 1662–1664, 2002.

[29] R. M. Warren, N. C. Gey van Pittius, M. Barnard et al., "Differentiation of *Mycobacterium tuberculosis* complex by PCR amplification of genomic regions of difference," *International Journal of Tuberculosis and Lung Disease*, vol. 10, no. 7, pp. 818–822, 2006.

[30] R. Brosch, S. V. Gordon, A. Pym, K. Eiglmeier, T. Garnier, and S. T. Cole, "Comparative genomics of the mycobacteria," *International Journal of Medical Microbiology*, vol. 290, no. 2, pp. 143–152, 2000.

[31] K. N. Lewis, R. Liao, K. M. Guinn et al., "Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guérin attenuation," *Journal of Infectious Diseases*, vol. 187, no. 1, pp. 117–123, 2003.

[32] A. S. Pym, P. Brodin, R. Brosch, M. Huerre, and S. T. Cole, "Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*," *Molecular Microbiology*, vol. 46, no. 3, pp. 709–717, 2002.

[33] S. Cole, R. Brosch, J. Parkhill et al., "Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence," *Nature*, vol. 393, pp. 537–544, 1998.

[34] J.-C. Camus, M. J. Pryor, C. Médigue, and S. T. Cole, "Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv," *Microbiology*, vol. 148, no. 10, pp. 2967–2973, 2002.

[35] I. Olivier and D. T. Loots, "A metabolomics approach to characterise and identify various Mycobacterium species," *Journal of Microbiological Methods*, vol. 88, no. 3, pp. 419–426, 2012.

[36] D. T. Loots, R. J. Meissner-Roloff, M. Newton-Foot et al., "A metabolomics approach exploring the function of the ESX-3 type VII secretion system of *M. smegmatis*," *Metabolomics*, vol. 9, pp. 631–641, 2013.

[37] I. du Preez and D. T. Loots, "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*, vol. 16, pp. 596–603, 2012.

[38] R. J. Meissner-Roloff, G. Koekemoer, R. M. Warren, and D. T. Loots, "A metabolomics investigation of a hyper- and hypo-virulent phenotype of Beijing lineage *M. tuberculosis*," *Metabolomics*, vol. 8, pp. 1194–1203, 2012.

[39] I. du Preez and D. Loots, "New sputum metabolite markers implicating adaptations of the host to Mycobacterium tuberculosis, and vice versa," *Tuberculosis*, vol. 93, pp. 330–337, 2013.

[40] P. R. Wheeler and C. Ratledge, "Metabolism of *Mycobacterium tuberculosis*," in *Tuberculosis: Pathogenesis, Protection, and Control*, B. R. Bloom, Ed., pp. 353–385, American Soceity for Microbiology Press, Washington, DC, USA, 1994.

[41] J. Tian, R. Bryk, M. Itoh, M. Suematsu, and C. Nathan, "Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of α-ketoglutarate decarboxylase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 30, pp. 10670–10675, 2005.

[42] M. Suematsu, T. Hishiki, T. Iwabuchi et al., "Biomedical application of metabolomics," *Cell*, vol. 25, pp. 1427–1431, 2006.

[43] K. Y. Rhee, L. P. S. D. Carvalho, R. Bryk et al., "Central carbon metabolism in *Mycobacterium tuberculosis*: an unexpected frontier," *Trends in Microbiology*, vol. 19, no. 7, pp. 307–314, 2011.

[44] L. P. S. de Carvalho, H. Zhao, C. E. Dickinson et al., "Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase," *Chemistry and Biology*, vol. 17, no. 4, pp. 323–332, 2010.

[45] D. J. V. Beste and J. McFadden, "Metabolism of *Mycobacterium tuberculosis*," *Systems Biology of Tuberculosis*, vol. 10, pp. 55–78, 2013.

[46] E. J. Muñoz-Elías and J. D. McKinney, "*Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence," *Nature Medicine*, vol. 11, no. 6, pp. 638–644, 2005.

[47] K. Liu, J. Yu, and D. G. Russell, "*pckA*-deficient *Mycobacterium bovis* BCG shows attenuated virulence in mice and in macrophages," *Microbiology*, vol. 149, no. 7, pp. 1829–1835, 2003.

[48] J. Marrero, K. Y. Rhee, D. Schnappinger, K. Pethe, and S. Ehrt, "Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for *Mycobacterium tuberculosis* to establish and maintain infection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 21, pp. 9819–9824, 2010.

[49] T. Ramakrishnan, P. S. Murthy, and K. P. Gopinathan, "Intermediary metabolism of mycobacteria," *Bacteriological Reviews*, vol. 36, no. 1, pp. 65–108, 1972.

[50] K. L. Olszewski, M. W. Mather, J. M. Morrisey et al., "Branched tricarboxylic acid metabolism in *Plasmodium falciparum*," *Nature*, vol. 466, pp. 774–778, 2010.

[51] M. J. Gardner, N. Hall, E. Fung et al., "Genome sequence of the human malaria parasite *Plasmodium falciparum*," *Nature*, vol. 419, no. 6906, pp. 498–511, 2002.

[52] D. Kaur, M. E. Guerin, H. Škovierová, P. J. Brennan, and M. Jackson, "Biogenesis of the cell wall and other glycoconjugates of *Mycobacterium tuberculosis*," *Advances in Applied Microbiology*, vol. 69, pp. 23–78, 2009.

[53] O. Neyrolles and C. Guilhot, "Recent advances in deciphering the contribution of *Mycobacterium tuberculosis* lipids to pathogenesis," *Tuberculosis*, vol. 91, no. 3, pp. 187–195, 2011.

[54] D. J. Murphy and J. R. Brown, "Identification of gene targets against dormant phase *Mycobacterium tuberculosis* infections," *BMC Infectious Diseases*, vol. 7, article 84, 2007.

[55] P. J. Brennan, "Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*," *Tuberculosis*, vol. 83, no. 1-3, pp. 91–97, 2003.

[56] J. K. Tang, L. You, R. E. Blankenship et al., "Recent advances in mapping environmental microbial metabolisms through 13C isotopic fingerprints," *Journal of the Royal Society Interface*, vol. 9, pp. 2767–2780, 2012.

[57] W. Salaemae, A. Azhar, G. W. Booker, and S. W. Polyak, "Biotin biosynthesis in *Mycobacterium tuberculosis*: physiology, biochemistry and molecular intervention," *Protein and Cell*, vol. 2, no. 9, pp. 691–695, 2011.

[58] J. Rabinowitz, J. Purdy, L. Vastag et al., "Metabolomics in drug target discovery," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 76, pp. 235–246, 2011.

[59] H. C. Wong, G. Liu, Y.-M. Zhang, C. O. Rock, and J. Zheng, "The solution structure of acyl carrier protein from *Mycobacterium tuberculosis*," *Journal of Biological Chemistry*, vol. 277, no. 18, pp. 15874–15880, 2002.

[60] P. Domenech, C. E. Barry III, and S. T. Cole, "*Mycobacterium tuberculosis* in the post-genomic age," *Current Opinion in Microbiology*, vol. 4, no. 1, pp. 28–34, 2001.

[61] C. Vilchèze, H. R. Morbidoni, T. R. Weisbrod et al., "Inactivation of the *inhA*-encoded fatty acid synthase II (FASII) enoyl- acyl carrier protein reductase induces accumulation of the FASI end products and cell lysis of *Mycobacterium smegmatis*," *Journal of Bacteriology*, vol. 182, no. 14, pp. 4059–4067, 2000.

[62] O. Zimhony, C. Vilchèze, and W. R. Jacobs Jr., "Characterization of *Mycobacterium smegmatis* expressing the *Mycobacterium tuberculosis* fatty acid synthase I (fas1) gene," *Journal of Bacteriology*, vol. 186, no. 13, pp. 4051–4055, 2004.

[63] E. Sacco, A. S. Covarrubias, H. M. O'Hare et al., "The missing piece of the type II fatty acid synthase system from *Mycobacterium tuberculosis*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 37, pp. 14628–14633, 2007.

[64] L. Kremer, L. G. Dover, S. Carrère et al., "Mycolic acid biosynthesis and enzymic characterization of the β-ketoacyl-ACP synthase A-condensing enzyme from *Mycobacterium tuberculosis*," *Biochemical Journal*, vol. 364, no. 2, pp. 423–430, 2002.

[65] C.-J. Zheng, M.-J. Sohn, and W.-G. Kim, "Atromentin and leucomelone, the first inhibitors specific to enoyl-ACP reductase (FabK) of Streptococcus pneumoniae," *Journal of Antibiotics*, vol. 59, no. 12, pp. 808–812, 2006.

[66] G. Lamichhane, "Novel targets in *M. tuberculosis*: search for new drugs," *Trends in Molecular Medicine*, vol. 17, no. 1, pp. 25–33, 2011.

[67] N. A. Kruh, J. Troudt, A. Izzo, J. Prenni, and K. M. Dobos, "Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome In vivo," *PLoS ONE*, vol. 5, no. 11, Article ID e13938, 2010.

[68] R. Van Crevel, J. Kleinnijenhuis, M. Oosting, L. A. B. Joosten, and M. G. Netea, "Innate immune recognition of *Mycobacterium tuberculosis*," *Clinical and Developmental Immunology*, vol. 2011, Article ID 405310, 12 pages, 2011.

[69] G. P. Bisson, C. Mehaffy, C. Broeckling et al., "Upregulation of the phthiocerol dimycocerosate biosynthetic pathway by rifampin-resistant, *rpoB* mutant *Mycobacterium tuberculosis*," *Journal of Bacteriology*, vol. 194, pp. 6441–6452, 2012.

[70] N. C. Gey Van Pittius, J. Gamieldien, W. Hide, G. D. Brown, R. J. Siezen, and A. D. Beyers, "The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria," *Genome Biology*, vol. 2, no. 10, 2001.

[71] A. M. Abdallah, N. C. Gey van Pittius, P. A. DiGiuseppe Champion et al., "Type VII secretion—mycobacteria show the way," *Nature Reviews Microbiology*, vol. 5, no. 11, pp. 883–891, 2007.

[72] A. Serafini, F. Boldrin, G. Palù, and R. Manganelli, "Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc," *Journal of Bacteriology*, vol. 191, no. 20, pp. 6340–6344, 2009.

[73] M. S. Siegrist, M. Unnikrishnan, M. J. McConnell et al., "Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 44, pp. 18792–18797, 2009.

