The role of the ESX-3 gene cluster and iron on mycobacterial viability

C Buys

21151326

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

Supervisor: Prof DT Loots

September 2013

Christa Buys

Honns B.Sc.

Dissertation submitted in fulfilment of the requirements for the degree Master of Science in Biochemistry School for Physical and Chemical Sciences, Centre for human metabonomics, North-West University (Potchefstroom Campus)

Supervisor: Prof Du Toit Loots

Potchefstroom

2013

Acknowledgements

Firstly, I would like to thank Prof. Du Toit Loots, for all his guidance and advice through the year.

Thanks to my family for all their support and motivation.

I would like to thank the following people personally, Nelus Schoeman and Fanie Kamfer, for their advice through the year and always finding the time to help me.

I would also like to thank the NRF for lessening the financial burden of my M.Sc. degree, by granting me a scarce skills scholarship during the course of this study.

To all my friends and fellow students, thank you for all the support.

Lastly, I would like to thank my Heavenly Father, for the ability and opportunity He gave me to be able to do this. " For this reason I remind you to fan into flame the gift of God, which is in you through the laying on of my hands" (2 Tim 1:6)

"I do all this for the sake of the gospel,

that I may share in its blessings" (1 Corinthians 9:23)

SUMMARY

According to the World Health Organization (WHO), *M. tuberculosis,* the causative agent of TB, accounts for approximately 1.7 million deaths annually. Further contributing causes to the worldwide TB incidence, is the widespread unavailability and ineffectiveness of TB vaccines, time consuming diagnostic methods and unsuccessful treatment approaches. Research for better characterising mycobacteria in general, or other *Mycobacterium* species, may help us to better understand *M. tuberculosis* and TB disease mechanisms, which will in turn lower the future TB disease prevalence, as this may lead to the development of better treatments, diagnostics and vaccines. Mycobacteria use various secretion pathways, including the ESX- or type VII secretion (T7S) system, to ensure transport across the complex cell wall. The genome of *M. tuberculosis* has five copies of a gene cluster known as the ESX gene cluster region, which is associated with virulence and viability of mycobacteria. The ESX-3 gene cluster is thought to be essential for growth of *M. tuberculosis* and proposed to be involved in iron / zinc homeostasis. Mycobacteria synthesise siderophores, which are proposed to be involved in the uptake of iron over their cell wall. *M. tuberculosis* are known to produce two types of siderophores, namely: carboxymycobactins and mycobactins. Loots and colleagues, however illustrated, that ESX-3 knockouts, show signs of iron overload, despite the absence of the mycobactins induced by knocking out the ESX-3 gene cluster. It was hypothesised, that this overload occurs due to an increase in exochelin synthesis, another iron uptake protein not associated with ESX-3, overcompensating for the perceived iron depletion in the knockout organism. A Metabolomics research approach was subsequently used in this study, to generate new information in order to better characterise the role of iron on the metabolism of these organisms, and additionally confirm the role of ESX-3 in iron uptake.

In this study, we firstly determined the most optimal extraction conditions for this metabolomics investigation. Two extraction methods were subsequently investigated and compared, considering their repeatability and their respective capacities to extract those compounds which best differentiate the *M. smegmatis* ESX-3 knockouts and wild-type parent strains. Considering the results generated, the total metabolome method was chosen for further analyses, for the following reasons: 1) it is simpler, 2) faster, 3) showed better repeatability, 4) extracts those

compounds best differentiating the compared groups and 5) has been previously described for metabolomics analyses characterising ESX-3 gene functionality, hence potentially allowing us to compare results to that previously generated and published data.

Subsequently, we used the chosen extraction method, followed by GCxGC-TOFMS analysis of the separately cultured *M. smegmatis* wild-type sample extracts, cultured in normal, low and high iron conditions, to determine the influence of varying iron concentrations on the metabolome of this organism, by metabolomics comparisons of these groups. Following this, an identical research approach was used to compare the metabolome of a *M. smegmatis* ESX-3 knock-out strain, to that of a *M. smegmatis* wild type parent strain, both cultured in normal / standardised iron concentrations. Considering the results generated when comparing the metabolome of a *M. smegmatis* ESX-3 knock-out strain to that of a *M. smegmatis* wild type parent strain, the altered metabolome of the *M. smegmatis* ESX-3 knockouts correlated well to that of the *M. smegmatis* wild type cultured in elevated iron growth conditions. This suggests ESX-3 is involved in iron uptake, and that knocking out the ESX-3 gene cluster of *M. smegmatis* does in fact result in a metabolome profile suggesting iron overload, as was proposed by Loots *et al* (2012), most probably due the exochelins overcompensating for the absence of mycobactins, in *M. smegmatis* ESX-3 knockouts.

Keywords: Metabolomics, ESX-3, *M. smegmatis*, Tuberculosis, Iron regulation

OPSOMMING

Volgens die Wêreld Gesondheids Organisasie (WHO), *M. tuberculosis*, die veroorsakende agent van TB, is verantwoordelik vir ongeveer 1,7 miljoen sterftes jaarliks. Verdere bydraende oorsake tot die wêreldwye TB voorkoms, is die wydverspreide onbeskikbaarheid en ondoeltreffendheid van TB entstowwe, tydrowend diagnostiese metodes en onsuksesvolle behandeling. Navorsing rakende 'n beter karakterisering van mycobacterieë kan ons help om *M. tuberculosis* en TB-siekte meganismes beter te verstaan, wat in die toekoms TB voorkoms sal vermider, wat ook kan lei tot die ontwikkeling van beter behandeling, diagnose en entstowwe. Mycobacterieë gebruik verskeie sekresie paaie, insluitende die ESX- of tipe VII sekresie (T7S) stelsel. Hierdie vervoer stelsel beheer die vervoer oor die komplekse mycobacterium sel muur. Die genoom van *M. tuberculosis* het vyf kopieë van 'n geen kompleks bekend as die ESX geen kompleks, wat geassosïeer word met virulensie en lewensvatbaarheid van mycobacterieë. Die ESX-3 gene kompleks is vermoedelik noodsaaklik vir die groei van *M. tuberculosis* en daar is voorgestel dat dit betrokke is in yster / sink homeostase. Mycobacterieë sintetiseer siderophores, wat help met die opname van yster oor hul selwand. *M. tuberculosis* produseer twee tipes siderophores naamlik: carboxymycobactins en mycobactins. Loots en kollegas, het egter geïllustreer dat ESX-3 ‘knockouts’ toon tekens van ysteroorlading, ten spyte van die afwesigheid van die mycobactins geïnduseer deur die uitslaan van die ESX-3 geen kpmpleks. Die hipotese is gevorm om die idee dat die oorlading voorkom as gevolg van 'n toename in exochelin sintese, 'n ander yster opname proteïen wat nie verband hou met ESX-3, wat oorkompenseer vir die waargenome yster uitputting in die ‘knockout’ organisme. 'n Metabolomika gebaseerde navorsing benadering is vervolgens gebruik in hierdie studie, om nuwe inligting te genereer om die rol van yster op die metabolisme van hierdie organismes beter te karakteriseer en ook die rol van die ESX-3 in die yster opname te bevestig.

In hierdie studie, het ons eerstens die mees optimale ekstraksie kondisies vir hierdie metabolomika-ondersoek bepaal. Twee ekstraksiemetodes is daarna ondersoek en vergelyk, met inagneming van hul herhaalbaarheid en hul vermoë om die veskeie verbindings te onttrek wat die *M. smegmatis* ESX-3 ‘knockouts’ en wilde-tipe beste onderskei. Na vergelyking van die resultate, is die totale metaboloom ekstraksie metode gekies vir verdere ontleding, vir die volgende redes: 1) dit is eenvoudiger, 2) vinniger, 3) het beter herhaalbaarheid getoon, 4) verbindings geëkstraeer differensieer die vergelykende groepe die beste en 5) die metode is voorheen beskryf vir metabolomika ontledings van die ESX-3 geen funksionaliteit, dus laat ons

toe om resultate te vergelyk met resultate wat voorheen gegenereer is en reeds gepubliseerde data.

Daarna het ons die gekose ekstraksie metode, gevolg deur GCxGC TOFMS analise van die afsonderlik gekweekte *M. smegmatis* wilde-tipe, gekweek in normale, lae en hoë yster, gebruik om die invloed te bepaal van verskillende yster konsentrasies op die metaboloom van hierdie organisme, deur metabolomiese vergelyking van hierdie groepe. Na aanleiding van hierdie, is 'n identiese navorsingsbenadering gebruik om die metaboloom van *M. smegmatis* ESX-3 'knockout' te vergelyk, teen dié van 'n *M. smegmatis* wilde-tipe, beide gekweek in normale / gestandaardiseerde yster konsentrasies. Vergelyking van die resultate wat gegenereer word wanneer die metaboloom van 'n *M. smegmatis* ESX-3 'knockout' teen dié van 'n *M. smegmatis* wilde tipe vergelyk word, die veranderde metaboloom van die *M. smegmatis* ESX-3 'knockout' korreleer goed met dié van die *M. smegmatis* wilde tipe gekweek in verhoogde yster groeitoestande. Dit dui daarop dat ESX-3 betrokke is in yster opname, en dat die uitslaan van die ESX-3 geen kompleks van *M. smegmatis* 'n metaboloom profiel van ysteroorlading voorstel, soos voorgestel deur Loots *et al* (2012), waarskynlik as gevolg van die exochelins wat oorkompenseer vir die afwesigheid van mycobactins, in *M. smegmatis* ESX-3 uitklophoue.

Sleutelwoorde: Metabolomika, ESX-3, *M. smegmatis*, Tuberkulose, Yster regulasie

Table of contents

Lists of tables i

Lists of figures ii

Lists of abbreviations ix

Chapter 1 Introduction 1

1.1 Introduction 2

1.2 Structure of the dissertation 3

1.3 Author contributions 5

Chapter 2 Literature study 7

2.1 General overview of tuberculosis epidemiology 8

2.2 *Mycobacterium* 8

2.2.1 Overview of the *mycobacterium* cell wall 9

2.3 The genome of *mycobacterium* 9

2.3.1 Evolution of *mycobacterium* and comparative genomics 10

2.3.2 Overview of the *mycobacterium tuberculosis* genome 10

2.3.3 Regions of difference 11

2.3.4 The ESX gene cluster 12

2.3.5 PE and PPE proteins 13

2.3.6 Genes encoding regulators 15

2.4 Iron acquisition and regulation in mycobacteria 15

2.4.1 Iron acquisition 16

2.4.2 Metal-dependent regulation 17

2.4.3 The role of ESX-3 in iron acquisition and growth 17

2.4.4 Role of ESX-3 in cellular metabolism 18

2.5 *Mycobacterium smegmatis* 19

2.6 Metabolomics 20

2.7 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS) 21

Chapter 3 Problem statement, aims, objectives and experimental design 22

3.1 Problem statement 23

3.2 Aims 23

3.3 Objectives 24

3.4 Experimental design 25

Chapter 4 Method development and validation 26

4.1 Introduction 27

4.2 Materials and methods 28

4.2.1 Reagents and chemicals 28

4.2.2 Preparation of *M.* smegmatis ESX-3 knockouts, culturing conditions and sample preparation 28

4.2.3 Extraction methods 30

4.2.3.1 The conventional total metabolome extraction method 30

4.2.3.2 The fractionated total metabolome (non-polar and polar phase) extraction method 31

4.2.3.3 Pellet extraction method 32

4.2.4 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS) 32

4.2.5 Data processing 33

4.2.5.1 Peak identification and alignment 33

4.2.6 Statistical data analysis 34

4.3 Results and discussion 34

4.3.1 Extraction efficiency 34

4.3.2 Extraction repeatability 38

4.3.3 Ability to extract differentiating compounds 42

4.3.4 Conclusion 44

Chapter 5 A metabolomics investigation of effects of varying iron on the metabolome of *M. smegmatis* 46

5.1 Introduction 47

5.2 Materials and methods 48

5.2.1 Reagents and chemicals 48

5.2.2 Bacterial cultures, preparation of *M.* smegmatis ESX-3 knockouts and sample preparation 48

5.2.3 Quality control samples 49

5.2.4 Extraction method (the conventional total metabolome method) 49

5.2.5 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS) 50

5.2.6 Data processing 50

5.2.6.1 Peak identification and alignment 50

5.2.7 Statistical data analysis 50

5.2.7.1 Data Pre-treatment 50

5.2.7.2 Multivariate and univariate statistics 51

5.3 Results 54

5.3.1. Data pre-treatment 54

5.3.2 PCA differentiation 54

5.3.3 Metabolite marker identification 55

5.4 Discussion 61

5.4.1 Metabolites markers in *M. smegmatis* cultured during comparatively higher iron concentrations 62

5.4.2 Metabolites markers in *M. smegmatis* cultured during comparatively reduced iron concentrations 67

Chapter 6 A metabolomics investigation of *M. smegmatis* wild type and ESX-3 knockouts 70

6.1 Introduction 71

6.2 Materials and methods 72

6.2.1 Reagents and chemicals 72

6.2.3 Quality controls 72

6.2.4 Extraction method (the conventional total metabolome method) 73

6.2.5 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS) 73

6.2.6 Data processing 73

6.2.6.1 Peak identification and alignment 73

6.2.7 Statistical data analysis 73

6.2.7.1 Data pre-treatment 74

6.2.7.2 Multivariate and univariate statistical analysis 74

6.3 Results 74

6.3.1 Data pre-treatment 74

6.3.2 PCA differentiation 74

6.3.3 Metabolite marker identification 75

6.4 Discussion 78

6.4.1 Comparison of *M. smegmatis* wild type grown in normal iron concentrations vs. *M. smegmatis* ESX-3 knockouts cultured in normal iron concentrations 78

Chapter 7 Conclusion 83

7.1 General conclusion 84

7.2 Future prospects 86

Chapter 8 References 87

LIST OF TABLES

Table 1.1:

Research team.....6

Table 4.1:

The number of peaks and internal standard areas detected after GCxGC-TOFMS analyses of 6 sample repeats extracted via each of the investigated extraction methods. Standard deviations are indicated in parenthesis……37

Table 4.2:

Coefficients of variation (CV) values for ten compounds, representative of various compounds classes and detected at retention times and percentage of compounds detected, after GCXGC-TOFMS analyses of the extracts obtained via the two extraction methods……41

Table 5.1:

Metabolite markers selected from the comparison of *M. smegmatis* cultured in normal iron vs. high iron concentrations, indicating their mean relative concentrations (standard deviations in parenthesis), VIP values, PCA modelling power, effect size d-values and t-test *p*-values……57

Table 5.2:

Metabolite markers selected from the comparison of *M. smegmatis* cultured in normal iron vs. low iron concentrations, indicating their mean relative concentrations (standard deviations in parenthesis), VIP values, PCA modelling power, effect size d-values and t-test *p*-values……60

Table 6.1:

Metabolite markers selected from the comparison of *M. smegmatis* wild type vs. *M. smegmatis* ESX-3 knockouts, cultured in normal iron, indicating their mean relative concentrations (standard deviations in parenthesis), VIP values, PCA modelling power, effect size d-values and *p*-values.....77

LIST OF FIGURES

Fig 3.1:

Schematic representation of the experimental design of this metabolomics experiment……..25

Fig 4.1:

Distribution of the coefficient of variation of all the compounds detected after GCxGC-TOFMS analyses of the samples extracted via the various extraction methods investigated. 38

Fig 4.2:

PCA scores showing the differentiation between the two sample groups for the GCxGC-TOFMS data obtained after the various extraction methods. The variance explained in the data sets for each PC is indicated in parenthesis. The *M. smegmatis* wild-type sample group are indicated in red and the ESX-3 knockout sample group are indicated in green. ……42

Fig 5.1:

Work-flow of the multi- and univariate statistical methods used in this study……...52

Fig 5.2:

PCA scores plots showing principal component 1 (PC1) versus principal component 2 (PC2) and the natural differentiation between the sample groups for the GCxGC-TOFMS data obtained after the total metabolome extraction method. The percentage variation explained by each PC is indicated in parenthesis…….....55

Fig 5.3:

Summary of the selection approach used for identifying the metabolite markers best differentiating the *M. smegmatis* cultured in normal iron vs high iron concentrations……...57

Fig 5.4:

Summary of the selection approach used for identifying the metabolite markers best differentiating the *M. smegmatis* cultured in normal iron vs low iron concentrations……...59

Fig 5.5: Schematic representation of the interaction between the glyoxylate, krebs cycle and associated pathways and metabolites altered when *M. smegmatis* are cultured in higher iron concentrations. The dotted arrows represents the Krebs cycle and other liked pathways, where the bold arrows represent the glyoxylate shunt in mycobacteria……64

Fig 6.1:

PCA scores showing the differentiation between the *M. smegmatis* wild type vs. *M. smegmatis* ESX-3 knockouts for the GCxGC-TOFMS data obtained after the total metabolome extraction method. (The *M. smegmatis* wild-type sample group are indicated in red and the ESX-3 knock-out sample group are indicated in green.)…….75

Fig 6.2:

Summary of the selection approach used for identifying the metabolite markers best differentiating the *M. smegmatis* cultured in normal iron vs low iron concentrations.…….....76

LIST OF ABBREVIATIONS

ATP – Adenosine triphosphate

BCG – Bacille de Calmette et Guèrin

CBTBR- Centre of excellence for biomedical TB research

CFP-10 – Culture filtrate protein of 10kDa

CV – Coefficient of variation

DtxR – Diphtheria toxin repressor

ESAT-6 – Early secreted antigenic target of 6kDa

FDA – Food and drug administration

FUR – Ferric uptake regulator

GC – Gas chromatography

GCxGC-TOFMS – Two dimensional gas chromatography time of flight mass spectrometry

HCl- Hydrochloric acid

HIV – Human Immune Virus

IdeR – Iron dependant regulator

LB – Luria-Bertani

LC – liquid chromatography

M. smegmatis – Mycobacterium smegmatis

M. tuberculosis – Mycobacterium tuberculosis

m/z – Mass to charge ratio

MPTR – Major polymorphic tandem repeats

MS – Mass spectrometry

MSTFA – Methoxyamine hydrochloride, n-methyl-n-(trimethylsilyl) trifluoroacetamide

NIST – National institute of standards and technology

PCA – Principal component analysis

PE – Proline-Glutamic acid

PGRS – Polymorphic GC-rich-repetitive sequence

PLS-DA- Partial least square discriminant analysis

PPE – Proline-Proline-Glutamic acid

QC- Quality control

RD – Region of difference

S/N – Signal to noise

SV – Similarity value

SVP – Serien-Valine-proline

T7SS – Type VII secretion system

TB – Tuberculosis

TMCS – Trimethylchlorosilane

TOFMS – Time of flight mass spectrometry

VIP- Variable important in projection

WHO – World Health Organization

Zur – Zinc uptake regulator

Chapter 1

Introduction

Chapter 1

INTRODUCTION

1.1 INTRODUCTION

More than 25% of all deaths globally, are caused by infectious diseases, with tuberculosis (TB) one of the major contributors (Morens *et al*., 2004). According to the World Health Organization (WHO), *Mycobacterium tuberculosis,* the causative agent of TB, accounts for approximately 1.7 million deaths annually, with a consequent prevalence of 2.6 mortalities per 100 000 due to this disease. This disease is especially prominent in poverty stricken regions and is therefore considered a disease strongly associated with various socio-economic issues including: overcrowding, poverty and malnutrition. Furthermore, co-infection with HIV/AIDS has led to an increased mortality, in addition to the multidrug-resistant strains of TB which have recently emerged, and are spreading rapidly, further complicating the control and prevention of this disease (Shiloh *et al.,* 2010). Additional contributing factors to the worldwide TB incidence, is thought to be the widespread unavailability and ineffectiveness of TB vaccines, time consuming diagnostic methods and unsuccessful treatment approaches. Considering this, large efforts are currently being made into research pertaining to the better characterisation of mycobacteria, in order to gain insight into curbing this epidemic.

Mycobacteria use various secretion pathways to ensure protein transport across the complex cell wall, including the ESX- or type VII secretion (T7S) system (Wiker *et al.,* 1992). The genome of *M. tuberculosis* has five copies of a gene cluster known as the ESX gene cluster region, and is associated with virulence and viability of mycobacteria (Gey van Pittius *et al.,* 2001). The ESX-3 gene cluster, is essential for *in vitro* growth of *M. tuberculosis* (Sassetti *et al*., 2003) and was proposed to be involved in iron/zinc homeostasis (Bitter *et al*., 2009). Mycobacteria synthesise siderophores, which are shown to assist in the uptake of iron over their cell wall. *M. tuberculosis* are known to produce two types of siderophores, namely: carboxymycobactins and mycobactins (Domenech *et al*., 2001). Previous research, however, illustrates that ESX-3 knockouts shows signs of iron overload, despite the absence of the mycobactins due to the ESX-3 knockouts (Loots *et al.*, 2013).

It was hypothesised, that this overload occurs due to an increase in exochelin synthesis, another iron uptake protein not associated with ESX-3, overcompensating for the perceived iron depletion in the knockout organism. In the current study, we like to use metabolomics approach to generate information which could further support this hypothesis, and possibly contribute to the understanding of ESX-3 in iron uptake and *Mycobacterium* cell viability.

1.2 STRUCTURE OF THE DISSERTATION

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

The current chapter, Chapter 1, gives a brief background to the conducted study. This chapter also discusses the structure of the dissertation and clarifies the contributions of each individual in the research team, towards the execution and completion of this study.

The literature study in Chapter 2, provides a brief introduction, followed by an extensive overview of mycobacteria in general, with specific focus on the ESX gene clusters of *M. tuberculosis*, the factors involved in regulating mycobacterial cell viability and virulence, and the role of iron, its uptake, and its regulation in these organisms, in addition to the use of metabolomics as a research approach for investigating these occurrences.

Chapter 3 describes the problem statement, aims, objectives and experimental design for attaining the goals as proposed by the aims and objectives.

In Chapter 4, two extraction methods were compared and validated, in order to determine which extraction method is the most suited to characterise the metabolome of *M. smegmatis* investigated in Chapter 5 and 6, using a metabolomics research approach. The extraction efficiency, repeatability and the methods ability to extract differentiating compounds, are investigated in this chapter. Chapter 4 includes a short introduction relevant to the metabolomics approach developed, followed by a description of the compared methods investigated and the results obtained by each, in order to make a final decision on the most

optimal metabolomics approach, for application to answering the biological question in Chapter 5 and 6.

In Chapter 5 and 6, the metabolome of *M. smegmatis* cultured in varying iron concentrations were analysed using the optimised methodology, and compared to that of an ESX-3 knockout grown in normal iron concentrations, in order to validate the previously generated hypothesis of the role of ESX-3 in iron uptake and the associated metabolome from which this information was generated. In Chapter 5, firstly those metaboliote markers best describing the variation observed when comparing the GCxGC-TOFMS analysed metabolome of the *M. smegmatis* wild type cultured in high iron conditions compared to the *M. smegmatis* wild type cultured in normal iron conditions, followed by the selection of those metabolite markers which best differentiate the *M. smegmatis* wild type cultured in low iron conditions comparative to the *M. smegmatis* wild type cultured in normal iron conditions. These metabolite markers were subsequently described in light of their associated role in iron (and / or zinc) metabolism.

In Chapter 6, the extraction method chosen in Chapter 4, was once again used to characterise the metabolome of *M. smegmatis* wild type parent strains cultured in normal iron conditions, comparative to that of the *M. smegmatis* ESX-3 knockouts cultured in normal iron conditions, the latter of which in theory, should correlate to that of the wild type parent strain previously culture in high iron conditions, previously investigated in Chapter 5, if the previously proposed hypotheses as published by Loots et al., (2013), are correct. The selected metabolite markers best describing the observed variation in the GCxGC-TOFMS metabolome of the compared groups were subsequently explained in light of their associated role of iron and / or zinc uptake, and the results generated in Chapter 5, and those previously described by Loots *et al* (2013).

In Chapter 7, the concluding remarks for the results generated in Chapters 4, 5 and 6, are presented, in addition to recommendations and future research prospects, potentially emanating from this metabolomics investigation.

1.3 AUTHOR CONTRIBUTIONS

The principle investigator and author of this thesis is Christa Buys. The contribution of the other contributing co-authors, co-workers and collaborators, towards this work, is given in Table 1.1.

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 thesis:

I declare that my role in this study, as indicated in Table 1.1, 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 Christa Buys.

....

Prof. D.T. Loots C. Buys

Table 1.1: Research team

Co-author	Co-worker	Collaborator	Contribution
C. Buys (B.Sc. Honns. 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.
	J.C. Schoeman (M.Sc Biochemistry)		Assisted with the sample extractions and sample injection on the GCxGC-TOFMS.
		DST/NRF centre of excellence for biomedical TB research (CBTBR), Stellenbosch University, South Africa.	Provided all *M. smegmatis* cultured samples

Chapter 2

Literature study

Chapter 2

LITERATURE STUDY

2.1 GENERAL OVERVIEW OF TUBERCULOSIS EPIDEMIOLOGY

Despite our fervent efforts for combating *Mycobacterium tuberculosis,* the causative agent of TB, the World Health Organization (WHO) reports approximately 1.7 million deaths annually. Furthermore, an estimated 0.4 million of these TB associated deaths were recorded for HIV-positive cases. Included in this 9.4 million, an estimated 440 000 were recorded to have multi-drug resistant *Mycobacterium tuberculosis* (WHO, 2010). Multidrug-resistant strains of TB have recently emerged and are spreading rapidly, complicating the control and prevention of the disease (Shiloh *et al.,* 2010). These statistics highlight the importance of the research that still needs to be done, in order to curb this TB epidemic, through the development of improved TB-medication, diagnostics and vaccines.

