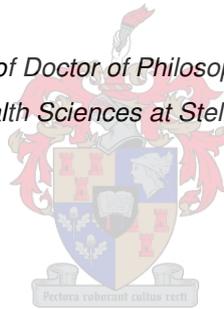


# Nucleoid Gene Regulation in Mycobacteria

by

Nastassja Lise Steyn

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



Supervisor: Prof Robin Mark Warren

Co-supervisor: Prof Nico Gey van Pittius

Co-supervisor: Prof Samantha Leigh Sampson

Co-supervisor: Dr Monique Joy Williams

December 2017

## Declaration

---

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

December 2017

## Abstract

---

*Mycobacterium tuberculosis*, the causative agent of tuberculosis disease, remains one of the leading causes of death worldwide. The ongoing tuberculosis epidemic and poor treatment outcomes can be attributed to the emergence of antibiotic resistant mycobacterial organisms as well as persistent *M. tuberculosis* infections. It is therefore imperative that we improve our understanding of *M. tuberculosis* pathogenesis to develop more effective anti-tuberculosis drugs and treatment strategies. *M. tuberculosis* is known to encounter adverse stress conditions such as nutrient starvation, hypoxia, nitrosative stress and low pH during infection and understanding how this pathogen adapts to its host environment is crucial in finding ways to circumvent tuberculosis disease. Conventional methods of investigating transcriptional responses, such as microarrays and ChIP-seq, are limited to the investigation of a single protein. Although valuable information has been obtained through this approach, adaptation to stress is likely mediated through several regulatory elements. Furthermore, new methodologies are required to identify these proteins. For example, nucleoid-associated proteins (NAPs), a group of global transcriptional regulators which have the unique ability to shape the bacterial chromosome, have been shown to be involved in the mycobacterial stress response. However, due to poor sequence homology with other bacterial organisms and conventional methodologies, only seven nucleoid associated proteins have been identified for mycobacterial organisms compared to 12 NAPs in *Escherichia coli*.

Nucleoprotein - Mass Spectrometry (NP-MS) makes use of affinity purification of formaldehyde cross-linked RNA polymerase transcriptional complexes to identify proteins involved in chromosomal structure, and transcriptional and translational processes using mass spectrometry. This approach successfully identified various DNA and RNA associated proteins as well as numerous proteins associated with energy, carbon, lipid and amino acid metabolism. These results suggested that NP-MS was effective in isolating not only proteins directly involved in transcription and translation but also proteins associated with the RNA polymerase transcriptional complex and DNA. MSMEG\_1060, MSMEG\_2695, MSMEG\_3754, MSMEG\_4306 and MSMEG\_5512 were identified as possible nucleic acid associated proteins through comparison of conserved proteins domains and were subjected to further investigation. Episomal expression of these proteins as FLAG-tagged fusion proteins in *M. smegmatis* revealed MSMEG\_1060, MSMEG\_2695, MSMEG\_4306 and MSMEG\_5512 to be putative DNA-associated proteins whilst no DNA association was found for MSMEG\_3754.

NP-MS was applied to investigate differences between DNA-associated proteins in exponential and stationary phase cultures. Investigation into the proteins which make up the RNA polymerase transcriptional complex and its associated proteins in stationary phase *M. smegmatis* cultures, revealed that NP-MS could effectively be used to identify proteins which are required for adaptation to stress in this organism. These

included the dormancy response regulator DevR, the ribosome hibernation promoting factor (hpf), heat shock protein HspX and the universal stress proteins MSMEG\_3811, MSMEG\_3945 and MSMEG\_3950.

These results demonstrated the ability of the developed NP-MS method to identify the proteins which mediate DNA structure, and transcriptional and translational changes in *M. smegmatis*. Furthermore, we propose that NP-MS can be used to investigate the proteins associated with the RNA polymerase complex in not only mycobacterial species but also other bacterial organisms.

## Opsomming

---

*Mycobacterium tuberculosis*, die organisme wat tuberkulose veroorsaak, bly steeds een van die hoofoorsake van dood in mense wêreldwyd. Die heersende tuberkulose epidemie en wanhopige behandelingsuitkomst van die siekte dra by tot die verskyning van antibiotika-weerstandige rasse asook nabywende infeksies. Dit is dus van belang om ons begrip van *M. tuberculosis* patogenese te verbeter om meer effektiewe anti-tuberkulose middels te ontwikkel. *M. tuberculosis* ervaar vele nadelige streskondisies tydens infeksie soos uithongering in nutriënte, suurstof, stikstof en lae pH. Deur te verstaan hoe die patogeen aanpas in die gasheeromgewing is van belang om maniere te vind om die siekte te omseil. Konvensionele maniere om transkripsie response te ondersoek met behulp van “microarrays” en “ChIP-seq” is beperk omdat net een proteïen op ‘n slag ondersoek kan word. Alhoewel waardevolle inligting verkry kan word met hierdie benaderings, is verskeie reguleerders betrokke tydens stressaanpassing. Nuwe metodologiese benaderings word dus benodig om hierdie proteïene te identifiseer.

As ‘n voorbeeld, nukleoïed-geassosieerde proteïene (NAPs) is ‘n groep transkripsie reguleerders wat die unieke vermoë het om die bakteriële chromosoom te vorm en is al gewys om ‘n rol te speel in die stresrespons van mikobakterieë. Vanweë swak DNS-volgorde homologie met ander bakterië en konvensionele metodieke is daar nog net sewe NAPS geïdentifiseer vir mikobakterieë in vergelyking met die twaalf NAPs in *Escherichia coli*.

Die tegniek Nukleoproteïen - Massa Spektrometrie (NP-MS) kan proteïene betrokke in die chromosoomstruktuur, transkripsie en translasiëprosesse identifiseer deur gebruik te maak van affiniteitsuiwering van formaldehyd-kruisbindende RNA polimerase transkripsiekomplekse met behulp van massa spektrometrie. Hierdie benadering kan DNS- en RNS-geassosieerde proteïene sowel as verskeie proteïene betrokke in energie, koolstof, lipied en aminosuurmetabolisme identifiseer. NP-MS is dus effektief in die isolasie van proteïene betrokke in transkripsie en translasië, maar ook proteïene direk geassosieer met die RNA polimerase kompleks en DNS. Die proteïene MSMEG\_1060, MSMEG\_2695, MSMEG\_3754, MSMEG\_4306 en MSMEG\_5512 is geïdentifiseer as nukleïensuur-geassosieerde proteïene deur vergelykings te maak met gekonserveerde proteïendomeine en is onderworpe aan verdere navorsing. Episomale uitdrukking van hierdie proteïene as FLAG-gehegte fusieproteïene in *Mycobacterium smegmatis* wys dat MSMEG\_1060, MSMEG\_2695, MSMEG\_4306 en MSMEG\_5512 as vermeende DNS-geassosieerde proteïene optree terwyl geen DNS assosiasie gevind kon word vir die proteïen MSMEG\_3754 nie.

NP-MS kan toegepas word om verskille te ondersoek tussen DNS-geassosieerde proteïene in eksponensieël en stasionêre fases. ‘n Ondersoek in *M. smegmatis* kulture in stasionêre fase se proteïene wat bydra tot die RNA polimerase kompleks asook die geassosieerde proteïene wys dat NP-MS gebruik

kan word met groot effek om proteïene te identifiseer wat benodig mag word vir die aanpassing tot stres in die organisme. Dit sluit in die dormansierespons reguleerder DevR, die ribosoomhibernasie aanwakkerfaktor hpf, die hittedkokproteïen HspX en die universele stres proteïene MSMEG\_3811, MSMEG\_3945 en MSMEG\_3950. Hierdie resultate toon die vermoë van 'n NP-MS metodiek om proteïene te identifiseer wat DNS struktuur, transkripsie en translasieranderinge in *M. smegmatis* kan fasiliteer. Verder stel ons voor dat NP-MS gebruik kan word om proteïene te identifiseer wat geassosieerd is met die RNA polimerase kompleks in nie net mikobakterieë nie, maar ook in ander bakterieë.

## Acknowledgements

---

I would like to acknowledge the following people and institutions for their contribution to this study:

My supervisors, Prof R.M. Warren, Prof N. C. Gey van Pittius, Prof S. L. Sampson and Dr M. J. Williams for their guidance, support, understanding and encouragement. Thank you for creating a stimulating environment which has allowed me valuable training opportunities and the ability to grow as a researcher.

The National Research Foundation (NRF), The Harry Crossley Foundation and Stellenbosch University for the financial aid which made this study possible.

To my colleagues and friends in the Division of Molecular Biology and Human Genetic for your friendship and support throughout this project. I would like to give special thanks to Dr T. Heunis for his valuable scientific input and support.

Dr A. Dippenaar, Dr. M. Newton-Foot, Dr. N. van Wyk, Dr. C. Beltran, J. Arries, and J. Gallant thank you for always listening, your constructive criticism, support and help.

My family, thank you for all the love and support.

My husband, Christopher, you were always there to listen to me, always on my side. Thank you for your love, understanding, support and encouragement.

## Table of Contents

---

<b>Declaration</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Opsomming</b>	<b>iv</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>Table of Contents</b>	<b>vii</b>
<b>List of Figures</b>	<b>x</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Abbreviations</b>	<b>xv</b>
<b>Chapter 1: Introduction</b>	<b>1</b>
Introduction	1
Problem Identification	2
Aim	3
Objectives	3
Structure of Thesis	3
<b>Chapter 2: Mycobacterial Nucleoid-associated Proteins</b>	<b>5</b>
Introduction	5
NapM	6
mIHF	8
HupB	9
MDP2	11
GroEL1	11
Lsr2	12
EspR	14
NAPs are regulators of the mycobacterial stress response	15
NAPs are attractive drug targets	17
	vii

Identification of NAPs in Mycobacteria	17
Knowledge Gaps and Future Work	19
Conclusion	21
<b>Chapter 3: Development of Nucleoprotein - Mass Spectrometry</b>	<b>22</b>
3.1. Introduction	22
3.2. Materials and Methods	24
3.2.1. Bacterial Strains	24
3.2.2. Media, Culture and Storage Conditions	24
3.2.3. Development of NP-MS	24
3.2.4. Nucleoprotein – Mass Spectrometry (NP-MS)	27
3.2.5. Validation of NP-MS	32
3.3. Results	43
3.3.1. Development of NP-MS	43
3.3.2. Identifying the Nucleic Acid Associated Proteins in <i>M. smegmatis</i>	49
3.3.2. Enrichment of Gene Ontologies Associated with Nucleic Acid Associated Proteins	55
3.3.3. Gene Regulatory Networks of <i>M. smegmatis</i>	60
3.3.4. DNA Association of Isolated Proteins	65
3.4. Discussion	73
3.4.1. NP-MS Identifies Proteins of the Gene Expression Proteome	73
3.4.2 DNA Association of NP-MS Identified Proteins	78
3.4.3 Limitations	79
3.5. Conclusion	81
<b>Chapter 4: Nucleoproteins associated with exponential and stationary growth phase in <i>Mycobacterium smegmatis</i></b>	<b>83</b>
4.1. Introduction	83
4.2. Materials and Methods	84
4.2.1. Bacterial strain and Culture Conditions	84
4.2.2. Chromatin Immunoprecipitation – Protein Mass Spectrometry	84
4.3. Results	88
4.3.1. Identifying Proteins Unique to the Exponential Phase and Stationary Phase	88
4.3.2. Gene Ontology Enrichment of Exponential and Stationary Phase Immunoprecipitations	94
	viii

4.3.3. Nucleoid-associated Proteins unique to each of the Growth Phases	101
4.4. Discussion	110
4.4.1. Uniquely identified proteins of the exponential phase	111
4.4.2. NP-MS Identifies Universal Stress Response Proteins in the Stationary Phase	112
4.4.3. Limitations	116
4.5. Conclusion	118
<b>Chapter 5: Summary</b>	<b>119</b>
5.1. Conclusion	119
5.2. Future directions	120
<b>References</b>	<b>122</b>
<b>Addendum A: Buffers</b>	<b>145</b>
<b>Addendum B: NP-MS results</b>	<b>149</b>
B.1. Formaldehyde induced modification frequency determination	149
B.2. Low confidence protein identification	150
B.3. High confidence protein identification	169
B.4. High Stringency vs. Lower Stringency Protein Identification	195
<b>Addendum C: Nucleoproteins of the exponential and stationary growth phase of <i>M. smegmatis</i></b>	<b>198</b>
C.1. Exponential Phase	198
C.2. Stationary Phase	209
C.3. Exponential and Stationary Phase Proteins	222

## List of Figures

---

### Chapter 1: Introduction

None

### Chapter 2: Literature Review

Figure 2. 1 Model of HupB and IdeR mediated regulation of siderophore production	10
Figure 2. 2 Lsr2 dimer oligomerization	13
Figure 2. 3 EspR dimers adjacently bound to DNA	14
Figure 2. 4 Identification of nucleoid-associated proteins	18
Figure 2. 5 ChIP-seq can be used to resolve the binding potential of NAPs	20