[74] G. M. Rodriguez, M. I. Voskuil, B. Gold, G. K. Schoolnik, and I. Smith, "*ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response," *Infection and Immunity*, vol. 70, no. 7, pp. 3371–3381, 2002.

[75] C. M. Sassetti and E. J. Rubin, "Genetic requirements for mycobacterial survival during infection," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 22, pp. 12989–12994, 2003.

[76] J. J. De Voss, K. Rutter, B. G. Schroeder, and C. E. Barry III, "Iron acquisition and metabolism by mycobacteria," *Journal of Bacteriology*, vol. 181, no. 15, pp. 4443–4451, 1999.

[77] A. Torosantucci, P. Chiani, F. De Bernardis, A. Cassone, J. A. Calera, and R. Calderone, "Deletion of the two-component histidine kinase gene (CHK1) of Candida albicans contributes to enhanced growth inhibition and killing by human neutrophils in vitro," *Infection and Immunity*, vol. 70, no. 2, pp. 985–987, 2002.

[78] C.-M. Kang, D. W. Abbott, T. P. Sang, C. C. Dascher, L. C. Cantley, and R. N. Husson, "The *Mycobacterium tuberculosis* serine/threonine kinases PknA and PknB: substrate identification and regulation of cell shape," *Genes and Development*, vol. 19, no. 14, pp. 1692–1704, 2005.

[79] Y. Av-Gay and M. Everett, "The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*," *Trends in Microbiology*, vol. 8, no. 5, pp. 238–244, 2000.

[80] A. Narayan, P. Sachdeva, K. Sharma, A. K. Saini, A. K. Tyagi, and Y. Singh, "Serine threonine protein kinases of mycobacterial genus: phylogeny to function," *Physiological Genomics*, vol. 29, no. 1, pp. 66–75, 2007.

[81] A. Koul, A. Choidas, A. K. Tyagi, K. Drlica, Y. Singh, and A. Ullrich, "Serine/threonine protein kinases PknF and PknG of *Mycobacterium tuberculosis*: characterization and localization," *Microbiology*, vol. 147, no. 8, pp. 2307–2314, 2001.

[82] N. C. Gey van Pittius, S. L. Sampson, H. Lee, Y. Kim, P. D. van Helden, and R. M. Warren, "Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions," *BMC Evolutionary Biology*, vol. 6, article 95, 2006.

[83] D. Bottai and R. Brosch, "Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families," *Molecular Microbiology*, vol. 73, no. 3, pp. 325–328, 2009.

[84] A. M. Abdallah, T. Verboom, E. M. Weerdenburg et al., "PPE and PE-PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5," *Molecular Microbiology*, vol. 73, no. 3, pp. 329–340, 2009.

[85] R. Simeone, D. Bottai, and R. Brosch, "ESX/type VII secretion systems and their role in host-pathogen interaction," *Current Opinion in Microbiology*, vol. 12, no. 1, pp. 4–10, 2009.

[86] S. E. Converse and J. S. Cox, "A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in Mycobacterium smegmatis," *Journal of Bacteriology*, vol. 187, no. 4, pp. 1238–1245, 2005.

[87] D. Chatterjee and K.-H. Khoo, "Mycobacterial lipoarabinomannan: an extraordinary lipoheteroglycan with profound physiological effects," *Glycobiology*, vol. 8, no. 2, pp. 113–120, 1998.

[88] I. W. Chen and C. F. Charalampous, "Biochemical studies on inositol. IX. D-Inositol 1-phosphate as intermediate in the biosynthesis of inositol from glucose 6-phosphate, and characteristics of two reactions in this biosynthesis," *Journal of Biological Chemistry*, vol. 241, no. 10, pp. 2194–2199, 1966.

[89] M. B. Reed, S. Gagneux, K. DeRiemer, P. M. Small, and C. E. Barry III, "The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated," *Journal of Bacteriology*, vol. 189, no. 7, pp. 2583–2589, 2007.

[90] A. Vercellone, J. Nigou, and G. Puzo, "Relationships between the structure and the roles of lipoarabinomannans and related glycoconjugates in tuberculosis pathogenesis," *Frontiers in Bioscience*, vol. 3, pp. e149–e163, 1998.

[91] M. Viveiros and L. Amaral, "Enhancement of antibiotic activity against poly-drug resistant *Mycobacterium tuberculosis* by phenothiazines," *International Journal of Antimicrobial Agents*, vol. 17, no. 3, pp. 225–228, 2001.

[92] Y. Zhang, "The magic bullets and tuberculosis drug targets," *Annual Review of Pharmacology and Toxicology*, vol. 45, pp. 529–564, 2005.

[93] G. D. Wright, "Bacterial resistance to antibiotics: enzymatic degradation and modification," *Advanced Drug Delivery Reviews*, vol. 57, no. 10, pp. 1451–1470, 2005.

[94] N. Zhang, J. B. Torrelles, M. R. McNeil et al., "The Emb proteins of mycobacteria direct arabinosylation of lipoarabinomannan and arabinogalactan via an N-terminal recognition region and a C-terminal synthetic region," *Molecular Microbiology*, vol. 50, no. 1, pp. 69–76, 2003.

[95] A. Telenti, D. Lowrie, L. Matter et al., "Detection of rifampicin-resistance mutation in *Mycobacterium tuberculosis*," *The Lancet*, vol. 341, no. 8846, pp. 647–650, 1993.

[96] Y. Zhang, B. Heym, B. Allen, D. Young, and S. Cole, "The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*," *Nature*, vol. 358, no. 6387, pp. 591–593, 1992.

[97] A. Somoskovi, L. M. Parsons, and M. Salfinger, "The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in *Mycobacterium tuberculosis*," *Respiratory Research*, vol. 2, no. 3, pp. 164–168, 2001.

[98] H. Rinder, K. T. Mieskes, E. Tortoli et al., "Detection of *embB* codon 306 mutations in ethambutol resistant *Mycobacterium tuberculosis* directly from sputum samples: a low-cost, rapid approach," *Molecular and Cellular Probes*, vol. 15, no. 1, pp. 37–42, 2001.

[99] P. Sheen, P. Ferrer, R. H. Gilman et al., "Effect of pyrazinamidase activity on pyrazinamide resistance in *Mycobacterium tuberculosis*," *Tuberculosis*, vol. 89, no. 2, pp. 109–113, 2009.

[100] C. Raynaud, M.-A. Lanéelle, R. H. Senaratne, P. Draper, G. Lanéelle, and M. Daffé, "Mechanisms of pyrazinamide resistance in mycobacteria: importance of lack of uptake in addition to lack of pyrazinamidase activity," *Microbiology*, vol. 145, no. 6, pp. 1359–1367, 1999.

[101] R. M. Anthony, A. R. J. Schuitema, I. L. Bergval, T. J. Brown, L. Oskam, and P. R. Klatser, "Acquisition of rifabutin resistance by a rifampicin resistant mutant of *Mycobacterium tuberculosis* involves an unusual spectrum of mutations and elevated frequency," *Annals of Clinical Microbiology and Antimicrobials*, vol. 4, article 9, 2005.

[102] R. Garrett and C. Grisham, "Biochemistry," in *The Synthesis and Degradation of Nucleotides*, pp. 853–878, Thomson Brooks/Cole, Belmont, Calif, USA, 2005.

[103] J. Nigou, M. Gilleron, and G. Puzo, "Lipoarabinomannans: from structure to biosynthesis," *Biochimie*, vol. 85, no. 1-2, pp. 153–166, 2003.

[104] M. S. Glickman and W. R. Jacobs Jr., "Microbial pathogenesis of *Mycobacterium tuberculosis*: dawn of a discipline," *Cell*, vol. 104, no. 4, pp. 477–485, 2001.

[105] L. Shi, C. D. Sohaskey, C. Pfeiffer et al., "Carbon flux rerouting during *Mycobacterium tuberculosis* growth arrest," *Molecular Microbiology*, vol. 78, no. 5, pp. 1199–1215, 2010.

[106] E. J. Muñoz-Elías, A. M. Upton, J. Cherian, and J. D. McKinney, "Role of the methylcitrate cycle in *Mycobacterium tuberculosis* metabolism, intracellular growth, and virulence," *Molecular Microbiology*, vol. 60, no. 5, pp. 1109–1122, 2006.

[107] D. T. Loots, "An altered *M. tuberculosis* metabolome induced by *katG* mutations resulting in isoniazid-resistance," *Antimicrobial Agents and Chemotherapy*, 2014.

[108] M. I. Voskuil, D. Schnappinger, K. C. Visconti et al., "Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program," *Journal of Experimental Medicine*, vol. 198, no. 5, pp. 705–713, 2003.

[109] S. Watanabe, M. Zimmermann, M. B. Goodwin, U. Sauer, C. E. Barry, and H. I. Boshoff, "Fumarate reductase activity maintains an energized membrane in anaerobic *Mycobacterium tuberculosis*," *PLoS Pathogens*, vol. 7, no. 10, Article ID e1002287, 2011.

[110] S. Malm, Y. Tiffert, J. Micklinghoff et al., "The roles of the nitrate reductase NarGHJI, the nitrite reductase NirBD and the response regulator GlnR in nitrate assimilation of *Mycobacterium tuberculosis*," *Microbiology*, vol. 155, no. 4, pp. 1332–1339, 2009.

[111] R. W. Honaker, A. Stewart, S. Schittone, A. Izzo, M. R. Klein, and M. I. Voskuil, "*Mycobacterium bovis* BCG vaccine strains lack narK2 and narX induction and exhibit altered phenotypes during dormancy," *Infection and Immunity*, vol. 76, no. 6, pp. 2587–2593, 2008.

[112] M. Y. Hirai, M. Yano, D. B. Goodenowe et al., "Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 27, pp. 10205–10210, 2004.

[113] M. Stitt and A. R. Fernie, "From measurements of metabolites to metabolomics: an "on the fly" perspective illustrated by recent studies of carbon-nitrogen interactions," *Current Opinion in Biotechnology*, vol. 14, no. 2, pp. 136–144, 2003.

[114] S. Hemaiswarya, A. K. Kruthiventi, and M. Doble, "Synergism between natural products and antibiotics against infectious diseases," *Phytomedicine*, vol. 15, no. 8, pp. 639–652, 2008.