2.2 *MYCOBACTERIUM*

The *Mycobacterium* genus, has a genome with a high G+C content, and is known to be aerobic, non-motile non-spore forming, slender, curved and rod-shaped in structure. The genus *Mycobacterium,* includes many pathogens that cause various diseases in mammals, as well as other non-pathogenic species such as; *M. smegmatis, M. fortuitum* and *M. phlei* (Beltan *et al*., 2000). The pathogenic mycobacteria species include *M. leprae*, the causative agent of leprosy, *M. avium*, *M. bovis*, *M. africanum* and most importantly *M. tuberculosis*. The success of *M. tuberculosis*, as well as other pathogenic mycobacteria, lies in their ability to survive inside the host macrophage (Nguyen *et al*., 2005)

2.2.1 Overview of the *Mycobacterium* cell wall

All *Mycobacterium* species have similar, characteristic cell envelopes, responsible for many of the medically challenging physiological characteristics associated with TB, including resistance to antibiotics and host defence mechanisms (Siegrist *et al*., 2009).

The *Mycobacterium* cell wall consists of an outer membrane, an arabinogalactan-peptidoglycan layer, and an inner cytoplasmic membrane (Bitter *et al.,* 2009). The peptidoglycan membrane consists of long-chain mycolic acids which are tightly packed and covalently linked, and non-covalently bound shorter lipids thought to play a role in the phatogenicity of these organisms, by functioning as an exceptionally efficient barrier against various environmental changes. Due to the resilient, tightly packed nature of this cell wall, mycobacteria have several specialised and general transport systems for transporting various important molecules, required for growth and survival. One of these is the transport 7 secretion system (T7SS), which will be discussed in greater detail in the latter sections of this literature study.

2.3 THE GENOME OF *MYCOBACTERIUM*

The information generated to date from various genome studies, are of great importance in the on-going fight to limit the spread of TB worldwide. DNA hybridization and comparative genomics are powerful techniques for identifying genetic variation between various mycobacteria (Domenech *et al*., 2001). These genomic variations could potentially explain differences in the biochemistry, physiology and virulence of these organisms. The complete genome sequencing of various mycobacterial species and strains, has provided the opportunity for intra-species genomic comparisons as well as comparisons between genomes of closely related species. After the sequencing of several mycobacteria species and the utilization of gene prediction methods, the whole genomes of several mycobacteria species have been defined to date (Mulder *et al*., 2009).

2.3.1 Evolution of *Mycobacterium* and comparative genomics

When the genome of *Mycobacterium* was studied using non-pathogenic mycobacteria, it became apparent that *M. tuberculosis* has undergone various evolutionary processes over time (Veyrier *et al*., 2011). The organisms' ability to evolve appears to be provided by the mechanism of gene duplication, and this evolutionary process results in the expansion of gene families, thereby providing the organism with the opportunity to evolve new functions (Mulder *et al*., 2009). A cross-species comparison is required to identify the initial stages of evolution. The genome sequence of *M. tuberculosis* and *M. bovis,* the causative agent of tuberculosis in humans and cattle respectively, are approximately 99.95% identical, but deletion of genetic information in the genome of *M. bovis* has led to a comparatively reduced genome size (Garnier *et al*., 2003). Based on the various similarities between *M. tuberculosis* and *M. bovis,* it was proposed that *M. tuberculosis* evolved from *M. bovis* (Veyrier *et al*., 2011). However, in contrast to what was previous believed, genomic comparisons of these two mycobacteria placed *M. bovis,* based on its smaller genome in a more derivative position and *M. tuberculosis* in a more ancestral position (Veyrier *et al*., 2011), and thereby revealing that *M. tuberculosis* did not in fact evolve from *M. bovis*. Comparing genomic sequence data of different mycobacteria, including *M. tuberculosis* helps one define the genetic differences between the abundance of species present. Considering this, comparisons of these genomic sequences in the light of the physiological differences, has shed light on specific genes and their proposed roles in growth and virulence, amongst others.

2.3.2 Overview of the *Mycobacterium tuberculosis* genome

The complete genome sequence of the H37Rv strain of *M. tuberculosis* was completed and published in 1998 (Cole *et al*., 1998). In order to identify the complete genome sequence of H37Rv, a combined approach was used, involving a systematic sequence analysis of selected large-insert and small-insert clones, from a whole-genome shotgun library (Cole *et al*., 1998). The genome of *M. tuberculosis* H37Rv has more than 4 million base pairs, contains approximately 4000 genes encoding proteins and 50 genes coding for stable RNA (Domenech *et al*., 2001). A unique feature of the of the *M. tuberculosis* genome, is the large

portion of genes coding for enzymes involved in lypolysis and lypogenesis (Cole *et al*., 1998).

Furthermore, the discovery that 9% of the genome encodes two families of proteins rich in glycine, the PPE family, which is characterised by the presence of a proline-proline-glutamic acid motif and the PE family, characterised by the presence of a proline-glutamic acid motif (Domenech *et al*., 2001), was also found to be important. The precise function of these proteins is, however, still unknown, although suggestions have been made that they play a role in antigenic variation and disease pathogenesis (Gey van Pittius *et al*., 2006). From the genome sequence, it is clear that in addition to many functions involved in lipid metabolism, the enzymes necessary for important metabolic pathways, for example glycolysis, tricarboxylic acid cycle, glyoxylate cycle and pentose phosphate pathway, are all present (Cole *et al*., 1998). Two haemoglobin-like protein encoding genes were also found (Cole *et al*., 1998). These proteins are suspected to play a role in protection against oxidative stress, which aids in the ability of the bacteria to adapt its metabolism to the host environment.

2.3.3 Regions of difference

Additional comparative analysis of the genomes of *M. tuberculosis*, virulent *M. bovis* and the attenuated Bacille de Calmette et Guérin (BCG) vaccine strain of *M. bovis,* revealed several regions of difference (RD) amongst these organisms (Ganguly *et al*., 2008). One of these regions, RD1 is absent in all *M. bovis* BCG strains (Lewis *et al*., 2003). RD1 includes the genes Rv3871 to Rv3879c (Gey van Pittius *et al.,* 2001), and include genes that encode for the early secreted antigenic target of 6 kDa (ESAT-6) and the culture filtrate protein of 10 kDa (CFP-10). Since deletion of RD1 seems to be the primary reason for the attenuation of *M. bovis,* this region of the genome became the subject of studies on the virulence of *M. tuberculosis* (Ganguly *et al*., 2008).

2.3.4 The ESX gene cluster

As previously mentioned, mycobacteria use various secretion pathways to ensure protein transport across the complex cell wall, including the ESX- or type VII secretion (T7S) system (Wiker *et al.,* 1992). The genome of *M. tuberculosis* has five copies of a gene cluster known as the ESX gene cluster region. This is comprised of five gene clusters, named: ESX-1 (Rv3866-Rv3883c locus), ESX-2 (Rv3884c-Rv3895c locus), ESX-3 (Rv0282-Rv0292 locus), ESX-4 (Rv3444c-Rv3450c locus) and ESX-5 (Rv1782-Rv1798 locus) (Gey van Pittius *et al*., 2001).

ESX-1 and ESX-5 are both implicated to play a role in virulence (Abdallah *et al.,* 2007). The ESX-1 system is responsible for the secretion of the ESAT-6 and CFP-10 proteins (McLaughlin *et al.,* 2007), (which forms a tight 1:1 complex (Ganguly *et al.,* 2008)), which are important T-cell antigenic targets, essential for the virulence of *M. tuberculosis* (Abdallah *et al.,* 2007). It is also believed that all the genes in the ESX gene cluster regions, act together to provide a system for the secretion of ESAT-6 and CFP-10 (Gey van Pittius *et al*., 2001). Most of ESX- 1 is situated in RD-1, possibly explaining the attenuation of BCG *M. bovis*. Two other regions of difference, namely: RD-5 and RD-8, also contain members of the ESAT-6 gene family (Domenech *et al.,* 2001) and contribute to the secretion of these proteins. Inactivation of these genes is associated with the attenuation in the secretion of ESAT-6 and CFP-10 (Simeone *et al*., 2009). A third ESX-1 substrate namely, EspA, was later identified (Fortune *et al*., 2005). Unlike ESAT-6 and CFP-10, EspA is encoded at a locus (Rv3614c-Rv3616c locus) distant from ESX-1 (Abdallah *et al.,* 2007), however is co-dependent on ESAT-6 and CFP-10 for secretion (McLaughlin *et al*., 2007). In the absence of EspA, the ESAT-6-CFP-10 dimer is produced efficiently, but is retained in the cell (Abdallah *et al.,* 2007).

Mycobacterial species can be divided in two groups according to the presence of the ESX-5 gene cluster in the species, namely fast- growing and slow-growing. The duplication of ESX-5, seems to correspond to the growth of the slow-growing species (Gey van Pittius *et al*., 2006). Additionally, ESX-5 also plays an important role in the secretion of PE and PPE proteins, as well as virulence (Simeone *et al*., 2009).

The ESX-3 gene cluster on the other hand, present in all mycobacterial species, was found to be involved in growth, in particular, though regulation of iron/zinc homeostasis (Bitter *et al*., 2009). It was also determined that the ESX-3 gene cluster is regulated by iron in both *M. smegmatis* and *M. tuberculosis,* and by zinc in *M. tuberculosis* (Macaig *et al.,* 2007). As this gene cluster in the focus for our investigation, the evidence regarding its functionality, will be discussed in greater detail in the relevant sections that follow.

The functions for ESX-2 and ESX-4 remain unknown, but it is believed that ESX-4 represents the most ancestral T7S system in mycobacteria (Gey van Pittius *et al*., 2001). Although the five ESX secretion systems can function separately, there does seems to be a certain level of interaction between the proteins encoded by the ESX gene clusters. This is confirmed by the inability of the four ESX systems (ESX-2 to 5) to make up for the loss of virulence, caused when ESX-1 is deleted. This is especially fascinating since both ESX-1 and ESX-5 are both involved in virulence (Abdallah *et al*., 2007). This might imply that these two systems have independent roles in the consecutive steps in the cellular infection cycle.

The ESX systems also differ in their regulation, for example ESX-2 genes are up-regulated upon starvation, while genes from ESX-1 are down-regulated under the same conditions. Furthermore, ESX-3 is regulated by the iron dependant regulator (IdeR) and the zinc uptake regulons (ZUR) in *M. tuberculosis*, while ESX-4 is regulated by the alternative sigma factor SigM (Abdallah *et al*., 2007). After comparison of the ESX gene cluster regions (ESX-1 to ESX-5), a set of six genes were identified that are found in all 5 regions (Tekaia *et al.,* 1999). These genes probably encode for the essential components of the T7S system. Some genes, for example the PE and PPE genes, are shared by some of the ESX gene cluster regions. But most importantly, some ESX regions have genes that are unique to the individual gene clusters (Abdallah *et al*., 2007).

2.3.5 PE and PPE proteins

There are two families of genes present within the ESX gene cluster, which have no clear function in the secretion system, namely the PE and PPE gene families (Gey van Pittius *et al*., 2006). These two gene families are unique to mycobacteria, and take up almost 9% of

the entire genome of *M. tuberculosis* (Cole *et al*., 1998). Although the precise function remains unknown, it is proposed that they are involved in antigenic variation and pathogenesis. PE and PPE proteins share a number of characteristics with ESAT-6 and CFP-10, in that they are secreted proteins that do not have a classical secretion signal (Gey van Pittius *et al*., 2001) and form a tight 1:1 complex with one another (Strong *et al*., 2006). The RD1 encoded PPE protein Rv3873, interacts with CFP-10 and ESAT-6 (Teutschbein *et al*., 2007), and this implies that Rv3873 may be secreted together with the ESAT-6-CFP-10 complex (Abdallah *et al*., 2007).

As mentioned, the PE proteins are characterised by the presence of a proline-glutamic acid (PE) motif and PPE proteins are characterised by a proline-proline-glutamic acid (PPE) motif (Cole *et al*., 1998). Both protein families can be divided into subfamilies. The largest subfamily of the PE family, the polymorphic GC-rich-repetitive sequence (PGRS), contains proteins with multiple tandem repeats of glycine-glycine-alanine or glycine-glycine-aspargine (Gordon *et al*., 1999). The PE-PGRS repeats, were shown, in some way, to influence antigen processing and presentation (Gey van Pittius *et al*., 2006).

The PPE protein family on the other hand, can be divided into 4 subfamilies, where the PPE-SVP (SVP: serien-valine-proline), is the largest of these. This subfamily contains the amino acids glycine-X-X-serien-valine-proline-X-X-tryptophan, between the position 300-350 in the amino acid sequence (Gey van Pittius *et al*., 2006). Another subfamily, the major polymorphic tandem repeat (MPTR) PPE subfamily, is the second largest of these, and it is suggested that the PPE-MPTR, is likely to be cell wall associated (Gey van Pittius *et al*., 2006). Other PE-PPE proteins, for example, the PPE36 and PPE68 proteins (Gey van Pittius *et al*., 2006), are also cell-wall constituents, which influence the cellular architecture, colony morphology, as well as the interaction of the organism with other cells (Brennan *et al*., 2001). As the PPE gene, Rv2123, is up-regulated under low iron conditions, it led to the hypothesis that this gene may encode a siderophore involved in iron uptake (Rodriquez *et al*., 1999). Although the exact link between the ESX cluster regions and the PE and PPE genes, is still unclear, it is suggested that the most PE and PPE proteins are secreted by ESX-5 (Abdallah *et al*., 2007). PE35 and PPE68, required for *in vivo* growth, are both situated in ESX-1, and involved in pathogenesis, while PE5 and PPE4, also required for *in vitro* growth, are situated in ESX-3 (Gey van Pittius *et al*., 2006). The fact that the PE and PPE genes encode for approximately 4% of the total proteins in the organism, suggests that they fulfil very important functions in mycobacteria (Gey van Pittius *et al*., 2006).

2.3.6 Genes encoding regulators

Gene expression is regulated by various DNA proteins and sigma factors, which repress or activate transcription of their target genes. The transcription of the genes encoding for these regulators, may in turn also be regulated (Dubnau *et al*., 2003). For example the iron dependent regulator (IdeR) is up-regulated during macrophage infection (Maciag *et al*., 2009) in response to low iron conditions. This will be explained in greater detail further in this literature study.

2.4 IRON ACQUISITION AND REGULATION IN MYCOBACTERIA

Metals such as iron and zinc have several biological roles, both as structural and catalytic co-factors for proteins, and considered essential for growth and survival.

These metals are also crucial for microbial invaders, as bacteria pathogens must acquire this nutrient metal in order to cause disease, hence, these metals are required for both bacterial metabolism and growth, as well as for virulence (Kehl-Fie *et al*., 2010). Shortages of zinc, have been shown to potentially disrupt a number of bacterial processes crucial to infection. Zinc however, may be toxic at a high concentration, since it competes with other metals for binding to active centres of enzymes (Maciag *et al*., 2007). Similarly, iron is involved in many enzymes as a vital co-factor, including various hydroxylases, oxygenases and oxygen transferring enzymes, and many cytochromes of the cells, where it participates in important reactions involved in oxidative phosphorylation and energy production (Ratlege, 2004). Without an adequate supply of iron, these cells are unable to function correctly, and cannot generate enough ATP to meet their energy demands. Although *M. tuberculosis* fails to grow in the absence of iron (Rodriguez *et al*., 2002), iron is not always readily available in its mammalian-host. Subsequently, these organisms have developed specialised mechanisms for acquiring iron from their environment. These organisms must also control the levels of intracellular iron, because excess iron, too, may be toxic, due to its participation in reactions that generate toxic oxygen radicals from normal aerobic metabolic products (Rodriguez, 2006). Hence, mycobacteria regulate the amount of intracellular iron by processes sensing

iron concentrations, and consequently controlling its uptake and storage (Rodriguez *et al*., 2002).

2.4.1 Iron acquisition

Siderophores, are high-affinity iron chelators, that bind to ferric iron (Rodriguez, 2006), and allow for its uptake into the bacterial cell. *M. tuberculosis* produces two types of siderophores: 1) carboxymycobactin and 2) mycobactins (Domenech *et al*., 2001). These two types of mycobactins differ in the length of their alkyl substitution, consequently influencing their polarities and solubility (Rodriguez, 2006) and function complementary to each other. The more polar of the two, carboxymycobactin, is released into the surrounding growth medium, where it competes with the iron-binding molecules of the host (Domenech *et al*., 2006). Carboxymycobactin also has the ability to remove iron from the host iron-binding proteins (Gobin *et al*., 1996). The less polar mycobactin remains cell-associated and acts as an ionophore to facilitate iron transport into the mycobacterial cell. The iron is subsequently released from these mycobactins once inside th cell, by the enzyme reductase (Ratledge, 2004).

A large cluster of genes, is predicted to encode polypeptide synthase and nonribosomal peptide synthase, required for the production of the mycobactin core (Cole *et al*., 1998). The synthesis of these mycobactins is a major point for the regulation of iron uptake in these organisms. These molecules are produced exclusively, under iron limiting conditions, and when iron concentrations return to normal, the synthesis of these mycobactins are down regulated (Raghu *et al*., 1993).

Non-pathogenic mycobacteria, for example *M. smegmatis,* also produce mycobactin, however, only a small amount of carboxymycobactin. They secrete an alternative peptidic siderophore, termed exochelins (Rodriguez, 2006), which function in the same manner as the carboxymycobactin described above. Metalloregulatory proteins in the mycobacterium cell, sense the intracellular levels of the specific metal and mediate a response to restore the homeostasis in the cell (Maciag *et al*., 2007).

2.4.2 Metal-dependent regulation

There are two families of transcriptional regulators found in bacteria that control iron homeostasis, the diphtheria toxin repressor (DtxR) and the ferric uptake regulator (FUR) (Rodriguez, 2006). Two genes, *furA* and *furB,* encode proteins for the FUR family, while IdeR and SirR, form part of the DtxR family (Schmitt *et al*., 1995). IdeR is the best characterised in terms of function and structure, and is a 230-amino acid, metal and DNA binding protein (Pohl *et* al., 1999). IdeR has been shown to be activated by several metals *in vitro*, however, iron was determined to be the preferred metal for IdeR functionality (Chou *et al*., 2004). In the presence of iron, these iron bound transcriptional regulators, bind to the specific iron-regulated promoters, and modulate the transcription of downstream genes.

In most cases, binding of the regulator to the promoter, represses transcription, however, in some cases, it may activate it (Gold *et al*., 2001). The function of IdeR as a positive or negative regulator, is determined by the location of the IdeR binding site in the promoter region (Rodriguez, 2006). *M. tuberculosis* also relies on IdeR to maintain iron-homeostasis. Additionally, IdeR has been determined to also be responsible for the iron-dependent siderophore repression in *M. smegmatis* (Dusserget *et al*., 1996) and is also required for expression of *katG* and *sodA*, which encodes catalase-peroxidase and superoxide dismutase, two important components in oxidative stress defence (Dusserget *et al*., 1998). However, it is still unclear, how IdeR influences the oxidative stress defence system of *M. tuberculosis* (Rodriguez, 2006).

The *furB* promoter is in turn induced by Zn^{++}, which suggests that FurB (Zur) senses Zn^{++} concentration and consequently controls Zn sensitive genes (Canneva *et al*., 2005).

2.4.3 The role of ESX-3 in iron acquisition and growth

Since the ESX-3 gene cluster is transcriptionally controlled by IdeR and ZUR, suggests that ESX-3 might be involved in metal ion homeostasis of mycobacteria (Serafini *et al*., 2009). ESX-3 has been shown to be essential for *in vitro* growth of *M. tuberculosis* (Sassetti *et al*.,

2003) and in particular, associated with iron acquisition. ESX-3 is thought to assist in iron-acquisition via the mycobactin pathway, as ESX-3 knockout of M. smegmatis, have previously been shown to be unable to used iron-bound mycobactins (Siegrist *et al*., 2009). These ESX-3 knockouts, however, still appear to produce carboxymycobactin and export carboxymycobactin appropriately.

Mycobacteria are known to up-regulate the expression of these siderophores when starved for iron. Similarly, the ESX-3 gene cluster also seems to be up-regulated (Hobson *et al.,* 2002) by iron starvation, and repressed by increased iron and IdeR (Maciag *et al*., 2009). In *M. tuberculosis*, however, the ESX-3 gene cluster is thought to be regulated by both iron and zinc (Serafini *et al.,* 2009), but only by iron in, *M. smegmatis* (Shiloh *et al*., 2010), until only recently, when Loots *et al*., (2013), indicated an association of ESX-3 with zinc homeostasis also.

2.4.4 Role of ESX-3 in cellular metabolism

A previous study done by Loots *et al* (2013), investigated the role of ESX-3 in cellular metabolism, by examining a *M. smegmatis* ESX-3 knockout strain in comparison to its isogenic wild type parent strain, using a metabolomics research approach.

In their study, it was found that the intracellular concentrations of several amino acids in particular, were seen to be significantly altered in the ESX-3 knockout strain comparatively (Loots *et al*., 2013). As the expression of several amino acid biosynthesis genes are known to be regulated by iron and / or zinc in these mycobacteria and other bacterial species (Rodriguez *et al.,* 2002 ; Maciag *et al.,* 2007 ; Maciag *et al.,* 2009), in addition to the previous associations of the ESX-3 gene cluster with iron uptake in mycobacteria, Loots *et al*. (2013), could clearly link the changes in these amino acid metabolites in the ESX-3 knockouts, to the proposed role of ESX-3 in iron and / or zinc homeostasis.

Due to the previously proposed role of ESX-3 in mycobactin synthesis and its role in the uptake of iron and zinc, it was initially expected that the *M. smegmatis* ESX-3 knockouts,

would experience both iron and zinc depletion. However, as described in their study, the metabolite profiles of the ESX-3 knockouts, in comparison to that of the wild type *M. smegmatis* parent strain investigated, proposed that the ESX-3 knockouts experience an iron overload. Siegrist *et al.* (2009), observed increased mycobactin expression in *M. bovis* BCG when ESX-3 expression was repressed. This they ascribed to the occurrence of a greater perceived iron starvation, resulting in the upregulation of various other iron acquisition genes, overcompensating for this. Considering this, Loots *et al*., (2013) suggested that the iron overload in the ESX-3 knockouts, is most likely also due to the fully functional exochelin iron uptake system, overcompensating for the perceived iron depletion, due to the inhibited mycobactin iron uptake system.

2.5 *MYCOBACTERIUM SMEGMATIS*

M. smegmatis, is a widely and acceptable model used for the better understanding of mycobacteria in general, and especially *M. tuberculosis* (Shiloh *et al*., 2010), and by directly comparing various experimental findings in this model to that of *M. tuberculosis,* have highlighted a number of similarities, and differences, between these species, and subsequently contributed to an increased understanding of *M. tuberculosis,* and its pathogenic characteristics. It has been also been especially useful, for investigating the ESX gene clusters. The genome of *M. smegmatis* contains three of the five ESX gene clusters, namely region 1, 3 and 4, with regions 2 and 5 absent (Gey van Pittius *et al*., 2001). Although ESX-1 is required for virulence of various pathogenic mycobacteria, studies in *M. smegmatis* have suggested that ESX-1 in this species is more focused towards DNA conjugation (Siegrist *et al*., 2009). As previously mentioned, the ESX-3 gene cluster region in this organism, is also associated with *in vitro* growth, however, unlike *M. tuberculosis,* it is not essential for *in vitro* growth of *M. smegmatis.* Thus, a deletion of the entire region, significantly slows its replication and growth, but does not totally prevent it, and hence the effects of this on the metabolome can be investigated. A further advantage of this species, is that it is, under normal circumstances, fast-growing, with a doubling time of 3h, (compared to 18-24h for *M. tuberculosis)*, is non-pathogenic (Shiloh *et al*., 2010) and can be cultured under Biosafety level 2 conditions (Kana *et al*., 2004), with no risk for infection to the investigating researchers.

2.6 METABOLOMICS

Metabolomics is a study of the unique chemical fingerprints that specific cellular processes leave behind and primarily the study of small-molecule metabolite profiles. The metabolome represents the collection of all metabolites in a sample, the intermediates and the end products of cellular processes. (Kaddurah-Daouk *et al*., 2008). This research approach has been proven to be a rapid and superior post-genomics technology, for pattern-recognition analyses of biological samples (Weckwerth *et al*., 2005). Another feature of this technique, is that it analyses all metabolites, thus allowing an open, unbiased approach to identifying biomolecules important for specific biological processes (Van der Werf *et al*., 2007). While unbiased approaches aims to cover the metabolome as broadly as possible, and at the same time maintain the ability to at least differentially quantify the metabolites, it also allows detection of previously unknown or poorly characterised metabolites (Orešič, 2009). Metabolomics studies have not only the capacity to identify metabolites that are endogenous to the organism, but also as those exogenously influencing the organism (Kaddurah-Daouk *et al*., 2008).

Sampling approaches, quenching methods and extraction methods are crucial in metabolomics. In microbial metabolomics it is critical that cellular lysis be prevented during quenching to be able to discriminate between extracellular and intracellular metabolites (Van der Werf *et al*., 2007). The global metabolome, however, cannot be measured using one single analytical approach, thus different analytical approaches are combined to measure the whole metabolome. Chromatography is an ideal analytical approach for metabolomics.

Various types of chromatographic setups are used for metabolomics, for example liquid chromatography (LC) and gas chromatography (GC), which can be coupled to various mass spectrometry (MS) systems. A variety of MS systems are available, for example the quadrupole MS, ion trap MS and the time of flight MS (TOF-MS).