### Chapter 3: Development of Chromatin Immunoprecipitation – Protein Mass Spectrometry (NP-MS)

Figure 3. 1 Data processing workflow	31
Figure 3. 2 Construction of pNFLAG	36
Figure 3. 3 Vector maps	41
Figure 3. 4 DNA fragmentation using micrococcal nuclease	43
Figure 3. 5 RNA polymerase identification and immunoprecipitation of crosslinked proteins	44
Figure 3. 6 Immunoprecipitation time optimisation	45
Figure 3. 7 Antibody elution following on-bead tryptic digestion	46
Figure 3. 8 Formaldehyde crosslinking concentration	47
Figure 3. 9 Formaldehyde concentration of 1% is required for the detection of proteins associated with the RNA polymerase complex	48
Figure 3. 10 2D PCA of NP-MS LFQ Intensity data	50
Figure 3. 11 Heatmaps representing less stringent NP-MS data	51
Figure 3. 12 Heatmap representing imputed low confidence protein data	52
Figure 3. 13 Metabolic pathway mapping of NP-MS identified proteins	54
Figure 3. 14 Hierarchical clustering of enriched biological processes GO identities	57
Figure 3. 15 Hierarchical clustering of enriched molecular function GO identities	58

Figure 3. 16 Hierarchical clustering of enriched cellular component GO identities	59
Figure 3. 17 Network representation of biological processes GO identities	61
Figure 3. 18 Network representation of molecular function GO identities	62
Figure 3. 19 Network representation of cellular component GO identities	63
Figure 3. 20 Protein-protein interaction network of NP-MS identified proteins	64
Figure 3. 21 Generation of pNFLAG0615 plasmid	66
Figure 3. 22 Generation of pNFLAG plasmid constructs	67
Figure 3. 23 Detection of N-terminally FLAG-tagged protein expression in <i>M. smegmatis</i> using Western blotting	68
Figure 3. 24 FLAG-MSMEG_1060, FLAG-MSMEG_2695 and FLAG-MSMEG_4306 are DNA-associated proteins	70
Figure 3. 25 Anti-FLAG Immunoprecipitation DNA concentrations	71
Figure 3. 26 Protein and DNA formaldehyde cross-linking	73
Figure 3. 27 Isolation of proteins that make up the <i>M. smegmatis</i> regulome	74
Figure 3. 28 High stringency vs lower stringency protein identification	77
<b>Chapter 4: Nucleoproteins of the exponential and stationary growth phase of <i>M. smegmatis</i></b>	
Figure 4. 1 Flow diagram illustrating data analysis pipeline	87
Figure 4. 2 <i>M. smegmatis</i> growth curve	88
Figure 4. 3 2D PCA of exponential and stationary phase immunoprecipitation data	89
Figure 4. 4 Heatmaps illustrating reproducibility of biological replicates for exponential and stationary phase immunoprecipitations.	90
Figure 4. 5 Exponential phase protein metabolic pathway mapping	92
Figure 4. 6 Stationary phase protein metabolic pathway mapping	93
Figure 4. 7 Hierarchical clustering of enriched biological processes GO identities for exponential phase NP-MS identified proteins	95
Figure 4. 8 Hierarchical cluster of enriched molecular function enriched GO terms for exponential phase proteins	96

Figure 4. 9 Hierarchical clustering of enriched cellular component GO terms for the exponential growth phase	97
Figure 4. 10 Hierarchical clustering of biological processes GO terms of stationary phase proteins	98
Figure 4. 11 Hierarchical clustering of molecular function GO identities of stationary phase proteins	99
Figure 4. 12 Hierarchical clustering of cellular component GO terms of stationary phase proteins	100
Figure 4. 13 Venn diagram of NP-MS identified exponential and stationary growth phase proteins	101
Figure 4. 14 Enriched GO identity classification of uniquely identified proteins from the exponential and stationary growth phases	109

### **Chapter 5: Summary**

None

### **Addendum A: Buffers**

None

### **Addendum B: NP-MS results**

Figure B. 1 Heatmaps representing high stringency NP-MS data	169
Figure B. 2 Heatmap representing imputed low confidence protein data	170
Figure B. 3 Visualization of enriched biological function GO identities	188
Figure B. 4 Hierarchical clustering of enriched molecular function GO identities	189
Figure B. 5 Treemap representation of enriched cellular component GO identities	190
Figure B. 6 Network representation of biological processes GO identities	191
Figure B. 7 Network representation of Molecular function GO identities	192
Figure B. 8 Network representation of cellular component GO identities	193
Figure B. 9 Protein association network of proteins identified using NP-MS	194

### **Addendum C: Nucleoproteins of the exponential and stationary growth phase of *M. smegmatis***

None

## List of Tables

---

### Chapter 1: Introduction

None

### Chapter 2: Literature Review

Table 2. 1 Nucleoid-associated proteins of the genus *Mycobacterium* 7

### Chapter 3: Development of Chromatin Immunoprecipitation – Protein Mass Spectrometry (NP-MS)

Table 3. 1 Washing of Dynabeads™ 28

Table 3. 2 Primer sequences 34

Table 3. 3 Master mix for Phusion High-Fidelity DNA Polymerase reaction 37

Table 3. 4 PCR amplification conditions for Phusion High Fidelity DNA polymerase 37

Table 3. 5 Plasmids 38

Table 3. 6 In-Fusion® HD cloning reaction 40

Table 3. 7 Percentage nucleic acid-associated proteins identified with different formaldehyde concentrations 47

Table 3. 8 DNA association validation candidate proteins 65

Table 3. 9 Anti-FLAG immunoprecipitation DNA concentrations 72

### Chapter 4: Nucleoproteins of the exponential and stationary growth phase of *M. smegmatis*

Table 4. 1 Unique proteins of the exponential growth phase 102

Table 4. 2 Unique proteins of the stationary growth phase 104

Table 3. 3 Transcriptomic data of uniquely identified exponential phase *M. smegmatis* proteins and their *M. tuberculosis* homologues 111

Table 4. 4 NP-MS stress proteins identified from the stationary growth phase 113

### Chapter 5: Summary

None

### Addendum A: Buffers

Table A. 1 NP-MS Buffers 145

xiii

Table A. 2 Culture Media	146
Table A. 3 SDS-PAGE and Silver Staining Buffers	147
Table A. 4 Separating gel composition	148
Table A. 5 Stacking gel composition	148
Table A. 6 Western Blotting buffers	148
<b>Addendum B: NP-MS results</b>	
Table B. 1 Formaldehyde induced modification frequency determination	149
Table B. 2 Less stringent protein identification list in <i>M. smegmatis</i>	150
Table B. 3 Lower stringency protein identification gene ontology enrichment data	164
Table B. 4 High stringency identified proteins in <i>Mycobacterium smegmatis</i>	171
Table B. 5 High stringency protein identification gene ontology enrichment data	183
Table B. 6 Proteins unique to high stringency protein identification	195
Table B. 7 Proteins unique to lower stringency identification	195
<b>Addendum C: Nucleoproteins of the exponential and stationary growth phase of <i>M. smegmatis</i></b>	
Table C. 1 Exponential Phase Protein Identifications	198
Table C. 2 Exponential phase protein identification gene ontology enrichment data	206
Table C. 3 Stationary Phase Protein Identification	209
Table C. 4 Stationary phase protein identification gene ontology enrichment data	218
Table C. 5 Proteins identified in both the exponential and stationary growth phase using NP-MS	222
Table C. 6 Gene ontology enrichment of uniquely identified exponential growth phase proteins	228
Table C. 7 Gene ontology enrichment of uniquely identified stationary growth phase proteins	229

## List of Abbreviations

---

$\beta$	Beta
$\Omega$	Ohm
%	Percentage
®	Registered trademark
°C	degrees Celcius
$\mu$ F	microFarad
$\mu$ Fd	microFaraday
$\mu$ g	microgram
$\mu$ l	microliter
6xHis	6 x histidine tag
<i>7H11</i>	BBLTM Seven H11 Agar Base
7H9	Difco™ Middlebrook 7H9 Broth
8xGly	8 x Glycine
ABC transporter	ATP-binding cassette transporters
AD	Bovine albumin fraction V and Dextrose
AE-MS	Affinity enrichment mass spectrometry
AFM	Atomic Force Microscopy
AmpR	Ampicillin resistance
ANOVA	Analysis of variance
AP-MS	Affinity purification - Mass spectrometry
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
AT-rich	Adenosine and Threonine rich
BD	Becton Dickson
bp	Base pair
BSA	Bovine Serum Albumin
CAF	Central Analytical Facilities
CateGORizer	GO Terms Classification Counter
CbpA	Curved DNA-binding protein A
ChIP	Chromatin Immunoprecipitation

ChIP-chip	Chromatin Immunoprecipitation on chip
ChIP-seq	Chromatin Immunoprecipitation sequencing
COG	Cluster of Orthologous Groups
CRAPome	Contaminant Repository for Affinity Purification
DNA	Deoxyribonucleic acid
DNAse I	Deoxyribonuclease I
<i>DPB</i>	<i>DNA binding</i> protein
<i>Dps</i>	DNA binding protein from starved cells
<i>E. coli</i>	<i>Escherichia coli</i>
ecc	esx conserved components
EDTA	Ethylenediamine tetra acetic acid
EMSA	Electro mobility shift assay
ESX	ESAT-6 secretion system
<i>et al.</i>	and others
F primer	Forward primer
FDR	False discovery rate
Fis	Factor for inversion stimulation
FLAG	FLAG-tag (GATTATAAAGATGACGACGATAAA)
g	gram
GC	Guanine and Cytosine
GO	Gene Ontology
GOEAST	Gene Ontology Enrichment Analysis Software Toolkit
GTP	Guanine triphosphate
HCHO	Formaldehyde
Hfq	Host factor for phage Q
HIV	Human Immunodeficiency Virus
H-NS	Histone-like nucleoid structuring protein
HPLC	High pressure liquid chromatography
HRP	Horseradish peroxidase
HU	Heat unstable protein
HygR	Hygromycin resistance
IdeR	Iron dependent repressor
IDT	Integrated DNA Technologies

IHF	Integration Host Factor
IP	Immunoprecipitation
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranoside
kan	Kanamycin
KanR	Kanamycin resistance
KanR	Kanamycin resistance
kb	kilobases
KCl	Potassium chloride
kDa	kiloDalton
kV	kiloVolt
L	Liter
LB	Luria-Bertani broth
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LFQ	Label Free Quantification
M	Molar
M.	<i>Mycobacterium</i>
m/z	Molecular mass to charge ratio
MDR-TB	Multi Drug Resistant Tuberculosis
MgCl <sub>2</sub>	Magnesium Chloride
mIHF	Mycobacterial Integration Host Factor
ml	Milliliter
mM	Millimolar
MST	Microscale thermophoresis
MTBC	<i>Mycobacterium tuberculosis</i> complex
MYH	Muscle myosin human
NAPs	<i>Nucleoid</i> associated proteins
NEB	New England Biolabs
ng	Nanogram
nm	Nanometer
NP-MS	Nucleoprotein - Mass Spectrometry
OD	Optical density
ORFs	Open Reading Frames
oriE	Origin of replication for <i>E.coli</i>

oriM	Mycobacterial origin of replication
PBS	Phosphate buffered saline
<i>PBS-T</i>	Phosphate buffered saline - tween
PCA	Principle Component Analysis
PCR	Polymerase chain reaction
PDB	Protein Databank
PDIM	Phthiocerol dimycocerosate
PE	Proline glutamine protein
PGL	Phenolic glycolipid
PGRS	Polymorphic GC-rich sequences
pH	Potential of hydrogen
<i>pmyc1tetO</i>	Conditional tetracycline expression promoter
PPE	Proline proline glutamine protein
<i>Psmyc</i>	Promoter for mycobacteria
R primer	Reverse primer
RBPs	RNA binding proteins
RBS	Ribosome binding site
REVIGO	Reduce and Visualize Gene Ontology
<i>revTetR</i>	Reverse TetR
RNA	Ribonucleic acid
RNPs	Ribonucleoproteins
rpm	Revolutions per minute
<i>rrnB</i>	Transcription terminator
RSLC	Rapid Separation LC
RT-PCR	Reverse Transcription - Polymerase Chain Reaction
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
STPK	Serine/Threonine Protein Kinases
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TAE	Tris-acetic acid-EDTA buffer
TBS	Tris-buffered saline
TEMED	N,N,N',N'-tetramethylethylenediamine
tet	Tetracycline

TetR	Tetracycline repressor protein
Thr	Threonine
Tm	Annealing temperature
TM	Trademark
topA	DNA topoisomerase I
Tris	Tris(hydroxymethyl)aminomethane
Tween-80	Polyoxyethylene sorbian monooleate
U	Units
UV	Ultraviolet
V	Volt
WCL	Whole cell lysate
WT	Wild-type
XDR-TB	Extensively Drug Resistant Tuberculosis
xg	G-force/relative centrifugal force
ZN	Zhief-Neelsen

---

## Chapter 1: Introduction

---

### Introduction

---

*Mycobacterium tuberculosis* is a serious global threat to human health with tuberculosis disease reportedly resulting in approximately 1.4 million deaths in 2015 (1). Even though the tuberculosis mortality rate has declined by 47% since 1990, the emergence of drug resistant forms of this pathogen, as well as the rise of tuberculosis and human immunodeficiency virus (HIV) co-infections, have hindered the eradication of this disease (2). *M. tuberculosis* has acquired resistance to many of the first-line (e.g. rifampicin) and second-line (e.g. fluoroquinolones) anti-tuberculosis drugs, giving rise to the emergence of multi- (MDR) and extensively- drug (XDR) resistant forms of *M. tuberculosis* (3). In 2015, there were an estimated 480 000 new MDR tuberculosis cases and a 100 000 individuals with rifampicin-resistant tuberculosis (1). Furthermore, tuberculosis and HIV co-infections contributed to an additional 0.4 million deaths in 2015 (1). In spite of all the advances made in combatting tuberculosis disease, it remained one of the top 10 causes of death worldwide in 2015 (1).