[115] K. L. Seib, G. Dougan, and R. Rappuoli, "The key role of genomics in modern vaccine and drug design for emerging infectious diseases," *PLoS Genetics*, vol. 5, no. 10, Article ID e1000612, 2009.

APPENDIX B

A metabolomics investigation of the function of the ESX-1 gene cluster in mycobacteria

Conrad C. Swanepoel, Mae Newton-Foot, Nicolaas

A metabolomics investigation of the function of the ESX-1 gene cluster in mycobacteria

Conrad C. Swanepoel[1], Mae Newton-Foot[2], Nicolaas C. Gey van Pittius[2], Du Toit Loots[1*]

[1]School for Physical and Chemical Sciences, Centre for Human Metabonomics, North-West University, Potchefstroom, Private Bag x6001, Box 269, South Africa, 2531.

[2]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa

*Corresponding author:

Tel: +27 (0)18 299 1818

Fax: +27 (0)18 299 1823

Email: dutoit.loots@nwu.ac.za

Abbreviated title: Metabolomics of ESX-1 in mycobacteria

Abstract

The ESX-1 gene cluster, encoding the Type-VII secretion (T7S) system and its virulence associated proteins, ESAT-6 and CFP-10, is thought to be responsible for the transport of extracellular proteins across the hydrophobic and highly impermeable, cell envelope of *Mycobacterium*, and is involved in virulence in *M. tuberculosis*, the causative agent of tuberculosis. This study identified metabolite markers associated with the function of the ESX-1 gene cluster in the fast-growing saprophyte *M. smegmatis*, using a GCxGC-TOFMS metabolomics approach. The metabolome of an *M. smegmatis* ESX-1 knock-out strain (ΔESX-1_{ms}) was compared to that of the *M. smegmatis* wild-type parent strain. Multi- and univariate statistical analyses were used to identify the metabolite markers which contribute most to the difference between the strains. A general increase in a variety of metabolites in the ΔESX-1_{ms} is associated with the roles described for ESX-1 in mycolic acid biosynthesis and cell wall integrity, and other changes not yet described in cell envelope structure, altering the levels of antioxidants and energy metabolism. The profiles of several metabolites correlated with the metabolomic variation observed between a hyper- and hypo-virulent Beijing strain of *M. tuberculosis,* suggesting that the pathways which modulate virulence in *M. tuberculosis* are also influenced by ESX-1, reaffirming the association of ESX-1 with virulence and cell envelope biogenesis.

KEY TERMS: metabolomics; tuberculosis; ESX-1; virulence; GCxGC-TOFMS

1 Introduction

Shortly after deciphering the genome of *Mycobacterium tuberculosis* H37Rv (Cole et al., 1998), five copies of an immunopathologically-important gene cluster were identified and named the ESX (or ESAT-6) gene cluster regions 1 to 5 (Gey van Pittius et al., 2001). These gene clusters (ESX-1 – ESX-5) contain genes encoding for members of the ESAT-6 and CFP-10 (Esx) protein family, mycosin proteases, various ATPases, membrane proteins and PE (Pro-Glu) and PPE (Pro-Pro-Glu) proteins (Gey van Pittius et al., 2001). The ESX gene clusters encode a novel Type-VII secretion machinery, proposed to be responsible for the secretion of these proteins (Esx, PE and PPE families) across the mycobacterial cell envelope (Stanley et al., 2003, Abdallah et al., 2007). Mycobacteria have a complex cell envelope structure, composed of an inner and outer membrane, cell wall and capsule. The inner plasma membrane is a phospholipid bilayer, while the highly impermeable outer mycomembrane is composed of hydroxylated branched-chain fatty acids called mycolic acids, and various free lipid components (Minnikin et al., 2002, Bayan et al., 2003, Sutcliffe and Harrington, 2004). A complex matrix of peptidoglycan and arabinogalactan, comprising the cell wall, spans the periplasm and is covalently linked to the mycomembrane. A polysaccharide capsule surrounds the mycomembrane. Phosphatidyl myo-inositol mannosides (PIM) and lipoarabinomannans (LAM) are inserted into the plasma membrane and extend into, or span the cell envelope (Hoffmann et al., 2008, Zuber et al., 2008). The ESX secretion systems are the only systems described to export proteins across the plasma membrane, cell wall matrix and mycomembrane, although the secretion mechanism and components involved in mycomembrane translocation remain elusive.

The ESX-1 gene cluster was found to span the RD1 deletion, the primary attenuating deletion in the *M. bovis* BCG vaccine strain (Mahairas et al., 1996, Behr et al., 1999), and has

been extensively researched due to its association with virulence in *M. tuberculosis* (Hsu et al., 2003, Lewis et al., 2003, Guinn et al., 2004). However, ESX-1 also occurs in various fast-growing saprophytic and non-pathogenic mycobacteria (Gey van Pittius et al., 2006) and the saprophyte, *M. smegmatis*, is able to secrete substrates of *M. tuberculosis* ESX-1, suggesting a level of functional conservation between pathogenic and non-pathogenic species (Converse and Cox, 2005, Wirth et al., 2012). This suggests that ESX-1 has additional functions or substrates in pathogenic mycobacteria, and/or interacts with additional pathways and mechanisms in pathogenic mycobacteria to optimize virulence. Various proteins, encoded outside of the ESX-1 gene cluster, have been shown to influence both ESX-1 secretion and virulence in *M. tuberculosis* (Fortune et al., 2005, McLaughlin et al., 2007, Xu et al., 2007, Bottai et al., 2011). The biological functions of the ESX-1 secretion systems are not clearly understood. In *M. smegmatis* ESX-1 has been linked to conjugal DNA transfer, where it negatively regulates DNA transfer from the donor strain, but is essential for DNA uptake by the recipient strain (Flint et al., 2004, Coros et al., 2008). ESX-1 and EspA, an ESX-1-associated protein, were shown to influence cell wall integrity in *M. tuberculosis* (Garces et al., 2010). In *M. marinum*, a component of ESX-1, $EccA_1$, was shown to impact mycolic acid synthesis by interacting with mycolic acid synthases, to shuttle them to the sites of mycolic acid synthesis, thereby contributing to the cell envelope structure and influencing virulence (Joshi et al., 2012). In *M. smegmatis, M. marinum* and *M. tuberculosis*, components of ESX-1 have been shown to be localized at the polar region of the cell (Carlsson et al., 2009, Wirth et al., 2012), the sites of cell growth, and hence maximal cell envelope synthesis.

This study uses a metabolomics-based approach to investigate the metabolic changes associated with the presence and absence of ESX-1 to better characterize the function of ESX-1

within the *Mycobacteria*. The metabolome of an ESX-1 knockout strain of *M. smegmatis* was compared to that of the wild-type parent strain, using two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS), in order to identify metabolite markers which expand our knowledge of the functions associated with this secretion system. The metabolite variation associated with the ESX-1 deletion functionally links ESX-1 to cell envelope biogenesis and associated energy demands and oxidative stress, and furthermore with pathways and functions associated with virulence in *M. tuberculosis*.

2 Materials and methods

2.1 Reagents and chemicals

Ampicillin and X-gal were purchased from Roche (Indianapolis, IN, USA). The BBL Seven H11 Agar Base used as solid culture medium for mycobacterial growth was purchased from Becton Dickinson & Company (Franklin Lakes, NJ, USA). Additional supplements were obtained from Roche (bovine serum albumin), Kimix (glucose) and Merck (glycerol). Derivatisation reagents: methoxyamine hydrochloride, N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) and pyridine, as well as, the internal standard 3-phenylbutyric acid, were purchased from Sigma Aldrich (Darmstadt, Germany). Furthermore, the organic solvents used for the study, (chloroform and methanol) were ultra-pure Burdick and Jackson grade (Honeywell International Inc., Muskegon, MI, USA).

2.2 Bacterial strains and culture conditions

Escherichia coli JM109 was used in the cloning procedures. *E. coli* was cultured in Luria-Bertani (LB) broth with shaking, and on LB agar plates, overnight at 37 °C. Solid and liquid media were supplemented with the antibiotics ampicillin (50 ug/ml, Roche) and kanamycin (50 ug/ml, Sigma) and solid media with X-gal (Roche) and 5% sucrose, as required. *M. smegmatis* mc^2155 was used to construct the *M. smegmatis* ESX-1 knock-out strain. For knock-out construction, *M. smegmatis* was grown in liquid LB medium with shaking, as well as on LB agar plates for 4-5 days at 37°C. LB is used as growth medium as Middlebrook 7H9 and 7H11 media contain sugars which are toxic to single crossover knock-out cultures containing a *sacB* marker gene. Media was supplemented with Tween-80 (0.1%, Sigma), kanamycin (25 μg/mL, Sigma), X-gal

(Roche) and 5% sucrose, as required. For metabolite sample preparation, *M. smegmatis* was grown on BBL Seven H11 Agar Base (Becton Dickinson & Company) supplemented with 0.2% glycerol (Merck), 0.5% bovine serum albumin (Roche) and 0.2% glucose (Kimix) and incubated at 37 ºC for 3 days.

2.3 Preparation of the M. smegmatis ESX-1 knock-out

The *M. smegmatis* ESX-1 knock-out strain (ΔESX-1_{ms}) was created using homologous recombination as described in Gordhan and Parish, (2001). Briefly, 824 bp and 845 bp pieces of DNA, located directly upstream and downstream of the ESX-1 gene cluster, respectively, were amplified from *M. smegmatis* mc^2155 genomic DNA, using the following polymerase chain reaction (PCR) primers containing specific restriction sites: UpReg1smeg f (5'-GGTACCAGTCGAACCTGGGCAAGCG-3', *Kpn*I) and UpReg1smeg r (5'-AGATCTTCGGCGTATTCGCAGCACC-3', *Bg*/II) and DownReg1smeg f (5'-AGATCTCCCCAACAAGAATTGATCCG-3', *Bg*/II) and DownReg1smeg r (5'-AAGCTTTTGTTCAGCGAGCCCGAG-3', *Hind*/III), respectively. UpReg1smeg and DownReg1smeg were cloned into the appropriate restriction sites of the p2NIL suicide vector (Parish and Stoker, 2000), via three way cloning, to create p2NIL_UpDownReg1Smeg. The *lacZ-sacB* marker gene cassette was digested out of pGOAL17 (Parish and Stoker, 2000), and inserted into the *Pac*l site of p2NIL_UpDownReg1Smeg to create the suicide vector p2NIL_ESX-1 KO.

p2NIL_ESX-1 KO was transformed into *M. smegmatis* mc^2155 and single recombinants were selected on kanamycin and X-gal, and passaged without selection to allow a second recombination event to occur. The resultant double recombinants were counter selected on sucrose in the absence of kanamycin. PCR was used to distinguish between wild-type and ESX-1 knock-out *M. smegmatis* using three combinations of primers. R1 F1 f (5'-GAGGTGCTCGACAGCGAGGC-3') and R1 F1 r (5'-TCGCCAGCGGAATCAGTGCC-3'), and R1 F2 f (5'-GAGGTGCTCGACAGCGAGGC-3') and R1 F2 r: (5'-GGTGTACCCGTATCCCTTTCCA-3') produce products in the wild-type strain, while R1 F1 f and R1 F2 r produce a product in ΔESX-1_{ms}. The ΔESX-1_{ms} was confirmed by sequencing across the deleted region.