The choice of analytical instrumentation is determined by die goal of the study, as each instrument has its own advantages and disadvantages. The application of biostatistics and novel mathematical frameworks, also plays a major role in identifying the biological meaningful information obtained from the huge datasets generated by these analytical approaches (Weckwerth *et al*., 2005).

2.7 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS)

Although there is a magnitude of analytical approaches available for applications to metabolomics studies, two dimensional gas chromatography coupled tandem mass spectrometry will be discussed due to its application to this metabolomics study. The coupling of a two dimensional GC (GCxGC), to time-of-flight (TOF) mass analysers is a relatively new and promising technology for metabolomics. The capacity of GCxGC-TOFMS provides an improvement over conventional GC-MS analysis, due to its improved capacity for analysing complex samples, as it overcomes the problem of compound co-elution (Weckwerth *et al*., 2005). GCxGC-TOFMS is currently considered the most sensitive analytical approach available for compound identification, in full-scan acquisition mode (Van der Werf *et al*., 2007). In principle, the method consists of two GC systems (GCxGC) equipped with columns of different polarity, connected by interface with a cryogenic trap (Kalinova *et al.,* 2006). The cryogenic trap repeatedly condenses compounds eluting from the primary polar column and release them at regular intervals to the secondary column. TOF-MS consists out of a flight tube which accelerates ions according to their different velocities, due to their different mass to charge ratio (m/z), which enables TOF-MS to analyse ions of any m/z value. TOF-MS has the capacity for extremely fast scan rates and further enhances the systems accuracy by improving sensitivity, without spectral shifting, allowing for better deconvolution and compound identification.

Chapter 3

Problem statement, Aims, Objectives and Experimental design

Chapter 3
PROBLEM STATEMENT, AIMS, OBJECTIVES AND EXPERIMENTAL DESIGN

3.1 PROBLEM STATEMENT

The *Mycobacterium* genome contains various ESX-gene clusters (ESX 1-5). The ESX-3 gene cluster has been associated with growth and viability of these organisms. These genes have been implicated in the coding of various proteins, mycobactins in particular, involved in iron uptake. A previous metabolomics study done by Loots *et al*., (2013), identified metabolites in ESX-3 knockouts of *M. smegmatis*, suggesting iron overload, despite the absence of the mycobactins. This is hypothesized to be due to an increased exochelin synthesis, another iron uptake protein not associated with ESX-3, overcompensating for the perceived iron depletion in the knockout organism. In this study we used metabolomics to generate information which could further support this hypothesis, and possibly contribute to the understanding of ESX-3 in iron uptake and *Mycobacterium* cell viability.

3.2 AIM

The aim of this study is to:

1) Determine the optimal extraction method for characterising the metabolome of the *M. smegmatis* wild-type and the *M. smegmatis* ESX-3 knockouts, using a GCxGC-TOFMS metabolomics research approach, and

2) Use this method to characterise the metabolome of *a M. smegmatis* wild-type, grown in normal, low and high iron conditions, using a GCxGC-TOFMS metabolomics approach, and compare the altered metabolome induced by varying iron to that of the ESX-3 knockout of *M. smegmatis,* previously hypothesised to experience iron overload, in order to further substantiate previously generated hypotheses suggesting the role of ESX-3 in iron uptake, and the associated metabolome.

3.3 OBJECTIVES

The above-mentioned aims will be accomplished by completing the following objectives:

1. The comparison of two extraction methods, prior to GCxGC-TOFMS analyses, for applications to this metabolomics investigation of *M. smegmatis* bacterial cultures. These methods will be compared with regards to their respective repeatability, extraction efficiency and ability to differentiate between a *M. smegmatis* ESX-3 knockouts and its wild-type parent strain.

2. Using the most optimal extraction method as determined in Objective 1, followed by GCxGC-TOFMS analysis of the separately cultured *M. smegmatis* wild-type, grown in normal, low and high iron conditions, determine the influence of varying iron concentrations on the metabolome of *this organism*, by metabolomics comparisons of these groups.

3. Using the above GCxGC-TOFMS metabolomics methodology, compare the separately cultured ESX-3 knockout *M. smegmatis* strain to that of the wild type parent strain, grown under normal iron concentrations, and interpret the metabolite markers in the light of the information generated in Objective 2, in order to confirm whether or not ESX-3 is associated with an increased iron uptake, as was proposed by a previous metabolomics investigation by Loots *et al*. (2013).

3.4 EXPERIMENTAL DESIGN

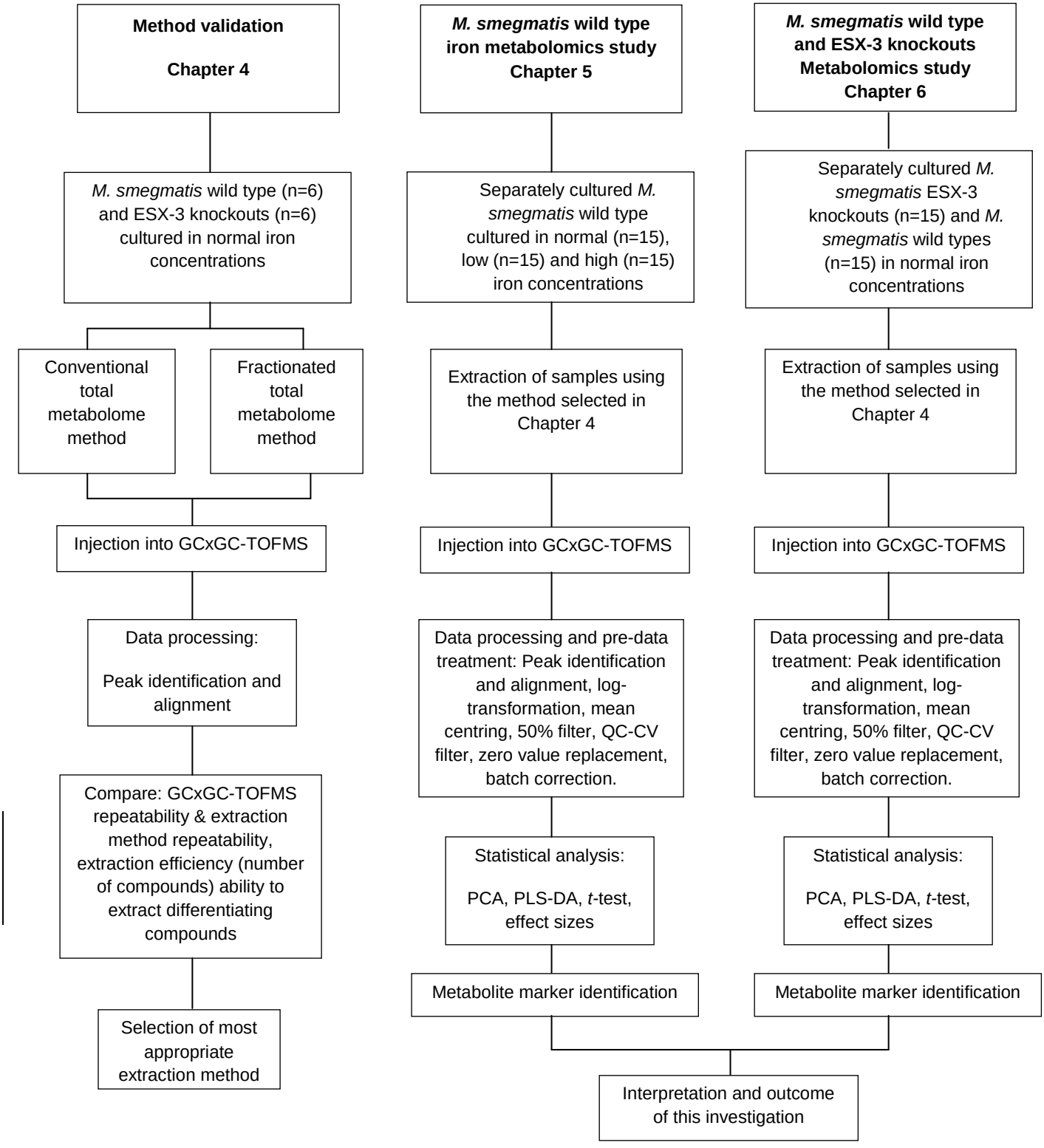


Fig 3.1: Schematic representation of the experimental design of this metabolomics experiment

Chapter 4

Method development and validation

Chapter 4

METHOD DEVELOPMENT AND VALIDATION

4.1 INTRODUCTION

The relatively new research field of metabolomics, can be defined as 'the non-biased identification and quantification of all the metabolites in a biological system', using very sensitive analytical procedures" (Dunn *et al*., 2005). Recently, metabolomics has effectively been used as a tool to characterise a variety of disease states, in order to better understand the underlying mechanisms of these diseases (Schoeman *et al.*, 2011). In the current investigation, we will use metabolomics to identify metabolite markers, better characterising the metabolome of *M. smegmatis* grown in a various iron concentrations, comparing these to one another, and then to an ESX-3 knockout of the same organism, in order to confirm previous metabolomics generated hypotheses, that ESX-3 is involved in iron uptake and that an ESX-3 knockout of M. smegmatis, results in iron overload and a the characteristic metabolite associated with this (Loots *et al*., 2013).

To accomplish our first Aim as described in 3.2, we compared two extraction methods, and investigated their repeatability and respective capacities to extract those compounds which best differentiate the *M. smegmatis* wild-type and ESX-3 knockouts. The first extraction method was chosen since it extracts the total metabolome (Olivier *et al*., 2012), considering that the definition of metabolomics is the "unbiased identification of all metabolites present in a biological system" (Dunn *et al.,* 2005). The second extraction method comparatively investigated, was a modification of this total metabolome method, however, fractionating the polar and non-polar phases from one another before separately derivatising and analysing these. The reason for this is that the total metabolome extracts, due to the fact that they potentially contain all compounds from all compound classes, may potentially contain more compounds and have a greater dimensionality than that of the analytical apparatus. Although the GCxGC-TOFMS already has a substantially larger dimensionality than a conventional single dimensional GC-MS, some information may potentially be lost due to peak overlap when analysing very complex samples, and hence the reason for this investigation and comparison of methods for GCxGC-TOFMS metabolomics.

4.2 MATERIALS AND METHODS

4.2.1 Reagents and chemicals

Ampicillin and X-gal was purchased at Roche. Kanamycin, methoxyamine hydrochloride, n-methyl-n-(trimethylsilyl) trifluoroacetamide (MSTFA) + 1% trimethylchlorosilane (TMCS), pyridine, methanolic HCl (3M), nonadecanoic acid, 3-phenylbutyric acid, and tricarballylic acid, were all purchased from Sigma Aldrich (St. Louis, MO). All organic solvents used (chloroform, methanol and hexane) were ultra-pure Burdick and Jackson grade (Honeywell International Inc., Muskegon, MI, USA).

4.2.2 Preparation of *M.* smegmatis ESX-3 knockouts, culturing conditions and sample preparation.

All *M. smegmatis* cultures were obtained from the DST/NRF centre of excellence for biomedical TB research (CBTBR), Stellenbosch University, South Africa.

This study was approved by the NWU ethics committee (NWU-00127-11-A1) and Stellenbosch ethics committee (N10/40/126), and performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

E. coli strain JM109 was used in all cloning procedures. *E. coli* was cultured in Luria-Bertani (LB) broth with shaking, and on LB agar plates, overnight at 37 ℃. Solid and liquid media were supplemented with the antibiotics ampicillin (50 µg/ml, Roche) and kanamycin (50 ug/ml, Sigma) and solid media with X-gal (Roche) and 5% sucrose, as appropriate. *M. smegmatis* mc^2155 was used as the parental strain to make the *M. smegmatis* ESX-3 knockout. For generation of the ESX-3 knockout, *M. smegmatis* was grown in liquid LB with shaking, and on LB agar plates, for 4-5 days at 37 ℃ (LB medium was used as the sugars in Middlebrook 7H9 medium are toxic to single crossover knockout cultures containing the counterselectable marker sacB gene). Liquid LB medium was supplemented with 0.1% Tween-80, and kanamycin (25µg/ml, Sigma) when required. Solid LB media were

supplemented with kanamycin (25µg/ml, Sigma), X-gal (Roche) and 5% sucrose, as appropriate.

The ESX-3 gene cluster was deleted from the genome of *M. smegmatis* mc^2155 using homologous recombination, as described by Gordhan and Parish (2001). Briefly, approximately 800bp of DNA directly upstream of MSMEG_0615 and downstream of MSMEG_0626 was amplified from *M. smegmatis* mc^2155 genomic DNA using primers containing the indicated restriction sites; R3 Upsmeg f (5'-GGGGTACCGGAGCATCCGCTGCAGACC-3', *KpnI*) and R3 Upsmeg r (5'-GGGGAGATCTCTCTCCCTTATGTATGCC-3', *BglII*) and R3 Downsmeg f (5'-GGGGAGATCTCGATCCCAGTGCTCCCACA-3, *BglII*) and R3 Downsmeg r (5'-GGGGAAGCTTCCCGAGCGATCCTTTCC, *HindIII*), respectively. The R3 Upsmeg and R3 Downsmeg fragments were simultaneously cloned into the *KpnI* and *HindIII* restriction sites of the p2NIL suicide vector (Parish and Stoker, 2000) via three-way cloning, to create p2NIL_R3UpDownSmeg. The *lacZ-sacB* marker gene cassette from pGOAL17 (Parish and Stoker, 2000) was then inserted into the *PacI* site of p2NIL_R3UpDownSmeg to create the suicide vector p2NIL_R3 KO. p2NIL_R3 KO was transformed into *M. smegmatis* mc^2155. Single recombinants were selected on kanamycin and X-gal, and passaged to allow a second recombination event to occur. Double recombinants were counterselected on sucrose. Wild type and ESX-3 knockout *M. smegmatis* were differentiated by PCR using primers specific for either the wild type strain (R3 F1 f: 5'-GCAGTGGTTCTCCGAGCGTGG-3' and R3 F1 r: 5'-ACGACGTCCGACCAGCGTTGG-3') or the ESX-3 knockout (R3 KO f: 5'-TCCTTCTTTGCGCTGGTCTT-3' and R3 KO r: 5'-TGTCGCTGCCGTGGTTCT-3'). The *M. smegmatis* ESX-3 knockout was confirmed by sequencing across the deleted region.

For stock preparation, *M. smegmatis* wild type (n = 6) and ESX-3 knockout strains (n = 6) were separately cultured in Difco Middlebrook 7H9 medium (BD) supplemented with 0.5% glucose, 0.2% glycerol, 0.05% Tween-80 and 10% ADC (bovine serum albumin-dextrose-catalase) to mid-log phase (OD_{600nm} = 0.8). These cultures were grown at 37℃ for 3 days. Colonies were scraped from each plate into a microtube, snap-frozen in liquid nitrogen, lyophilized and stored at -80ºC until extraction and GCxGC-TOFMS metabolomic analysis.

In order to compare the various extraction methods described below, 6 individually cultured sample repeats of both the wild-type and ESX-3 knockouts (5 mg ~ 1 X 10^8 cells) were extracted via each of the compared methods.

These numbers of individually prepared samples per group are sufficient for investigating the repeatability of the compared extraction methods for metabolomics applications (Schoeman *et al*., 2012). For the metabolite marker identification experiments using these developed methods, however, larger groups were used (Chapter 5).

4.2.3 Extraction methods

Two extraction methods: 1) the conventional total metabolome extraction method as described by (Olivier *et al*., 2012) and 2) a fractionated approach to this total metabolome extraction method (allowing for separate analysis of the non-polar and polar fractions), were compared with regards to their respective repeatability, extraction efficiency, and ability to extract those compounds best differentiating between *M. smegmatis* wild-type and ESX-3 knockout bacterial strains on the basis of the extracted metabolome / metabolite data. The remaining pellet was also extracted in order to determine if any valuable biological data can be generated from these fractions in order to differentiate the groups.

4.2.3.1 The conventional total metabolome extraction method

The conventional total metabolome extraction method was specifically chosen due to its successful applications to many previous metabolomics investigations (Loots *et al.,* 2013), and capacity to extract a large variety of compounds, including amongst others; organic acids, fatty acids, amino acids, alkenes, alkanes, and sugars, with the resulting metabolite profile providing a holistic overview of the "total metabolome", maximising the chance of identifying metabolite markers (Olivier *et al*., 2012) characterising the alterations to the metabolome of *M. smegmatis* due to the ESX-3 knockouts. As internal standards, 100µL of

3-phenylbutyric acid (0.0265 g.50 mL^{-1}) and 100µL nonadecanoic (0.0197 g.50 mL^{-1}) acid were added to the 5 mg (~1 X 10^8 cells) of the above mentioned cell culture samples. This was followed by the addition of 750µL methanol, 150µL chloroform and 100µL ddH_2O. A 3-mm carbide tungsten bead (Retsch GmbH & co. KG, Haan, Germany) was subsequently added to the mixture followed by shaking in a vibration mill (Retsch GmbH & co. KG, Haan, Germany) at 30 Hz, for 5 minutes.

After centrifugation of the sample suspension at 12000 rpm (Boeco Centrifuge M-24 A, 128.79g-force), the supernatant was transferred to a GC-MS sample vial and dried under a light stream of nitrogen at room temperature. The remaining pellet (cell debris), was also extracted as described in section 4.2.3.3. The dried supernatant was subsequently derivatised using 50µl of methoxyamine hydrochloride in pyridine (15 $mg.mL^{-1}$) at 40˚C for 90 minutes, and trimethylsilylated with 50µl MSTFA with 1% TMCS at 40˚C for 60 minutes (Kanani *et al*., 2007).

4.2.3.2 The fractionated total metabolome (non-polar and polar phase) extraction method

As an internal standard, 100µL (0.0197 g.50 mL^{-1}) of nonadecanoic acid was added to 5 mg (~1 X 10^8 cells) of the above mentioned samples, followed by the addition of 750µL methanol, 150µL chloroform and 100µL ddH_2O. After adding a 3-mm carbide tungsten bead (Retsch GmbH & co. KG, Haan, Germany), the solvent mixture was shaken in a vibration mill (Retsch GmbH & co. KG, Haan, Germany) for 5 minutes at 30 Hz. Hereafter, 250µL ddH_2O and 250µL chloroform were added to the sample suspensions, and mixed in a rotary shaker for 30 minutes. After centrifugation, the non-polar (chloroform) phase was collected and transferred to new tube. The polar (water) phase was re-extracted with 500µl of chloroform, and similarly, the non-polar phase was re-extracted with 500µl ddH_2O. At this point, a second internal standard, 100µL of 3-phenyl butyric acid (0.0265g/50ml), was added to each of the extracted phases (polar and non-polar). This internal standard was only added at this point because it can be extracted in both phases, and therefore its addition before the phase separation may lead to an inconsistent distribution between the phases. The remaining pellet (cell debris), was also extracted as described in section 4.2.3.3. The extracts were then dried under a light stream of nitrogen at room temperature, derivatised with 50µL methoxyamine hydrochloride in pyridine (15 $mg.mL^{-1}$) at 40˚C for 90 minutes, and

trimethylsilylated with 50µl MSTFA with% TMCS at 40˚C for 60 min. The remaining pellet was collected and analysed as described below.

4.2.3.3 Pellet extraction method

The remaining pellets from both the conventional total metabolome method and the fractionated total metabolome method were also further investigated. A lipid extraction method was done in order to determine if the remaining cell wall components and cell debris, contains lipid metabolites which may contribute to this metabolomics investigation. As an internal standard, 50µL of nonadecanoic acid (0.0197g.50mL^{-1}) was added to the pellet, followed by 1 mL methanolic HCl, and incubation at 80˚C for 90 min. After the sample suspension was cooled to room temperature, it was extracted twice with 1 mL hexane, after which the combined hexane phases were combined and dried under a light stream of nitrogen. The dried extract was then derivatised with 50µL MSTFA with 1% TMCS at 40˚C for 60 minutes.

4.2.4 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS)

All sample extracts (1µL) were injected in a 1:5 split mode on an Pegasus GCxGC-TOFMS supplied by LECO (LECO Africa (PTY) LTD) made up of an Agilent 7890A 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).

The primary column used was a 30 m Rxi-5Sil MS capillary column with an internal diameter of 0.25 mm and a film thickness of 0.25 µm (Retsch GmbH & co. KG, Haan, Germany). The secondary column was a 0. 900 m Rxi-17 with a 100 µm internal diameter and a 0.1 µm film thickness. Modulation of the primary column effluent onto the secondary column was performed using cryo-modulation and a hot pulse of nitrogen gas for 0.70 seconds, every 3

seconds, was used to initialize each new secondary column separation. The injector temperature was held constant at 280 ℃ for the entire run. The GC oven temperature for the primary column was initially held at 70 ℃ for 2 minutes and the temperature was then increased at a rate of 4 ℃.min^{-1} to a final temperature of 300 ℃, where it was kept constant for 2 minutes. The secondary column temperature programme was identical to that of the primary column, only with an offset of + 5 ℃. Helium was used as the carrier gas at a constant flow rate of 1 mL.min^{-1}.

The first 500 seconds of each run was considered a solvent delay and no mass spectra were recorded during this time. All GC×GC-TOFMS separations, however, were plotted with the 5 min included on the Column 1 time axis to reflect the correct retention times. The transfer line was held at 270 ℃ and the ion source at 230 ° C for the entire run time. The detector voltage was set at 1590 V and the filament bias at −70 eV. Mass spectra were collected from 50–600 *m*/*z* at an acquisition rate of 200 spectra per second.

4.2.5 Data processing

4.2.5.1 Peak identification and alignment

LECO Corporation ChromaTOF software (version 4.32) was used for peak identification and mass spectral deconvolution (signal to noise of 200). Compound identification was done by comparison of both mass spectra and the relative retention index of the detected compounds, to that of libraries compiled from previously injected standards. In order to eliminate the effect of retention time shifts, peaks (across all samples) with similar mass spectra, were aligned before statistical data analyses, using an optional function of the ChromaTOF software, Statistical Compare. The peak alignment function of ChromaTOF aligns peaks with matching mass spectra between various samples within the corresponding retention time window, in order to eliminate potential retention time shifts between samples and subsequently generate a data matrix which includes compound names, retention times and detected areas, which can be used for statistical data processing. All peak areas were normalised relative to the respective internal standards by calculating the relative concentration of each compound.

For the total metabolome and polar phase extraction methods, 3-phenylbutyric acid was used to normalise the peaks, whereas nonadecanoic acid was used for the non-polar phase and pellet extractions. The final data matrixes were then subjected to statistical data analysis.

4.2.6 Statistical data analysis

The coefficient of variation values (CV values) were calculated using the relative concentrations of all the compounds detected after GCxGC-TOF/MS analysis of the extracted and injected samples. The distribution of the CV values of all the detected compounds was then used as a measure to compare the repeatability of the various extraction methods.

The multivariate statistical data analyses were done using MetaboAnalyst, a web server for metabolomics data analysis and interpretation based on the statistical program "R" version 2.10.0 (Xia *et al.*, 2009). This included principal component analysis (PCA) which is an unsupervised method used to summarise data and determine if a natural separation exists between sample groups. In this case, PCA was performed to determine if natural separation exists between the sample groups, *M. smegmatis* wild-type and ESX-3 knockouts, and in so doing ascertain whether or not the compared methods extract those compounds best differentiating and potentially characterising the groups.

4.3 RESULTS AND DISCUSSION

4.3.1 Extraction efficiency

The extraction efficiency of each of the investigated extraction methods was investigated, in order to evaluate the compared extraction methods application to *M. smegmatis* bacterial metabolomics characterisation.

This was done by comparatively considering the number of peaks detected, as well as the intensities of the internal standard areas, after GCxGC-TOFMS analyses of the separately extracted sample repeats, for each extraction method (Schoeman *et al*., 2012) (Table 4.1). The number of compounds extracted by each of the evaluated methods were ranked as follows: non-polar phase extracts, extracted via the fractionated total metabolome method (303 ± 26) > the polar phase extracts, extracted via the fractionated total metabolome method (244 ± 12) > the conventional total metabolome extracts (241 ± 11). When comparing the extractions on the remaining pellet of each of the above mentioned methods, the fractionated total metabolome method's pellet extraction (251 ± 40) resulted in comparatively more compounds when compared to that of the remaining pellet extracts from the conventional total metabolome method (226 ± 29). For the *M. smegmatis* ESX-3 knockout sample group, GCxGC-TOFMS analyses of the non-polar phase extracts, extracted via the fractionated total metabolome method, once again led to the detection of the most compounds (306 ± 38) > the combined total metabolome extracts (287 ± 95) > the polar phase extracts (269 ± 24). On the other hand, the pellet extracted via the total metabolome method (280 ± 47) > the pellet phase extracts, extracted via the fractionated total metabolome method (233 ± 25) respectively. Due to the fact that the polar extracts and the non-polar extracts represent two separate analyses of the total metabolome fraction, the compounds detected via this method should be combined, in order to comparatively evaluate this against the conventional total metabolome analysis of these samples. When combining the number of peaks detected in the polar and non-polar fraction of the fractionated total metabolome method, it is important to note that both fractions may contain many of the same compounds. The total number of combined unique peaks identified for the fractionated total metabolome method were 340 ± 30 for the *M. smegmatis* wild-type sample group, and 335 ± 41 for the *M. smegmatis* ESX-3 knockout sample group, indicating this approach to potentially extract and identify comparatively more compounds then the conventional total metabolome extraction method, as would be expected. However, although more compounds were detected using the fractionated total metabolome extraction approach, potentially more "noise" or rubbish may have been detected also i.e. compounds of no biological value to this metabolomics study.