During the course of tuberculosis infection, *M. tuberculosis* adapts to various stresses such as oxygen deprivation, nutrient limitation, nitrosative stress and low pH within the host environment (4–7). Understanding how this pathogenic organism responds to the stresses it encounters within the host is crucial in developing new therapies for the treatment of tuberculosis disease. Investigations into the adaptation of the pathogen to the hostile host environment often targets only one of the components which are known to make up the gene regulome, which can be defined as DNA, RNA, proteins, and the associated metabolites. The knowledge gained from these studies have helped us to better understand this pathogenic organism, however, the infection of the host, as well as the adaptation to the host environment is a complex response which involves all these regulatory components.

Whole genome sequencing and *in silico* analysis of *M. tuberculosis* have been used to identify 11 two-component systems, approximately 214 genes encoding DNA binding proteins and an estimated 50 RNA associated proteins (8–10). The large number of ORFs for which no function is assigned or predicted allows us to speculate that a number of these genes might encode proteins that are DNA or RNA associated. Microarray studies have successfully been used in the past to describe the transcriptomic responses of *M. tuberculosis* to external stimuli such as carbon starvation, dormancy, drug metabolism and anoxia (11, 12). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) has also recently been used to map the binding sites of known and predicted DNA-binding proteins in *M. tuberculosis*, with the aim of elucidating the DNA binding network (9). These molecular techniques are however limited to studying one transcriptional regulator at a time and this focused approach is problematic when correlating whole transcriptome data with these studies as a transcriptional response is likely to be caused by several

regulators or regulatory mechanisms. RNA binding proteins (RBPs) or ribonucleoproteins (RNPs) are known to influence the transport, localisation, translation, and stability of RNA molecules in various organisms (13). The diverse functions of these proteins are accommodated by multiple copies of RNA binding domains presented in various structural arrangements to facilitate the variety functions exhibited by RBPs (14, 15). RBPs, like small RNAs, have been shown to control gene expression in bacterial organisms through post-transcriptional modification of RNA (16). Unfortunately, very little information about mycobacterial RBPs are available and the approximate number of RBPs for mycobacterial species still needs to be quantified.

### **Problem Identification**

---

To understand the transcriptional changes that are required for the adaption of *M. tuberculosis* to its host environment, it is crucial to first identify the proteins which are required by the organism regulate the desired transcriptional response. These effector proteins are DNA and RNA associated proteins, of which some are notoriously difficult to identify. For example, transcriptional regulators are often short-lived and produced in low quantities. Methodologies frequently utilized by researchers for the identification of these proteins include sucrose density centrifugation of nucleoids followed by mass spectrometry. Albeit effective in the identification of nucleic acid associated proteins, this method is prone to the identification of protein contaminants not associated the region in the cell which contains all of the genetic material, also known as the nucleoid (17). Novel methodologies are therefore required for the identification of nucleic acid associated proteins but more specifically, new methodologies are required for the recognition of proteins associated with nucleic acids under environmental stress conditions. The development of such a method will possibly enable the identification of novel DNA or RNA associated proteins. Additionally, the development of a global high-throughput technique which will allow for the investigation of the protein occupancy changes associated with nucleic acids under environmental stress conditions will allow us to better understand how pathogenic organisms like *M. tuberculosis* adapts to the challenges presented by the host environment. Identifying proteins that are associated with the transcription machinery and are required by *M. tuberculosis* to adjust to the stresses presented by the host, could potentially identify novel drug therapy targets for the treatment and prevention of the tuberculosis disease.

## Aim

---

This project aims to investigate nucleic acid associated proteins of *Mycobacterium smegmatis* through affinity purification of nucleoprotein complexes followed by mass spectrometry.

## Objectives

---

1. To develop a method which combines affinity purification and protein mass spectrometry to identify proteins associated with the RNA polymerase transcription complex in *M. smegmatis*.
2. To validate the efficiency of identifying DNA-associated proteins by demonstrating DNA association of selected proteins identified through mass spectrometry.
3. To describe DNA- and RNA-binding proteins found in *M. smegmatis* using bioinformatic analyses and to demonstrate the ability of the developed methodology to isolate and identify members of a protein defined networks.
4. To use the developed method to investigate proteins associated with the RNA polymerase transcription complex in the exponential and stationary growth phases of *M. smegmatis*.
5. To identify proteins uniquely found in the stationary growth phase of *M. smegmatis* and to describe their possible roles in the adaptation of the organism to a dormant growth state.

## Structure of Thesis

---

Each chapter presented in this thesis is structured for potential publication. All chapters remain unpublished to date, however, the references were formatted according to instructions of the Journal of Bacteriology.

### Chapter 1: Introduction

The general introduction chapter highlights the knowledge gaps of understanding how complex transcriptional changes are induced by mycobacterial organisms following exposure to environmental stress. This chapter draws attention to the requirement of new methodologies to investigate the adaptation of mycobacterial organisms to their stressful environments.

### Chapter 2: Mycobacterial Nucleoid-associated Proteins

This chapter focusses on the proposed mycobacterial nucleoid-associated proteins, also known as histone-like proteins, which possess the ability to physically shape the mycobacterial chromosome and act as global transcriptional regulators. This chapter draws attention to the role of these nucleoid-associated proteins as

possible regulators of the mycobacterial stress response and emphasizes the attractiveness of these proteins as possible drug development targets.

### **Chapter 3: Development of Nucleoprotein - Mass Spectrometry**

This chapter discusses the development of Nucleoprotein – Mass Spectrometry (NP-MS), an affinity purification technique coupled with mass spectrometry, aimed at isolating proteins associated with the RNA polymerase transcriptional complex. This chapter demonstrates the effective use of AP-MS to enrich for proteins associated with the transcriptional complex.

### **Chapter 4: Nucleoproteins of the exponential and stationary growth phase of *M. smegmatis***

In this application chapter, NP-MS is used to investigate the proteins associated with the RNA polymerase transcriptional complex from exponential and stationary growth phase cultures. The data in this chapter demonstrates the effectiveness of NP-MS to describe known stress proteins expressed by mycobacterial organisms which are required for adaptation to stress.

### **Chapter 5: Summary**

This chapter serves as a general conclusion chapter and highlights the development of NP-MS and how this technique can be used to investigate proteins associated with the RNA polymerase transcriptional complex and its associated proteins. This chapter also discusses the future development and application of NP-MS, whilst also discussing the functional investigations to be undertaken for proteins identified using NP-MS. This thesis has added significant value to understanding the workings of the transcriptional complex and its associated proteins in *M. smegmatis*.

---

## Chapter 2: Mycobacterial Nucleoid-associated Proteins

---

### Introduction

---

The pathogen *Mycobacterium tuberculosis* encodes approximately 214 DNA binding proteins which allow for adaptation to many adverse environmental conditions such as nutrient starvation, antibiotic exposure, and hypoxia (4, 9, 12, 18). DNA binding proteins include polymerases, gyrases, topoisomerases, nucleases and various transcription factors. Distinct from these are the nucleoid-associated proteins (NAPs), which have the unique ability to shape the bacterial nucleoid and influence gene expression (19, 20). NAPs are a set of highly abundant proteins that are positively charged, low molecular weight, dimeric proteins (21). Structurally NAPs are made up of an  $\alpha$ -helical body with two  $\beta$ -arms that assist in DNA binding (22–25). NAPs are known to modulate DNA structure by means of bending, wrapping and bridging of DNA which results in the looped formation of the nucleoid (19, 26). The formation of these looped domains have been linked to an increase in transcriptional activity and fast replicating bacteria often have more looped domains (19). NAPs are unique in that they not only organize and compact bacterial chromatin, but also play a role in a variety of DNA related processes such as DNA recombination, repair, replication and transcription (19, 27). Apart from structurally altering the shape of the nucleoid, NAPs also serve as a physical barrier which is able to protect DNA from reactive oxygen intermediates and nucleases such as DNase I (28–31).

The majority of NAPs are believed to be promiscuous with respect to DNA binding however, most NAPs favour binding to AT-rich regions in DNA sequences (26, 32). The preference for AT-rich DNA implies that NAPs have a natural tendency to bind within promoter regions, suggesting that these proteins are global regulators of gene expression (19). Chromatin Immunoprecipitation (ChIP) followed by sequencing (ChIP-seq) studies and DNase I protection assays have recently been used to identify recognition sequences for some NAPs, suggesting that the DNA binding abilities of NAPs are less promiscuous than previously thought (33, 34). ChIP-chip studies demonstrated that Fis, H-NS and IHF, known NAPs in *Escherichia coli*, are located in intergenic regions of the genome. Interestingly, many of their binding regions overlapped with those of RNA polymerase as well as with each other (32). The shared DNA binding sites of these NAPs with RNA polymerase suggests that NAPs influence transcriptional regulation by competing with RNA polymerase for the same DNA positions. A number of NAPs have also been shown to favour binding to DNA with nicks, gaps, single-stranded breaks as well as four-way junctions (35–38).

Quantitative western blot analysis demonstrated that the expression of NAPs from *E. coli* was dependent on growth phase (39). This study showed that some NAPs were highly expressed during exponential growth (Fis, Hfq, HU), whereas others were predominantly expressed during the early stationary phase (Dps) and late stationary phase (CbpA) (39). The differential abundances of individual NAPs during different growth conditions have been hypothesized to control the structure of the bacterial nucleoid during growth, as can

be seen with the compaction of the bacterial nucleoid and silencing of genome function in stationary phase (39, 40). Although NAPs have differential expression depending on growth phase, they do have overlapping DNA binding regions, suggesting some level of redundancy. NAPs have also been shown to autoregulate their own expression as well as cross-regulate each other's expression (33, 41, 42).

In *E. coli* twelve NAPs have been characterized, however, due to poor sequence identity similarities very few homologues of these have been identified in the genus *Mycobacterium*. Currently, *M. tuberculosis* is believed to encode seven NAPs, Lsr2, MDP2, HupB, EspR, mIHF, NapM and GroEL1 (Table 2. 1) (31, 33, 43–47). This review will discuss the various NAPs identified in mycobacterial species with emphasis on their roles in transcriptional regulation and nucleoid structure.

## **NapM**

---

NapM, encoded by Rv0047c, is a small alkaline protein which is not required for optimal growth of *M. tuberculosis* (47–50). This predicted NAP is conserved in all sequenced mycobacterial species and *M. smegmatis* and *M. tuberculosis* homologues share 91.9% sequence identity of this predicted PadR family transcriptional regulator (47).

### *NapM is a DNA binding protein*

Co-localisation studies demonstrated that ectopically expressed NapM in *E. coli* localised with the bacterial chromosomal region thereby implying that NapM associates with bacterial DNA (47). Electrophoretic mobility shift assays (EMSAs) were used to show that NapM binds DNA in a length-dependent but sequence independent manner with a preference for AT-rich DNA. The binding of NapM was also shown to protect DNA from DNase I digestion (47). Furthermore, NapM was demonstrated to interact with the major groove of the DNA molecule following DNA binding competition assays with methyl green and actinomycin D, which are known to interact with the DNA major and minor grooves, respectively (31, 47).

The ability of NapM to structurally alter the shape of DNA was revealed when increasing amounts of NapM was shown to supercoil plasmid DNA, even in the presence of DNA topoisomerase I (TopA, MSMEG\_6157). This observation led to the hypothesis that NapM could possess an inhibitory function against the DNA relaxation activity of TopA. Furthermore, atomic force microscopy (AFM) was used to demonstrate that NapM was able to condense large DNA fragments into compact DNA structures through DNA bridging (47).