2.4 Sample preparation

Individually cultured replicates of ΔESX-1_{ms} (n=16) and the wild-type parent strain (n=18) were grown on BBL Seven H11 Agar Base in the absence of Tween-80, oleic acid and catalase for 3 days at 37 °C. The cells were scraped off the plates and immediately snap frozen in liquid nitrogen, lyophilized and stored at -80 °C prior to extraction for GCxGC-TOFMS metabolomic analysis.

2.5 Sample extraction and derivatization

A modified approach as described in Meissner-Roloff et al., (2012), was used for the extraction of the metabolome samples. Five *M. smegmatis* wild-type and five ΔESX-1_{ms} samples were

chosen at random and 0.01 g of each transferred to a microcentrifuge tube, followed by mixing of the combined cell fractions, using a carbide tungsten bead in a vibration mill for 3 minutes at 30 $Hz.s^{-1}$. From this master mix, four samples were prepared by weighing off amounts of 0.01 g into sterile microcentrifuge tubes to use as quality control (QC) samples. One QC sample was extracted and repeatedly injected at regular intervals throughout each experimental sample batch. Ten milligrams (mg) of each sample (QC, wild-type and $\Delta ESX\text{-}1_{ms}$) was weighed into a microcentrifuge tube and 50 µL phenylbutyric acid (0.525 mg/mL H_2O) added as internal standard. To each sample, a 1.25 mL mixture containing chloroform, methanol and water in a ratio of 1:3:1 was added. Each sample was briefly vortexed, followed by the addition of a carbide tungsten bead (Retsch), and then placed in a vibration mill for 10 minutes at 30 $Hz.s^{-1}$. The samples were centrifuged for 5 minutes at 12 000 rpm, and the supernatants were transferred to a GC-MS vial and dried under a light stream of nitrogen gas. These samples were derivatized using 50 µL of methoxyamine hydrochloride (containing 15 mg/mL pyridine) at 50 °C for 90 minutes. Thereafter, 80 µL MSTFA with 1% TMCS was added, and the samples were further derivatized at 50 °C for 60 minutes (Kanani and Klapa, 2007). These derivatized samples were transferred to a 0.1 ml insert and placed into a GC vial prior to GCxGC-TOFMS analysis.

2.6 GCxGC-TOFMS analysis

A Leco Pegasus 4D GCxGC-TOFMS was used to analyze the prepared samples. One microliter of the prepared samples was injected in a splitless mode on an Agilent 7890A two dimensional gas chromatograph (GCxGC) (Agilent, Atlanta, GA), coupled to a time of flight mass spectrometer (TOFMS) (LECO Africa (PTY) LTD) equipped with a Gerstel Multi Purpose

Sampler (MPS) (Gerstel GmbH & co. KG, Eberhard-Gerstel-Platz 1, D-45473 Mülheim an der Ruhr). A Rxi-5Sil MS primary column (30 m, 0.25 mm internal diameter, 0.25 µm film thickness) (Restch GmbH & co. KG, Haan, Germany) enabled first dimensional separation, and a Rxi-17 capillary column (1 m, 0.1 mm internal diameter, 0.1 µm film thickness) was fitted as the secondary column. The cryomodulator concentrated the effluent of the primary column before subjecting the compounds to the secondary column, with a hot pulse of nitrogen gas for 0.7 seconds, at 3 second intervals. The front inlet temperature was held at a constant 280 °C for the entire run, ensuring rapid vaporization. The GC ramp for the primary oven was initially held at 70 °C for 2 minutes, followed by a rise in temperature of 4 °C per minute until the oven reached a final temperature of 300 °C, where it was held for 2 minutes. The secondary column oven temperature was initially held at 75 °C for 2 minutes, followed by an increase of 4 °C per minute, until a final temperature of 305 °C was achieved and maintained for 2 minutes.

Every sample run started with a 560 second solvent delay, where the filament was switched off, and consequently no mass spectrum was recorded during this period. These 560 seconds, where however included in the primary column run time, to ensure accurate retention times of the compounds of interest. The transfer line temperature was held at a constant 275 °C, with the ion source temperature at a constant 230 °C. The detector voltage was adjusted to 1800 V with the filament bias -70 eV. Spectra were collected from 50-600 m/z at an acquisition rate of 200 spectra per minute.

2.7 Data processing

Deconvolution and peak identification were done using Leco ChromaTOF software version 4.44 (Leco, St. Joseph, MI, USA). This software provides in-line acquisition and analysis of chromatograms collected by the GCxGC-TOFMS. The mass spectra were deconvoluted at a signal to noise (S/N) ratio of 200 and the compounds were aligned via ChromaTOF statistical compare, using the characteristic mass fragmentation patterns and respective GC retention times generated. The compounds were identified and named by comparing the GCxGC-TOFMS generated information to commercial and in-house libraries constructed from previously injected standards. Retention time shifts disables comparative data which is eliminated via peak alignment. Using the internal standard, peak areas were normalized by calculating the relative concentration of each individual compound, which yielded the final data matrix, consisting out of all the compounds aligned to one another and their respective relative concentrations.

2.8 Statistical data analysis

Due to the complexity of GCxGC-TOFMS data generated (the amount of compounds detected, their varying concentrations and the different compound classes present in an individual sample) data pre-treatment (log transformation) is necessary (Meissner-Roloff et al., 2012, Schoeman et al., 2012). Data clean-up included a zero value replacement (replacing the missing values in the data set with half of the lowest detected concentration for that specific compound across the samples, assuming the missing values were due to that compound concentration being lower than the detection limit of the GCxGC-TOFMS) and 50% filter (deleting the compounds not detected in at least 50% of the naturally grouped samples). A QC-CV filter followed, using the previously mentioned QC samples and calculating the coefficient of variation (CV) values for all

the detected compounds. CV can be defined as: $CV = s/\tilde{x} \times 100$, where $\tilde{x}$ is the sample mean with s being the sample standard deviation. Using the calculated concentration of a specific compound across all the samples, the determined CV values can serve as a tool to quantify the measurement or detection error. Utilizing these pre-determined CV values for the QC samples, all compounds detected in the QC's with a CV value which exceeded 50% were removed (Schoeman et al., 2012). This allows for a "cleaner" data matrix, by the removal of those compounds which inherently have large analytical variation and in so doing, simplifies metabolite marker identification, in the subsequent multivariate and univariate statistical analysis.

The compiled data matrix was used for further metabolomics investigations and group comparisons, using a variety of multivariate and univariate statistical approaches. Multivariate statistical analyses included principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). PCA was used to determine if a natural grouping exists between the two strains (wild-type and ΔESX-1_{ms}) on the basis of an altered metabolite profile due to the absence or presence of ESX-1. PCA is an example of an unsupervised statistical data analysis procedure, and attempts to elucidate the relationship between groups, without any prior knowledge that these groups should differ. It accomplishes this by transforming a large set of related variables (metabolites) into a set of new, unrelated variables smaller in size termed, principle components (PCs). The main focus of the PCs is to express the maximum variation in the data set by characterizing each component in a multidimensional space. Finally the modeling power of each variable in the PCA was determined and used to select those variables of importance in the projection (Meissner-Roloff et al., 2012, Schoeman et al., 2012), by selecting only those which had a modeling power which exceeded 0.5 (Chong and Jun, 2005).

PLS-DA on the other hand, is an example of a supervised data analysis method, used to identify compounds which differ most when comparing the groups. PLS-DA is used to determine variables important in projection (VIPs), which is defined as a combined sum of squares of collected PLS-DA weights, ranking the importance of a specific metabolite, when differentiating between two assigned groups (Schoeman et al., 2012). Variables with a VIP > 1 were identified and selected as potential markers (Chong and Jun, 2005).

To identify the univariate importance of individual metabolites, an unpaired t-test and effect size calculations were used. Unpaired T-test involves *t*-value determination followed by the calculation of a *p*-value, which is then used to establish if the difference in various averages is statistically significant (Xia et al., 2012). A value of: $p < 0.05$, was selected as the level of significance (Ellis and Steyn, 2003) and all compounds which did not adhere to this cut-off were excluded as possible metabolite markers. Effect size is defined as the smallest change in the data set which can reliably be detected (Goodpaster et al., 2010) or alternatively as a measurement of practical significance, in order to determine relevance. The effect size (d) can be illustrated by the formula*:* $d = |X_1 - X_2|/S_{max}$, where X_1 and X_2 signify the group means and S_{max}, the standard deviation of two groups analyzed. For non-parametric data, an effect size of $d > 0.1$ can be considered as being of small practical importance, whereas an effect size of $d > 0.3$ and $d > 0.5$ signifies a medium and large practical significance respectively (Ellis and Steyn, 2003). In this study, an effect size of $d > 0.3$ was used to identify compounds of importance. MetaboAnalyst a web server designed for metabolomics data analysis and interpretation (Xia et al., 2012) was used to identify the metabolites adhering to the cut-offs following univariate statistics. Another statistics and analytics software package named Statistica (version 10), was implemented to

identify metabolites which adhered to their respectable cut-offs following multivariate statistics and finally "R" (version 2.13.0) was used for all CV value determinations.