Subsequently, a more detailed look at these compounds are required, and among the 340 combined unique peaks previously mentioned in the *M. smegmatis* wild-type sample group, only 235 compounds could be annotated by comparison to the in-house and commercially

available libraries, of which 25 of these 235 compounds, were annotated as inorganic compounds, not of biological origin. Therefore, only 210 of these compounds are of potential biological importance to this study. Similar evaluations of the 335 compounds extracted via the same fractionated approach for the *M. smegmatis* ESX-3 knockout sample group, indicated only 230 compounds, of these to be of potential biological relevance to this investigation. Comparatively, similar evaluations of the compounds detected via the conventional total metabolome method however, in the *M. smegmatis* wild-type sample group, indicated 212 compounds were of potential biological importance, and in the *M. smegmatis* ESX-3 knockout sample group, 240 compounds.

Considering this, although the fractionated total metabolome extraction method detected comparatively more compounds, far more metabolite "noise" or inorganic contaminants were detected, and the conventional total metabolome method, subsequently detected more compounds of potential biological importance to this study.

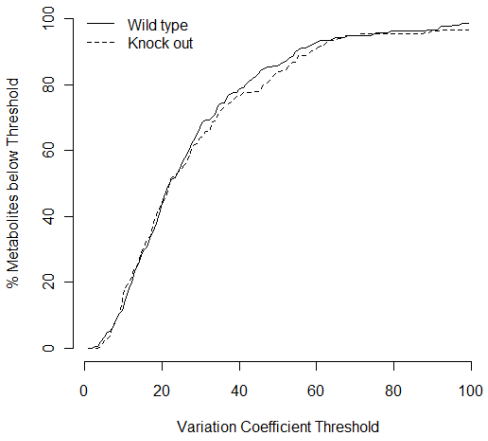
As explained by Schoeman *et al*., (2012), apart from the number of compounds detected, the comparative efficiency of the extraction methods can also be evaluated by comparing the intensities at which these extracted compounds are detected. In order to accomplish this, the average area of the internal standard, detected after applying each extraction method can be compared (Table 4.1). These areas should, in theory, be identical, due to the fact that the exact same amount and concentration of the internal standard was added to all the analysed samples. As seen in Table 4.1, nonadecanoic acid was detected at a higher intensity in the conventional total metabolome extracts than in the other fractionated extracts, extracted via the fractionated total metabolome method. Similarly, 3-phenylbutyric acid was also detected at a higher intensity in the combined total metabolome phase extracts than in the other fractionated extracts, extracted via the fractionated total metabolome method. This indicates that, in both sample groups, that apart from detecting more compounds of biological importance, the conventional total metabolome method extracts / identifies these with a better sensitivity or compound intensity.

Table 4.1: The number of peaks and internal standard areas detected after GCxGC-TOFMS analyses of 6 sample repeats extracted via each of the investigated extraction methods. Standard deviations are indicated in parenthesis

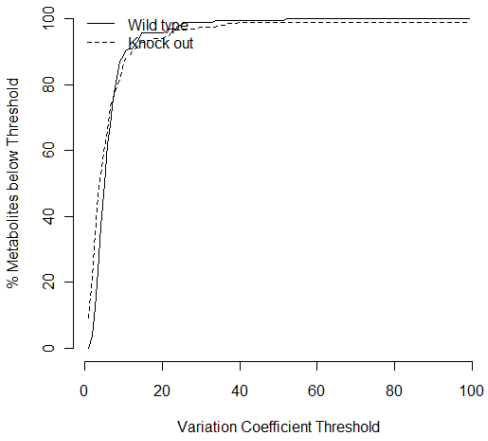
			Conventional total metabolome method		Fractionated total metabolome method		
	Internal Standard		**Combined total metabolome extract**	**Pellet**	**Non-polar Phase**	**Polar phase**	**Pellet**
Number of peaks identified (± standard deviation)		***M. smegmatis* Wild type**	241 (11)	226 (29)	303 (26)	244 (12)	251 (40)
					340 (30) (combined unique peaks)		
		***M. smegmatis* ESX-3 Knockout**	287 (95)	280 (47)	306 (38)	269 (24)	233 (25)
					335 (41) (combined unique peaks)		
Average internal standard area for all analysed samples (± standard deviation)	**3-Phenylbutyric acid (100 µl)**	***M. smegmatis* Wild type**	6819684 (1469026)	57058 (21456)	1827776 (2324346)	1917038 (1498458)	86633 (58915)
	Nonadecanoic acid (100 µl)		2729503 (617748)	57058 (21456)	319620 (193251)	-	36633 (8915)
	3-Phenylbutyric acid (100 µl)	***M. smegmatis* ESX-3 Knockout**	8526757 (2093965)	102321 (19739)	1782207 (1327392)	2191894 (6709405)	104076 (87477)
	Nonadecanoic acid (100 µl)		2863395 (624229)	102321 (19739)	284466 (131037)	-	100076 (87477)

4.3.2 Extraction repeatability

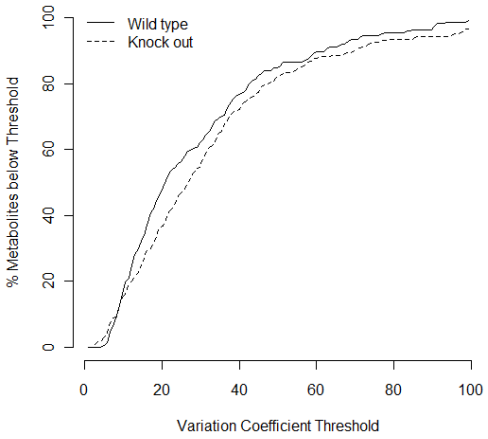
The comparative repeatability of each method was determined by calculating the CV values of the relative concentrations of all the compounds detected after GCxGC-TOFMS analysis of the 6 sample repeats, extracted via each of the investigated extraction procedures (Fig 4.1).



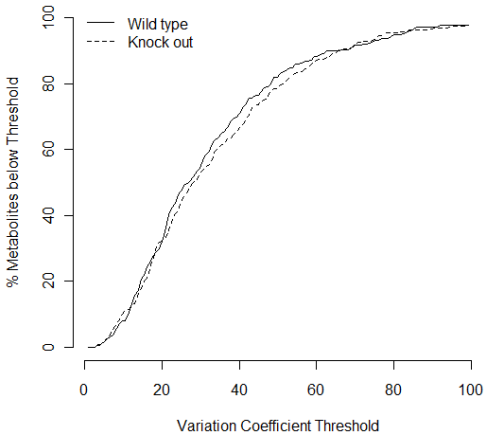
a) Conventional total metabolome extraction method



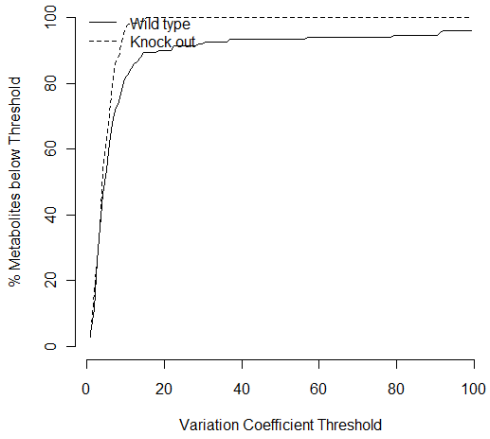
b) Pellet resulting from total metabolome extraction method



c) Fractionated total metabolome method: polar phase



d) Fractionated total metabolome method - non-polar phase



e) Pellet resulted from fractionated total metabolome method

Fig 4.1: Distribution of the coefficient of variation of all the compounds detected after GCxGC-TOFMS analyses of the samples extracted via the various extraction methods investigated.

When comparing these results, it is important to keep in mind that a compound's CV value is directly correlated to its detected concentration, and hence, compounds with lower concentrations, are prone to have higher CV values (Kindt *et al*., 2009). However, as metabolomics is an unbiased approach, involved in detecting "all" compounds in a mixture (including being unbiased to concentration), compounds occurring in lower concentrations are potentially also of interest. However, these may be accompanied by larger CV values. For target compound analyses, the Food and Drug administration (FDA) recommends a CV value of 20% (Kindt *et al*., 2009), but as this is an unbiased metabolomics approach, a hypothetical CV value of 50% was shown to work well for these evaluation purposes and data in the data base pre-treatment steps which follow (Schoeman *et al*., 2012). These results are indicated in Fig 4.1 and also summarised in Table 4.2, and indicates that 86% of all the compounds detected via GCxGC-TOFMS analyses after performing the conventional total metabolome method of the *M. smegmatis* wild type, has a CV value below this hypothetical 50% cut-off, followed by 85% of all compounds detected in the polar phase of the fractionated approach, 83% of all compounds detected in the non-polar phase of the fractionated approach, 57% of all compounds detected in the pellet extract of the fractionated approach and 53% of all compounds detected in the pellet extracted via the conventional total metabolome method, respectively. Similarly, for the *M. smegmatis* ESX-3 knockout sample group 84%, 82% and 79% of all compounds detected after GCxGC-TOFMS analyses of the extracts, extracted via the conventional total metabolome method,

the polar and non-polar fractions from the fractionated total metabolome method, respectively, had CV values below 50%, confirming the above hierarchical order.

For further comparative purposes, using more conventional CV interpretations, the calculated CV values for 20 compounds, representative of various compounds classes and detected at regular retention time intervals throughout the total chromatographic run, are also given in Table 4.2. As can be seen, all 20 compounds selected were detected in the conventional total metabolome extracts. The pellet only extracted a select few of these compounds, which would be expected as the extraction method was focussed to fatty acids. The non-polar phase, prepared using the fractionated total metabolome method, extracted mostly non-polar compounds, which was also to be expected. Similarly, the polar phase, extracted via the fractionated total metabolome method, extracted mostly polar compounds. However, the CV values for these compounds were considerably better for the conventional total metabolome extraction method for all 20 compounds. This could be ascribed to a number of reasons, including the comparative ease of the method, requiring less analytical steps to complete. Perhaps even a bigger contributor to the lower repeatability of the fractionated approach, is that some of these compounds may be extracted in both the polar and non-polar fractions, however, not consistently in the same amounts every time, negatively influencing their CV values.

Considering these results it is clear that the conventional total metabolome method is the more repeatable of the two approaches being compared, for both the *M. smegmatis* wild type and ESX-3 knockout sample groups analysed in this metabolomics investigation.

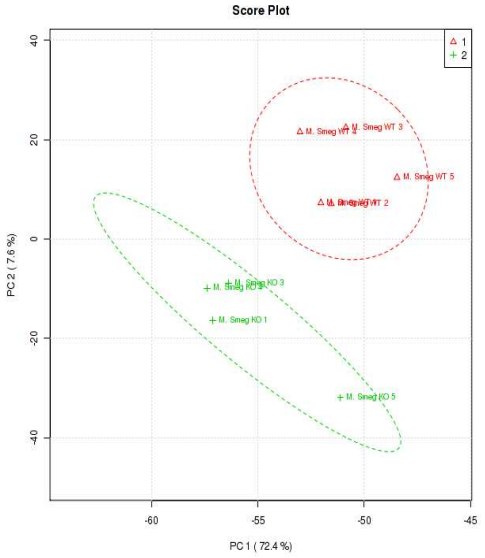
Table 4.2: Coefficients of variation (CV) values for ten compounds, representative of various compounds classes and detected at retention times and percentage of compounds detected, after GCXGC-TOFMS analyses of the extracts obtained via the two extraction methods.

		Extraction method						
				Conventional total metabolome method		Fractionated total metabolome method		
		Retention time (sec)		Combined total metabolome phases	Pellet	Non polar Phase	Polar Phase	Pellet
		1st dimension	2nd dimension					
% of compounds with a CV <50 %	*M. smegmatis* wild type			86%	53%	83%	85%	57%
	M. smegmatis ESX-3 Knockout			84%	58%	79%	82%	65%
CV values (%) calculated for the relative concentrations of compounds detected in *M. smegmatis* wild type	Oxalic acid	611	0.66	3.97	29.34	ND	18.78	33.66
	Hydroxypyruvic acid	650	0.905	19.40	ND	ND	27.74	ND
	Propanoic acid	716	1.16	16.24	39.91	6.64	24.71	24.95
	Acetic acid	755	0.85	19.88	ND	20.95	14.64	ND
	Hexanoic acid	788	0.945	16.29	18.85	50.32	48.51	26.91
	Hydroxybutyric acid	923	0.845	22.54	ND	ND	ND	ND
	L-Norleucine	1160	0.9	25.48	ND	ND	23.04	ND
	L-Isoleucine	1205	0.9	24.33	ND	21.99		ND
	Glycine	1217	0.845	19.32	ND	ND	27.09	28.12
	L-Aspartic acid	1516	0.92	11.09	ND	ND	35.17	ND
	Alanine	1679	1.4	29.76	61.73	30.85		ND
	5-oxo-Proline	1617	1.26	8.19	ND	31.15	20.59	ND
	Glutamine	1799	1	8.19	ND	ND	25.83	ND
	Mannose	1868	0.835	7.94	ND	ND	ND	ND
	Adipic acid	2114	1.325	4.23	ND	23.21	6.76	ND
	Oleic acid	2669	1.005	11.41	ND	16.32	ND	ND
	Octadecanoic acid	2705	1.2	19.46	30.53	8.64	28.79	57.31
	Propenoic acid	2774	1.3	16.62	ND	40.72	15.81	ND
	Dehydroabietic acid	2888	1.9	11.17	ND	29.31	4.84	ND
	Hexadecanoic acid	3092	0.985	27.87	32.32	29.88	43.21	13.81

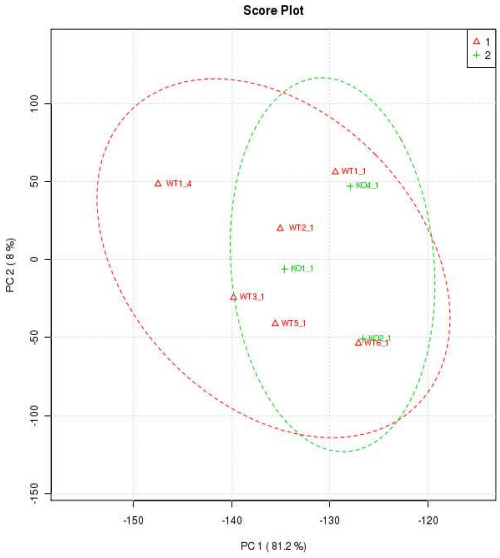
4.3.3 Ability to extract differentiating compounds

Another important factor in comparative metabolomics studies, in addition to the repeatability and extraction capacity of an extraction method, is the method's ability to extract those compounds most important for differentiating the investigated sample groups. The identification of these compounds will allow for a better understanding of the variation between the various sample groups, ultimately answering the specific biological question at hand.

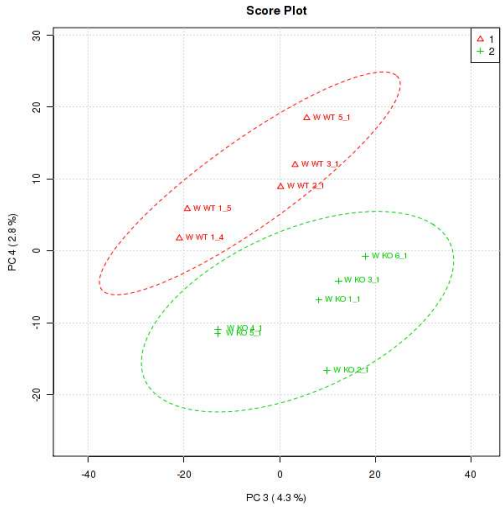
To determine the ability of the extraction methods described above to extract those compounds most important for distinguishing between the *M. smegmatis* wild type and the ESX-3 knockouts, a PCA was performed on the data obtained after GCxGC-TOFMS analyses of the various extracts. As seen in Fig 4.2, the PCA scores plot generated from the data obtained after extraction and GCxGC-TOFMS analyses, resulted in a complete differentiation between the wild type and ESX-3 knockout sample groups, for the conventional total metabolome extracts as well as the polar and non-polar extracts, extracted via the fractionated total metabolome method.



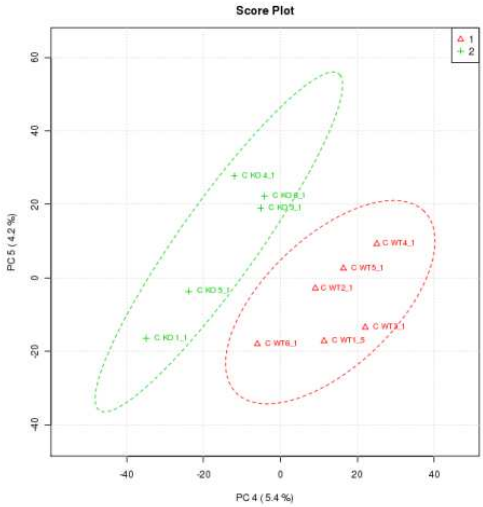
a) Combined total metabolome phase extracts, extracted via the conventional total metabolome method



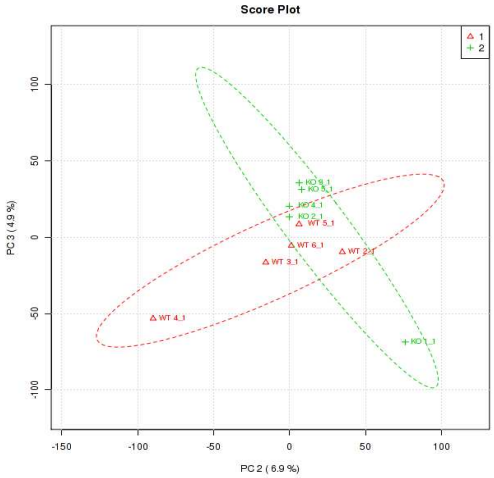
b) Pellet phase extracts, extracted via conventional total metabolome method



c) Polar phase extracts, extracted via the fractionated total metabolome method



d) Non- polar phase extracts, extracted via the fractionated total metabolome method



e) Pellet phase extracts, extracted via the fractionated total metabolome method

Fig 4.2: PCA scores showing the differentiation between the two sample groups for the GCxGC-TOFMS data obtained after the various extraction methods. The variance explained in the data sets for each PC is indicated in parenthesis. The *M. smegmatis* wild-type sample group are indicated in red and the ESX-3 knockout sample group are indicated in green.

The percentage cumulative variance for the combined total metabolome extraction method for the first two PCs (R^2 X cum) was 50%, of which PC 1 contributed 72.4% and PC 2 contributed 7.6%. The percentage cumulative variance for the pellet extracts from the combined total metabolome extraction method for the first two PCs (R^2 X cum) was 89.2%, of which PC 1 contributed 81.2% and PC 2 contributed 8%. The percentage cumulative variance for the polar phase extracts from the fractionated total metabolome extraction method for the first two PCs (R^2 X cum) was 7.1%, of which PC 1 contributed 4.3% and PC 2 contributed 2.8%. The percentage cumulative variance for the non-polar phase extracts

from the fractionated total metabolome extraction method for the first two PCs (R^2 X cum) was 47.4%, of which PC 1 contributed 5.4% and PC 2 contributed 42%. The percentage cumulative variance for the pellet extracts from the fractionated total metabolome extraction method for the first two PCs (R^2 X cum) was 11.8%, of which PC 1 contributed 6.9% and PC 2 contributed 4.9%.

The conventional total metabolome extraction method, visibly showed the lowest degree of within group variation, leading to a clearer differentiation between the groups (inter group variation), as compared to the GCxGC-TOFMS data of the polar and non-polar fraction extracts. This may be due to a variety of reasons, one being the better repeatability comparatively, another being its superior extraction efficiency, and / or that the type of compounds it extracts better differentiates the groups.

4.3.4 Conclusion

Considering all the results, the conventional total metabolome method showed the best comparative repeatability, extraction efficiency and capacity to extract those compounds which best differentiate between the *M. smegmatis* wild-type and ESX-3 knockouts. The slightly poorer repeatability of the fractionated total metabolome method could potentially be ascribed to the fact that this method involves comparatively more steps to complete, increasing the chance for analytical errors, or perhaps due to the fact that many of the extracted compounds, may be extracted in both phases, but not always in equal amounts. Although the conventional total metabolome method has the ability to best differentiate between the *M. smegmatis* wild-type and ESX-3 knockouts, the fractionated total metabolome method also showed an acceptable differentiation between the sample groups, and hence, possible applications when more specific groups of compounds need to be analysed.

Considering this, the conventional total metabolome method as described by Olivier and Loots (2012), was subsequently chosen for further analyses and comparisons, for answering the biological questions proposed in Aim 2, and Objectives 2 and 3, for the following reasons: 1) it is comparatively simpler, 2) comparatively faster, 3) showed comparatively better repeatability, 4) extracts those compounds better differentiating the compared groups

and 5) has been previously described for metabolomics analyses characterising ESX-3 gene functionality (Loots *et al.,* 2013), hence potentially also allowing us to compare results obtained in this investigation, to that of the previous metabolomics study done on ESX-3 *M. smegmatis* knockouts, described by Loots *et al*. (2013).

Chapter 5

A metabolomics investigation of effects of varying iron on the metabolome of *M. smegmatis*

Chapter 5

A metabolomics investigation of effects of varying iron on the metabolome of *M. smegmatis*

5.1 INTRODUCTION

As extensively described in the literature study, iron is crucial for a variety of bacteria pathogens, including *M. tuberculosis*, for bacterial metabolism / growth, and virulence (Kehl-Fie *et al*., 2010). Iron is involved in many enzymes as a vital co-factor, including various hydroxylases, oxygenases and oxygen transferring enzymes, and many cytochromes of the cells, where it participates in important reactions involved in oxidative phosphorylation and energy production (Ratlege, 2004). Without an adequate supply of iron, these cells are unable to function correctly, and cannot generate enough ATP to meet their energy demands. Although *M. tuberculosis* fails to grow in the absence of iron (Rodriguez *et al*., 2002), iron is not always readily available in its mammalian-host. Subsequently, these organisms have developed specialised mechanisms for acquiring iron from their environment. These organisms must also control the levels of intracellular iron, because excess iron may be toxic, due to its participation in reactions that generate toxic oxygen radicals from normal aerobic metabolic products (Rodriguez, 2006). Hence, mycobacteria regulate the amount of intracellular iron by processes sensing iron concentrations, and consequently controlling its uptake and storage (Rodriguez *et al*., 2002).

As described earlier, mycobacteria use various secretion pathways to ensure transport across the complex cell wall, including the ESX- or type VII secretion (T7S) system (Wiker *et al.,* 1992). Mycobacteria synthesize siderophores to capture iron and transport this into the cell. *M. tuberculosis* produces two types of siderophores, carboxymycobactins and mycobactins (Domenech *et al*., 2001).

Considering this, in the current investigation, we will attempt to determine the effect of varying iron concentrations (high and low concentrations in the growth media of these organisms), on the metabolome of a wild type parent strain of *M. smegmatis*, in order to

better understand its role in mycobacterial metabolism, growth and virulence, as proposed in Objective 2. This will be accomplished using the total metabolome extraction approach previously described by Olivier and Loots (2012) and validated for this study in Chapter 3, followed by GCxGC-TOFMS analysis of the separately cultured *M. smegmatis* wild-type samples, grown in normal (n = 15), low (n = 15) and high iron (n = 15) concentrations, and comparing the groups using a metabolomics research approach.

5.2 MATERIALS AND METHODS

5.2.1 Reagents and chemicals

All reagents and chemicals used are described in 4.2.1.

5.2.2 Bacterial cultures, preparation of *M.* smegmatis ESX-3 knockouts and sample preparation.

All *M. smegmatis* cultures were obtained from the DST/NRF centre of excellence for biomedical TB research (CBTBR), Stellenbosch University, South Africa. The preparation of these bacterial samples are described in section 4.2.2. For this investigation however, the *M. smegmatis* wild type parent strain were cultured in Difco Middlebrook 7H9 media using 3 different iron concentrations as described below.

From the original culture media, 100µl of the *M. smegmatis* wild-type was transferred to Difco Middlebrook 7H9 medium, with either Chelex-100, creating an iron concentration of about half of the normal iron concentration ±75µM (low iron) (n = 15), or 300µM ferric ammonium citrate (7H9 Fe, creating a high iron condition of 450µM Fe) (n = 15), or no added iron creating a normal iron concentration of 150µM (n=15), supplemented with 0.2% glycerol and 0.05% Tween-80. The high iron concentration was sufficient to overcome the essentiality of ESX-3 in *M. tuberculosis* (Serafini *et al* 2009), suggesting that expression of ESX-3 may be completely repressed under these conditions. The bacteria were incubated at 37 ºC with shaking at 180 rpm until the culture growth reached mid-log phase (OD_{600nm} = 0.6 – 0.8).

The cells, cultured in all three iron concentrations, were subsequently washed twice with phosphate buffered saline, collected via centrifugation, snap frozen in liquid nitrogen, lyophilized and stored at -80ºC until extraction and GCxGC-TOFMS metabolomic analysis.