**Table 2. 1 Nucleoid-associated proteins of the genus *Mycobacterium***

NAP	<i>M. tb</i> Rv number	<i>M. smegmatis</i> MSMEG number	<i>M. bovis</i> Mb number	<i>M. marinum</i> MMAR number	<i>M. leprae</i> ML number	% Identity of <i>E. coli</i> homologue <sup>A</sup>	Essential in <i>M. tb</i> <sup>B</sup>	Molecular Mass (kDa)	Isoelectric point	Native Protomer	Binding Groove	DNA binding properties
Lsr2	Rv3597c	MSMEG_6092	Mb3628c	MMAR_5101	ML0234	-	Yes	12.09	10.69	dimer	Minor Groove	Bridging (43, 51, 52)
EspR	Rv3849	MSMEG_6431	Mb3879	MMAR_5399	ML0069	30%	No <sup>C</sup>	14.7	9.04	dimer	Major Groove	Bending, Bridging (33, 53, 54)
mIHF	Rv1388	MSMEG_3050	Mb1423	MMAR_2201	ML0540	-	Yes	20.81 (12.10) <sup>D</sup>	11.02 (10.06)	dimer	Unknown	Bending, Bridging, Wrapping (55, 56)
HupB (Hlp)	Rv2986c	MSMEG_2389	Mb3010c	MMAR_1728	ML1683	43%	Yes	22.18	12.48	dimer	Minor Groove	Bending (57)
MDP2 (H-NS)	Rv3852	No homologue	Mb3882	MMAR_5402	ML0067	-	No	13.82	11.34	dimer	Unknown	Unknown
GroEL 1	Rv3417c	MSMEG_0880	Mb3451c	MMAR_1126	ML0381	52%	Yes	55.87	4.74	dimer	Minor Groove	Unknown
NapM	Rv0047c	MSMEG_6903	Mb0048c	MMAR_0066	ML2691	-	No	20.41	10.39	-	Major Groove	Bridging (47)

<sup>A</sup> BLASTP was used to search *M. tuberculosis* H37Rv protein sequence against *E. coli* homologue sequences

<sup>B</sup> Based on the findings of (48–50)

<sup>C</sup> Slow growth mutant (48)

<sup>D</sup> Alternative start codon (55)

*NapM is a regulator of gene expression*

Microarray studies comparing wild-type *M. smegmatis* and an *M. smegmatis*  $\Delta napM$  mutant revealed that 156 genes were differentially expressed. One hundred and twenty-two genes were found to be upregulated and the remaining 35 genes were downregulated in the absence of NapM. These results imply that NapM, like Lsr2, has a negative effect on gene expression. A Cluster of Orthologous Groups of proteins (COG) categories found to be represented within the group of differentially expressed genes included secondary metabolite biosynthesis, transport and catabolism and lipid, carbohydrate and amino acid transport and metabolism, implicating NapM in the regulation of genes involved in various cellular processes within *M. smegmatis* (47).

In *M. tuberculosis*, ChIP-seq was used to show that NapM bound to 348 loci throughout the genome and microarray data revealed that ectopic expression of NapM in *M. tuberculosis* resulted in the differential expression of 62 genes. NapM was found to bind directly within its own coding region as well as within genes encoding transcriptional regulators such as WhiB3, SigD, MoaR1, and Mce3R. Analysis of the functional categories of NapM bound genes in *M. tuberculosis* revealed that they are involved in cell wall and cell processes (*eccB3*, *eccC3*, *eccE5*, *eccA5*, *espL* and *mmpL2*), intermediary metabolism and respiration (*ccsA*, *purD*, *mycP5*, *moaD1*, and *moaC1*) and PE/PPE protein families (*pe\_pgrs54*, *ppe68*, *ppe57*, *ppe58* and *ppe59*). Other functional categories represented included virulence, detoxification, and adaption as well as lipid metabolism (9).

**mIHF**


---

Mycobacterial integration host factor (mIHF), encoded by *Rv1388* in *M. tuberculosis*, is the third most abundant protein in the *M. tuberculosis* proteome (58). The expression of mIHF is growth phase dependent with maximal expression at stationary phase (56). This essential gene is required for *in vitro* growth in both *M. tuberculosis* and *M. smegmatis* (48, 50, 59).

The size of the predicted open reading frame for mIHF is known to vary between mycobacterial species. Previous studies suggested that mIHF encoded by *M. tuberculosis* is between 105-111 amino acids in length, however, in a recent study *M. tuberculosis* mIHF was shown to be 190 amino acids (55, 58, 60–62). The 79 amino acid N-terminally truncated mIHF (mIHF-80) was previously shown to bind to DNA at multiple positions in a sequence independent manner, however, full length mIHF displayed an increased binding affinity for curved DNA, *attB* and *attP* sequences (55, 56). The high mIHF binding affinity for *attB* and *attP* sites was suggested to aid in the circular shaping of the DNA molecule through the generation of a salt-stable protein-protein complex (56).

*mIHF is a DNA bending, wrapping and bridging protein*

AFM determined that mIHF-80 compacts DNA by a DNA bending mechanism. Interestingly, at low concentrations of mIHF-80 an opening of negatively supercoiled DNA was seen, however, at higher concentrations increased compaction was observed (55). Full length mIHF was determined by AFM to compact linear DNA by wrapping DNA into nucleoid-like structures. Bridging of linear DNA into linear multimers were also observed with full length mIHF, although to a lower extent (56).

## HupB

---

HupB (also known as MDP1, Hlp, MtbHU) is an essential gene for *in vitro* and *in vivo* growth of *M. tuberculosis* encoded by *Rv2986c* (34, 48, 50). HupB is a two domain protein from which the N-terminal regions exhibit sequence homology with *E. coli* HU. The C-terminal domain (although unique between members of the *Mycobacterium tuberculosis* complex) has been shown to contain several PAKK/KAAG/AKKA tetrapeptide repeats commonly seen in eukaryotic histone H1/H5 family proteins (63–66).

*HupB is a DNA binding protein*

DNA binding assays using truncated forms of *Mycobacterium smegmatis* Hup showed that both the N- and C-terminal regions contributed to high affinity DNA binding and Hup lacking the entire C-terminal region displayed diminished DNA binding (38). In contrast, the C-terminal region was shown to be completely devoid of DNA binding activity, while the N-terminal region was shown to have reduced DNA binding when compared to full length *M. tuberculosis* HupB (67). The C-terminal region has been proposed to impart greater selectivity in DNA binding and it was recently shown that the DNA binding property of HupB was regulated by posttranslational acetylation of lysine residues in the C-terminal region (65, 68). Additionally, the N-terminal region of HupB has been shown to be a substrate of Serine/Threonine Protein Kinases (STPKs), with phosphorylation of N-terminal residues resulting in diminished DNA binding.

HupB was also shown to bind to a variety of DNA repair and replication intermediates and was revealed to inhibit RecA mediated strand exchange (63). The tetrapeptide repeats in the C-terminal region of *M. smegmatis* Hup were shown to mediate DNA end-joining in the presence of T4 DNA ligase (38).

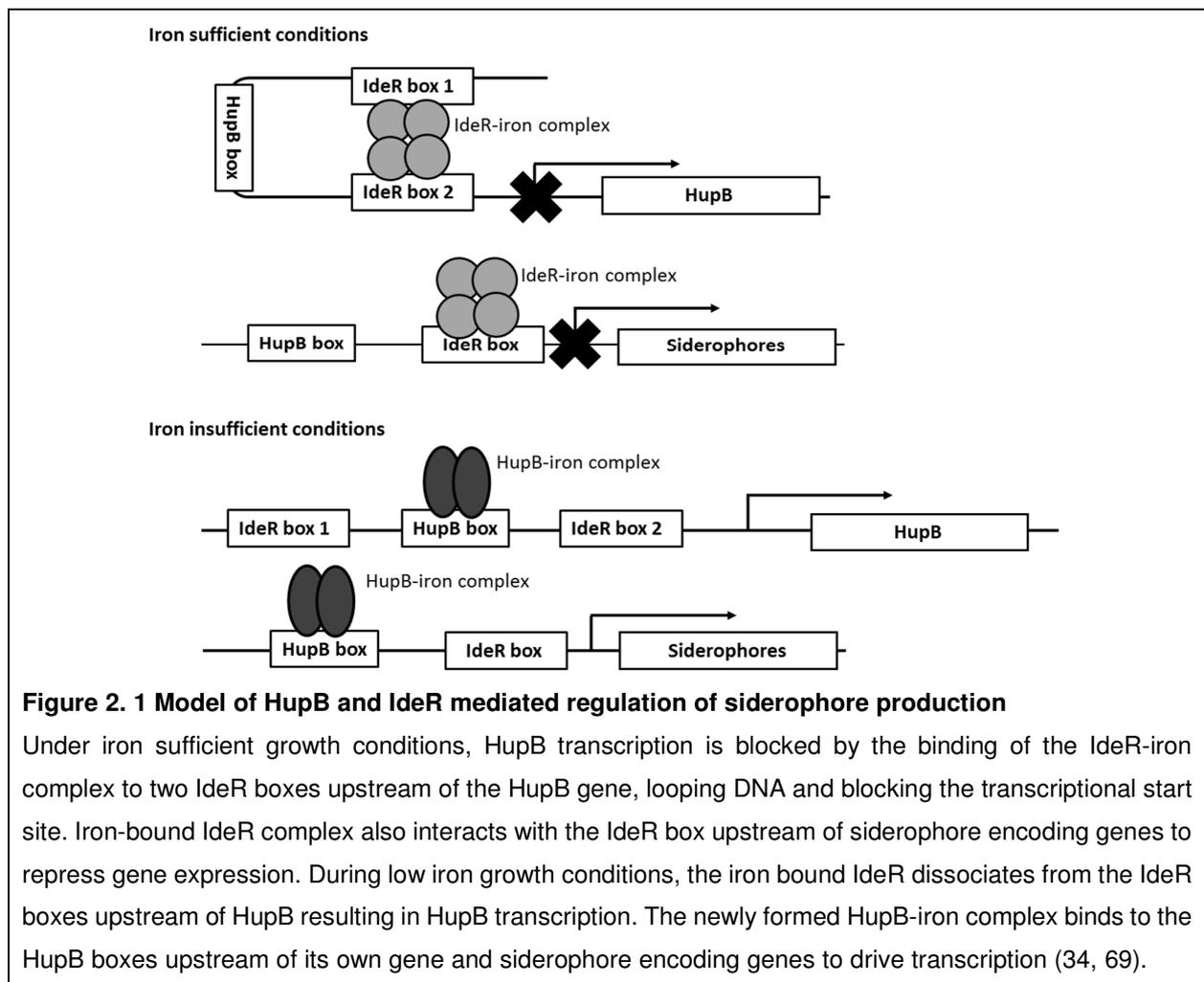
*HupB is a DNA bending protein*

Structurally HupB was also shown to constrain DNA in negative supercoils and to introduce negative super helical turns into relaxed DNA (63). Crystal structures revealed that the DNA binding  $\beta$ -saddle of HupB dimers is made up of two regions, the core region which is responsible for the docking with DNA or another HupB dimer and a region which forms mobile DNA embracing  $\beta$ -arms. Upon binding to the minor groove of

the DNA molecule, the  $\beta$ -saddle induces a change in the tips of the  $\beta$ -arms, resulting in the bending of DNA (57).

#### *HupB is required for iron metabolism*

DNA footprinting analysis recently revealed a 10 bp high affinity HupB binding domain (5'-CACTAAAATT-3') upstream of *IdeR*, a repressor of siderophore transcription (34). It has been hypothesized that siderophore transcription is repressed by the *IdeR*-Fe complex under iron sufficient conditions and upregulated by HupB under low iron conditions (Figure 2. 1) (34, 69). This hypothesis is supported by the observation that the *M. tuberculosis*  $\Delta$ *hupB* mutant expresses low levels of siderophores, mycobactin and carboxymycobactin under iron limiting conditions (34). It has also been proposed that the inability of *M. tuberculosis*  $\Delta$ *hupB* to survive within macrophages may be attributed to the insufficient expression of mycobactin and carboxymycobactin (69).



## MDP2

---

Mycobacterial DNA binding protein 2 (MDP2,H-NS) is a small protein encoded by *Rv3852* that contains several tetrapeptide repeats (PAKK/ KKAP/KAAK) often found in eukaryotic histone H1/H5 family proteins (67, 70). MDP2 is believed to be a novel NAP of unknown function in *M. tuberculosis* due to its inability to complement the function of *E. coli* H-NS (44, 71). In *M. tuberculosis* MDP2 was found to be most abundant in the exponential growth phase with abundance levels declining to 30% of the initial level in the stationary phase (72).