3 Results

A total of 870 compounds were detected in the 34 extracted samples analyzed via GCxGC-TOFMS, using the Chromatof software and data processing methods described above. Four hundred and fifty eight of the compounds were identified by comparing their mass fragmentation patterns and retention times to various in house and commercially available compound libraries (subsequently yielding a 53% compound identification rate). The 870 aligned compounds in each of the analyzed samples were subjected to multi- and univariate statistical analyses.

PCA analysis of the data generated from the ΔESX-1_{ms} and *M. smegmatis* wild-type parent strain samples illustrates clear differentiation between the groups (Fig 1), which is achieved due to characteristic differences in the metabolomes of these strains. The total amount of variance explained by the first three PCs (R^2 X cum) was 99.9%, of which PC 1 contributed 99.8%, PC 2 contributed 0.1%, and PC 3 contributed 0%.

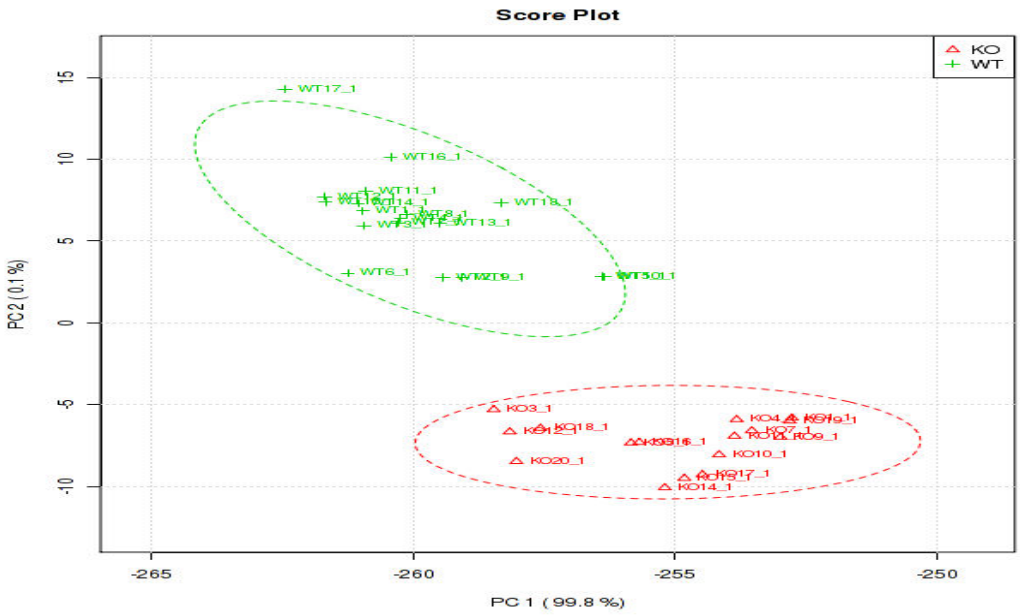


Figure 1: PCA scores plot (PC 1 vs. PC 2) of the GCxGC-TOFMS acquired data shows a natural grouping of the individual samples into distinct *M. smegmatis* wild-type (WT) and ΔESX-1_{ms} groups, based on their varying metabolite profiles.

The PLS-DA indicated that the variance between the two components can be explained as: the modeling parameter R^2Y (cum) being 92.9%, indicative of the total explained variation of the response Y. Q^2 (cum), the cross-validated variation explained by the response Y, was 89.5%.

Metabolite markers were subsequently selected using the results generated from the PCA, PLS-DA, effect sizes and the unpaired *t*-test, and the predetermined cut-offs for each of these, as described above and illustrated in the "Design of the experiment" (Supplementary file 1). Of the 30 metabolites identified as important, 22 could be identified (Table 1), and 8 remained unknown.

Table 1: Metabolite markers identified best describing the variation between the *M. smegmatis* wild-type and ΔESX-1_{ms} samples together with their respective relative mean concentrations, retention times at which they eluted, PLS-DA VIP values, PCA powers, effect size d-values and *t*-test p-values.

Compound (Chemspider)	Mean relative concentration (µg/g cells) and standard deviation		Retention time (s)	PCA Power	PLS-DA VIP	Effect sizes	*t*-Test
	***M. smegmatis* wild-type**	***M. smegmatis* ΔESX-1_{ms}**					
Glycerol (733)	2407.243 (1716.3)	4773.515 (2214.2)	330, 0.424	0.623	5.164	0.398	<0.001
Citric acid (305)	448.276 (287.4)	1061.352 (217.8)	2119, 5.070	0.97	5.051	2.655	<0.001
Mannose (1419823)	64.454 (45.1)	401.41 (232.4)	2234, 2.442	0.811	4.799	1.33	<0.001
Galactose (76670)	92.133 (42.8)	380.483 (73.9)	2041, 0.189	0.969	4.26	3.573	<0.001
Cadaverine (13866593)	9.023 (6.1)	54.407 (17.9)	665, 9.132	0.985	3.099	2.655	<0.001
L-Alanine (5735)	38.278 (20.1)	154.582 (58.4)	821, 5.906	0.952	2.854	2.033	<0.001
D-Gulonic acid (134243)	19.31 (10.8)	86.084 (19.4)	1592, 0.878	0.968	2.771	3.546	<0.001
γ-Aminobutyric acid (GABA) (116)	5.018 (4.4)	47.339 (8.6)	1649, 0.975	0.953	2.417	4.804	<0.001
L-Valine (6050)	57.945 (19.6)	184.038 (46.3)	797, 4.542	0.949	2.355	2.423	<0.001
L-Leucine (5880)	23.014 (9.1)	74.373 (18.1)	932,1.552	0.937	1.656	3.43	<0.001
D-Gulonolactone (588)	41.948 (9.4)	112.867 (32.3)	850, 5.478	0.921	1.979	2.494	<0.001

L-Lysine (5747)	32.496 (10.1)	79.84 (10.5)	2274, 0.953	0.985	1.859	5.305	<0.001
L-Glutamine (5746)	175.204 (52.3)	195.79 (60.4)	1802, 1.050	0.917	1.736	1.772	<0.001
Glucuronic acid (58552)	24.794 (5.4)	77.333 (20.8)	2282, 4.035	0.941	1.718	2.748	<0.001
L-Isoleucine (6067)	17.396 (6.7)	50.748 (10.0)	1205, 0.439	0.949	1.521	3.339	<0.001
Phosphoric acid (979)	34.372 (12.1)	95.677 (21.9)	2798, 1.091	0.833	1.43	2.045	<0.001
L-Norvaline (58608)	3.166 (1.1)	14.086 (5.7)	1049, 0.925	0.935	1.371	2.19	<0.001
L-Methionine (5907)	5.338 (1.7)	14.553 (5.1)	1630, 4.121	0.931	1.348	2.228	<0.001
Dodecanoic acid (3756)	10.244 (4.5)	33.709 (8.6)	2820, 0.585	0.861	1.281	2.008	<0.001
L-Phenylalanine (5910)	2.367 (0.9)	12.764 (3.2)	1814, 1.168	0.963	1.22	2.359	<0.001
D-Ribose-5-phosphate (3560875)	24.575 (11.8)	64.569 (17.0)	1558, 3.600	0.558	1.169	1.625	<0.001
3,4,5 Trihydroxy-pentanoic acid (9703151)	2.088 (0.6)	5.444 (1.2)	600, 7.795	0.97	1.168	3.535	<0.002

4 Discussion

The functions of ESX-1 in *M. smegmatis*, a non-pathogenic *Mycobacterium*, are not well defined, and although it has been shown to be required for conjugal DNA transfer in this organism (Flint et al., 2004, Coros et al., 2008), potential involvement in general metabolism has not been determined. Several metabolites, including amino acids, carbohydrates and fatty acids, were found to be elevated in ΔESX-1_{ms} relative to wild-type, and appear to indicate alterations to cell envelope structure, and additional effects on energy demand and redox stress (Figure 2), correlating with pathways influencing virulence in *M. tuberculosis* (Meissner-Roloff et al., 2012), with 9 of 22 metabolite markers identified in this study identical to that of those described by Meisnerr-Roloff et al., (2012), who used a metabolomics approach to compare a hyper- and hypo-virulent Beijing strain of *M. tuberculosis*, whilst the remaining 13, are intermediates in the same/related metabolic pathways. The majority of those were seen to be elevated in the ΔESX-1_{ms} parent strain which is indicative of reduced utilization for various essential processes in these organisms (Meissner-Roloff et al., 2012).

Several of the metabolites which were influenced by the presence or absence of ESX-1 (Table 1) contribute to cell envelope structure, possibly highlighting a role for ESX-1 in cell envelope biogenesis in *M. smegmatis*. A component of ESX-1, $EccA_1$, has been shown to contribute to mycolic acid synthesis in *M. marinum,* thereby contributing to cell envelope biogenesis (Joshi et al., 2012). The characteristic structure of mycobacterial mycolic acids relies on the incorporation of methyl, cyclopropane and oxygenated groups into the fatty acid chains. These reactions are catalyzed by S-adenosyl methionine (SAM)-dependant methyl-transferases, which transfer a methyl group from the methionine to the unsaturated fatty acid (Asselineau et

al., 2002). An overabundance of methionine was detected in ΔESX-1_{ms} relative to wild-type and this might be attributed to a reduced requirement for methionine for mycolic acid methylation, as it was shown that deletion of $EccA_1$ in *M. marinum* resulted in the synthesis of 30-40% less mycolic acid (Joshi et al., 2012). $EccA_1$ was predicted to bind to and transport various enzymes involved in mycolic acid synthesis to the sites of mycolate synthesis in *M. marinum*. These results suggest of a similar role for ESX-1 component(s) in *M. smegmatis.*

The mycolic acids of the mycomembrane are covalently linked to the arabinogalactan in the cell wall (Crick et al., 2001). Arabinogalactan and arabinomannan, both cell envelope components, are comprised of galactose and arabinose (1.0:2.7), and galactose and mannose (1.0:1.1), respectively (Azuma et al., 1970). The elevated levels of mannose and galactose in ΔESX-1_{ms} may be indicative of reduced arabinogalactan and arabinomannan production in this strain, relative to wild-type, resulting in accumulation of these constituents within the cell. Various amino acids also contribute to the synthesis of structural components of the mycobacterial cell wall. Miyakawa et al., (1972), indicated large amounts of alanine and glutamic acid to be present in the bacterial cell wall, and it is believed that these two amino acids, bound to muramic acid, are essential components of peptidoglycoan. Alanine and glutamic acid bind to muramic acid in the cell wall, forming essential components of peptidoglycan, and comprise most of the roughly 22.8% amino acid content of the cell wall (Acharya and Goldman, 1970). Increased concentrations of glutamine, GABA (derived from glutamate), alanine and finally valine, the precursor to alanine, are possibly as a result of the reduced peptidoglycan synthesis. Furthermore, increased concentrations of glutamine, lysine and phenylalanine in ΔESX-1_{ms} may indicate that aspartic acid is preferentially being used to produce these amino acids, rather than meso-diaminopimelate (meso-DAP), which is also used in

peptidoglycan biosynthesis (Umbarger, 1978, Pavelka and Jacobs, 1996), further confirming these observations. Secretion, together with proper structure and function of the ESX-1-associated protein, EspA, have previously been shown to be required for suitable cell wall integrity of *M. tuberculosis* (Garces et al., 2010). These results suggest that this function is conserved in *M. smegmatis* and that ESX-1 is required for proper cell wall synthesis and/or maintenance.