5.2.3 Quality control samples

Quality control (QC) samples were introduced in the current metabolomics study for the following reasons: 1) to condition or equilibrate the analytical platform before sample analysis, ensuring repeatable data, 2) to provide a means to calculate technical precision within each analytical batch run (quality assurance procedures), and 3) to provide data for signal correction or analytical variation within and between analytical batches (Dunn *et al.,* 2011). These QC samples were made-up of a mixture of 5 mg (~1 X 10^8 cells) of all individually cultured groups, which was subsequently aliquoted in 6 identical samples of 5 mg each. One QC sample aliquot was extracted per batch using the extraction method discussed in 5.2.3. Furthermore each batch contained 15 samples and the QC sample, injected seven times throughout each batch. The following run order was used for experimentation:

[QQQQQ] $S_1S_2S_3S_4S_5S_6S_7$ [Q] $S_8S_9S_{10}S_{11}S_{12}S_{13}S_{14}S_{15}$ [Q]

5.2.4 Extraction method (the conventional total metabolome method)

In order to compare the wild type in normal, high and low iron conditions, the 13 individually cultured sample repeats of the wild type *M. smegmatis*, cultured in normal, high and low iron concentrations were extracted via the conventional total metabolome extraction method discussed in 4.2.3.1, and two samples, of all three different growth medium, selected at random were extracted twice via the conventional total metabolome extraction method discussed in 4.2.3.1 (thus creating a sample set of 15 samples).

5.2.5 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS)

The same GCxGC-TOFMS conditions were used for the injection of these samples, as described in 4.2.4

5.2.6 Data processing

5.2.6.1 Peak identification and alignment

The same peak identification and alignment criteria were used as described in 4.2.5.1.

5.2.7 Statistical data analysis

The statistical data analysis (PCA, PLS-DA and *t*-test) were performed using MetaboAnalyst, a metabolomics web server, based on the statistical package "R" version 2.10.0.Statistica (version 10) were used for all the PCA power data.

5.2.6.1 Data Pre-treatment

Typical to metabolomic studies, data pre-treatment is required before the necessarily statistical comparisons can be made. Firstly, data were normalised relative to the internal standard by determining each compounds relative concentration, in order to compensate for any potential sample loss during the extraction or chromatographic injection. Those compounds which were not detected in at least 50% of the samples, in one or more of the sample groups, together with those compounds with no detected variation between the groups, were then removed. A "QC correction" step was then applied to the data set. In this step, the QC samples of the various batches are used to correct for any non-biological

systematic differences which may exist among the data of the various batches analysed. These differences may be attributed to small analytical change which may occur over the duration the analysis. This data transformation method corrects for any linear and nonlinear differences in distribution which may be present among batches of semi quantitative data, obtained from the same analytical method (Draisma *et al*., 2010). This was followed by "zero value replacement", i.e. those compounds which were not detected in a specific sample, were substituted with a value calculated as half of the minimum concentration in the original data, assuming that most missing values were a reflection of low abundance metabolites (i.e. those not detected as they fell below the detection limit of the apparatus). A "QC-CV" filter was then applied to the data, by removing all compounds with a CV value of above 50% (Schoeman *et al*., 2012) from the sample data set, as determined using in the QC samples.

5.2.7.2 Multivariate and univariate statistics

The statistical data analysis (PCA, PLS-DA and *t*-test) were performed using MetaboAnalyst, a metabolomics web server, based on the statistical package “R” version 2.10.0. Effect sizes were determined, by calculating its *d*-value and a t-test was done by calculating the compounds *p*-value, both described later in this section.

Using MetaboAnalyst, supported by The Metabolomics Innovation Centre (TMIC), the dataset was log transformed, as the scales of the features measured differ drastically. These non-parametric transformation functions both have a large effect on the eventual data analysis. In both cases, the data was also centred. Centring is done in order to convert all of the concentrations to fluctuate on zero instead of on the mean of the metabolite concentrations. Thus, it is used to focus on fluctuations in the data, leaving relevant variation for analysis.

As mentioned, both multivariate and univariate statistical approaches were used in this study, as illustrated in Fig.5.1.

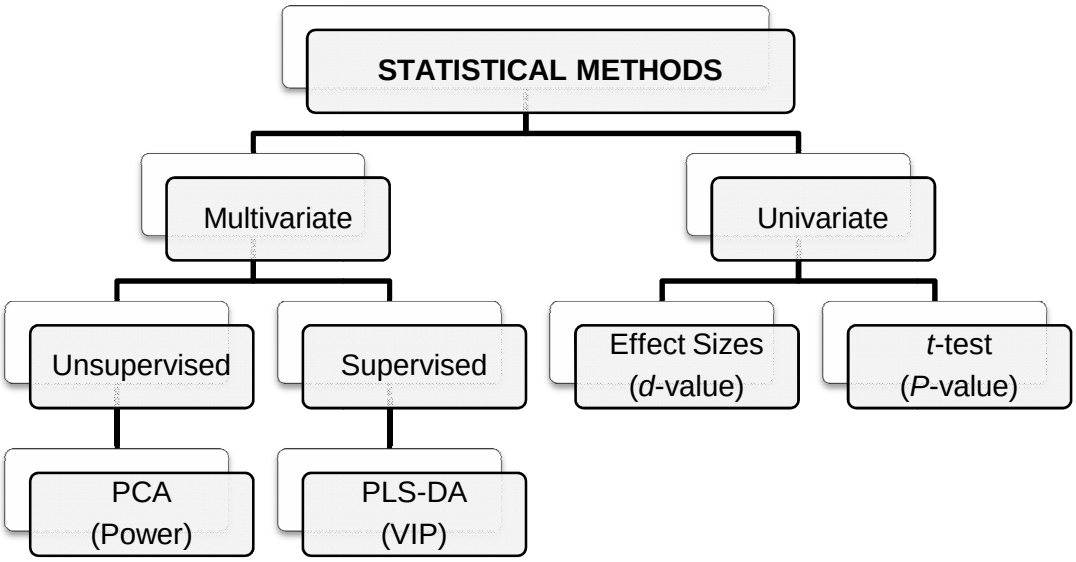


Fig 5.1: Work-flow of the multi- and univariate statistical methods used in this study

Principle component analysis (PCA) was performed in order to determine whether or not a natural grouping or differentiation exists between the various sample groups based on their detected metabolite profiles. PCA involves a mathematical procedure that transforms a number of possibly related variables (in this case metabolites) into a smaller number of unrelated variables known as principal components (PC's). PC 1 accounts for the most variance in the data and each subsequent PC (PC 2, PC 3 etc.) accounts for the next highest variance of the remaining data. Using one PC per axis, the PCA can then be visualized as a scores plot. Using the data generated from the species differentiation analyses, the metabolites were ranked according to their respective PCA modelling powers and those with a modelling power of greater than 0.5 were considered important (Brereton *et al.,* 2003).

Additionally, a partial least squares discriminant analysis (PLS-DA) model was built in order to identify those compounds that contribute most to the separation between the sample groups. PLS-DA uses various linear regression techniques in order to find the direction of maximum covariance between a data set and a class membership (Xia *et al*., 2009). PLS-DA summarizes the original variables into fewer variables, also known as scores, using their weighted averages. The PLS-DA thus provides a second set of important metabolites i.e. variables important in projection (VIPs), defined as the weighted sum of squares of the PLS

loadings, which takes into account the amount of explained class variance of each component. Those metabolites with a VIP value of greater than 1 were considered important as metabolite markers (Chong & Jun, 2005).

Univariate analyses, which included Effect Sizes (ES) and an unpaired *t*-test, are used to obtain an estimated raking of potentially important features. It examines each variable separately, without taking the effect of multiple comparisons into account. ES is a simple way of quantifying the difference between two groups, by emphasising the size of the effect, rather than confounding this with sample size, and therefore it is a true measurement of the significance of the difference. Thus, it is a sophisticated, particularly valuable tool for quantifying, reporting and interpreting the effectiveness of a particular intervention, relative to some comparison, by determining how well it works in a range of context. The ES of each compound was determined by calculating its *d*-value, using the standardised mean difference between the cured and failed groups, or rather $d = |X_1 - X_2|/S_{max}$, where X_1 and X_2 signify the group means and S_{max}, the maximum standard deviation of two groups analysed. The value of *d*, for non-parametric data, should be interpreted as $d < 0.2$ having a small effect, $d > 0.5$ having a medium effect, and $d > 0.8$ having a large effect.

A *t*-test, however, refers to a statistical examination of the means of two groups, in order to determine whether or not the averages of two groups are different, by establishing the statistical significance. Thus, as opposed to ES, *t*-test takes sample size into account. First, a *t*-value is determined, followed by the calculation of the statistical significance by determining a *P*-value. The *P*-value is the probability that a difference of at least the same size would have occurred by chance, even if there is no difference between the two groups. Conventionally, if the *P*-value is less than 0.05, the compound is considered to be significant. Although the *t*-test supports both paired and unpaired analysis options, the unpaired option was chosen for this study. This is due to the fact that the patients of the cured and failed groups differ, and therefore the urine samples were not obtained from the same individuals, and thus, was independent of one another.

5.3 RESULTS

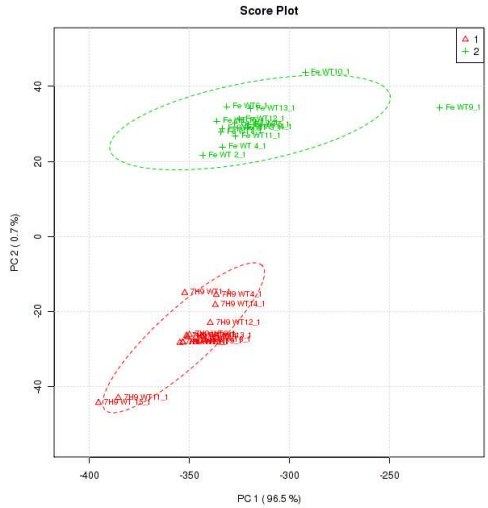
5.3.1. Data pre-treatment

In the current chapter, two comparisons, using the prepared sample groups, were done. The first included a comparison of the metabolite profiles of the *M. smegmatis* wild types grown in normal iron conditions, to that of the *M. smegmatis* wild type grown in high iron conditions. The second comparison was of the *M. smegmatis* wild type grown in normal iron conditions, to that of the *M. smegmatis* wild type grown in low iron conditions.

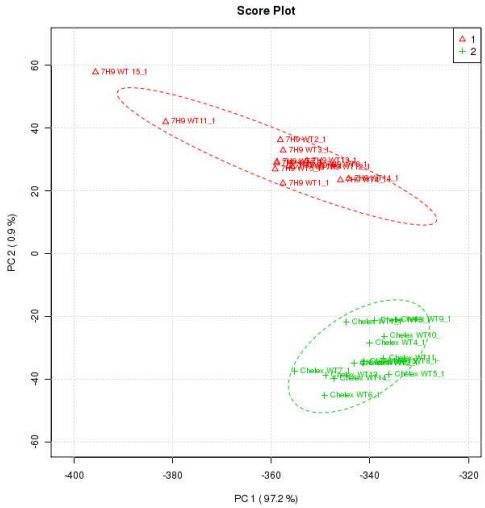
The data pre-treatment steps described in section 5.2.7.1, led to a reduction of the total number of compounds from 699 to 372, of which 295 could be annotated by comparison of their mass spectra and retention times, to libraries prepared from previously injected standards, for the first group comparison (*M. smegmatis* cultured in normal iron vs. high iron concentrations*)*, and from 699 to 401 (of which 302 were annotated), for the second group comparison (*M. smegmatis* cultured in normal iron vs. low iron concentrations).

5.3.2 PCA differentiation

A PCA analysis was done on the GCxGC-TOFMS pre-treated data set described above, generated from the extracted samples collected from cultures grown in varying iron concentrations, in order to determine whether or not a natural differentiation occurs between the groups, based on the differences in their underlying metabolite profiles. As seen in Fig 5.2, for both comparisons, a clear differentiation between the sample groups was achieved when using the first two PCs of the PCA, and this natural differentiation between the groups can be attributed to underlying differences in the analysed metabolite profiles of these groups.



a) Group 1 - The *M. smegmatis* wild-type samples cultured in normal iron concentrations are indicated in red and *M. smegmatis* wild-type samples cultured in high iron concentrations in green.



b) Group 2 - The *M. smegmatis* wild-type samples cultured in normal iron concentrations are indicated in red and *M. smegmatis* wild-type samples cultured in low iron concentrations in green.

Fig 5.2: PCA scores plots showing principal component 1 (PC1) versus principal component 2 (PC2) and the natural differentiation between the sample groups for the GCxGC-TOFMS data obtained after the total metabolome extraction method. The percentage variation explained by each PC is indicated in parenthesis.

For the *M. smegmatis* wild type sample group cultured in normal iron concentrations vs. the *M. smegmatis* wild type sample group cultured in high iron, the total amount of variance explained by the first two PCs (R^2X cum) was 97.2%, of which PC 1 contributed 96.5% and PC 2 contributed 0.7%. For *M. smegmatis* wild type sample group cultured in normal iron vs. the *M. smegmatis* wild type sample group cultured in low iron compared, the total amount of variance explained by the first two PCs (R^2 X cum) was 98.1%, of which PC 1 contributed 97.2% and PC 2 contributed 0.9%.

5.3.3 Metabolite marker identification

One of the primary functions of a metabolomics research approach, is not only determining if a differentiation of various sample groups exists on the basis of varying metabolite profiles, but also to identify those metabolite markers which vary most between the compared groups. These markers, in turn, may be used to better characterise the sample groups

investigated, and in so doing answer a biological question pertaining to this objective, i.e. the effect of varying iron concentrations on the metabolome of *M. smegmatis*.

A number of supervised univariate (*t*-test and effect size) and multivariate (PLS-DA) statistical data analyses methods were used, in addition to the previously described unsupervised, multivariate test (PCA), in order to select those metabolites which differed the most (considering their novelty and relative concentrations) or best described the variation observed between the groups, when comparing the GCxGC-TOFMS data obtained from the extracted *M. smegmatis* wild types, cultured in varying iron concentrations.

For the first comparison done (*M. smegmatis* cultured in normal iron vs. high iron) the PLS-DA model used for identifying the characteristic marker metabolites, had a modelling parameter R^2Y (cum) of 69.7%, with a Q^2 (cum) of 62.6%. For the second comparison (*M. smegmatis* cultured in normal iron vs. low iron), the PLS-DA model used for identifying the characteristic marker metabolites, had a modelling parameter R^2Y (cum) of 69.3%, with a Q^2 (cum) of 62%.

Metabolites were selected as markers if they had a VIP value (identified via PLS-DA) > 1 (Chong & Jun, 2005), a PCA power > 0.5 (Brereton *et al.,* 2003) and effect size > 0.5 (Ellis *et al.,* 2003).

A t-test was not used for marker selection in this instance. Due to the fact that a *t*-test considers sample size in its calculation, when used in this comparison, it excluded a number of compounds with potential biological importance, possibly due to the comparatively small sample size, accompanied by the a larger standard deviation for many compounds in this data set, which can be expected, as biological samples of this nature are accompanied by large variation. Considering this, the effect size calculations were considered a far better measure for selecting metabolite markers from a univariate perspective in the current data set. The *p*-values as determined by the t-test, are however indicated for each of the selected compounds, in Table 5.1

As indicated in Fig 5.3, metabolite markers for the comparison of the *M. smegmatis* cultured in normal iron vs. high iron concentrations, were selected by including those compounds having both a PLS-DA VIP > 1 and effect size d value of > 0.5, in addition to those

compounds with a PCA modelling power > 0.5 and effect size d value > 0.5, totalling 148 compounds, of which 35 were annotated / identified using libraries prepared from previously injected standards, by comparison of their mass spectra and retention times. These metabolites will subsequently be used to explain the biological variation between the normal and high iron cultured sample groups.

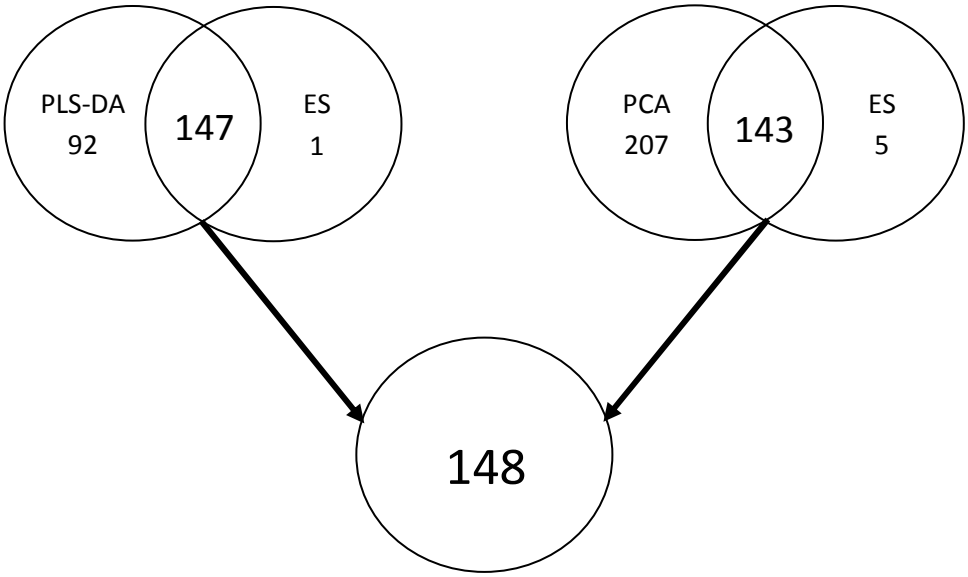


Fig 5.3: Summary of the selection approach used for identifying the metabolite markers best differentiating the *M. smegmatis* cultured in normal iron vs high iron concentrations.

Table 5.1: Metabolite markers selected from the comparison of *M. smegmatis* cultured in normal iron vs. high iron concentrations, indicating their mean relative concentrations (standard deviations in parenthesis), VIP values, PCA modelling power, effect size d-values and t-test *p*-values.

			Normal iron	High Iron		
VIP	PCA Power	Compound name (Chemspider ID)	Mean relative concentration (µg/mg)	Mean relative concentration (µg/mg)	Effect sizes	*p*-value
0.363	0.996	Palmitelaidic acid (444572)	0.071 (0.017)	0.088 (0.02)	>0.8	0.809
1.048	0.995	Glycoside (56585)	0.152 (0.123)	0.001 (0.002)	>0.5	0.567
0.977	0.991	Cadaverine (13866593)	0.005 (0.006)	0.017 (0.017)	>0.5	0.458
0.377	0.991	Propionic acid (1005)	0.193 (0.120)	3.318 (1.750)	>0.5	0.184

0.527	0.991	Hexadecane (10540)	0.016 (0.004)	0.835 (2.836)	>0.5	0.245
0.950	0.989	Arabinofuranose (9052560)	0.008 (0.004)	0.09 (3.4)	>0.5	0.015
0.471	0.988	Phosphoric acid (979)	0.086 (0.039)	0.15 (0.072)	>0.5	0.421
0.513	0.988	à-l-Galactofuranoside (24776978)	0.003 (0.002)	0.411 (1.52)	>0.5	0.0001
1.026	0.987	Octadecene (7925)	0.003 (0.001)	0.465 (1.586)	>0.5	0.342
0.913	0.984	Butyric acid (259)	0.059 (0.042)	1.434 (0.034)	>0.5	0.191
0.410	0.983	Anhydro-á-d-glucose (71358)	0.069 (0.025)	1.302 (4.4)	>0.5	0.003
0.625	0.978	Undecane (13619)	0.008 (0.006)	0.534 (1.876)	>0.5	0.008
0.494	0.978	Mannonic acid (588)	0.014 (0.008)	0.012 (0.005)	>0.5	0.104
0.318	0.975	Acetamide (173)	0.250 (0.011))	2.897 (9.5)	>0.5	0.001
0.689	0.975	Glucose (71358)	1.925 (0.0002)	169.707 (57.680)	>0.5	0.356
0.550	0.974	Galactose (5814)	0.113 (0.116)	0.582 (1.515)	>0.5	0.172
0.168	0.970	Nonanoic acid (7866)	0.013 (0.003)	0.0322 (0.060)	>0.5	0.156
0.436	0.968	Valine (6050)	0.019 (0.006)	0.0257 (0.009)	>0.5	0.067
0.371	0.966	Lysine (5747)	0.008 (0.007)	0.631 (2.200)	>0.5	0.091
0.483	0.963	Ornithine (6026)	0.027 (0.025)	0.009 (0.01)	>0.8	0.835
0.370	0.954	Tryptophane (1116)	0.006 (0.0008)	0.007 (0.003)	>0.5	0.001
0.734	0.943	Alanine (582)	0.086 (0.083)	0.303 (0.358)	>0.5	0.002
0.305	0.924	Nonadecane (11895)	0.007 (0.006)	0.008 (0.004)	>0.5	0.181
0.453	0.889	Glycerol (733)	5.913 (3.527)	142.502 (66.184)	>0.5	0.006
0.692	0.878	Tridecane (11882)	0.013 (0.009)	0.131 (0.22)	>0.5	0.401
0.275	0.870	Hydroxypyruvic (939)	0.025 (0.011)	0.201 (0.5)	>0.5	0.012
2.589	0.870	Ribulose (133316)	0.011 (0.023)	0.031 (0.050)	>0.5	0.005
3.523	0.863	Ethylene (6085)	0.006 (0.012)	0.136 (0.422)	>0.5	0.001
0.400	0.855	Cyclohexylamine (7677)	0.06 (0.056)	0.094 (0.0001)	>0.5	0.002
2.235	0.816	Glycine (730)	0.002 (0.001)	0.007 (0.003)	>0.5	0.489
2.303	0.675	Isoleucine (769)	0.003 (0.0006)	0.007 (0.0009)	>0.5	0.016
1.011	0.609	Fumarate (4445588)	0.002 (0.003)	0.003 (0.002)	>0.5	0.128

2.769	0.600	Propene (7954)	0.039 (0.030)	0.356 (1.302)	>0.5	0.048
1.909	0.566	Ala-Gly (145194)	0.0001 (0.0002)	0.004 (0.011)	>0.5	0.037
0.992	0.565	Leucine (834)	0.006 (0.0009)	0.0122 (0.015)	>0.5	0.002

Similarly, as indicated in Fig 5.4, metabolite markers for the comparison of the *M. smegmatis* cultured in normal iron vs. low iron concentrations, were selected by including those compounds having both a PLS-DA VIP > 1 and effect size d value of > 0.5, in addition to those compounds with a PCA modelling power > 0.5 and effect size d value > 0.5, totalling 194 compounds, of which 41 were annotated / identified using libraries prepared from previously injected standards by comparison of their mass spectra and retention times. These metabolites will subsequently be used to explain the biological variation between the normal and low iron cultured sample groups..

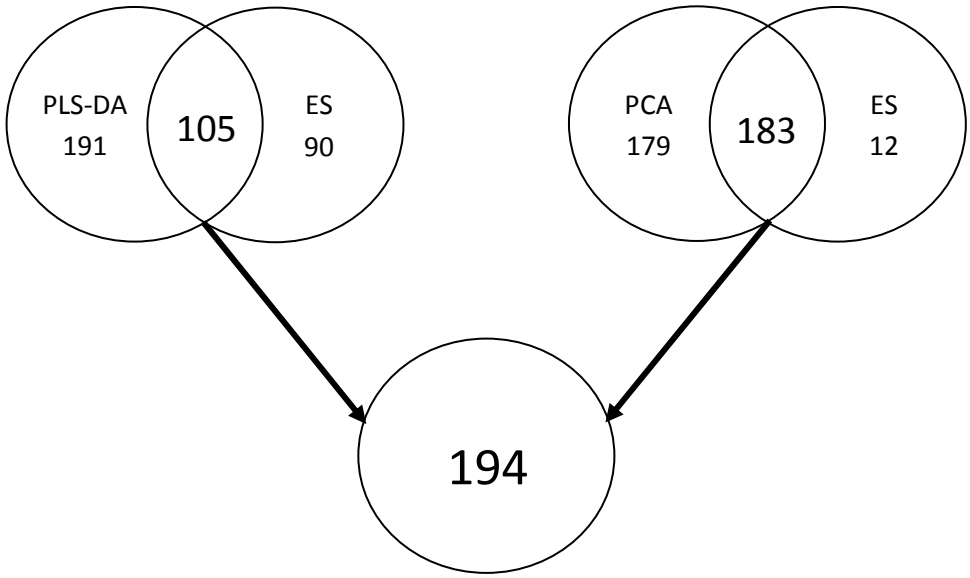


Fig 5.4: Summary of the selection approach used for identifying the metabolite markers best differentiating the *M. smegmatis* cultured in normal iron vs low iron concentrations.

Table 5.2: Metabolite markers selected from the comparison of *M. smegmatis* cultured in normal iron vs. low iron concentrations, indicating their mean relative concentrations (standard deviations in parenthesis), VIP values, PCA modelling power, effect size d-values and t-test *p*-values.