### *MDP2 is a DNA binding protein*

MDP2 binds poorly to DNA fragments rich in GC content and exhibits high affinity for curved DNA and holliday junctions (44, 67). *In vitro*, MDP2 was shown to constrain DNA into negative supercoils and like *E. coli* H-NS, was revealed to reduce RecA-mediated strand exchange in a concentration dependent manner (67). This NAP also was determined to negatively regulate the expression of lipid biosynthetic genes, *groEL1*, *kasA* and *kasB* (71). In contrast to what has previously been reported, no deregulation of *kasA* and *groEL1* was observed in a *rv3852* deletion mutant, however, deletion of MDP2 did result in the differential regulation of 22 genes. Interestingly, deletion of MDP2 resulted in the downregulation of the intergenic region upstream of *rv3852*, suggesting autoregulation. Furthermore, deletion of MDP2 resulted in the upregulation of the transcriptional regulator *mce1R* and *rv0164* (TB18.5) (72).

### *MDP2 expression influences cell membrane stability*

The C-terminal domain of MDP2 contains a transmembrane helix which anchors this protein to the mycobacterial membrane. Overexpression of *Rv3852* within *M. smegmatis* resulted in a decompacted nucleoid and a 3-fold increase of DNA within the membrane fraction. Loss of the C-terminal region abrogated decompaction of the nucleoid suggesting that the decompaction of the nucleoid is a result of MDP2 tethering DNA to the membrane. The ectopic expression of *Rv3852* in *M. smegmatis* also displayed defects in biofilm formation, sliding motility and a change in colony morphology (71). Similarly, in *M. tuberculosis* *Rv3852* was shown to be associated with the cell membrane fraction, however, deletion of *rv3852* did not result in any structural changes to the mycobacterial nucleoid (72).

## GroEL1

---

Group I Chaperonins facilitate ATP-dependent protein folding and are encoded from two genes, *groEL* and *groES* (73). In *E. coli* these two chaperonins form a tetradecameric structure comprising of rings of subunits stacked back to back (74). Some mycobacterial species are unique in that they encode two paralogues of GroEL. *M. tuberculosis* encodes for GroEL2 (*Rv0440*) and GroEL1 (*Rv3417c*) with the latter positioned in the *groESL* operon with *groES* (*Rv3418c*) (10, 75, 76). The GroEL1, GroEL2 and *E. coli* GroEL apical

domains have been shown to closely resemble one another, however, uncharacteristically to chaperonins GroEL1 has a histidine-rich C-terminal region (77, 78). In contrast, GroEL2 has a glycine-methionine rich C-terminal region characteristic of chaperonin proteins (78). Interestingly, in initial studies GroEL1 and GroEL2 were unable to form higher order oligomers and presented with the loss of ATPase activity and poor protein refolding abilities (79). A more recent study determined that the oligomerization of GroEL1 from its heptameric form to its tetradecameric form was dependent on phosphorylation. Oligomerization of GroEL1 into its tetradecameric form allowed this protein to act as a functional chaperone within *E. coli* (80).

GroEL2 is the most abundant protein in the *M. tuberculosis* proteome, and GroEL1 is the second most abundant. Interestingly, deletion of GroEL1 does not result in any growth defects within culture or murine derived macrophages, whereas GroEL2 appears to be essential (58, 81). The inability to generate a GroEL2 deletion mutant, together with the known differences in the C-terminal regions of these proteins, suggests that these two paralogues have different biological functions (81, 82).

#### *Is GroEL1 a NAP?*

EMSA and ChIP-seq have been used to show that GroEL1 is a DNA binding protein (9, 31). GroEL1 binds DNA with low specificity but high affinity, to protect DNA from DNase I and hydroxyl-radicals, as well as condense large DNA molecules into compact structures (31). GroEL1 like many other NAPs binds upstream of itself, resulting in the upregulation of transcription (9). GroEL1 is considerably larger than other NAPs however as mentioned above, possesses several characteristics that are associated with this protein class.

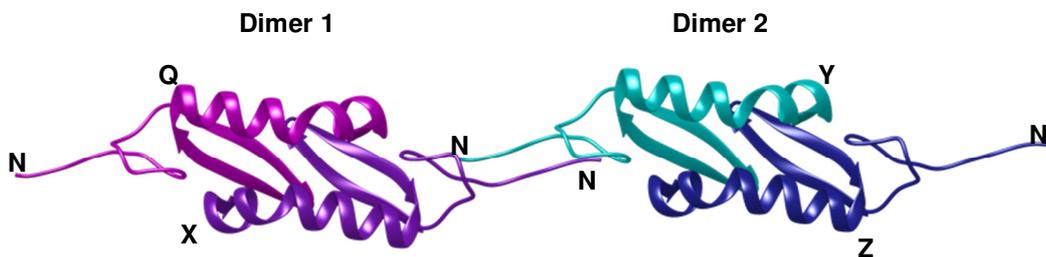
## **Lsr2**

---

In *M. tuberculosis*, Lsr2 is encoded by *Rv3597c* and is essential for *in vitro* growth (48, 50). Lsr2 was first identified as a T-cell antigen of *Mycobacterium leprae* and homologues of this gene can be found in related actinomycetes *Streptomyces*, *Nocardia*, and *Rhodococcus* (43, 51, 83). Following N-terminal proteolytic processing, the N-terminal region facilitates oligomerization of Lsr2 into a dimer (84). The C-terminal region of Lsr2 contains a DNA binding domain, "Q/RGR", which interacts with the minor DNA groove and facilitates the binding to AT-rich DNA (52, 84). Lsr2 has been proposed to be a functional analogue of *E. coli* H-NS, despite the low level of sequence homology between these two genes (43).

#### *Lsr2 is a DNA bending and bridging protein*

Lsr2 was found to bind to linear DNA, circular DNA and supercoiled DNA with equal affinity unlike *E. coli* H-NS (51). Crystallization of Lsr2 revealed tight association of the  $\beta$ -strands of neighbouring dimers to form linear chains of dimers (Figure 2. 2). These higher order assemblies facilitate the condensation of DNA into compact structures (84). AFM showed that Lsr2 compacts the nucleoid by bending and bridging DNA (51, 85).



**Figure 2. 2 Lsr2 dimer oligomerization**

The figure shows the N-terminal region of Lsr2 subunit X (dark purple) forms an anti-parallel  $\beta$ -sheet with the N-terminal region of Lsr2 subunit Y (light blue). The formation of these  $\beta$ -sheets between neighbouring dimers allows the formation of linear chains of dimers. PDB: 4E1R. (84)

#### *Lsr2 is a global regulator of gene expression*

Lsr2 binds to DNA in a relatively sequence independent manner but shows a preference for AT-rich DNA regions (51, 52, 86). DNA binding by Lsr2, like *E. coli* H-NS, was revealed to result in transcriptional repression of bound genes, however, in *M. smegmatis* upregulation of transcription has also been shown for Lsr2 (86, 87). Members of the *Mycobacterium tuberculosis complex* are believed to have acquired 76 genes through horizontal gene transfer (88). These horizontally acquired genes have a lower percentage of GC-content in relation to the rest of the *M. tuberculosis* genome and Lsr2 binding to 44 of these genes have been suggested to result in transcriptional repression (86, 88, 89). Homologues of *lsr2* have also been identified in many mycobacteriophages and it has been proposed that these phages use Lsr2 as a gene silencing weapon upon infection (84).

ChIP-chip studies in *M. tuberculosis* and *M. smegmatis* demonstrated that Lsr2 co-precipitates with 21% (840 of 4009 protein encoding ORFs) and 13% (904 of 6716 protein encoding ORFs) of these genomes, respectively (86). Genes bound by Lsr2 are involved in energy metabolism (ATP synthase *atpB*, *-E*, *-F* and *-H*), aerobic respiration (cytochrome C oxidase subunits *ctaC*, *-D* and *-E*), cell-wall peptidoglycan synthesis (*dacB2*, *murA*, *murI*), mycolic acid synthesis (*fabG1*, *fbpC*, *pcaA*), chromosomal DNA replication (*dnaA*, *recF*, *dnaE1*), transcription (*sigA*, *rho*) and protein synthesis (*rrf*, *rrs*, *rrl*) (86). Additionally, Lsr2 also binds genes involved in stress response (*ahpC*, *cspA*, and *dnaJ1*) as well as virulence related genes (multiple genes within the ESAT-6 gene cluster regions, *espA* and members of the *pe/ppe* gene family) (86). The regulation of diverse gene categories by Lsr2 shows that this NAP is indeed a global regulator.

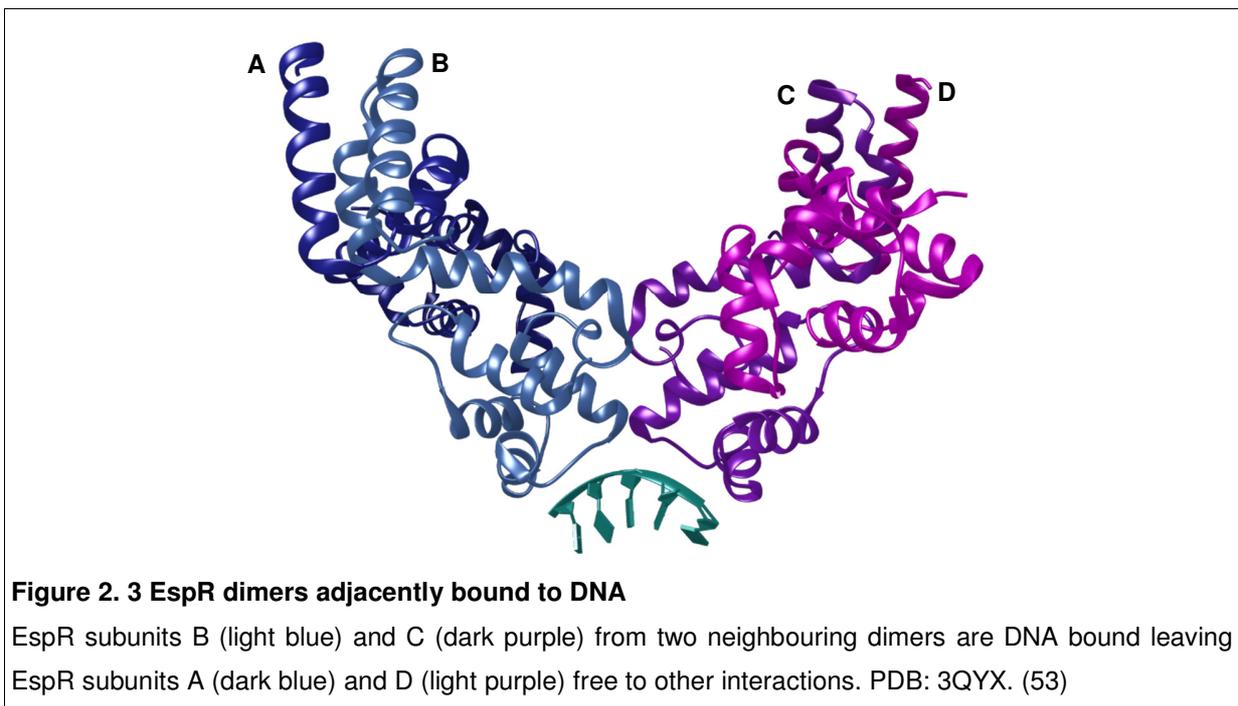
## EspR

---

EspR is a small, basic protein encoded by *Rv3849* and although this NAP is not required for *M. tuberculosis* survival, the transposon mutant did present with a slow growth phenotype (48, 50). Interestingly, the overexpression of EspR also presented with a decreased growth rate despite the intracellular concentration of EspR increasing through the bacterial growth cycle (33). The production of EspR appears to be autoregulated as EspR can bind to three DNA binding sites upstream of its own promoter (33). EspR has previously been implicated as a transcription factor necessary for the activation of ESX-1, a virulence associated secretion system in *M. tuberculosis* (90). Until recently, EspR was believed to be a product of ESX-1 secretion, however more recent studies have shown that this protein is not secreted (33, 90).

### *EspR is a DNA bending and bridging protein*

A dimer of dimers model has been proposed for EspR DNA binding. Monomers of two EspR dimers are thought to bind to two consecutive DNA major grooves leaving the second EspR subunit free to form higher order oligomers (53, 54) (Figure 2. 3). The formation of these higher order oligomers allows bending and bridging of DNA into looped domains by EspR (53).



### *EspR is a regulator of mycobacterial virulence*

ChIP-seq demonstrated that EspR binds to 165 loci within the *M. tuberculosis* genome (45% intergenic and 55% intragenic) and data from this study along with DNase I protection assays were used to generate a DNA binding consensus sequence for EspR, TTTGC[TC][GA] (33). This consensus sequence identified 199 possible EspR binding sites within putative intergenic regions and 827 within intragenic regions. Interestingly 163 of the 165 binding sites identified using ChIP-seq was also identified using the consensus sequence. Quantitative RT-PCR demonstrated that EspR acted as both a positive and negative regulator of gene transcription (33).