Cadaverine was also overabundant in ΔESX-1_{ms} relative to wild-type. Cadaverine is a polyamine antioxidant, synthesized from lysine, which is involved in membrane permeability, acid tolerance and RNA synthesis in various bacteria (Kamio et al., 1981, Neely and Olson, 1996, Samartzidou and Delcour, 1999). Cadaverine has been shown to induce porin closure and reduce membrane permeability (Samartzidou and Delcour, 1999) and is also found covalently linked to peptidoglycan (Kamio et al., 1981). Increased cadaverine levels within the cell may represent increased synthesis, or a reduced utilization. ΔESX-1_{ms} may increase cadaverine production to reduce membrane permeability to compensate for the underdeveloped cell envelope structure. Alternatively, defective cell envelope structure may prevent the efficient incorporation of cadaverine into the structure, resulting in accumulation within the cell. Taken together the above results suggest that in the absence of ESX-1, *M. smegmatis* is unable to efficiently generate and maintain its cell envelope structure, and that these defects may involve the mycolic acids in the mycomembrane, the peptidoglycan and arabinogalactan in the cell wall, as well as the lipoarabinomannans which span the capsule. Type VII, or ESX secretion is dependent on an intact cell wall structure, however these results show that ESX-1 may reciprocally be required for the efficient synthesis of this cell envelope. The conserved localization of various ESX-1 proteins in *M. smegmatis, M. marinum* and *M. tuberculosis* at the

polar regions, where growth, and therefore maximum cell envelope synthesis occurs, is supportive of a role for ESX-1 in cell envelope biogenesis in *M. smegmatis*.

Elevated levels of various antioxidants and their intermediates were detected in ΔESX-1_{ms}. Gulonic acid, glucuronic acid and gulonolactone, intermediates in the bacterial biosynthesis of ascorbic acid (Metzler and Metzler, 2003), an antioxidant in mycobacterial species (Taneja et al., 2010), were elevated in ΔESX-1_{ms}. Cadaverine can also function as an antioxidant to scavenge superoxide radicals and regulate external pH by protecting the cells from acid stress (Kim et al., 2006). Inadequate cell envelope formation will result in increased permeability of the cell in ΔESX-1_{ms} and hence elevated oxidative stress, hence the need to generate more antioxidants.

Several metabolites which represent carbon and nitrogen sources were also elevated in ΔESX-1_{ms}, suggestive of reduced energy utilization. The elevated concentrations of citric acid in the ΔESX-1_{ms} comparatively, suggest partial or full inhibition of the TCA cycle in this organism. Supporting this is the overabundant norvaline detected in this strain, as the successful catabolism thereof yields α-ketogluturate, an essential intermediate in the TCA-cylce (Ramakrishnan et al., 1972). Glycerol, mannose and galactose, the primary energy substrates for mycobacteria (Merrill, 1931, Ramakrishnan et al., 1972, Patterson et al., 2003) were all elevated in ΔESX-1_{ms}, relative to wild-type, indicating a reduction in the utilization of these simple sugars for energy production. Ribose-5-phosphate is also an essential component in the synthesis of nucleotides (Wood et al., 1973), additional important factors such as ATP and NAD (which regulate growth) (Roos et al., 2004), the amino acids histidine and tryptophan, and is utilized in the pentose phosphate pathway (Pilz et al., 1984), further supporting the hypothesis of decreased energy

utilization in ΔESX-1_{ms}. Lipid catabolism becomes essential for mycobacterial viability after depletion of simple sugars as a carbon source (Micklinghoff et al., 2009), producing energy via beta-oxidation and the glyoxylate cycle (Upton and McKinney, 2007). Increased concentrations of glycerol and dodecanoic acid in ΔESX-1_{ms} strain may indicate reduced utilization of these fatty acids for energy in the absence of ESX-1. Nitrogen is furthermore essential for mycobacterial growth and cell division (Amon et al., 2008), and glutamic acid (provided in the culture medium and likely the main nitrogen source in this study) can be used to synthesise both glutamine and GABA (γ-Aminobutyric acid), (Jones, 2002) which are elevated in ΔESX-1_{ms}. Glutamine is a well described nitrogen reservoir for various bacterial species and GABA is capable of supporting bacterial growth as either a carbon or nitrogen source (Seth and Connell, 2000). The increased levels of these two metabolites in ΔESX-1_{ms} may result from a reduced nitrogen and carbon requirement. The elevated levels of simple sugars, nitrogen metabolites and fatty acids in ΔESX-1_{ms}, relative to wild-type *M. smegmatis*, may indicate reduced utilization due to a lower energy requirement, possibly associated with a reduction the metabolic state, growth and cell envelope components.

Interestingly, as previously mentioned, 9 of the 22 metabolite markers identified in this study are identical to those described by Meissner-Roloff et al., (2012), who used a metabolomics approach to compare a hyper- and hypo-virulent Beijing strains of *M. tuberculosis*, and the remaining 13 are intermediates in the same, or related, metabolic pathways. Most strikingly, seven of the nine metabolites attributed to altered cell envelope biogenesis in this study, were similarly elevated in the hypo-virulent strain, relative to the hyper-virulent strain. These include mannose, galactose, alanine, glutamine, lysine, phenylalanine and cadaverine. Increased levels of antioxidants and energy metabolites (mannose, galactose,

glycerol, fatty acids and glutamate) in the hypo-virulent *M. tuberculosis* strain also correlate with the observations in the absence of ESX-1 in *M. smegmatis*.

Defective cell envelope biogenesis could affect virulence in a number of ways. The effect of EspA inactivation/deletion on virulence was shown to be independent of ESX-1 mediated secretion, suggesting that the effect was due to the resultant altered cell wall composition (Joshi et al., 2012). The altered cell wall structure may affect virulence by disrupting the bacterial interface, or by increasing its susceptibility to innate host defences (Garces et al., 2010). Various ESX-1-associated proteins have been identified in the capsular layer, including Esx, Esp and PPE proteins, which are secreted by the secretion system (Sani et al., 2010). Disruption of this capsular layer may release some of these proteins, or prevent their insertion into it, thereby preventing efficient host-pathogen interactions with these virulence-mediating proteins, affecting bacterial virulence. ESX-1 may also be involved in transporting other proteins to the cell envelope in order for them to perform their functions, or to assist in cell envelope synthesis and maintenance. As ESX-1 codes for ESAT-6 and CFP-10 (which are also secreted over the cell envelope), a subsequent loss of the ESX-1 gene cluster encoding these proteins may serve as a possible contributor to the elevation in leucine, isoleucine, valine, methionine, alanine and phenylalanine, as the side chains of these proteins consisting predominantly of these amino acids (Renshaw et al., 2005). Furthermore, certain substrates and components themselves could be involved in cell wall biogenesis. This was proposed by Joshi et al., (2012) to be the case for $EccA_1$, which is predicted to transport enzymes responsible for mycolic acid synthesis to the site of mycolic acid biogenesis. Various cell surface lipids, proteins and carbohydrates, are implicated in virulence via regulation of phagocytosis (Cywes et al., 1997, Stokes et al., 2004), immune modulation (Gagliardi et al., 2007, Geurtsen et al., 2009)

and persistence (Sambou et al., 2008). Even small changes to the mycolic acid profile can affect virulence (Yuan et al., 1998, Dubnau et al., 2000, Glickman et al., 2000). An altered cell envelope structure, associated with ESX-1 deletion, is therefore likely to influence virulence, and hence the strong correlations in the metabolite profiles of the ESX-1 wild-type and the hyper-virulent *M. tuberculosis* compared to their respective controls.

No significant variation in ESX-1 protein expression was detected between the hyper- and hypo-virulent *M. tuberculosis* strains (de Souza et al., 2010), suggesting that the perturbations in these strains are not directly associated with ESX-1. The absence of anomalies in methionine levels in the differentially virulent *M. tuberculosis* strains, further suggests that the altered cell wall structure described by Meissner-Roloff et al., (2012) results from mechanisms other than the defective mycolic acid synthesis described for ΔESX-1_{ms} (Joshi et al., 2012). However, the same pathways influenced by ESX-1 in *M. smegmatis* appear to be influenced by alternate mechanisms influencing virulence in *M. tuberculosis*. ESX-1 is, however, required for full virulence of *M. tuberculosis*, therefore the combined roles of ESX-1 and other mechanisms contributing to the variation between the hyper- and hypo-virulent *M. tuberculosis* strains, in mycolic acid and cell wall biosynthesis, as well as other related pathways, promote mycobacterial virulence.