			Normal iron	Low Iron		
VIP	PCA Power	Compound name (Chemspider ID)	Mean relative concentration (µg/mg)	Mean relative concentration (µg/mg)	Effect sizes	*p*-value
0.630	0.984	Lysine (5747)	0.008 (0.007)	0.016 (0.01)	>0.5	0.038
0.682	0.984	Propanoic acid (1005)	0.812 (0.12)	0.173 (0.046)	>0.5	0.428
0.740	0.983	Cadaverine (13866593)	0.005 (0.006)	0.01 (0.001)	>0.8	2.83x10e-7
1.024	0.971	Mannonic acid (1005)	0.014 (0.008)	0.095 (0.027)	>0.5	0.043
0.636	0.964	Nonadecane (11895)	0.007 (0.006)	0.012 (0.002)	>0.5	0.66
0.824	0.960	Glycoside (56585)	0.152 (0.123)	0.33 (0.128)	>0.5	0.021
0.642	0.958	Butyric acid (259)	0.059 (0.042)	0.221 (0.02)	>0.8	1.24x10e-5
0.206	0.953	Adipic acid (191)	0.0034 (0.001)	0.002 (0.0006)	>0.5	0.205
1.071	0.953	Benzoic acid (238)	0.566 (0.053)	0.062 (0.039)	>0.5	0.003
3.683	0.939	1-Cyclopentyl-ethane (14686)	0.004 (0.003)	0.00004 (0.0001)	>0.8	2.01x10e-7
0.465	0.939	D-Xylofuranose (79694)	0.034 (0.016)	0.082 (0.024)	>0.5	0.135
0.131	0.932	Palmitelaidic acid (444572)	0.071 (0.017)	0.094 (0.012)	>0.5	0.974
0.388	0.924	Valine (6050)	0.019 (0.006)	0.02 (0.007)	>0.5	0.011
0.476	0.923	Pyroglutamic acid (7127)	0.02 (0.013)	0.045 (0.019)	>0.5	0.052
0.425	0.920	à-l-Galactofuranoside (24776978)	0.003 (0.002)	0.008 (0.002)	>0.5	0.981
0.104	0.913	Malonic acid (844)	0.032 (0.005)	0.025 (0.004)	>0.5	0.369
0.297	0.909	Proline (594)	0.15 (0.077)	0.288 (0.1)	>0.5	0.258
0.079	0.901	Propane (6094)	0.047 (0.014)	0.038 (0.085)	>0.5	0.789
0.266	0.900	Succinic acid (1078)	0.757 (0.082)	0.422 (0.01)	>0.5	0.658
0.115	0.899	Hexanoic acid (8552)	0.058 (0.021)	0.072 (0.013)	>0.5	0.064
0.084	0.876	Levoglucosan (9587432)	0.001 (0.0002)	0.012 (0.002)	>0.5	0.721
0.287	0.872	Decane (14840)	0.014 (0.002)	0.008 (0.003)	>0.5	0.706

0.610	0.872	Alanine (582)	0.086 (0.083)	0.239 (0.069)	>0.5	0.157
0.230	0.848	Ribitol (10254628)	0.207 (0.047)	0.355 (0.094)	>0.5	0.159
3.948	0.842	Methyl Malonamate (2831361)	0.044 (0.001)	0.0001 (0.0002)	>0.5	0.348
0.213	0.806	3-Buten-2-one (6322)	0.012 (0.004)	0.009 (0.005)	>0.5	0.082
0.242	0.799	Acetic acid (171)	0.1 (0.030)	0.174 (0.053)	>0.8	0.0001
3.158	0.788	Crotonic acid (552744)	1.46x10e-7 (5.1x10e-23)	0.001 (0.001)	>0.5	0.111
4.128	0.781	Mandelic acid (1253)	0.424 (4.4x10e-4)	0.004 (0.007)	>0.5	0.358
0.138	0.778	Phosphoric acid (979)	0.086 (0.039)	0.111 (0.020)	>0.5	0.766
1.262	0.764	Aspartic acid (411)	0.003 (0.007)	0.001 (0.0007)	>0.5	0.458
0.844	0.758	Octadecene (7925)	0.003 (0.001)	0.006 (0.002)	>0.5	0.045
0.266	0.749	2-Furoic acid (10251740)	0.007 (0.002)	0.004 (0.0007)	>0.5	0.03
0.084	0.745	Octadecanoic acid (5091)	0.389 (0.074)	0.473 (0.080)	>0.5	0.609
2.047	0.729	Glycine (730)	0.002 (0.001)	0.008 (0.001)	>0.5	0.986
2.320	0.698	Isoleucine (769)	0.003 (0.0006)	0.01 (0.0004)	>0.5	0.043
0.092	0.631	Nonanoic acid (7866)	0.013 (0.003)	0.01 (0.002)	>0.5	0.118
0.355	0.580	Tridecane (1182)	0.013 (0.009)	0.019 (0.005)	>0.5	0.234
1.088	0.534	Leucine (834)	0.006 (0.0009)	0.019 (0.0007)	>0.5	0.486
1.363	0.432	Fumaric acid (10197150)	0.002 (0.0003)	0.008 (0.003)	>0.5	0.166
1.828	0.415	Mannose (17893)	0.015 (0.025)	0.042 (0.03)	>0.8	1.06x10e-10

5.4 DISCUSSION

An interesting fact to note at this point, is that 13 metabolite markers identified in Table 5.1 (comparing *M. smegmatis* wild type cultured in normal vs. high iron concentrations) and Table 5.2 (comparing *M. smegmatis* wild type cultured in normal vs. low iron concentrations), were identical, with only differences to their concentrations, and the direction in which these concentrations changed, due to the varying iron concentrations in which these organisms

were cultured. The above results will subsequently be discussed under the following headings:

5.4.1 Metabolites markers in *M. smegmatis* cultured during comparatively higher iron concentrations

When considering the results in Table 5.1, various amino acids, sugars, alkanes, alkenes, fatty acids and carboxylic acids were identified as markers in the *M. smegmatis* cultured in higher iron comparatively.

As stated earlier, the expression of various metabolite biosynthesis genes has been shown to be regulated by iron in mycobacteria and various other bacterial species (Rodriguez *et al.,* 2002; Maciag *et al.,* 2007; Maciag *et al.,* 2009). Additionally, iron is an essential cofactor for many of the enzymes involved in the metabolism in these organisms. Considering this, we will explain the metabolic differences observed in *M. smegmatis* cultured in the higher iron concentration comparatively, in light of the associated role of iron in these metabolic pathways, and summarised in Fig. 5.4.

Excessive iron is known to lead to oxidative stress in all living organisms (Touati *et al*., 2000), and is ascribed to the Fenton reactions, in which ferrous iron and hydrogen peroxide, produce ferric iron, hydroxide anions and hydroxide radicals. During these oxidative stress conditions, the Krebs cycle's activity has been shown to diminish, with an increased capacity of the glyoxylate shunt (Tretter *et* al., 2000), in addition to an accumulation of pyruvic acid and lactic acid (Abordo *et* al., 1999). Phosphoenolpyruvate carboxykinase (pckA), is an iron regulated enzyme catalysing the conversion of oxaloacetate to phosphoenolpyruvate, and is also up-regulated when iron increases inside the cell (Sauer *et al.,* 2004). Furthermore, phosphoenolpyruvate is converted to pyruvate via the enzyme pyruvate kinase, also believed to be up-regulated by intracellular iron (Chadvadi *et al*., 2009). Although pyruvate was not directly detected in our metabolomics investigation, a derivative of this, hydroxypyruvic acid (Esakova *et al*., 2008), was detected in elevated concentrations in the high iron cultured *M. smegmatis,* indicative of potential iron induced oxidative stress. Further confirmation of this, are the elevations in alanine, isoleucine, leucine, valine and glycine, in

the *M. smegmatis* cultured in the higher iron concentrations, all synthesised from pyruvate (Schulze-Siebert *et al*., 1984). The elevated isoleucine and valine, could also explain the elevated propionic acid detected, which is the natural breakdown product of these amino acids (Elsden *et al*., 1976). Pyruvate also acts as a substrate for the glyoxylate shunt, thus, the proposed increased pyruvate, would also be expected to result in an increased glyoxylate shunt activity, further confirmed by the elevated glycine detected (Tang *et al*., 2009) in the *M. smegmatis* cultured in elevated iron concentrations in our investigation. Glycine is also known to serve as a substrate for the hydroxypuruvate previously mentioned (Dagley *et al*., 1962), further contributing to its synthesis.

IdeR binding sites and iron-dependant regulation, were previously identified upstream of *trpE2,* in various mycobacteria, including *M. smegmatis* (Gold *et al.,* 2001; Rodriguez *et al.,* 2002; Yellaboina *et al.,* 2006). TrpE2 is thought to function as an anthranilate synthase component, converting chorismate to anthranilate during the synthesis of tryptophan. Expression of TrpE2 is known to be down-regulated when intracellular iron is elevated, leading to a reduced synthesis of tryptophan (Yellaboina *et al.,* 2006). However, as seen in Table 5.1, tryptophan was found in elevated levels in the high iron cultured *M. smegmatis*. According to Watanabe and Snell (1972), tryptophan can be synthesised from pyruvate, indole and amomonia, when pyruvate is present in elevated concentrations, potentially explaining this occurrence in our investigation.

Further support for an elevated glyoxylate cycle activity, was the elevated concentrations of the glyoxylate intermediate, fumarate, detected in the *M. smegmatis* cultured in the high iron media. During anaerobic conditions, or conditions of oxidative stress, mycobacteria are known to synthesise fumarate, from the previously discussed phosphoenolpyruvate. Iron also acts as a known co-factor in the conversion of succinate to fumurate though the enzyme succinate dehydrogenase, which consists out of four subunits, sdhA, sdhB, sdhC and sdhD (Yankovskaya *et al*., 2003), of which sdhB is an iron-sulphur protein, which further explains the increased fumarate detected in the high iron cultured samples in our investigation. As seen in Figure 5.5, malate, derived from fumurate and glyoxylate, which are both proposed to be increased during growth in high iron concentrations, is a known substrate for gluconeogenis, a metabolic pathway that results in the generation of glucose from non-carbon substrates (Sauer *et al*., 2004,) especially during anaerobic conditions or conditions of oxidative stress (Mizock 1995), explaining the elevations in glucose in the high iron group.

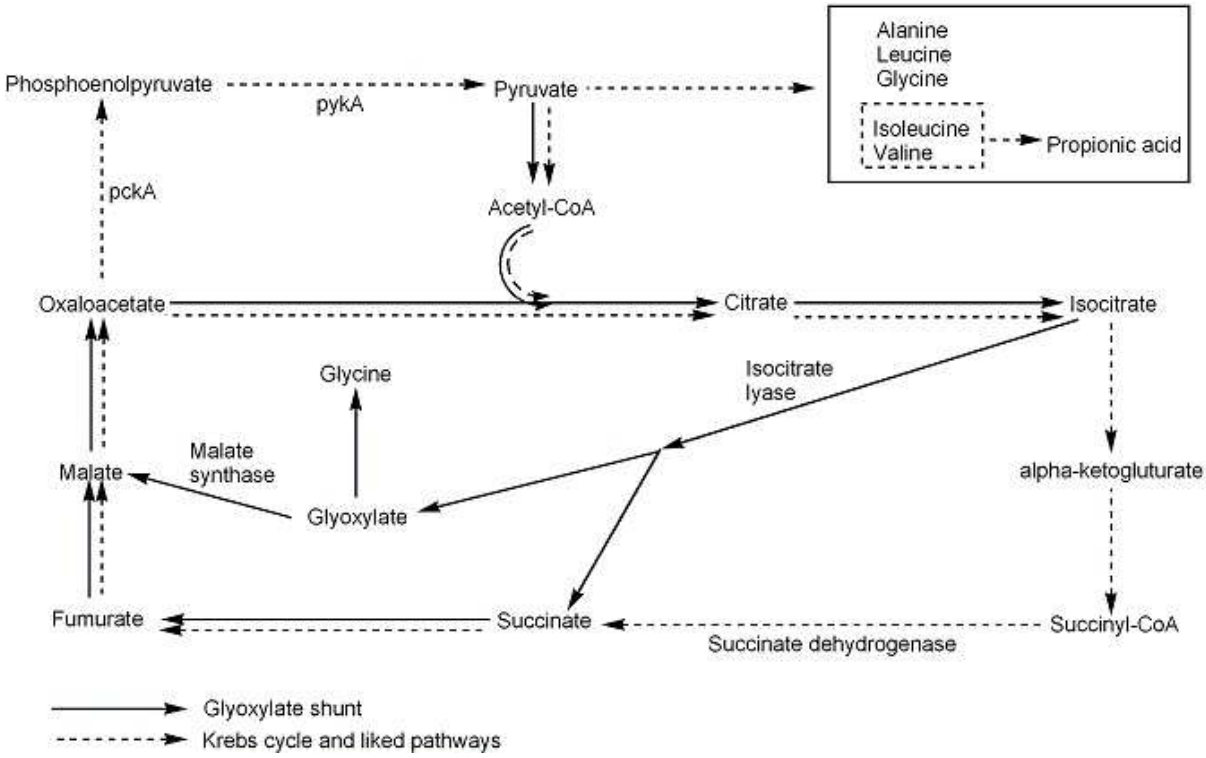


Fig 5.5: Schematic representation of the interaction between the glyoxylate, krebs cycle and associated pathways and metabolites altered when *M. smegmatis* are cultured in higher iron concentrations. The dotted arrows represents the Krebs cycle and other liked pathways, where the bold arrows represent the glyoxylate shunt in mycobacteria.

The oxidative response of *M. bovis* BCG was recently tested by Jang et al. (2009), who showed an up-regulation of amino acid transport and certain amino acid synthesis proteins, as well as a decreased lipid transport, lipid metabolism and membrane formation, during hydrogen peroxide induced oxidative stress. Not only do these results coincide with the altered amino acid metabolism as determined in our metabolomics investigation, but we additionally detected a general reduction in the use of fatty acids in the higher iron cultured *M. smegmatis*, indicated by intracellular elevations of these fatty acids. A and are seen to be elevated in the high iron cultured *M. smegmatis*, comparatively. According to Shi et al. (2010), reductions in these glycolysis precursors are not only as a result of their roles as substrates for energy production during periods of increased growth, but also due to their roles as precursors for the biosynthesis of other macromolecules. Furthermore, in a metabolomics investigation done by Meissner-Roloff et al.,(2012), a reduction in these metabolites, were ascribed to their increased utilisation for growth, metabolism and cell wall synthesis in hyper-virulent Beijing *M. tuberculosis*. Considering this, the accumulation of these glycolysis substrates in this study, suggests hindered growth, metabolism and cell wall synthesis in this group, possibly due to the iron induced oxidative stress. Furthermore, as

described by Meissner-Roloff *et al*., (2012), other carbohydrate molecules associated with the cell wall of these organisms, include glycerol-galactofuranoside, anhydro-á-d-glucose, arabinofuranose and mannose (Brennan 2003). Arabinofuranose, for instance, acts as a bridge between the peptidoglycan and the mycolic acids of the cell wall (Brennan 2003) and is known as the cell wall core or the arabinogalactin-peptidoglycan complex (mAGP). These carbohydrates form part of long length chain molecules which assist in maintaining cell wall structure and even provide immunogenic variation (Chatterjee 1997). All the above mentioned intracellular carbohydrates were found to be comparatively elevated in the high iron *M. smegmatis* group investigated in this study, supporting the hypothesis that these compounds are being used at a reduced rate for cell wall synthesis due to the iron overload.

Cadaverine, a polyamine antioxidant synthesised from lysine, also plays an important function in membrane permeability, acid tolerance and RNA synthesis in various bacteria (Kamio et al. 1981; Neely and Olson 1996; Samartzidou and Delcour, 1998). In Escherichia Coli, endogenous and exogenous cadaverine has been shown to induce porin closure, reducing the cellular uptake of various compounds (Samartzidou and Delcour 1998) including iron (Jones and Niederweis et al., 2012). Considering this, the comparatively increased intracellular cadaverine seen in the high iron cultured *M. smegmatis* in our investigation, may be indicative of a similar mechanism by which this organism attempts to regulate the amount of iron uptake through its porin structures of the cell. Additionally, due to its antioxidant functions, these organisms may also up-regulate cadaverine production to help manage the oxidative stress induced by the elevated iron concentrations.

Siderophore synthesis would also be expected to be down-regulated during conditions of increased iron, considering their role in iron uptake and homeostasis (Rodriguez *et al*., 2002). Considering that these siderophores are made up of amongst others, a large amount of lysine (Cole *et al,* 1998), its accumulation in the *M. smegmatis* cultured in high iron media, can be explained by a decreased utilisation of this amino acid for siderophore synthesis. Additionally, the PE and PPE proteins previously described, have also been associated with siderophore synthesis, and their production being up-regulated during conditions of reduced iron and repressed during elevated iron (Rodriguez *et al*., 1999). These have been divided into subfamilies, one of which, the polymorphic GC-rich repetitive sequence (PGRS), encodes for multiple tandem repeats of glycine and alanine (Gordon et al. 1999). Considering this, the elevated concentrations of alanine and glycine, detected in our metabolomics study, may be due to a decrease in PGRS protein formation in the high iron cultured group comparatively, due to down-regulation of its synthesis. The increased concentrations of glycylalanine, a dipeptide of glycine and alanine detected in the high iron

group comparatively, further supports the notion that these amino acids are also being used in combination for PGRS formation.

Miethke *et al*., (2006), indicated that glutamate synthase is also iron-cofactor-dependant, and that during conditions of iron depletion, down-regulation of its expression in *B. subtilis* occurs. High intracellular iron concentrations, on the other hand, would be expected to increase intracellular glutamate and glutamine concentrations in this organism. The up-regulation of glutamine / glutamate synthesis, would affect the concentrations of other metabolite intermediates also, as glutamine metabolism is directly associated with the biosynthesis of many other compounds in the metabolism of *M. tuberculosis* (Kyoto Encyclopaedia of Genes and Genomes - Kanehisa *et al*., 2002), including ornithine, detected in elevated levels in the *M. smegmatis* cultured in high iron concentrations comparatively, as well as the butyric acid detected, which is synthesised via pyruvate (Henderson *et al*., 1997) which was also proposed to be elevated, as previously described.

Hydrocarbons, such as alkanes and alkenes, are common compounds in lipids of plants and insects and are also often found in bacteria. The two most considered pathways for *n*-alkane biosynthesis, are the elongation-decarboxylation pathway and the condensation-reduction pathway (Major *et al*., 1978). In bacteria, a modification of the condensation-reduction pathway, in which even-numbered fatty acids are decarboxylated, with a loss of a carboxyl group, resulting in an odd-numbered alkene or alkane, was proposed by Major *et al.,* (1978). In the high iron cultured *M. smegmatis,* elevated concentrations of alkenes and alkanes were found as seen in Table 5.1. Little information is known about the uptake of alkanes and alkenes, but is speculated that the uptake increases when these mycobacteria are under stress, such as oxidative stress (Ron *et al.,* 2002). The increased of these alkanes and alkenes, are also be due to decarboxylation of the elevated fatty acids, previously described.

5.4.2 Metabolites markers in *M. smegmatis* cultured during comparatively reduced iron concentrations

A comparison of *M. smegmatis* cultured in normal vs. comparatively lower iron concentrations, was additionally done, in order to validate the altered metabolome detected when comparing *M. smegmatis* cultured in comparatively higher iron concentrations, as described previously. When considering the markers identified using this metabolomics research approach, and indicated in Table 5.2, as was the case in the previous comparison, various amino acids, sugars, alkanes, alkenes, fatty acids and carboxylic acids, amongst others, were once again determined to be altered when *M. smegmatis* are cultured in comparatively reduced iron concentrations.

These metabolic differences observed in *M. smegmatis* cultured in comparatively lower iron concentrations, will be discussed in light of the associated role of iron in these metabolic pathways, and our previous discussion of the altered metabolome of the same organism cultured in comparatively elevated iron. One would expect, that in this comparison, similar metabolite differences would be detected, however, with changes in the opposite direction as to that previously detected in the *M. smegmatis* cultured in higher iron concentrations. This however, was not the case for all metabolite markers detected.

As mentioned earlier, excessive iron is known to lead to oxidative stress in all living organisms (Touati *et* al., 2000), however, oxidative stress also occurs when intracellular iron is too low (Ricci *et al*., 2002). As described in our previous comparison, and indicated by Husain *et al*., 2012, oxidative stress results in modulation of several metabolites involved in central energy metabolism, with a general increase in the abundances of those metabolites involved in glycolysis and its associated pathways. During these oxidative stress conditions, the Krebs cycle's activity has been shown to diminish, with an increased capacity of the glyoxylate shunt (Tretter *et* al., 2000), in addition to an accumulation of pyruvic acid and lactic acid (Abordo *et* al., 1999). They also observed an increase in most of the glycolytic intermediates from glucose-6-phosphate to pyruvate. Subsequently, alanine, isoleucine, leucine, valine glycine (Schulze-Siebert *et al*., 1984) and butyric acid (Henderson *et* al., 1997), all synthesised from the elevated pyruvate, induced by oxidative stress, were also elevated in the *M. smegmatis* cultured in low iron concentrations. Further proof for this, was

the elevated crotonic acid detected, which is derived from elevated crotonyl-CoA, an intermediate of the butyric acid synthesis pathway (Robinson *et al*., 1963). According to Rui *et al* (2010), acetic acid also increases significantly during oxidative stress, which we additionally observed in the *M. smegmatis* cultured in the comparatively lower iron concentrations. Due to the increased glyoxylate shunt, associated metabolites, including fumurate, were once again increased, accompanied by reductions in succinic acid, confirming the proposed inhibition of the Krebs cycle. Furthermore, as previously mentioned, oxidative stress results in up-regulation of amino acid transport and certain amino acid synthesis proteins, as well as a decreased lipid transport, lipid metabolism and membrane formation (Jang *et al*., 2009). Not only do these results coincide with the elevated amino acids detected in our study, but once again a general reduction in the use of fatty acids in the lower iron cultured *M. smegmatis*, indicated by intracellular elevations of these fatty acids. Furthermore, mannose, ribitol and mannonic acid, well-known primary substrates for energy metabolism via glycolysis (Ramakrishnan *et al*. 1972), are seen to be elevated in the high iron cultured *M. smegmatis*, comparatively, indicating reduced glycolysis and growth due to oxidative stress. Furthermore, mannonic acid hydratase, catalysing the conversion of mannonic acid to 3-deoxy-2-D-glucosonate, belongs to a class of iron requiring enzymes (Dreyer, 1986). Thus, during low iron conditions, the activity of mannonic acid hydratase will decrease, which in turn will cause mannonic acid to increase.

Galactofuranoside, D-xylofuranose and levoglucosan, are all cell wall component of these bacteria (Brennon 2003), were also found to be comparatively elevated in the low iron *M. smegmatis* group investigated in this study, supporting the hypothesis that these compounds are being used at a reduced rate for cell wall synthesis, due to the oxidative stress inhibition of growth, as was the case in the oxidative stress induced during the high iron growth conditions.

Cadaverine was also observed in elevated concentrations comparatively, in the *M. smegmatis* cultured in low iron concentrations. As previously mentioned, cadavarine is a polyamine antioxidant synthesised from lysine. Due to its antioxidant functions, these organisms most likely up-regulate cadaverine production, to help manage the oxidative stress induced by the decreased iron concentrations as was the case during the oxidative stress induced during the high iron growth conditions.

Once again, increased levels of odd chained alkanes was detected in the low iron cultured *M. smegmatis*. As previously mentioned, it is speculated that this increased alkanes are due to a decarboxylation of even-chained fatty acids.

Considering the above, those metabolites that coincided in the *M. smegmatis* cultured in high iron and low iron concentrations respectively, when compared to those cultured during normal iron concentrations, can be explained by a similar oxidative stress occurrence, which is common to both growth conditions. However, a number of differences also occurred. One of these would be to siderophore synthesis, due to its role in iron uptake. Siderophore synthesis would be expected to be up-regulated during conditions of reduced iron, considering their role in iron uptake and homeostasis (Rodriguez *et al*., 2002). As previously mentioned, siderophores are made up of amongst others, a large amount of lysine (Cole *et al*., 1998), which is derived from aspartic acid (Abelson *et al*., 1953), subsequently explaining the reduced levels of this amino acid, and elevated lysine detected, in the *M. smegmatis* cultured during reduced iron conditions.

As previously mentioned, Miethke *et al*., (2006), indicated that glutamate synthase, is iron-cofactor-dependant, and that during conditions of iron depletion, down-regulation of its expression in *B. subtilis* occurs, subsequently resulting in a decrease in glutamate, and it's associated metabolites, explaining the reduction in the glutamate derived metabolite marker pyroglutamic acid in our investigation (Reitzer, 2003).

Another interesting observation was the comparative reduction of propionic acid, methylmalonic acid, and the previously mentioned succinic acid. Propionic acid (or propionyl-CoA) is formed either by the breakdown of odd-chain-length fatty acids via β-oxidation, or the previously mentioned branched chain amino acids, or methionine. This in turn, can be used to synthesise the Krebs cycle intermediate succinic acid, via methylmalonic acid. The apparent reduction in all three of these related compounds seems to indicate the that iron is required for the synthesis of propionyl-CoA via the above mentioned catabolic pathways, due to the fact that in the high iron samples previously discussed, propionic acid was elevated. Considering this, it seems that during low iron growth conditions, there is an impairment of either β-oxidation, which is a common occurrence during times of oxidative stress.

Chapter 6

A metabolomics investigation of *M. smegmatis* wild type and esx-3 knockouts

Chapter 6

A METABOLOMICS INVESTIGATION OF *M. SMEGMATIS* WILD TYPE AND ESX-3 KNOCKOUTS

6.1 INTRODUCTION

The genome of *M. tuberculosis* has five copies of a gene cluster known as the ESX gene cluster region, and is associated with virulence and viability of mycobacteria (Gey van Pittius *et al.,* 2001). The ESX-3 gene cluster is essential for *in vitro* growth of *M. tuberculosis* (Sassetti *et al*., 2003) and was found to be involved in iron / zinc homeostasis (Bitter *et al*., 2009). Mycobacteria synthesize siderophores to capture iron and transport this into the cell. *M. tuberculosis* produces two types of siderophores: carboxymycobactins and mycobactins (Domenech *et al*., 2001).