The majority of genes bound by EspR were determined to be involved in cell envelope functions and specifically, EspR was found bound within the region that encodes the enzyme system required for the synthesis of phthiocerol dimycocerosate (PDIM) and phenolic glycolipid (PGL) (33). PDIM and PGL have both been shown to contribute to virulence in *M. tuberculosis* as well as *M. marinum* (91–93). Other virulence related genes bound by EspR included members of the PE-PGRS protein family (*Rv1067c* and *Rv4068c*) and EspR was found to bind to multiple sites within ESAT-6 gene cluster regions (ESX-1 and ESX-5).

### **NAPs are regulators of the mycobacterial stress response**

---

A number of mycobacterial NAPs are upregulated under various stress responses. Lsr2, miHF, and MDP2 are upregulated under nutrient starvation and heat-shock conditions whereas HupB was only found to be induced under nutrient starvation conditions and GroEL1 under heat-shock conditions (4, 94). Lsr2 has also been shown to be upregulated under long term hypoxia, antibiotic exposure (Isoniazid) and high iron conditions (12, 95, 96). Additionally, Lsr2 in *M. tuberculosis* was recently shown to be required for survival in high oxygen environments and for the adaptation to anaerobic environments (97). In *M. smegmatis*, Lsr2 has been implicated in resistance to oxidative stress through interaction with MSMEG\_4334, a FAD-binding flavoprotein (98). HupB in *M. tuberculosis* was found to be upregulated under the stringent response as well as under iron limiting conditions and in *M. smegmatis* Hup was determined to be induced during dormancy and cold shock conditions (69, 99–102). Additionally, deletion of *hup* from *M. smegmatis* increases sensitivity to UV exposure, freezing and thawing as well as isoniazid (103). GroEL1 has been shown to have increased levels of expression during macrophage infection, oxidative stress, osmotic stress and dehydration (76, 104, 105). Furthermore, mice infection studies showed that a *M. tuberculosis* GroEL1 deletion mutant was defective in inducing a cytokine response resulting in the failure to produce granulomas (81).

The upregulation of these global regulators under various stress conditions has the potential to set in motion a broad transcriptional response. For example, the upregulation of Lsr2 under nutrient limiting conditions

and hypoxia could result in the down regulation of genes involved in energy metabolism, aerobic respiration, cell wall synthesis, DNA replication, transcription and protein synthesis suggests. These results implicate Lsr2 as a possible regulator of dormancy in *M. tuberculosis* (86).

Apart from generating a transcriptional response, mycobacterial NAPs have also been shown to protect DNA from physical damage. Lsr2 and GroEL1 both physically protect DNA from reactive oxygen intermediates through DNA binding (30, 31). HupB was shown to have a dual mechanism of DNA protection by reducing the damage caused by Fe<sup>3+</sup> through the ferroxidase activity and by DNA binding (106). Lsr2, HupB, mIHF, NapM and GroEL1 also physically protect DNA from DNaseI (31, 47, 55, 65, 84).

#### *STPK regulation of NAPs*

Protein phosphorylation through protein kinases translates extracellular stimuli into cellular processes, thereby allowing adaptation of the bacterial cell to adverse conditions experienced (107). *M. tuberculosis* encodes 11 STPKs which are predicted to phosphorylate 7.5% of all *M. tuberculosis* proteins (10, 108). In *M. tuberculosis*, Serine/Threonine protein kinases (STPKs) are known to regulate themselves, ABC transporter proteins, transcription factors, proteins involved in regulating cell division and pathogenesis (109). Recently, HupB and GroEL1 have both been shown to be substrates of STPKs (64, 110). *In vitro*, PknE, PknF, and PknB were shown to modify Thr<sup>65</sup> and Thr<sup>74</sup> of the N-terminal of HupB, with phosphorylation resulting in diminished DNA binding. HupB is maximally expressed during stationary phase in *M. tuberculosis*, however, the kinases regulating HupB DNA interaction were found to be most abundant during the exponential growth phase. Phosphorylation of HupB during the exponential growth phase would limit DNA interaction, however, during the stationary phase the abundant non-phosphorylated HupB would be capable of nucleoid compaction (64). *M. tuberculosis* GroEL1 was shown to be phosphorylated by PknF on Thr<sup>25</sup> and Thr<sup>54</sup>, however, the *M. smegmatis* homologue was not found to be a substrate of PknF phosphorylation (110). PknF phosphorylation has been shown to influence glucose transport, cell growth and septum formation in *M. tuberculosis* (111). Interestingly, the overexpression of PknF in *M. smegmatis* resulted in slowed growth and an altered colony morphology which was found to negatively impact the sliding motility and biofilm formation of this organism (112). Coincidentally, *groEL1* deletion from *M. smegmatis* prevented the formation of mature biofilms and in *M. tuberculosis* overexpression of MDP2 resulted in a 3-fold decrease of *groEL1* transcription which in turn resulted in reduced sliding motility and biofilm formation (71, 78). These results therefore suggest that PknF phosphorylation of GroEL1 results in the repression of GroEL1 mediated activities.

NAPs have the potential to regulate the transcriptional response of a large number of genes. The phosphorylation of these proteins through STPKs serves as a mechanism by which environmental changes can be transmitted to a large number of genes by a global regulator to bring about adaptive changes required. As is seen with the phosphorylation of HupB, modification by STPKs not only results in a

transcriptional response but phosphorylation has the potential to play a role in compacting the mycobacterial nucleoid.

### **NAPs are attractive drug targets**

---

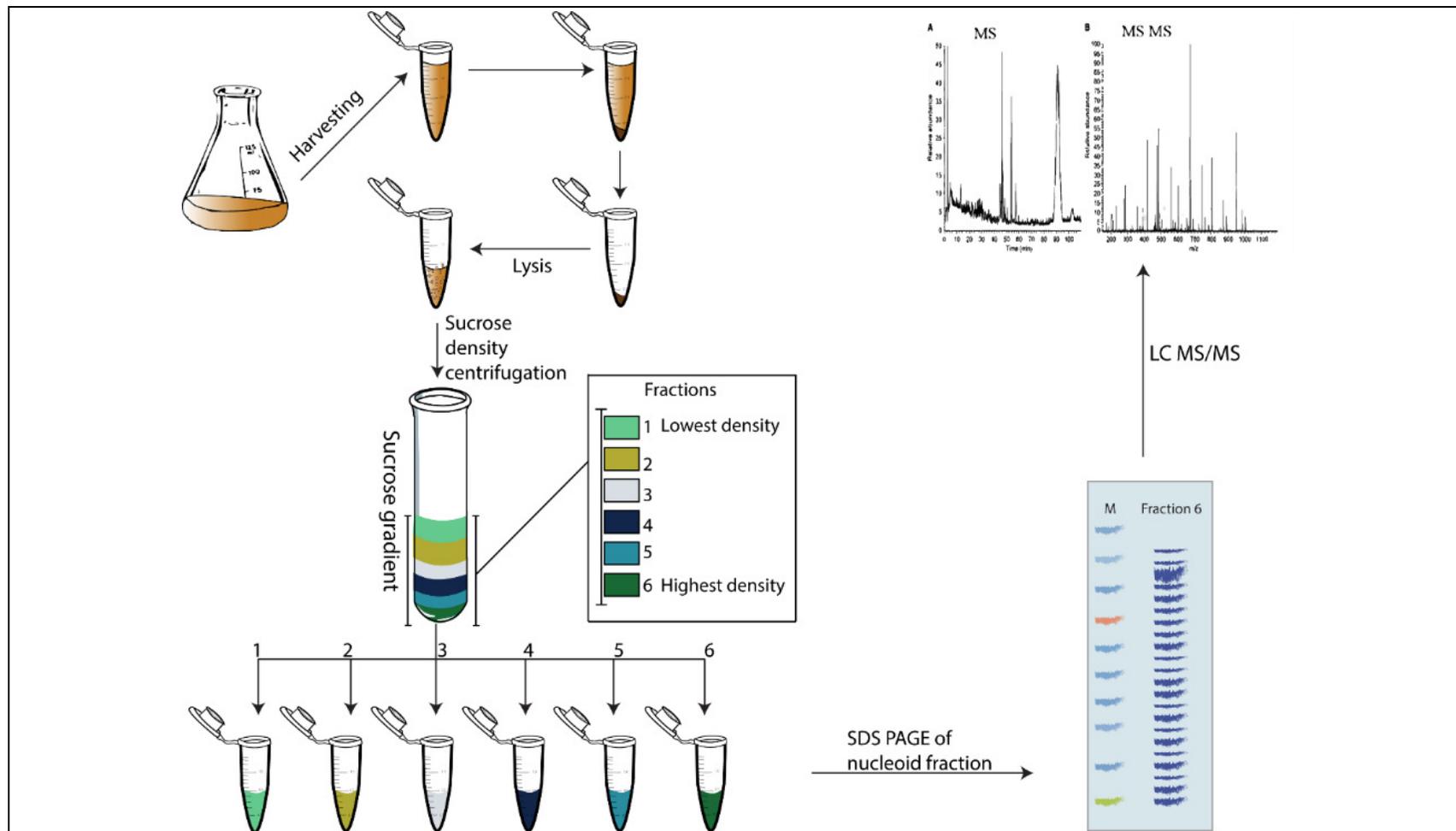
The emergence of drug resistant strains of *M. tuberculosis* has created a need for new drug targets and effective treatments to curb the spread of this disease. The essentiality of many NAPs for mycobacterial survival and their ability to regulate the expression of a large number of genes across numerous functional categories makes members of this unique class of proteins attractive drug targets. Zafirlukast, a cysteinyl leukotriene receptor antagonist currently used to treat asthma, has been demonstrated to inhibit the complexation between Lsr2 and DNA, resulting in the inhibition of mycobacterial growth in a concentration dependent manner. Zafirlukast was shown to permeate the mycomembrane, be active at the site of *M. tuberculosis* infection and to disrupt transcriptional regulation. This suggests that Zafirlukast has the potential to be repurposed as an anti-tuberculosis treatment (113). SD1, a *trans*-stilbene derivative, was identified through a structure-based approach to bind to the core region of HupB, resulting in diminished DNA binding. SD1 treated *M. tuberculosis* cells were shown to have reduced compaction of the nucleoid and inhibited growth (57). These studies demonstrated that by reducing DNA interaction with NAPs that are essential for mycobacterial growth, a diminished growth phenotype can be seen.

### **Identification of NAPs in Mycobacteria**

---

The under representation of identified NAPs in *M. tuberculosis* when compared to other organisms is of interest as conventional methods used to identify homologues of these proteins fail to do so in mycobacteria. The *M. tuberculosis* genome encodes a number of proteins with unknown functions and given the low number of identified NAPs in mycobacterial species it is possible that a number of these could encode NAPs.

Functional studies used in the past to identify members of this protein class made use of sucrose density gradient centrifugation followed by mass spectrometry (Figure 2. 4) (17). Albeit a very effective way of identifying proteins associated with the bacterial nucleoid, this approach is prone to protein contamination from other fractions. Proteins identified using this approach may be DNA binding but not necessarily NAPs and the ability of the proteins identified to shape DNA remains to be determined. AFM, a high resolution scanning probe microscopy technique, is commonly used to visualize the method of DNA compaction by NAPs (31, 51, 53, 55). Affinity purification of formaldehyde cross-linked mitochondrial nucleoids has been used to identify a number of proteins associated with mitochondrial DNA but also proteins which are involved in mitochondrial RNA metabolism and translation. Unfortunately, like with sucrose density centrifugation, this approach was found to be prone to protein contamination through protein-protein cross-linking (114).