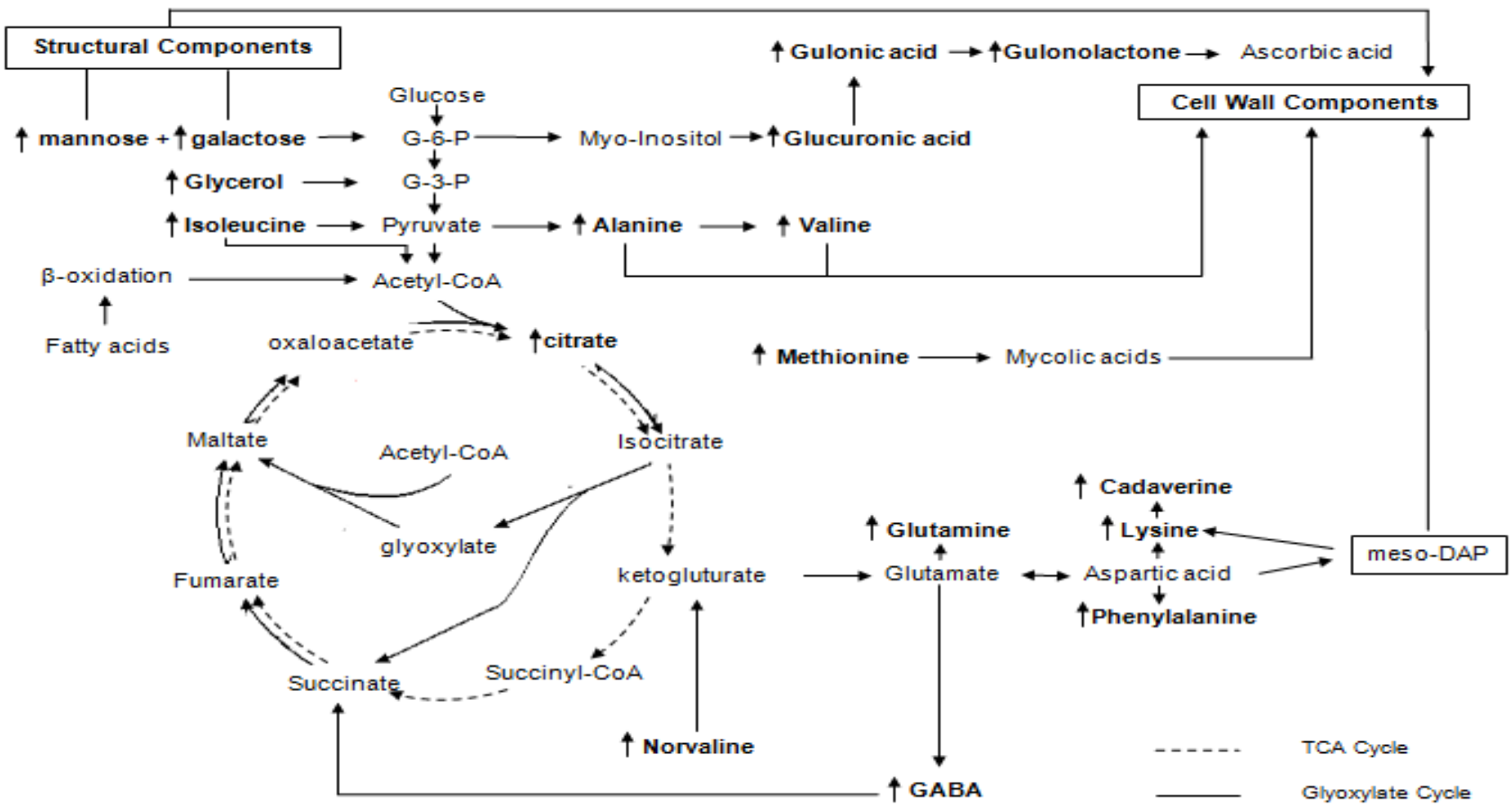


Figure 2: The metabolic pathways altered in ΔESX-1_{ms} when compared to *M. smegmatis* wild-type, showing an increase in various metabolites associated with glycolysis, the TCA and glyoxylate cycles, structural components of the cell envelope, and oxidative stress/antioxidant effect.

5 Conclusion

This is the first study to use a metabolomics approach to describe the metabolome associated with the ESX-1 gene cluster in the non-pathogenic specie *M. smegmatis*. The metabolite profiles of an ESX-1 knock-out strain (ΔESX-1_{ms}) and its wild type parental strain were compared and various metabolites, predominantly carbohydrates, amino acids and fatty acids were found to be overabundant in the absence of ESX-1. The altered metabolite profiles were attributed to defects in cell envelope structure, and associated effects on redox stress and energy metabolism. These results correlate the function of ESX-1 in *M. smegmatis* with the roles in mycolic acid synthesis and cell wall biogenesis previously described for ESX-1 in the pathogenic species, *M. marinum* and *M. tuberculosis*. The metabolite profile of ΔESX-1_{ms} relative to wild-type, was similar with that described for a hypo-virulent Beijing strain of *M. tuberculosis* relative to a related hyper-virulent strain, suggesting that ESX-1 influences the same pathways which influence virulence in *M. tuberculosis*. ESX-1-associated secretion is dependent on an intact cell wall structure, however it appears to also be involved in generating such structure. This role for ESX-1 in cell envelope biogenesis may contribute to virulence in *M. tuberculosis* by ensuring the incorporation and maintenance of various immune-modulating components at the interface between the host and pathogen. Lastly it should be mentioned, that although these results have shed light on a number of new insights regarding the functionality of ESX-1, especially considering its previously described role in virulence. Metabolomics is more focused on hypothesis generation, and hence, future studies investigating these findings using a more direct approach should be considered.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Abdallah AM, van Pittius, Nicolaas C Gey, Champion P A D, Cox J, Luirink J, Vandenbroucke-Grauls C M, Appelmelk B J, Bitter W (2007). Type VII secretion—mycobacteria show the way. Nature reviews microbiology **5**(11): 883-891.

Acharya PN, Goldman D S (1970). Chemical composition of the cell wall of the H37Ra strain of Mycobacterium tuberculosis. Journal of Bacteriology **102**(3): 733-739.

Amon J, Bräu T, Grimrath A, Hänβler E, Hasselt K, Höller M, Jeβberger N, Ott L, Szököl J, Titgemeyer F (2008). Nitrogen control in Mycobacterium smegmatis: nitrogen-dependent expression of ammonium transport and assimilation proteins depends on the OmpR-type regulator GlnR. Journal of Bacteriology **190**(21): 7108-7116.

Asselineau C, Asselineau J, Lanéelle G, Lanéelle M (2002). The biosynthesis of mycolic acids by Mycobacteria: current and alternative hypotheses. Progress in lipid research **41**(6): 501-523.

Azuma I, Ajisaka M, Yamamura Y (1970). Polysaccharides of Mycobacterium bovis Ushi 10, Mycobacterium smegmatis, Mycobacterium phlei, and Atypical Mycobacterium P1. Infection and immunity **2**(3): 347-349.

Bayan N, Houssin C, Chami M, Leblon G (2003). Mycomembrane and S-layer: two important structures of< i> Corynebacterium glutamicum</i> cell envelope with promising biotechnology applications. Journal of Biotechnology **104**(1): 55-67.

Behr M, Wilson M, Gill W, Salamon H, Schoolnik G, Rane S, Small P (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284**(5419): 1520-1523.

Bottai D, Majlessi L, Simeone R, Frigui W, Laurent C, Lenormand P, Chen J, Rosenkrands I, Huerre M, Leclerc C (2011). ESAT-6 secretion-independent impact of ESX-1 genes espF and espG1 on virulence of Mycobacterium tuberculosis. Journal of Infectious Diseases **203**(8): 1155-1164.

Carlsson F, Joshi S A, Rangell L, Brown E J (2009). Polar localization of virulence-related Esx-1 secretion in mycobacteria. PLoS pathogens **5**(1): e1000285

Chong I, Jun C (2005). Performance of some variable selection methods when multicollinearity is present. Chemometrics and Intelligent Laboratory Systems **78**(1): 103-112.

Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S, Eiglmeier K, Gas S, Barry C 3 (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature **393**(6685): 537-544.

Converse SE, Cox J S (2005). A protein secretion pathway critical for Mycobacterium tuberculosis virulence is conserved and functional in Mycobacterium smegmatis. Journal of Bacteriology **187**(4): 1238-1245.

Coros A, Callahan B, Battaglioli E, Derbyshire K M (2008). The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. Molecular microbiology **69**(4): 794-808.

Crick DC, Mahapatra S, Brennan P J (2001). Biosynthesis of the arabinogalactan-peptidoglycan complex of Mycobacterium tuberculosis. Glycobiology **11**(9): 107R-118R.

Cywes C, Hoppe H C, Daffé M, Ehlers M (1997). Nonopsonic binding of Mycobacterium tuberculosis to complement receptor type 3 is mediated by capsular polysaccharides and is strain dependent. Infection and immunity **65**(10): 4258-4266.

de Souza GA, Fortuin S, Aguilar D, Pando R H, McEvoy C R, 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. Molecular & Cellular Proteomics **9**(11): 2414-2423.

Dubnau E, Chan J, Raynaud C, Mohan V P, Lanéelle M, Yu K, Quémard A, Smith I, Daffé M (2000). Oxygenated mycolic acids are necessary for virulence of Mycobacterium tuberculosis in mice. Molecular microbiology **36**(3): 630-637.

Ellis S, Steyn H (2003). Practical significance (effect sizes) versus or in combination with statistical significance (p-values). Management dynamics **12**(4): 51-53.

Flint JL, Kowalski J C, Karnati P K, Derbyshire K M (2004). The RD1 virulence locus of Mycobacterium tuberculosis regulates DNA transfer in Mycobacterium smegmatis. Proceedings of the National Academy of Sciences of the United States of America **101**(34): 12598-12603.

Fortune S, Jaeger A, Sarracino D, Chase M, Sassetti C, Sherman D, Bloom B, Rubin E (2005). Mutually dependent secretion of proteins required for mycobacterial virulence. Proceedings of the National Academy of Sciences of the United States of America **102**(30): 10676-10681.

Gagliardi MC, Lemassu A, Teloni R, Mariotti S, Sargentini V, Pardini M, Daffé M, Nisini R (2007). Cell wall-associated alpha-glucan is instrumental for Mycobacterium tuberculosis to block CD1 molecule expression and disable the function of dendritic cell derived from infected monocyte. Cellular microbiology **9**(8): 2081-2092.

Garces A, Atmakuri K, Chase M R, Woodworth J S, Krastins B, Rothchild A C, Ramsdell T L, Lopez M F, Behar S M, Sarracino D A (2010). EspA acts as a critical mediator of ESX1-dependent virulence in Mycobacterium tuberculosis by affecting bacterial cell wall integrity. PLoS pathogens **6**(6): e1000957.

Geurtsen J, Chedammi S, Mesters J, Cot M, Driessen N N, Sambou T, Kakutani R, Ummels R, Maaskant J, Takata H (2009). Identification of mycobacterial α-glucan as a novel ligand for DC-SIGN: involvement of mycobacterial capsular polysaccharides in host immune modulation. The Journal of immunology **183**(8): 5221-5231.