A previous metabolomics investigation done by our group, illustrates metabolic changes in the ESX-3 knockouts, suggesting iron overload, despite the absence of the mycobactins due to the ESX-3 knockout (Loots *et al*., 2013). It was hypothesized that this overload occurs due to an increased exochelin synthesis, another iron uptake protein not associated with ESX-3, overcompensating for the perceived iron depletion in the knockout *M. smegmatis*. Considering this, in the current investigation, we will attempt to confirm these findings, by comparing the metabolome of the wild type parent strain of *M. smegmatis* to the altered metabolome of an ESX-3 knockout strain, cultured in normal iron conditions, and in so doing confirm whether or not a knockout of the ESX-3 gene cluster, is associated with a variation in iron uptake.

To accomplish our third objective as described in 3.3, and subsequently our second aim of this thesis, using the previously described GCxGC-TOFMS metabolomics methodology, we will compare the individually cultured ESX-3 knockout *M. smegmatis* organisms, to that of the wild type parent strain, grown under normal iron concentrations. The metabolite markers identified, will be discussed in the light of the information generated in Objective 2 (*M. smegmatis* wild types cultured in elevated or reduced iron), in order to confirm whether or not ESX-3 is associated with an increased iron uptake, as was proposed by a previous metabolomics investigation by Loots *et al*., (2013).

6.2 MATERIALS AND METHODS

6.2.1 Reagents and chemicals

All reagents and chemicals used are described in 4.2.1.

6.2.2 Bacterial cultures, preparation of *M.* smegmatis ESX-3 knockouts and sample preparation.

All *M. smegmatis* cultures were obtained from the DST/NRF centre of excellence for biomedical TB research (CBTBR), Stellenbosch University, South Africa.

The preparation of all bacterial cultures is described in 4.2.2.

For the culturing of *M. smegmatis* wild-type and *M. smegmatis* ESX-3 knockouts in normal iron conditions, 100µl of the *M. smegmatis* wild-type and *M. smegmatis* ESX-3 knockouts were transferred to Difco Middlebrook 7H9 medium, with a normal iron concentration of 150µM, supplemented with 0.2% glycerol and 0.05% Tween-80. The bacteria were incubated at 37 ºC with shaking at 180 rpm until the culture growth reached mid-log phase (OD_{600nm} = 0.6 – 0.8). The QC samples introduced in this study are the same as described in 5.2.2.4.

6.2.3 Quality controls

The same approach was used as described in 5.2.3, for preparing QC samples for this metabolomics investigation.

6.2.4 Extraction method (the conventional total metabolome method)

In order to compare the *M. smegmatis* wild-type and the *M. smegmatis* ESX-3 knockouts cultured in normal iron, 13 individually cultured sample repeats of the wild type and ESX-3 knockouts were extracted via the conventional total metabolome extraction method discussed in 4.2.3.1, and two samples, selected at random were extracted twice via the conventional total metabolome extraction method discussed in 4.2.3.1 (thus creating a sample set of 15 samples).

6.2.5 Two dimensional gas chromatography time of flight mass spectrometry (GCxGC-TOFMS)

The same GCxGC-TOFMS program was used for the injection of these samples, as described in 4.2.4.

6.2.6 Data processing

6.2.6.1 Peak identification and alignment

The same peak identification and alignment method was used as described in 4.2.5.1.

6.2.7 Statistical data analysis

All statistical data analysis was done as described in 5.2.6.

6.2.7.1 Data pre-treatment

The same data pre-treatment approach was used in this metabolomics investigation as described in 5.2.7.1.

6.2.7.2 Multivariate and univariate statistical analysis

The same data statistical analysis approach was used in this metabolomics investigation as described in 5.2.7.2.

6.3 RESULTS

6.3.1 Data pre-treatment

In the current chapter *M. smegmatis* wild type grown in normal iron conditions were compared to *M. smegmatis* ESX-3 knockouts grown in normal iron conditions.

The data pre-treatment steps described in section 6.2.7.1, led to a reduction of the total number of compounds from 699 to 374, of which 215 could be annotated by comparison of their mass spectra and retention times, to libraries prepared from previously injected standards.

6.3.2 PCA differentiation

A PCA analysis was done on the GCxGC-TOFMS pre-treated data set described above, generated from the extracted samples, in order to determine whether or not a natural differentiation occurs between the groups, based on the differences in their underlying metabolite profiles. As seen in Fig 6.1, for both comparisons, a clear differentiation between the sample groups was achieved when using the first two PCs of the PCA, and this natural differentiation between the groups can be attributed to underlying differences in the analysed metabolite profiles of these groups.

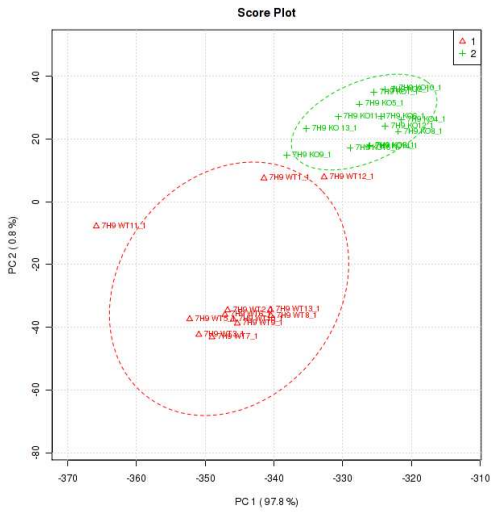


Fig 6.1: PCA scores showing the differentiation between the *M. smegmatis* wild type vs. *M. smegmatis* ESX-3 knockouts for the GCxGC-TOFMS data obtained after the total metabolome extraction method. (The *M. smegmatis* wild-type sample group are indicated in red and the ESX-3 knock-out sample group are indicated in green.)

The total amount of variance explained by the first two PCs (R^2 X cum) was 98.6%, of which PC 1 contributed 97.8% and PC 3 contributed 0.8%.

6.3.3 Metabolite marker identification

The same approach as described in 5.3.2 was used to select the metabolite markers which vary the most between these groups.

The PLS-DA model used for identifying the characteristic marker metabolites extracted using the total metabolome extraction procedure, had a modelling parameter R_2Y (cum) of 72.4%, with a Q_2 (cum) of 66.6%.

As indicated in Fig 5., metabolite markers for the comparison of the *M. smegmatis* wild type cultured in normal iron vs. *M. smegmatis* ESX-3 knockouts also cultured in normal iron, were selected by including those compounds having both a PLS-DA VIP > 1 and effect size d value of > 0.5, in addition to those compounds with a PCA modelling power > 0.5 and effect size d value > 0.5, totalling 129 compounds, of which 27 were annotated / identified using libraries prepared from previously injected standards by comparison of their mass spectra and retention times . These metabolite markers identified together with their mean relative concentrations, VIP value, PCA power and effect size are shown in Table 6.1. These metabolites will subsequently be used to explain the biological variation between the cure and failed treatment outcome groups.

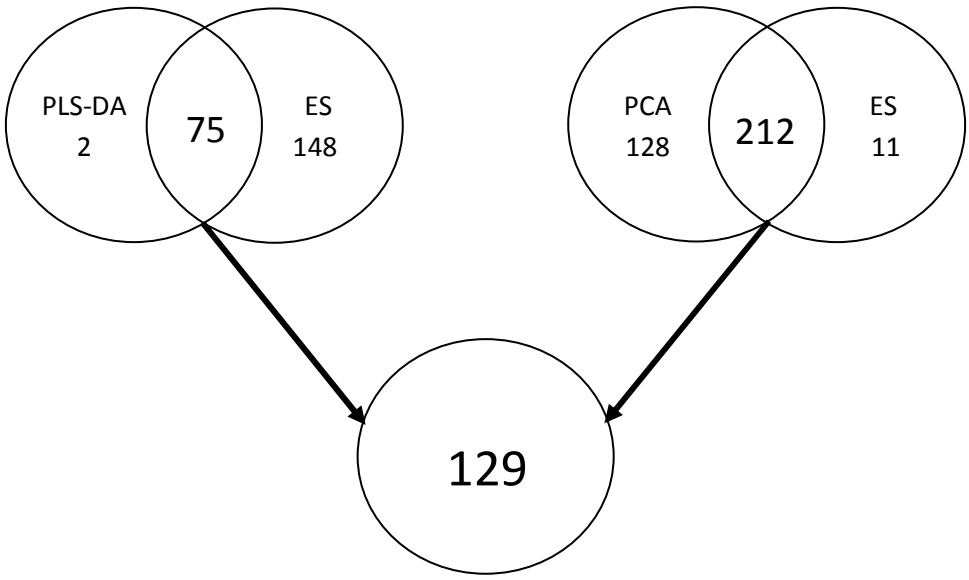


Fig 6.2: Summary of the selection approach used for identifying the metabolite markers best differentiating the *M. smegmatis* cultured in normal iron vs low iron concentrations.

Table 6.1: Metabolite markers selected from the comparison of *M. smegmatis* wild type vs. *M. smegmatis* ESX-3 knockouts, cultured in normal iron, indicating their mean relative concentrations (standard deviations in parenthesis), VIP values, PCA modelling power, effect size d-values and *p*-values.

			Wild type	ESX-3 Knockout		
VIP	PCA Power	Compound name (Chemspider ID)	Mean relative concentration (µg/mg)	Mean relative concentration (µg/mg)	Effect sizes	*p*-value
1.480	0.990	Proline (594)	0.005 (0.003)	0.026 (0.007)	>0.5	0.007
1.488	0.989	Glycoside (565858)	0.121 (0.057)	1.117 (0.264)	>0.5	0.106
0.621	0.978	Heptadecanoic acid (10033)	0.005 (0.001)	0.017 (0.001)	>0.5	7.42x10-16
1.167	0.975	Galactose (5814)	0.068 (0.108)	0.266 (0.090)	>0.5	1.88x10-5
0.149	0.968	Beta-Carotene (4444129)	0.023 (0.004)	0.017 (0.005)	>0.5	0.128
0.250	0.954	Malonic acid (844)	0.032 (0.004)	0.02 (0.005)	>0.5	0.038
0.814	0.935	Valine (6050)	0.012 (0.010)	0.057 (0.015	>0.5	0.012
0.398	0.931	Pyroglutamic acid (7127)	0.122 (0.036)	0.261 (0.105)	>0.5	0.005
0.861	0.922	Butyric acid (259)	0.013 (0.008)	0.072 (0.032)	>0.5	0.182
0.161	0.916	Ethane (6084)	0.224 (0.033)	0.168 (0.041)	>0.5	0.002
0.134	0.897	Octadecane (11145)	0.033 (0.010	0.044 (0.007)	>0.5	0.761
0.368	0.893	Hexadecanoic acid (960)	0.182 (0.076)	0.086 (0.035)	>0.5	0.937
1.238	0.883	Threonine (6051)	0.0007 (0.0001)	0.006 (0.001)	>0.5	0.38
0.824	0.881	Alanine (582)	0.067 (0.064)	0.288 (0.055)	>0.5	2x10e-7
0.650	0.858	Tridecane (11882)	0.011 (0.009)	0.031 (0.008)	>0.5	0.116
0.317	0.855	Acetic acid (171)	0.14 (0.016)	0.077 (0.029)	>0.5	0.368
0.456	0.848	Octadecanoic acid (5091)	0.112 (0.036)	0.047 (0.022)	>0.5	0.328
1.975	0.844	Xylose (559198)	0.0002 (0.0001)	0.002 (0.0006)	>0.5	0.125
0.499	0.809	Benzoic acid (238)	0.005 (0.003)	0.011 (0.002)	>0.5	0.028
2.326	0.772	Ribulose (133316)	0.01 (0.025)	0.002 (0.002)	>0.5	0.002

2.033	0.764	Aspartic acid (411)	0.0004 (0.001)	0.002 (0.001)	>0.5	0.358
0.402	0.754	Mannonic acid (588)	0.008 (0.004)	0.019 (0.008)	>0.5	0.0004
0.289	0.742	Levoglucosan (9587432)	0.01 (0.002)	0.019 (0.007)	>0.5	0.268
0.355	0.723	Oleic acid (393217)	0.036 (0.005)	0.019 (0.008)	>0.5	0.399
2.370	0.714	Glycine (730)	0.001 (0.001)	0.008 (0.002)	>0.5	0.0005
2.787	0.698	Isoleucine (769)	0.001 (0.001)	0.01 (0.001)	>0.5	0.0001
2.974	0.667	Pipecolic acid (826)	0.0001 (0.00003)	0.004 (0.003)	>0.5	0.289
2.314	0.643	Propionic acid (1005)	0.011 (0.009)	0.032 (0.006)	>0.5	0.002
1.460	0.541	Propene (7954)	0.058 (0.019)	0.007 (0.006)	>0.5	0.145
1.289	0.530	Leucine (864)	0.003 (0.003)	0.019 (0.002)	>0.5	0.001

6.4 DISCUSSION

6.4.1 Comparison of *M. smegmatis* wild type grown in normal iron concentrations vs. *M. smegmatis* ESX-3 knockouts cultured in normal iron concentrations

In 2013, Loots *et al.* investigated the response of *M. smegmatis* ESX-3 knockouts cultured in normal iron growth conditions. It was expected that intracellular iron would decrease, however, the ESX-3 knockouts showed an altered metabolome inductive of iron overload, despite the absence of the mycobactins. This was hypothesized to be due to an increased exochelin synthesis, another iron uptake protein not associated with ESX-3, overcompensating for the perceived iron depletion in the knockout organism. In this study, we also used metabolomics research approach to generate information which could support this hypothesis.

We will explain the metabolic differences observed in *M. smegmatis* ESX-3 knockouts cultured in normal / standard iron concentrations in light of the associated role of iron in these metabolic pathways, considering the previous results generated in Chapter 5, comparing growth of *M. smegmatis* wild type strains cultured in either elevated or reduced iron concentrations, and to that previously described by Loots *et al*. (2013).

An important point to mention at this stage is a major difference in this study, compared to that of Loots *et al.* (2012), is the absence of the enzyme catalase in the growth media used for culturing the organisms investigated. Catalase is an enzyme catalysing hydrogen pyroxide to water and oxygen, and is thus considered an important enzyme in its response to oxidative stress. Thus any oxidative stress that occurred in this study, will be expected to be more pronounced as compared to that observed by Loots *et al*. (2013), were catalase was present in the growth media.

As extensively described in the previous chapter, excessive iron is known to lead to oxidative stress (Touati *et al*., 2000), due to its participation in the Fenton reactions. During these oxidative stress conditions, the Krebs cycle's activity has been shown to diminish, with an increased capacity of the glyoxylate shunt (Tretter *et* al., 2000), in addition to an accumulation of pyruvic acid and lactic acid (Abordo *et* al., 1999). Phosphoenolpyruvate carboxykinase (pckA), an iron regulated enzyme catalysing the conversion of oxaloacetate to phosphoenolpyruvate, is also up-regulated when iron increases inside the cell (Sauer *et al.,* 2004). Furthermore, phosphoenolpyruvate, is converted to pyruvate via the enzyme pyruvate kinase, also believed to be up-regulated by intracellular iron (Chadvadi *et al*., 2009). Confirmation of the elevated pyruvate due to iron overload, similarly, as to what we detected in the *M. smegmatis* wild type strains grown in elevated iron conditions in Chapter 5, are the elevations in alanine, isoleucine, leucine, valine, threonine and glycine in the *M. smegmatis* ESX-3 knockouts, all synthesised from pyruvate (Schulze-Siebert *et al*., 1984). Further evidence for the elevated iron in the ESX-3 knock-outs, is the elevated propionic acid detected, which is the natural breakdown product of isoleucine and valine (Elsden *et al*., 1976), and the elevated butyric acid detected, which is synthesised via pyruvate (Henderson *et* al., 1997), which were also elevated in the wild type *M. smegmatis* cultured in elevated iron conditions in Chapter 5. As previously discussed, pyruvate also acts as a substrate for the glyoxylate shunt, thus, the increased pyruvate would also be expected to result in an increased glyoxylate shunt activity, further confirmed by the elevated glycine detected (Tang *et al*., 2009) in the *M. smegmastis* ESX-3 knockouts in our investigation, the *M. smegmatis* cultured in elevated iron in Chapter 5, as well as in the previous study by Loots *et al*., (2013). According to Rui *et al* (2010), acetic acid also increases significantly during oxidative stress, which we observed in the *M. smegmatis* ESX-3 knockouts.

Similarly as to that previously describe in the ESX-3 knockouts described by Loots *et al*., (2013) and in the elevated iron cultured *M. smegmatis* group discussed in Chapter 5, various fatty acids, in this instance: oleic acid, octadecanoic acid, hexandecanoic acid and

heptadecanoic acid, were all seen in reduced intracellularly in the *M. smegmatis* ESX-3 knockouts, relative to the *M. smegmatis* wild type strains. As described by Loots *et al*., (2013), this may be attributed to the previously described oxidative stress induced by the elevated iron in these organisms (Jang *et al.,* 2009). Additionally, iron has very poor solubility, thus iron must form a suitable complex for translocation into a cell. Saturated fatty acids subsequently serve as a substrates for iron translocation and forms fatty acid-iron complexes for these purposes (Qian *et al*., 1991). Considering this, the increased intracellular iron concentrations, would be expected to result in an increased synthesis of these fatty acid-iron complexes, and subsequently result in the reduction in the free fatty acids seen in the *M. smegmatis* ESX-3 knockouts relative to the *M. smegmatis* wild type strains.

Similarly as to what we detected in the elevated iron cultured *M. smegmatis* group in Chapter 5, glucose, xylose, galactose, ribulose and mannonic acid, the well-known primary substrates for energy metabolism via glycolysis (Ramakrishnan *et al*. 1972), are seen to be elevated in the *M. smegmatis* ESX-3 knockout strain comparatively. According to Shi *et al*. (2010), reductions in these glycolysis precursors are not only as a result of their roles as substrates for energy production during periods of increased growth, but also due to their roles as precursors for the biosynthesis of other macromolecules. Furthermore, in a metabolomics investigation done by Meissner-Roloff *et al*.,(2012), a reduction in these metabolites where also ascribed to their increased utilization for growth, metabolism and cell wall synthesis in hyper-virulent Beijing *M. tuberculosis*. Considering this, the accumulation of these glycolysis substrates in this study, suggests hindered growth, metabolism and cell wall synthesis in this group, possibly due to the iron overload and the iron induced oxidative stress. Furthermore, as described by Meissner-Roloff *et al*., (2012), another carbohydrate molecule associated with the cell wall of these organisms, includes levoglucosan (Brennan 2003). This carbohydrate forms part of long chain molecules, which assist in maintaining cell wall structure and even provide immunogenic variation (Chatterjee, 1997). Thus the elevated levels of levoglucosan detected in the high iron cultured *M. smegmatis* ESX-3 knockout group investigated in this study, supports the hypothesis that these compounds are being used at a reduced rate for cell wall synthesis, further confirming reduced growth, implicating iron overload, considering a similar occurrence was reported in Chapter 5. Similarly to the high iron cultured *M. smegmatis* in Chapter 5, various alkanes and alkenes were also altered in the *M.* smegmatis ESX-3 knockout group. Little information is however known about the uptake or synthesis of these alkanes and alkenes, but it is speculated that these alterations may be induced by stress, including oxidative stress (Ron *et al.,* 2002)

As previously mentioned and reported in the previous Chapter, high intracellular iron concentrations would be expected to increase intracellular glutamate and glutamine concentrations in the organism. Although glutamine / glutamate was not directly detected, various metabolites associated with an elevation in this amino acid were, including elevated proline, threonine and pyroglutamic acid. In the previous study done by Loots *et al*., (2013), an increased intracellular concentration of pyroglutamic acid was also observed, in addition to a reduction in this metabolite in the *M. smegmatis* cultured in reduced iron concentrations in Chapter 5, further supporting iron overload induced by knocking out the ESX-3 gene cluster.

As described, siderophores are derived from salicylate by the addition of serine, two lysine's and various fatty acids (Cole *et al,* 1998). Since the *M. smegmatis* ESX-3 knockouts are not able to synthesise these siderophores due to the ESX-3 gene cluster being absent, lysine would be expected to accumulate, similarly as to what was seen in the high iron cultured *M. smegmatis* wild types in Chapter 5. Although lysine wasn't detected as one of the prominent markers using the above mentioned marker selection strategy, pipecolic acid, a derivative of lysine was detected in elevated concentrations (Fujii *et al*., 2002).

Lastly, β-carotene was also identified as a metabolite marker, reduced in the ESX-3 knockouts. β-carotene 15,15'-monooxynase (BCMO) requires iron as a co-factor for activation (Wang, 2008). BCMO catalyses the cleavage of β-carotene, on the central bond into two vitamin A molecules. Elevated iron would subsequently be expected to result in a reduction in β-carotene, as was the case in the *M. smegmatis* ESX-3 knockouts relative to the *M. smegmatis* wild type.

In summary, all of the metabolic alterations detected and described above, it does in fact support the previous hypotheses postulated by Loots *et al.* (2013), that iron increases in the ESX-3 knockouts, most likely due to an increased exochelin synthesis, another iron uptake protein not associated with ESX-3, overcompensating for the perceived iron depletion in the knockout organism.

As a final observation, when comparing the metabolite markers identified in the *M. smegmatis* wild type strains cultured in normal iron vs. the *M. smegmatis* wild type cultured in high iron (Table 5.1), comparatively to the metabolite markers identified in the *M. smegmatis* wild type strains vs. *M. smegmatis* ESX-3 knockouts, both cultured in normal / standard iron concentrations (Table 6.1), 11 markers are identical. All 11 markers changed in concentration in the same direction, however, the metabolic perturbation / concentration changes in the *M. smegmatis* wild type cultured in high iron, are more pronounced than that

of the *M. smegmatis* ESX-3 knockouts cultured in normal iron. This more sever metabolic perturbation, may simply be explained by the *M. smegmatis* wild types cultured in the largely elevated iron, are suffering from a more sever iron induced stress, simply because there is more iron in the growth media to cause these alterations. Although the ESX-3 knockouts take up more iron due to the exochelins overcompensating for the lacking mycobactins, the extent of the stress is limited by the available iron in the growth media, which was at normal / standard levels, and hence comparatively less stress was endured, although comparatively more in the latter than in the *M. smegmatis* wild types cultured in normal / standard iron concentrations. This observation only further confirms our findings.

Chapter 7
Conclusion

Chapter 7

CONCLUSION

In conclusion, I will summarise the major findings and conclusions made in Chapter 4, 5 and 6, in the context of the aims proposed in Chapter 3. Additionally, I will discuss final remarks and future research prospects based on the outcomes of the study.

7.1 GENERAL CONCLUSION

According to the World Health Organization (WHO), *M. tuberculosis,* the causative agent of TB, accounts for approximately 1.7 million deaths annually. Further contributing causes to the worldwide TB incidence, is the widespread unavailability and ineffectiveness of TB vaccines, time consuming diagnostic methods and unsuccessful treatment approaches. Research better characterising mycobacteria on general, or other mycobacterium species, may help us to better understand *M. tuberculosis* and TB disease mechanisms, which will in turn lower the future TB disease prevalence.

A Metabolomics research approach was subsequently used to generate new information, which could potentially better explain the underlying metabolic processes associated with the role of ESX-3 and iron in *Mycobacterium,* using *M. smegmatis* as a model.

In Chapter 4, we determined the most optimal extraction method for this metabolomics investigation, as described in Aim 1 (section 3.2). Two extraction methods were investigated, considering their comparative repeatability and their respective capacities to extract those compounds which best differentiate the *M. smegmatis* ESX-3 knockouts and wild-type parent strains. Considering the results generated, the conventional total metabolome method showed best comparative repeatability and capacity to extract those compounds which best differentiate the compared groups. Furthermore, the slightly poorer comparative repeatability of the fractionated total metabolome method, was ascribed to the fact that this method involves extra steps, increasing the chance for analytical errors, and due to the fact that many of the extracted compounds, were extracted in both phases, but not always in equal amounts. Although the conventional total metabolome method has the ability to best differentiate between the *M. smegmatis* wild-type and ESX-3 knock-outs, the fractionated total metabolome method also showed an acceptable differentiation between

the sample groups. Considering this, the total metabolome method was chosen for further analyses, for the following reasons: 1) it is simpler, 2) faster, 3) showed better repeatability, 4) extracts those compounds better differentiating the compared groups and 5) has been previously described for metabolomics analyses characterising ESX-3 gene functionality, hence potentially allowing us to compare results to that previously generated and published previously by ourselves and other groups.

Subsequently, in Chapter 5 and 6, we used the chosen extracted method to accomplish our Aim 2 as described in section 3.2.

This involved using the chosen extraction method, as investigated in Chapter 4, followed by GCxGC-TOFMS analysis of the separately cultured *M. smegmatis* wild-type sample extracts, cultured in normal, low and high iron conditions, to determine the influence of varying iron concentrations on the metabolome of this organism, by metabolomics comparisons of these groups (Chapter 5).

In Chapter 6, an identical research approach was used to compare the metabolome of a *M. smegmatis* ESX-3 knock-out strain, to that of a *M. smegmatis* wild type parent strain, both cultured in normal / standardised iron concentrations, and then comparing the metabolomics results generated to that in Chapter 5. In doing so, we aimed to correlate whether or not the *M. smegmatis* ESX-3 knockout metabolome, correlates to that of *M. smegmatis* cultured in either elevated or reduced iron, in order to prove, or disprove the previous hypotheses postulated by Loots *et al*., (2013).