**Figure 2. 4 Identification of nucleoid-associated proteins**

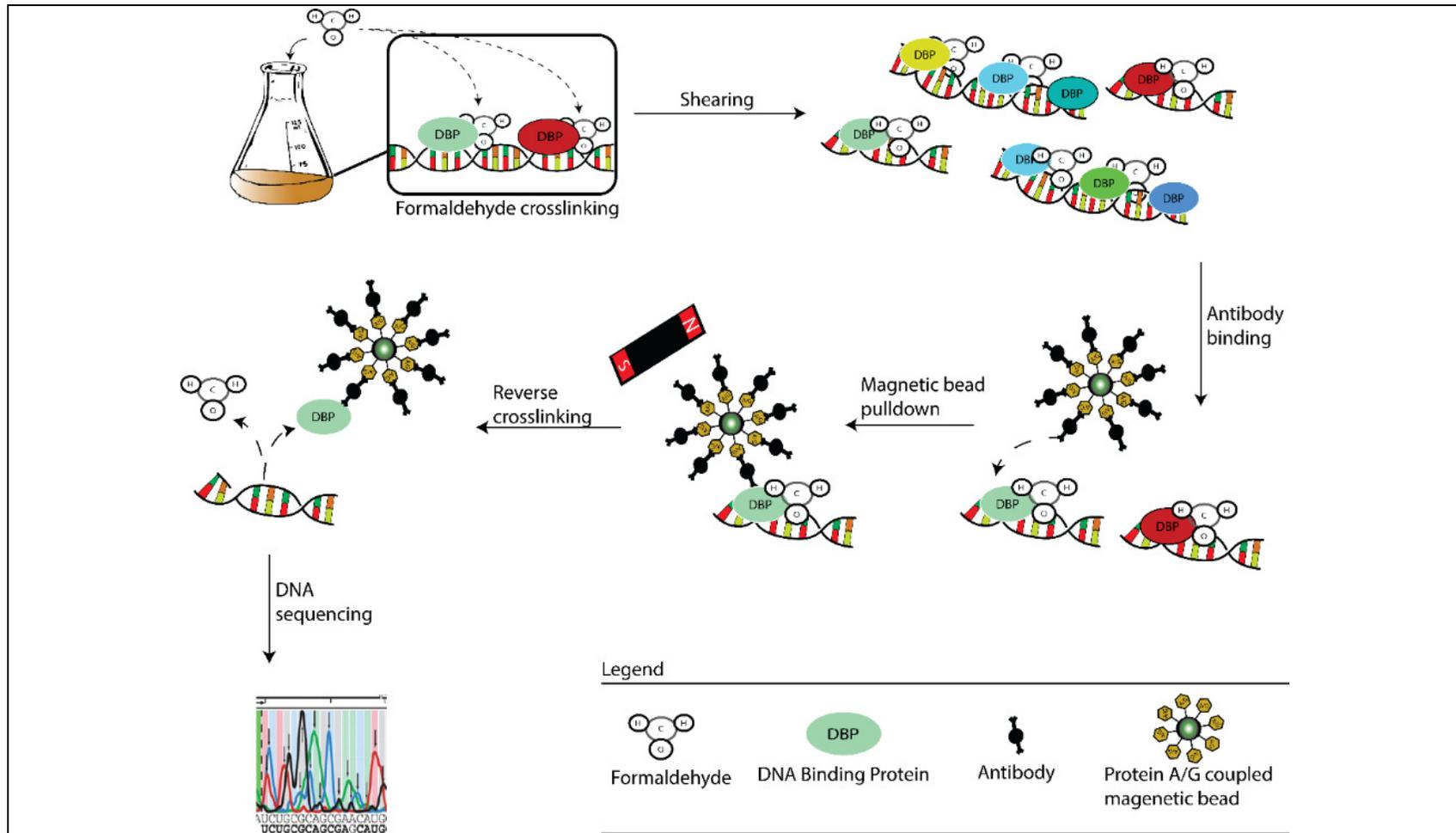
Cell lysates were subjected to sucrose density centrifugation to separate the lysate into fractions of different densities. The fraction with the highest concentration is considered to be the nucleoid fraction, and is subjected to analysis using LC-MS/MS to identify proteins associated with the bacterial nucleoid.

In recent years, ChIP (ChIP-chip and ChIP-seq) based techniques (Figure 2. 5) have been used to identify genes targeted by NAPs (9, 33, 86). The data generated by these high-throughput techniques and DNA footprinting studies have been used to describe DNA binding sequences for these proteins previously described as non-specific DNA interactors (33, 69). The identification of these DNA binding sequences has subsequently been used to predict a vast number of possible DNA binding sites throughout the genome, emphasizing the role of NAPs as global regulators of transcription. Additionally, data generated from a ChIP-seq study was used to show that the DNA binding of EspR was unconventional for a transcriptional regulator resulting in the reclassification of EspR as a NAP (33). Unfortunately, ChIP-seq experiments require previous knowledge of DNA binding and with the low availability of antibodies for mycobacterial proteins, this complicates the using of ChIP-seq to identify novel NAPs in mycobacteria.

### **Knowledge Gaps and Future Work**

---

ChIP-Chip, ChIP-seq and microarray analysis of several NAPs have effectively demonstrated the widespread ability of NAPs to bind to various positions in the mycobacterial genome and influence gene expression across various functional categories (9, 33, 47, 86, 87). The research presented in this review does, however, highlight the continued need to investigate the proposed mycobacterial NAPs identified in literature. Due to the lack of antibodies specific to the proteins being investigated, researchers have made use of episomal expression of tagged DNA-binding proteins in order to elucidate the DNA binding capacity of NAPs (9). Although effective, this approach could provide false DNA binding information as the proteins are not being expressed from their native promoters, possibly resulting in abnormal protein levels. Furthermore, better visualisation methodologies and technologies are required to investigate the effect of NAP binding on the structure of the mycobacterial nucleoid. The development of sub-diffraction limited resolution techniques such as STORM (stochastic optical reconstruction microscopy) has the potential to study NAPs in live cells and could aid in elucidating the role of these proteins in maintaining chromosome structure (115).



**Figure 2. 5 ChIP-seq can be used to resolve the binding potential of NAPs**

Formaldehyde is used to generate reversible crosslinks between DNA-binding proteins and DNA. Subsequent to cross-linking, DNA is sheared and immunoprecipitation of DNA-protein complexes are done. DNA-protein cross-linking is reversed before DNA fragments are sequenced to identify the binding capabilities of the targeted DNA binding protein.

## Conclusion

---

Mycobacterial pathogens such as *M. tuberculosis*, *M. leprae*, and *M. ulcerans* pose a threat to human health. Understanding how these bacterial pathogens maintain their nucleoids and regulate gene responses required to adapt to the adverse growth conditions experienced during infection is crucial to the development of novel treatment strategies to combat these infections. NAPs are a group of highly conserved transcriptional regulators, of which some are required for bacterial survival. The identification of NAPs in mycobacterial organisms has been hindered by poor sequence homology to the NAPs of other organisms, which suggest that several mycobacterial NAPs remain unidentified. Furthermore, the research presented in this literature review emphasizes the need for new methodologies to identify these NAPs as well as investigate not only their influence on transcriptional regulation but also their ability to shape the nucleoid. Investigation of these proteins holds the promise of identifying possible drug development targets with several studies already demonstrating the notable effects of targeting NAPs.

---

## Chapter 3: Development of Nucleoprotein - Mass Spectrometry

---

### 3.1. Introduction

---

Affinity purification coupled with mass spectrometry (AP-MS) has become an attractive approach for the investigation of protein-protein interaction studies and has contributed significantly to the interactome mapping of several organisms (116–119). Apart from the characterization of molecular networks, AP-MS has also successfully been used to study transcription factors in a variety of organisms (114, 116, 117, 120, 121). Affinity purification entails the capturing of biological material through specific interaction with a ligand (DNA, RNA, peptide, protein or antibody) immobilized onto a solid matrix (often agarose or magnetic beads) (122, 123). Following elution and targeted digestion (often using trypsin) of isolated proteins, high pressure liquid chromatography (HPLC) is used in conjunction with tandem mass spectrometry analysis for the identification of affinity enriched proteins. Peptide mixtures are loaded onto HPLC columns where they bind according to their physical properties, before being eluted with a hydrophobicity gradient into the mass spectrometer. Tandem mass spectrometry measures the molecular mass to charge ratio ( $m/z$ ) of ionized peptides in a two-step process. Firstly, intact ionized peptides are surveyed before being selected based in their relative abundance. In the second step, isolated intact ionized peptides are fragmented and  $m/z$  ratios are measured (123). Various bioinformatic tools (for example: MaxQuant, MSGF, SEQUEST, and XTandem) have been developed for the analysis of mass spectrometry data, and are effectively used for the identification of isolated proteins through automated database searching (124–127).

The overexpression of tagged proteins is frequently used to improve the recovery of proteins in AP-MS studies given the limited range of antibodies that are commercially available. This approach is however problematic as the overexpression of a protein can alter the physiological state of the cell and result in the identification of non-specific interactors (128). The identification of non-specifically bound proteins is however not unique to AP-MS experiments making use of overexpressed tagged proteins. Non-specific protein identifications are often as a result of untargeted proteins interacting with the solid matrix, the immunoglobulins used for the targeting of proteins and non-specific interaction with the targeted proteins themselves. Experimental efforts undertaken to limit the isolation of non-specific interactors includes the pre-clearance of the cell lysate using the unbound solid matrix prior to the addition of the ligand coupled matrix. Negative control immunoprecipitations using the empty matrix and non-related protein or antibody coupled matrices have been used to identify non-specific interactors during affinity purification experiments. Computationally, the Contaminant Repository for Affinity Purification (CRAPome) has aided in the identification of non-specific interactors for given experimental conditions (129).

Recently, an improvement on the conventional antibody-matrix AP-MS approach has been proposed. Affinity enrichment mass spectrometry (AE-MS) makes use of antibody purification of the protein of interest,

followed by quantification based mass spectrometry to identify possible protein interactors. AE-MS overcomes some of the limitations of AP-MS such as ectopic protein expression through the targeting of native proteins with the use of an antibody. As mentioned before, this will likely result in a reduction of identifying non-specific interactors however, due to the inadequate range of commercially available antibodies, antibody purification remains a limitation. Furthermore, AE-MS makes use of quantification mass spectrometry data to establish signature signal frequencies for non-specific identifications. These established signatures for non-specific interactors are used to identify proteins that are possibly true interactors which may also be present in the negative controls, thereby increasing the number of possible “true interactors” (119).

In this study, we aimed to develop a method which would enable the identification of proteins associated with the RNA polymerase transcription complex in *M. smegmatis*. We hypothesize that by combining whole cell formaldehyde cross-linking with affinity purification of protein-DNA and -RNA complexes it will be possible to identify DNA- and RNA- associated proteins using mass spectrometry. We anticipate the identification of nucleic acid associated proteins such as DNA polymerases, transcription factors, ribosomal proteins, sigma factors and translation factors. We believe that the results from this study will aid in our understanding of the proteins required to mediate transcriptional changes which are likely the result of multiple effector proteins. Conventional molecular techniques such as ChIP-seq or microarrays are limited through the autonomous investigation of transcription factors and microarray analysis frequently requires the generation of gene knockout or expression strains to measure the transcriptional regulation induced by a protein. Likewise, the limited range of commercially available antibodies for ChIP-seq analysis results in the need of ectopic protein expression which could ultimately result in the false positive identifications.

To achieve our objective, we developed a Nucleoprotein – Mass Spectrometry (NP-MS) approach. We anticipate that the targeting of nucleoprotein complexes through affinity purification of an anti-RNA polymerase antibody coupled to protein G Dynabeads™ will allow for the identification of the proteins which make up the RNA polymerase transcription complex and its associated proteins. We believe that this approach will allow us to identify multiple proteins required by the organism to adapt to its stressful environmental conditions.

## 3.2. Materials and Methods

---

All buffers used in this study are described in Table A. 1 and can be found in addendum A.

### 3.2.1. Bacterial Strains

*Escherichia coli* XL-1 Blue was used for all cloning procedures.

*Mycobacterium smegmatis* mc<sup>2</sup>155 was used for the establishment of NP-MS and for the validation of proteins identified using NP-MS (130).

### 3.2.2. Media, Culture and Storage Conditions

All culture media compositions are described in Table A. 2.

*E. coli* was cultured in Luria-Bertani liquid broth (LB) with shaking or on LB agar plates at 37°C, overnight. Media was supplemented with kanamycin (50 µg/ml) or hygromycin (150 µg/ml) when appropriate.

*M. smegmatis* was grown in Difco™ Middlebrook 7H9 broth supplemented with Albumin-Dextrose (AD) and Tween-80 at 37°C with shaking as required or on BBL™ Seven H11 Agar Base plates supplemented with AD at 37°C for 2-3 days. Kanamycin (25 µg/ml) and/or hygromycin (50 µg/ml) was added when appropriate. *M. smegmatis* cultures were screened for contamination using Ziehl-Neelsen (ZN) staining (131).

*E. coli* and *M. smegmatis* stocks were stored at -80°C.

### 3.2.3. Development of NP-MS

#### 3.2.3.1. Formaldehyde Cross-linking

*M. smegmatis* was cultured in a total volume of 50 ml to an optical density of ~0.4 at 600 nm prior to formaldehyde cross-linking. Thereafter, a 37% molecular grade formaldehyde solution (Merck) was added to a final concentration of 1% and cultures were incubated for 10 minutes at 37°C with gentle shaking (100 rpm). Glycine was then added to final concentration of 125 mM to quench the cross-linking reaction and cultures were incubated at 37°C for 10 minutes with no shaking. Cells were pelleted by centrifugation at 3220xg for 10 minutes at 4°C. Cell pellets were washed in ice-cold tris-buffered saline (TBS) (equal volume equal to that of the cell culture) and centrifuged at 3220xg for 10 minutes at 4°C. This wash step was repeated before freezing cell pellets at -80°C overnight.

### 3.2.3.2. Whole cell lysate preparation

Cell pellets were thawed on ice and resuspended in 4 ml immunoprecipitation (IP) buffer I containing protease inhibitors (cOmplete™, mini, EDTA-free protease inhibitor cocktail, Roche) before being transferred to a 15 ml conical centrifuge tube. Cells were sonicated using a QSonica Q700 probe sonicator, four times for 20 seconds at an amplitude of 30 with 2 minute intervals on ice.