Gey van Pittius NC, Gamieldien J, Hide W, Brown G D, Siezen R J, Beyers A D (2001). The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G C Gram-positive bacteria. Genome Biol **2**(10): 44.1-44.18.

Gey van Pittius NC, Sampson S L, Lee H, Kim Y, Van Helden P D, Warren R M (2006). Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evolutionary Biology **6**(1): 95.

Glickman MS, Cox J S, Jacobs W R (2000). A Novel Mycolic Acid Cyclopropane Synthetase Is Required for Cording, Persistence, and Virulence of< i> Mycobacterium tuberculosis</i>. Molecular cell **5**(4): 717-727.

Goodpaster AM, Romick-Rosendale L E, Kennedy M A (2010). Statistical significance analysis of nuclear magnetic resonance-based metabonomics data. Analytical Biochemistry **401**(1): 134-143.

Gordhan BG, Parish T (2001). Gene replacement using pretreated DNA. In: Anonymous Mycobacterium tuberculosis Protocols. Springer, pp.77-92.

Guinn KM, Hickey M J, Mathur S K, Zakel K L, Grotzke J E, Lewinsohn D M, Smith S, Sherman D R (2004). Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis. Molecular microbiology **51**(2): 359-370.

Hoffmann C, Leis A, Niederweis M, Plitzko J M, Engelhardt H (2008). Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. Proceedings of the National Academy of Sciences **105**(10): 3963-3967.

Hsu T, Hingley-Wilson S M, Chen B, Chen M, Dai A Z, Morin P M, Marks C B, Padiyar J, Goulding C, Gingery M (2003). The primary mechanism of attenuation of bacillus Calmette–Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proceedings of the National Academy of Sciences **100**(21): 12420-12425.

Jones EA (2002). Ammonia, the GABA neurotransmitter system, and hepatic encephalopathy. Metabolic brain disease **17**(4): 275-281

Joshi SA, Ball D A, Sun M G, Carlsson F, Watkins B Y, Aggarwal N, McCracken J M, Huynh K K, Brown E J (2012). EccA1, a Component of the< i> Mycobacterium marinum</i> ESX-1 Protein Virulence Factor Secretion Pathway, Regulates Mycolic Acid Lipid Synthesis. Chemistry & biology **19**(3): 372-380.

Kamio Y, Itoh Y, Terawaki Y, Kusano T (1981). Cadaverine is covalently linked to peptidoglycan in Selenomonas ruminantium. Journal of Bacteriology **145**(1): 122-128.

Kanani HH, Klapa M I (2007). Data correction strategy for metabolomics analysis using gas chromatography–mass spectrometry. Metabolic engineering **9**(1): 39-51.

Kim J, Choi S H, Lee J K (2006). Lysine decarboxylase expression by Vibrio vulnificus is induced by SoxR in response to superoxide stress. Journal of Bacteriology **188**(24): 8586-8592.

Lewis KN, 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-Guerin attenuation. Journal of Infectious Diseases **187**(1): 117-123.

Mahairas GG, 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. Journal of Bacteriology **178**(5): 1274-1282.

McLaughlin B, Chon J S, MacGurn J A, Carlsson F, Cheng T L, Cox J S, Brown E J (2007). A mycobacterium ESX-1–secreted virulence factor with unique requirements for export. PLoS pathogens **3**(8): e105.

Meissner-Roloff RJ, Koekemoer G, Warren R M (2012). A metabolomics investigation of a hyper-and hypo-virulent phenotype of Beijing lineage M. tuberculosis. Metabolomics **8**(6): 1194-1203.

Metzler DE, Metzler C M (2003). Biochemistry: the chemical reactions of living cells. Academic Press.

Merrill MH (1931). Studies on Carbon Metabolism of Organisms of the Genus Mycobacterium: II. Utilization of Organic Compounds in a Synthetic Medium. Journal of Bacteriology **21**(5): 361.

Micklinghoff JC, Breitinger K J, Schmidt M, Geffers R, Eikmanns B J, Bange F (2009). Role of the transcriptional regulator RamB (Rv0465c) in the control of the glyoxylate cycle in Mycobacterium tuberculosis. Journal of Bacteriology **191**(23): 7260-7269.

Minnikin DE, Kremer L, Dover L G, Besra G S (2002). The Methyl-Branched Fortifications of< i> Mycobacterium tuberculosis</i>. Chemistry & biology **9**(5): 545-553.

Miyakawa T, Matsuzawa H, Matsuhashi M, Sugino Y (1972). Cell wall peptidoglycan mutants of Escherichia coli K-12: existence of two clusters of genes, mra and mrb, for cell wall peptidoglycan biosynthesis. Journal of Bacteriology **112**(2): 950-958.

Neely MN, Olson E R (1996). Kinetics of expression of the Escherichia coli cad operon as a function of pH and lysine. Journal of Bacteriology **178**(18): 5522-5528.

Parish T, Stoker N G (2000). Use of a flexible cassette method to generate a double unmarked Mycobacterium tuberculosis tlyA plcABC mutant by gene replacement. Microbiology **146**(8): 1969-1975.

Patterson J, Waller R, Jeevarajah D, Billman-Jacobe H, McConville M (2003). Mannose metabolism is required for mycobacterial growth. Biochem.J **372**: 77-86.

Pavelka M, Jacobs W R (1996). Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of Mycobacterium smegmatis. Journal of Bacteriology **178**(22): 6496-6507.

Pilz R, Willis R, Boss G (1984). The influence of ribose 5-phosphate availability on purine synthesis of cultured human lymphoblasts and mitogen-stimulated lymphocytes. Journal of Biological Chemistry **259**(5): 2927-2935.

Ramakrishnan T, Murthy P S, Gopinathan K (1972). Intermediary metabolism of mycobacteria. Bacteriological Reviews **36**(1): 65.

Renshaw PS, Lightbody K L, Veverka V, Muskett F W, Kelly G, Frenkiel T A, Gordon S V, Hewinson R G, Burke B, Norman J (2005). Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. The EMBO journal **24**(14): 2491-2498.

Roos AK, Andersson C E, Bergfors T, Jacobsson M, Karlen A, Unge T, Jones T A, Mowbray S L (2004). < i> Mycobacterium tuberculosis</i> Ribose-5-phosphate Isomerase has a Known Fold, but a Novel Active Site. Journal of Molecular Biology **335**(3): 799-809.

Samartzidou H, Delcour A H (1999). Excretion of endogenous cadaverine leads to a decrease in porin-mediated outer membrane permeability. Journal of Bacteriology **181**(3): 791-798.

Sambou T, Dinadayala P, Stadthagen G, Barilone N, Bordat Y, Constant P, Levillain F, Neyrolles O, Gicquel B, Lemassu A (2008). Capsular glucan and intracellular glycogen of Mycobacterium tuberculosis: biosynthesis and impact on the persistence in mice. Molecular microbiology **70**(3): 762-774.

Sani M, Houben E N, Geurtsen J, Pierson J, de Punder K, van Zon M, Wever B, Piersma S R, Jiménez C R, Daffé M (2010). Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. PLoS pathogens **6**(3): e1000794.

Schoeman JC, du Preez I, Loots D T (2012). A comparison of four sputum pre-extraction preparation methods for identifying and characterising< i> M. tuberculosis</i> using GCxGC-TOFMS metabolomics. Journal of microbiological methods.

Seth A, Connell N D (2000). Amino Acid Transport and Metabolism in Mycobacteria: Cloning, Interruption, and Characterization of anl-Arginine/γ-Aminobutyric Acid Permease inMycobacterium bovis BCG. Journal of Bacteriology **182**(4): 919-927.

Stanley SA, Raghavan S, Hwang W W, Cox J S (2003). Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. Proceedings of the National Academy of Sciences **100**(22): 13001-13006.

Stokes RW, Norris-Jones R, Brooks D E, Beveridge T J, Doxsee D, Thorson L M (2004). The glycan-rich outer layer of the cell wall of Mycobacterium tuberculosis acts as an antiphagocytic capsule limiting the association of the bacterium with macrophages. Infection and immunity **72**(10): 5676-5686.

Sutcliffe IC, Harrington D J (2004). Lipoproteins of Mycobacterium tuberculosis: an abundant and functionally diverse class of cell envelope components. FEMS microbiology reviews **28**(5): 645-659.

Taneja NK, Dhingra S, Mittal A, Naresh M, Tyagi J S (2010). Mycobacterium tuberculosis transcriptional adaptation, growth arrest and dormancy phenotype development is triggered by vitamin C. PLoS One **5**(5): e10860.

Umbarger H (1978). Amino acid biosynthesis and its regulation. Annual Review of Biochemistry **47**(1): 533-606.

Upton AM, McKinney J D (2007). Role of the methylcitrate cycle in propionate metabolism and detoxification in Mycobacterium smegmatis. Microbiology **153**(12): 3973-3982.

Wood AW, Becker M A, Seegmiller J E (1973). Purine nucleotide synthesis in lymphoblasts cultured from normal subjects and a patient with Lesch-Nyhan syndrome. Biochemical genetics **9**(3): 261-274.

Wirth SE, Krywy J A, Aldridge B B, Fortune S M, Fernandez-Suarez M, Gray T A, Derbyshire K M (2012). Polar assembly and scaffolding proteins of the virulence-associated ESX-1 secretory apparatus in mycobacteria. Molecular microbiology **83**(3): 654-664.

Xia J, Mandal R, Sinelnikov I V, Broadhurst D, Wishart D S (2012). MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. Nucleic acids research **40**(W1): W127-W133.

Xu J, Laine O, Masciocchi M, Manoranjan J, Smith J, Du S J, Edwards N, Zhu X, Fenselau C, Gao L (2007). A unique Mycobacterium ESX-1 protein co-secretes with CFP-10/ESAT-6 and is necessary for inhibiting phagosome maturation. Molecular microbiology **66**(3): 787-800.

Yuan Y, Zhu Y, Crane D D, Barry III C E (1998). The effect of oxygenated mycolic acid composition on cell wall function and macrophage growth in Mycobacterium tuberculosis. Molecular microbiology **29**(6): 1449-1458.

Zuber B, Chami M, Houssin C, Dubochet J, Griffiths G, Daffé M (2008). Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. Journal of Bacteriology **190**(16): 5672-5680.