As stated earlier, the expression of various metabolite biosynthesis genes has been shown to be regulated by iron in *Mycobacteria* and various other bacterial species (Rodriguez *et al.,* 2002; Maciag *et al.,* 2007; Maciag *et al.,* 2009). Additionally, iron is an essential cofactor for many of the enzymes involved in metabolism in these organisms. Considering this, it came as no surprise that an altered metabolite profile occurred as a result of the varying iron concentrations in the growth media. These alterations were subsequently explained in light of the associated role of iron in these altered metabolic pathways, not only for gaining a better understanding of the role of iron in the metabolism of these organisms, but also as a validation for the markers detected (Chapter 5).

Considering the results generated in Chapter 6, using the same metabolomics research approach, the altered metabolome of the *M. smegmatis* ESX-3 knockouts, correlated well to that of the *M. smegmatis* cultured in elevated iron growth conditions (Chapter 5). This

suggests ESX-3 is involved in iron uptake, and that knocking out the ESX-3 gene cluster of *M. smegmatis* does in fact result in a metabolome profile suggesting iron overload, as was proposed by Loots *et al* (2013), most probably due the exochelins overcompensating for the absence of mycobactins, in *M. smegmatis* ESX-3 knockouts.

7.2 FUTURE PROSPECTS

In the light of the above observations, future research studies further confirming these observed alterations to the metabolome of *M. smegmatis* and the role of ESX-3 in iron uptake would be: comparatively analysing the ESX-3 knockout, grown under a variety of varying iron concentrations (low, normal and high iron media), to that of the wild type parent strain. Considering the results of this study, reduced extracellular iron would be expected to result in metabolite profiles in the ESX-3 knockouts, similar to that observed in wild type *M. smegmatis* grown in normal extracellular iron concentrations. High extracellular iron concentrations on the other hand, may have opposing effects on the growth of *M. smegmatis* ESX-3 knockouts, resulting in a vastly distinct metabolic profile, induced by severe iron overload and oxidative stress. Furthermore, the current investigation can be repeated using other species of *Mycobacterium,* which also utilize both exochelins and mycobactins, such as *M. bovis,* for further confirmation of these conclusions, and in order to ensure that these changes are not due to some other species specific characteristic of *M. smegmatis*.

More importantly however, now that we have more conclusive evidence as to the role of ESX-3 in growth, through regulating iron uptake through exochelin synthesis, we can focus on ways of inhibiting iron uptake of *M. tuberculosis*, by investigating means of sideophore synthesis inhibition, or alternatively, disruption of those metabolic pathways involved in iron utilisation of this organism.

Chapter 8

References

Chapter 8

REFERENCES

- ABELSON, P.H., BOLTON, E., BRITTEN, R., COWIE, D.B. & ROBERTS, R.B. 1953. synthesis of the aspartic and glutamic families of amino acids in *Escherichia coli*. *Proceedings of the. National Acadademy of Sciences of the USA*, 130:120-126.

- ABDALLAH, A.M., GEY VAN PITTIUS, N.C., DiGUISEPPE CHAMPION, P.A., COX,J., LUIRINK, J., VANDENBROUCKE-GRAULS, C.M.J.E., APPELMELK, B.J. & BITTER, W. 2007. Type VII secretion-mycobacteria show the way. *Nature reviews Microbiology,* 5:883-891.
- ABORDO, E.A., MINHAS, H.S. & THORNALLEY, P.J. 1999. Accumulation of alpha-Oxoaldehydes during Oxidative Stress: A role in Cytotoxicity. *Biochemical Pharmacology*, 58:641-648.

- BELTAN, E., HORGEN, L. & RASTOGI, N. 2000. Secretion of cytokines by human macrophage upon infection by phatogenic and non-pathogenic mycobacteria. *Microbial Pathogenesis,* 25:313-318.

- BEGONA, B.G., BAPTISTADE, J.V., TOMA, V. & JORGE, B.V. 1999. Histamine and Cadaverine Production by Bacteria Isolated from Fresh and Frozen Albacore (*Thunnus alalunga*). *Journal of food protection,* 62(8):933-937

- BITTER, W., HOUBEN, E.N.G., LUIRINK, J., APPELMELK, B.J. 2009. Type VII secretion in mycobacteria: classification in the line with cell envelope structure. *Trends in Microbiology*, 17(8):337-338.

- BRENNAN, M.J., DELOGU, G., CHEN, Y., BARDAROV, S., KRIAKOV, J., ALAVI, M. & JACOBS W.R. 2001. Evidence that mycobacterial PE-PGRS proteins are cell surface constituents that influence interactions with other cells. *Infection and Immunity*, 69:7326-7333.

- BRERETON, R.G., SHEN, H., CARTER,J.F. & ECKERS, C. 2003. Discrimination between tablet production methods using pyrolysis-gas chromatography-mass spectrometry and pattern recognition. *Analyst*, 128:287-292.

- CANNEVA, F., BRANZONI, M, RICCARDI, G., PROWEDI, R. & MILANO, A. 2005. Rv2358 and FurB: two transcriptional regulators from *Mycobacterium tuberculosis* which respond to zinc. *Journal of Bacteriology,* 187:5837-5840.

- CHADVADI, S., WOOFF, E., COLDHAM, N.G., SRITHARAN, M., HEWINSON, R.G., GORDON, S.V. & WHEELER, P.R. 2009. Global effect of inactivation of pyruvate kinase gene in the *Mycobacterium tuberculosis* complex. *Journal of Bacteriology*, 191(24):7545-7553.

- CHATTERJEE, D. 1997. The mycobacterial cell wall: Structure, biosynthesis and sites of drug action. *Current Opinion in Chemical Biology*, 1(4)579–588.

- CHONG, I.G. & JUN, C.H. 2005. Preformance of some variable selection methods when multicollinearity is present. *Chemometics and Intelligent Laboratory Systems*, 78:103-112.

- CHOU, C.J., WISEDCHAISRI, G., MANFELI, R.R., ORAM, D.M., HOLMES, R.K., HOL, W.G.J. & BEESON, C. 2004. Functional studies of the *Mycobacterium tuberculosis* iron-dependent regulator. *Journal of Biogical Chemistry,* 279:53554-53561.

- 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., CONOR, 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., OLIVIER, K., OSBORNE, J., QUAIL, M.A., RAJANDREAM, M.A., ROGERS, J., RUTTER, S., SEEGER, K., SKELTON, J., SQUARES, R., SQUARES, S., SULSTON, J.E., TAYLOR, K., WHITEHEAD, S. & BARELL, B.G. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature,* 393:537-544.

- COLE, S.T. 2002. Comparative mycobacterial genomics as a tool for drug target and antigen discovery. *European Respiratory Journal,* 20(36):785-865.

- DAGLEY, S., TRUDGILLANDA, P.W. & CALLELY,G. 1961. Synthesis of Cell Constituents from Glycine by a Pseudomonas. *Biochemistry Journal*, 81:623-631.

- DOMENECH, P., BARRY, C.E. & COLE, S. 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., SPES-FABER, E., VOGELS, J.T.WE., MEULMAN, J.J., BOOMSMA, D.I., VAN DER GREEF,J. & HANKMEIER, T. 2010. Equating, or correction for between-block effects with aplication to body fluid LC-MS and NMR metabolomics data sets. *Analytical Chemistry*, 82:1039-1046.

- DREYER, J.L. 1986. The role of iron in the activation of mannonic and altronic acid hydratase, two Fe-requiring hydro-lyases. *European Journal of Biochemistry,* 166:623-630.

- DUNAU, E. & SMITH, I. 2003. *Mycobacterium tuberculosis* gene expression in macrophages. *Microbes and Infection,* 5:629-637.

- DUNN, W.B., BAILEY, N.J.C & JOHNSON, H.E. 2005. Measuring the metabolome: current analytical technologies. *Analyst*, 30:606-625.

- DUSSERGET, O., RODRIGUEZ, G.M. & SMITH, I. 1996. An *ideR* mutant of *Mycobacterium smegmatis* has a depressed siderophore production and an altered oxidative-stress response. *Molecular Microbiology*, 22:535-544.

- DUSSURGET, O., RODRIGUEZ, M. & SMITH, I. 1998. Protective role of mycobacterial IdeR against reactive oxygen species and isoniazid toxicity. *Tuberculosis Lung Disease*, 79:99-106.

- ELLIS, S.M. & STEYN, H.S. 2003. Practical significance (effect size) versus or in combination with statistical significance (p-value). *Management Dynamics*, 12(4):51-53.

- ELSDEN, S.R., HILTON, M.G. & WALLER, J.M. 1976. The End Products of the Metabolism of Aromatic Amino Acids by Clostridia. *Arch. Microbiology*,170:283-288.

- ESAKAOVA, O.A., MESHALKINA, L.E., KOCHETOV, G.A. & GOLBIK, R. 2009. Halogenated Pyruvate Derivatives as Substrates of Transketolase from *Saccharomyces cerevisiae*. *Biochemistry (Moscow),* 74(11):1518-1523.

- FORTUNE, S.M., JAEGER, A., SARACINO, D.A. & CHASE, M.R. 2005. Mutually dependent secretion of proteins required for mycobacterial virulence. *Proceedings of the. National Acadademy of Sciences of the USA*, 102:10676-10681.

- FUJII, T., MUKAIHARA, M., AGEMATU, H. & TSUNEKAWA, H. 2002. Biotransformation of L-lysine to L-pipecolic acid catalyzed by L-lysine 6-aminotransferase and pyrroline-5-carboxylate reductase. *Bioscience Biotechnology Biochemistry*, 66(3):622-627.

- 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, J. MEDINA, N., MANSOOR, H., PRYOR, M., DUTHOY, S., GRODIN, S., LACROIX, C., MONSEPE, C., SIMON, S., HARRIS, B., ATKIN, R., DOGGETT, J., MAYES, R., KEATING, L., WHEELER, P.R., PARKHILL, J., BARELL, B.G., COLE, S.T., GORDON, S.V. & HEWISON, R.G. 2003. The complete gnome sequence of *Mycobacterium bovis. Proceedings of the. National Acadademy of Sciences of the USA.* 100(13):7877-7882.

- GEY VAN PITTIUS, N.C., 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. *BioMed Central*, 6(95):1471-2148.

- 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 biology*, 2(10):0044.1-0044.18.

- GOBIN, J. & HORWITZ, M. 1996. Exochelins of *Mycobacterium tuberculosis* remove iron from human iron-binding proteins and donate iron to mycobactins in *M. tuberculosis* cell wall. *Journal Experimental Medicine,* 183:1527-1532.

- GOLD, B., RODRIGUEZ, G. M., MARRAS, M.P., PENTECOST, M. & SMITH, I. 2001. The *Mycobacterium tuberculosis* IdeR is a dual functional regulator that controls transcription of genes involved in iron acquisition, iron storage and survival in macrophages. *Molecular Microbiology*, 42:851-865.

- GOODRACE, R., VAIDYANATHAN, S., DUNN, W.B., HARRINGAN, G.G. & KELL, D.B. 2004. Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends in biotechnology*, 22:245-252.

- GORDON, S.V., EIGLMEIER, K., BROSCH, R., GARNIER, T., HONORE, N., BARELL, B. & COLE, S.T. 1999. Genomics of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. *Mycobacteria:molecular biology and virulence,* 93-109.

- GORDHAN, B.G. & PARISH, T. 2001. Gene Replacement using Pretreated DNA. In *Mycobacterium tuberculosis* Protocols. Eds Parish, T. and Stoker, N.G. (Humana Press, New Jersey), 77-92.

- HENDERSON, R.A., JONES, C.W. 1997. Poly-3-hydroxybutyrate production by washed cells of Alcaligenes eutrophus; purification, characterisation and potential regulatory role of citrate synthase. *Arch. Microbiology.* 1686486-492.

- HOBSON, R.J., McBRIDE, A.J., KEMPSELL, K.E., DALE, J.W. 2005. Use of an arrayed promoter-probe library for the identification of macrophage-regulated genes in *Mycobacterium tuberculosis*. *Microbiology*, 184(5):1571-1579.

- HUSAIN, A., SANTO, D., JEELANI, G., SOGA, T. & NOZAKI, T. 2012. Dramatic increase in glycerol biosynthesis upon oxidative stress in the anaerobic protozoan parasite Entamoeba histolytica*. PLOS Neglected Tropical Diseases*, 6(9):1-11.

- JANG, H.J., NDE, C., TOGHROL,F. & BENTLEY, W.E. 2009. Global transcriptome analysis of the *Mycobacterium bovis* BCG response to sodium hypochlorite. *Applied Microbiology and biotechnology*, 85(1):127-140.

- JONSSON, P., GUILLBERG, J., NODSTROM, A., KUSANO, M., KOWALCZYK, M., SJOSTROM, M. & MORITS, T. 2004. A strategy for identifying differences in large series of metabolome samples analyzed by GC/MS. *Analytical chemistry*, 76:1738-1745.

- JONES, C.M. & NIEDERWEISS, M. 2012. Mycobacterium tuberculosis can utilize heme as an iron source. *Journal of Macteriology*, 7:1767-70.

- KADDURAH-DAOUK, R., KRISTAL, B.S. & WEINSHILBOUM, R.M. 2008. Metabalomics: A global Biochemical approach to drug response and disease. *Annual Review of Pharmacology and Toxicology,* 48:653-683.

- 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.

- KANEHISA, M., GOTO, S., KAWASHIMA, S., NAKAYA, A. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Res*. 30(1): 42-46

- KEHL-FIE, T.E., SKAAR, E.P. 2010. Nutritional immunity beyond iron: a role for manganese and zinc. *Current opinion in Chemical Biology*, 14(2):218-224

- LOOTS, D. T., MEISSNER-ROLOFF, R. J., NEWTON-FOOT, M., GEY VAN PITTIUS, N. C. 2013. A metabolomics approach exploring the function of the ESX-3 type VII secretion system of *M. smegmatis.* Metabolomics, in press.

- MACIAG, A., DAINESE, E., RODRIGUEZ, G.M., MILANO, A., PROVVEDI, R., PASCA, M.R., SMITH, I., PALù, G., RICCARDI, G. & MANGANELLI,R. 2007. Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *Journal of Bacteriology,* 189(3):730-740.

- MACIAG, A., PIAZZA, A., RICARDI, G. & MILANO, A. 2009. Transcriptional analysis of ESAT-6 cluster 3 in *Mycobacterium smegmatis. BMC Microbiology*, 9(48):1471-2180.

- MAJOR, M.A. 1978. Biosynthesis of hydrocarbons in Insects: Decarboxylation of Long Chain acids to n-alkanes in Poriplaneto. *Lipids*, 13(5):323-328

- 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.

- MEISNER-ROELOFF, R., KOEKEMOER, G., WARREN, R.M., LOOTS, D.T. 2012. A metabolomics investigation of a hyper- and hypo-virulent phenotype of Beijing lineage M. tuberculosis. *Metabolomics Springer*, 10.1007/s11306-012-0424-6

- MIETHKE, M., WESTERS, H., BLOM, E., KUIPERS, O.P., MARAHIEL, M.A. 2006. Iron starvation triggers the stringent response and induces amino acid biosynthesis for bacillibactin production in *Bacillus subtilis. J Bacteriol*. 188(24): 8655-8657

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

- MULDER, N., RABIU, H., JAMIESON, G. & VUPPU, V. 2009. Comparative analysis of microbial genome to study unique and expanded gene families in *Mycobacterium tuberculosis*, 9:314-321.

- 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.

- NGUYEN, L. & PIETERS, J. 2005. The Trojan horse: survival tactics of pathogenic mycobacteria in macrophages. *TRENDS in Cell Biology,* 15(5):267-276.

- ORESIC, M. 2009. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutrition, Metabolism & Cardiovascular Diseases*, 19:816-824.

- OLIVIER, I & LOOTS, D.T. 2012. A comparason of two extraction methods for differentiating and characterising various *Mycobacterium* species and *Pseudomonas auruginosa* using GC-MS metabolomics. *African Journal of Microbiology Research*, 6:3159-3172.

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

- POHL, E., HOLMES, R.K. & HOL, W.G.J. 1999. Crystal structure of the iron-dependent regulator (IdeR) from *Mycobacterium tuberculosis* shows both metal binding sites fully occupied. *Journal of Molecular Biology,* 285:1145-1156.

- QIAN, M.W. & EATON, J.W. 1991. Iron translocation by free fatty acids. *American journal of pathology*, 139(6):1425-1434.

- RAGHU, B. *et al*. 1993. Effect of iron on the growth and sidephore production of mycobacteria. *Biochemistry and Molecular Biology International,* 31:341-348.

- RAMAKRISHNAN, T., MURTHY, P. S., & GOPINATHAN, K. P. 1972. Intermediary metabolism of mycobacteria. *Bacteriology Review,* 36(1), 65–108.

- RATLEDGE, C. 2004. Iron, mycobacteria and tuberculosis. *Tuberculosis,* 84:110-130.

- REITZER, L. 2003. Nitrogen assimilation and global regulation in Escherichia coli. *Annual Review of Microbiology*. 57, 155-76.

- RICCI, S., JANULCZYK, R. & BJORCK, L., 2002. The Regulator PerR Is Involved in Oxidative Stress Response and Iron Homeostasis and Is Necessary for Full Virulence of *Streptococcus pyogenes*. *Infectious immunology*, 10:4968-4976.

- ROBINSON, J.D., BRADLEY, R.M. &. BRADY, R.O. 1963. Biosynthesis of fatty acids. *The Journal of Biological Chemistry*, 238:528-532.

- RODRIGUEZ, G.M. 2006. Control of iron metabolism in *Mycobacterium tuberculosis*. *TRENDS in Microbiology*, 14(7):320-327.

- RODRIGUEZ, G.M., GOLD, B., GOMEZ, M., DUSSERGET, O. & SMITH I. 1999. Identification and characterization of two divergently transcribed iron regulated genes in *Mycobacterium tuberculosis. Tuberculosis Lung disease*, 79:287-298.

- RODRIGUEZ, G.M., VOSKUIL,M.I., GOLD,B., SCHOOLNIK, G.K. & SMITH, I. 2002. *ideR,* an Essential gene in *Mycobacterium tuberculosis*: Role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infection and Immunity*,

70(7):3371-3381.
- RON, E.L. & ROSENBERG, E. 2002. Biosurfactants and oil bioremediation. *Current opinion in Biology*. 13:249-252.

- RUI, B., SHEN, T., ZHOU, H., LUI, J.,CHEN, J., PAN, X., LIU, H., WU, J., HAORAN, Z. & SHI, Y. 2010. A systematic investigation of *Escеherichia coli* central carbon metabolism in response to superoxide stress. *BMC Systems Biology*. 4:122

- SAMARTZIDOU, H., & DELCOUR, A. H. 1998 . *E.coli* PhoE porin has an opposite voltage-dependence to the homologous OmpF. *EMBO Journal*, 17(1) 93–100.

- SASSETTI, C.M., BOYD, D.H. & RUBIN, E.J. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48:77-84.

- SAUER, U. & EIKMANNS, B.J. 2004. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiology reviews*, 29:765-794.

- SHI, L., SOHASKEY, C. D., PFEIFFER, C., DATTA, P., PARKS, M., MCFADDEN,
- J. 2010. Carbon flux rerouting during Mycobacterium tuberculosis growth arrest. *Molecular Microbiology*, 78(5)1199–1215.

- SCHOEMAN, J.C. & Loots, D.T. 2011. Improved disease characterisation and diagnostics using metabolomics: a review. *Journal of Cell tissue Research*, 11:2673-2683.

- SCHOEMAN, J.C., DU PREEZ, I. & LOOTS, D.T. 2012. A comparison of four sputum pre-extraction preparation methods for identifying and characterising *M. tuberculosis* using GCxGC-TOFMS metabolomics. *Journal of Microbiological Methods*.. *91(2):301-11*

- SCHMITT, M.P., PREDICH, M., DOUKHAN, L., SMITH, I. & HOLMES, R.K. 1995. Characterization of an iron-dependent regulatory protein (IdeR) of *Mycobacterium tuberculosis* as a functional homolog of the diphtheria toxin repressor (DtxR) from *Corynebacterium diphtheria*. *Infection and Immunity*, 63:4284-4289.

- SERAFINI, A., BODRIN, 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): 6430-6344.

- SHILOH, M.U. & DiGUISEPPE CHAMPION, P.A. 2010. To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis? *Current Opinion in Microbiology*, 13:86-92.

- SIEGRIST, M.S., UNNIKRISHNAN, M., MCCONELL, M.J., BOROWSKY, M., CHENG, T., SIDDIQI, N., FORTUNE, S.M., MOODY, D.B. & RUBIN, E.J. 2009. Mycobacterial ESX-3 is required for mycobactin-mediated iron acquisition. *Proceedings of the. National Acadademy of Sciences of the USA,* 106(44):18792-18797.

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

- SORESEN, A.L., NAGAI, S., HOUEN, G., ANDERSEN, P. & ANDERSE, A.B. 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis. Infect Imun,* 63:1710-1717.

- STRONG, M., SAWAYA, M., WANG, S., PHILIPS, M., CASCIO, D. & EISENBER, D. 2006. Toward the structural genomics of complexes: crystal structure of PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proceedings of the. National Acadademy of Sciences of the USA*, 103:8060-8065.

- SAUER, U. & EIKMANNS, B.J. 2006. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiology Reviews*. 29: 765-794.
- SAUER, U., BERNHARD, J. & EIKMANS, B. 2004. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMSMicrobiology Reviews,* 29:765-794.

- SCiULZE-SIEBERT, D., HEINEKE, D., SCHARF, H. & SCHULTZ, G. 1984. Pyruvate-derived amino acids in spinach chloroplasts synthesis and regulation during photosynthetic carbon metabolism. *Plant physiology*, 76:465-471.

- TANG, Y.J., W. SHUI, S. MYERS, X. FENG, C. BERTOZZI, AND J.D. KEASLING. 2009. Central metabolism in *Mycobacterium smegmatis* during the transition from O2-rich to O2-poor conditions as studied by isotopomer-assisted metabolite analysis. *Biotechnol. Lett.* 31: 1233-1240.

- TEUTSCHBEIN, J., SCHUMANN, G., MOLLMANN, U., GRABLEY, S. & COLE, S.T. 2007. A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*. *Research in Microbiology,* 164(3)253-259.

- TOUATI, D. 2000. Iron and oxidative stress in bacteria. *Archives of Biochemistry and Biophysics.* 373(1):1-6.

- TRETTER, L. & ADAM-VIZI, V. 2000. Role of a-Ketoglutarate Dehydrogenase in Limiting NADH Production under Oxidative Stress. *The Journal of Neoroscience.* 20(24):8972-8979.

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

- VEYRIER, F.J., DUFORT, A. & BEHR, A.M. 2011. The rise and fall of the *Mycobacterium tuberculosis* genome. *Trends in Microbiology,* 781:1-6.

- WANG, Z. 2008. Iron dependency of beta-carotene 15, 15'-monooxygenase in Caco-2 TC7 cells. Graduate thesis and dissertation. Iowa State.

- WATANABE, T. & SNELL,E.E. 1972. Reversibility of the Tryptophanase Reaction: Synthesis of Tryptophan from Indole, Pyruvate, and Ammonia. *Proc. Nat. Acad. S&I USA,* 69(5):1086-1090.

- WECHWERTH, W.& MORGENTAL, K. 2005. Metabolomics: from pattern recognition to biological interpretation. *Drugs discovery today*, 10(22):1551-1558.

- WIKER, H.G. & HARBOE, M. 1992. The antigen 85 complex: a major secretion product of *Mycobacterium tuberculosis*. *Microbiolgy Reviews,* 56:645-661.

- WORLD HEALTH ORGANIZATION 2010. WHO Report 2010: Global tuberculosis Control. Switserland. WHO. 1 p.

- KALINOVA, B., JIRO, P., ZD'AREK, J., WEN, X. & HOSKOVEC, M. 2006. GCxGC/TOF MS technique- A new tool in identification of insect pheromones: Analysis of the permsimmon bark borer sex pheromone gland. *Talanta*, 69:542-547.

- KANA, B.D. & MIZRAHI, V. 2004. Molecular genetics of *Mycobacterum tuberculosis* in relation to the discovery of novel drugs and vaccines. *Tuberculosis*, 84:63-75.

- KANEHISA, M., GOTO, S., KAWASHIMA, S. & NAKAYA, A. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30(1), 42-46.

- KEHL-FIE, T.E. & SKAAR, E.P. 2010. Nutritional immunity beyond iron: a role for manganese and zinc. *Current opinion in Chemical Biology*, 14:218-224.

- KINDT, R., MOREEL, K., DEFORCE, D., BOERJAN, W. & VAN BOCXLAER, J. 2009. Joint GC-MS and LC-MS platforms for comprehansive plant metabolomics: Repeatability and sample pre-treatment. *Journal of Chromatography*, 877:3572-3580.

- LEWIS, K.N., LIAO, R., GUIN, M., HICKEY, M.J., SMITH, S., BEHR, M.A. & SHERMAN, D.R. 2003. Deletion of RD1 from *Mycobacterium tuberculosis* mimics Bacill Calmette-Guerin attenuation. *Journal of Infectious disease,* 187:117-123.

- XIA, J., PSYCHOGIOS, N., YOUNG, N. & WISHART, D.S. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic acid Research*, 37:652-660.

- YANKOVSKAYA, V., HORSEFIELD, R., TORNROTH, S., LUNA-CHAVEZ, C., MIYOSHI, H. & LEGER, C. 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science*, 299:700-704.

- YELLABOINA, S., RANJAN, S. & RANJAN, A. 2006. IdeR in mycobacteria: from target recognition to physiological function. *Critical review Microbiology*, 32(2):69-75.