### 3.2.3.3. DNA fragmentation

DNA fragmentation was done using micrococcal nuclease (Roche). Briefly, CaCl<sub>2</sub> (9 mM), RNase A (0.002 mg/ml) (Roche) and 50 - 800 U of micrococcal nuclease was added to whole cell lysates followed by incubation at 4°C for one hour with rotation. EDTA was added to a final concentration of 10 mM to stop the micrococcal nuclease reaction followed by centrifugation at 3220xg for 15 minutes at 4°C. Supernatants were recovered and kept on ice for further experimental procedures.

### 3.2.3.4. Proteinase K digestion and DNA precipitation

Proteinase K was used to digest proteins present in whole cell lysates in order to determine crosslinking and DNA fragmentation efficiency. Briefly, 25 µg proteinase K was added to 100 µl of whole cell lysate and incubated at 50°C for 2 hours, followed by over-night incubation at 65°C. DNA was precipitated through the addition of 10 µl 3M sodium acetate and 330 µl 95% ethanol followed by incubation on ice for 3 hours. Reactions were centrifuged at 16 873xg for 30 minutes before discarding the liquid. DNA precipitation pellets were washed with the addition of 500 µl 70% ethanol and centrifugation at 16 873xg for 5 minutes. Following the removal of ethanol, DNA pellets were allowed to dry at room temperature before being resuspended in nuclease free water.

### 3.2.3.5. Immunoprecipitation

For optimisation reactions, an anti -RNA polymerase β-subunit antibody (Santa Cruz) was coupled to Protein G Dynabeads™ (Thermo Fisher Scientific) to generate the resin for the immunoprecipitation of protein complexes associated with nucleic acids.

Dynabeads™ were prepared as per manufacturer's instructions. Briefly, beads were rotated for approximately 5 minutes before transferring 50 µl to 2 ml Eppendorf® microcentrifuge tubes. Beads were incubated with 200 µl phosphate buffered saline with tween (0.02%) (PBS-T) and 5 µg anti-RNA polymerase β-subunit antibody for 10 min rotating at room temperature. Following antibody binding, beads were gently washed using PBS-T before the addition of 2ml cross-linked whole cell lysate. Antibody coupled beads were incubated with cross-linked whole cell lysates for 2 hour intervals at 4°C with rotation before being subjected two washes using PBS-T and two washes using IP buffer I with protease inhibitors.

### 3.2.3.6. Agarose Gel Electrophoresis

Gel electrophoresis was used to separate precipitated DNA, PCR products and restriction enzyme digestions. Briefly, 1% agarose gels were subjected to electrophoresis in TAE buffer with ethidium bromide (0.00005%) at approximately 100V. Agarose gels were visualized using UV light. GeneRuler™ 100 bp Plus DNA Ladder (Thermo Fisher Scientific), GeneRuler™ 1 kb Plus DNA Ladder (Thermo Fisher Scientific) or KAPA™ Express Ladder (Roche) was used where appropriate to determine DNA fragment sizes.

### 3.2.3.7. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

All buffers used for SDS-PAGE and silver staining are described in Table A. 3.

Protein samples were prepared for separation by being incubated in 2X reducing SDS sample buffer at 100°C for 10 minutes before being separated on 12% SDS-PAGE gels. Separating gels were prepared as stipulated in Table A. 4 and stacking gels as indicated in Table A. 5. PageRuler™ Prestained Protein Ladder (Fermentas) was used as a molecular weight marker and SDS-PAGE gels were stained using silver staining. Briefly, gels were incubated in a fixing solution for 20 minutes at room temperature with gentle shaking. SDS-PAGE gels were rinsed with water prior to sensitising using sodium thiosulphate (0.005%) for 20 minutes followed by incubation with a 0.1% silver nitrate solution for 20 minutes. Gels were washed with water prior to the addition of developing buffer for 1-2 minutes followed by a 50 mM EDTA solution to stop the staining reaction.

### 3.2.3.8. Western Blotting

Separating and stacking gel compositions are described in Table A. 4 and Table A. 5, respectively. All buffers used for Western Blotting are described in Table A. 6.

Following SDS-PAGE, immunoblotting was done followed by blocking of the membrane in 5% Bovine Serum Albumin (BSA) overnight with gentle shaking.

#### *Western blot detection*

The blocked membrane was rinsed twice with TBS-T wash buffer before being washed 10 times with TBS-T for 10 minutes with gentle shaking at room temperature. The membrane was incubated with the primary antibody (mouse anti-RNA polymerase, 1:1000 dilution with TBS-T) for 60 minutes at room temperature with gentle shaking where after the membrane was rinsed twice with TBS-T wash buffer. The membrane was subsequently washed three times for 10 minutes at room temperature with gentle shaking. Following, the membrane was incubated with the secondary antibody (goat anti-mouse, 1:2000 dilution with TBS-T) for 60 minutes at room temperature with gentle shaking. Lastly, the membrane was rinsed twice with TBS-T before being washed three times for 10 minutes at room temperature with gentle shaking.

The membrane was placed in a sealable plastic bag and incubated with ECL detection fluid (Clarity™ Western ECL Blotting Substrate, Bio-Rad) in a dark room. The detection liquid was drained from the sealable bag and the membrane was placed within an X-ray cassette. Signal was detected with radiographic film.

### **3.2.4. Nucleoprotein – Mass Spectrometry (NP-MS)**

#### **3.2.4.1. Formaldehyde Cross-linking**

Two 50ml *M. smegmatis* cultures were grown to an optical density of ~0.4 at 600 nm prior to cross-linking. Formaldehyde cross-linking and quenching was done as described in section 3.2.3.1. Cells were pelleted by centrifugation at 3220xg for 10 minutes and resuspended in an equal volume ice-cold TBS before centrifugation at 3220xg for 10 minutes at 4°C. This wash step was repeated before freezing cell pellets at -80°C overnight.

#### **3.2.4.2. Whole Cell Lysate Preparation**

Cell pellets were thawed on ice and resuspended in 4 ml immunoprecipitation (IP) buffer I containing protease inhibitors (cOmplete™, mini, EDTA-free protease inhibitor cocktail, Roche) before being transferred to a 15 ml conical centrifuge tube. Cells were sonicated using a QSonica Q700 probe sonicator, four times at an amplitude of 30 for 20 seconds with 2 minute intervals on ice. Subsequently, CaCl<sub>2</sub> (9 mM), RNase A (0.002 mg/ml) (Roche) and 100 U micrococcal nuclease (Roche) was added to the cell lysates and incubated at 4°C for one hour with rotation. EDTA (Sigma-Aldrich) was added to a final concentration of 10 mM to stop the micrococcal nuclease reaction before centrifugation of whole cell lysates at 3220xg for 15 minutes at 4°C. Whole cell lysates obtained from two 50 ml cultures were pooled.

#### **3.2.4.3. Immunoprecipitation**

An anti-RNA polymerase  $\beta$ -subunit antibody (Santa Cruz) was coupled to Protein G Dynabeads™ (Thermo Fisher Scientific) to generate the resin for the immunoprecipitation of protein complexes associated with nucleic acids.

Dynabeads™ were prepared as per manufacturer's instructions. Briefly, beads were rotated for approximately 5 minutes before transferring 50  $\mu$ l to 2 ml Eppendorf® Protein LoBind microcentrifuge tubes. Beads were incubated with 200  $\mu$ l phosphate buffered saline with tween (0.02%) (PBS-T) and 50  $\mu$ l anti-RNA polymerase  $\beta$ -subunit antibody or 50  $\mu$ l anti-muscle myosin (human) (MYH7) antibody where appropriate for 10 min rotating at room temperature. Following antibody binding, beads were gently washed using PBS-T before the addition of 2ml cross-linked whole cell lysate. Antibody coupled beads were incubated with cross-linked whole cell lysates for 2 hours at 4°C with rotation before being subjected to

series of washes at room temperature to remove non-specifically bound proteins (as detailed in Table 3.1).

**Table 3.1 Washing of Dynabeads™**

Buffer	Number of washes	Time	Rotation	Pipetting
IP Buffer I + protease inhibitor	2	3 min	Yes	No
IP Buffer II + 500 mM NaCl	2	2 min	Yes	No
IP Buffer II + 750 mM NaCl	2	2 min	Yes	No
LiCl wash	1	3 min	Yes	No
PBS-T	1	3 min	Yes	No
Ammonium bicarbonate (NH <sub>4</sub> HCO <sub>3</sub> ) (50 mM)	2	N/A	No	Yes

#### 3.2.4.4. On-bead Tryptic Digestion

Immunoprecipitated proteins were subjected to an on-bead tryptic digestion followed by stage-tip purification before mass spectrometry analysis. Briefly, following immunoprecipitation and washing beads were resuspended in 200 µl 50 mM ammonium bicarbonate with 20 µl sequence modified trypsin (100 ng/µl) (Promega). Proteins were digested in a ThermoMixer® at 37°C with shaking at 700 rpm for 18 hours. Tryptic digests were recovered by separating the Protein G Dynabeads™ using a DynaMag™ magnetic separator and stored in 1.5 ml Eppendorf® Protein LoBind microcentrifuge tubes at -80°C.

#### 3.2.4.5. Concentration and Desalting of Tryptic Digests

Stage-tip purification of peptides was done to remove any salts and other contaminants before samples were loaded onto the mass spectrometer for analysis. Briefly, samples were concentrated using the Concentrator*plus* (Eppendorf) before being loaded onto activated stage-tips for desalting. Stage-tips were prepared by wedging Empore™ Octadecyl C18 beads/matrices into P200 pipette tips before activating them with 30 µl methanol followed by equilibration using 2% acetonitrile: H<sub>2</sub>O. Concentrated tryptic digests diluted in 2% acetonitrile and 0.1% formic acid and then loaded onto the stage tips. Stage-tips were washed with 30 µl solution of 2% acetonitrile and 0.1% formic acid before peptides were eluted using a solution of 50% acetonitrile and 0.1% formic acid. Eluted peptides were concentrated using the Concentrator*plus* before being resuspended in a 2% acetonitrile and 0.1% formic acid solution.

#### 3.2.4.6. Mass Spectrometry Analysis

LC-MS/MS analysis was done on the Orbitrap Fusion™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific) in collaboration with Dr M. Vlok at the Mass Spectrometry Proteomics unit at the Central

Analytical Facility (CAF) at Stellenbosch University. Dr. M. Vlok performed the LC-MS/MS analysis as well as the initial data acquisition.

#### *Liquid chromatography*

Liquid chromatography was performed using a Dionex-UltiMate 3000 Rapid Separation LC (RSLC) (Thermo Fisher Scientific) equipped with a 2 cm x 100  $\mu\text{m}$  C18 trap column and a custom 35 cm x 75  $\mu\text{m}$  C18 analytical column (Luna C18, 5  $\mu\text{m}$ , Phenomenex). Solvent A (2% acetonitrile and 0.1% formic acid and water) and solvent B (99.8% acetonitrile and 0.1% formic acid solution) were used for liquid chromatography. Samples were loaded onto the trap column using 100% solvent A at a flow rate of 5  $\mu\text{l}/\text{min}$  using a temperature controlled autosampler fixed at 7°C. The column was washed for 10 minutes before being eluted from the analytical column at a flow rate of 350 nL/min using the following gradient: 2% - 10% solvent B over 5 minutes, 10% - 25% solvent B over 45 minutes, 25% - 45% solvent B over 15 minutes using Chromeleon non-linear gradient 6. The column was subsequently washed for 10 minutes with an 80% solvent B solution followed by equilibration using solvent A. Chromatography was done at 50°C and the outflow was delivered to the mass spectrometer through a stainless steel nano-bore emitter.

#### *Mass spectrometry*

Mass spectrometry was done using a Thermo Scientific Fusion mass spectrometer equipped with a Nanospray Flex ionization source. Samples were introduced through a stainless steel emitter and data was collected in a positive mode with spray voltage set to 2kV and ion transfer capillary set to 275°C. Spectra were internally calibrated using polysiloxane ions at  $m/z = 445.12003$  and  $371.10024$ . MS1 scans were done using the orbitrap detector set at a resolution of  $R = 120\,000$  over the scan range 350-1650 with the AGC target at  $3\text{ E}5$  with a maximum injection time of 40 ms. Data was acquired in profile mode. MS2 acquisitions were done using monoisotopic precursor selection for an ion with charges between +2 and +6 with an error tolerance set to  $\pm 10$  ppm and precursor ions were excluded from fragmentation at 30 seconds. Precursor ions were selected for fragmentation in HCD mode using the quadrupole mass analyzer with HCD energy set to 32.5%. Fragment ions were detected within the ion trap mass analyzer using a rapid scan rate. The AGC target was set at  $1\text{ E}4$  with a maximum injection time of 45 ms. Data was acquired in centroid mode.

#### **3.2.4.7. Data Analysis**

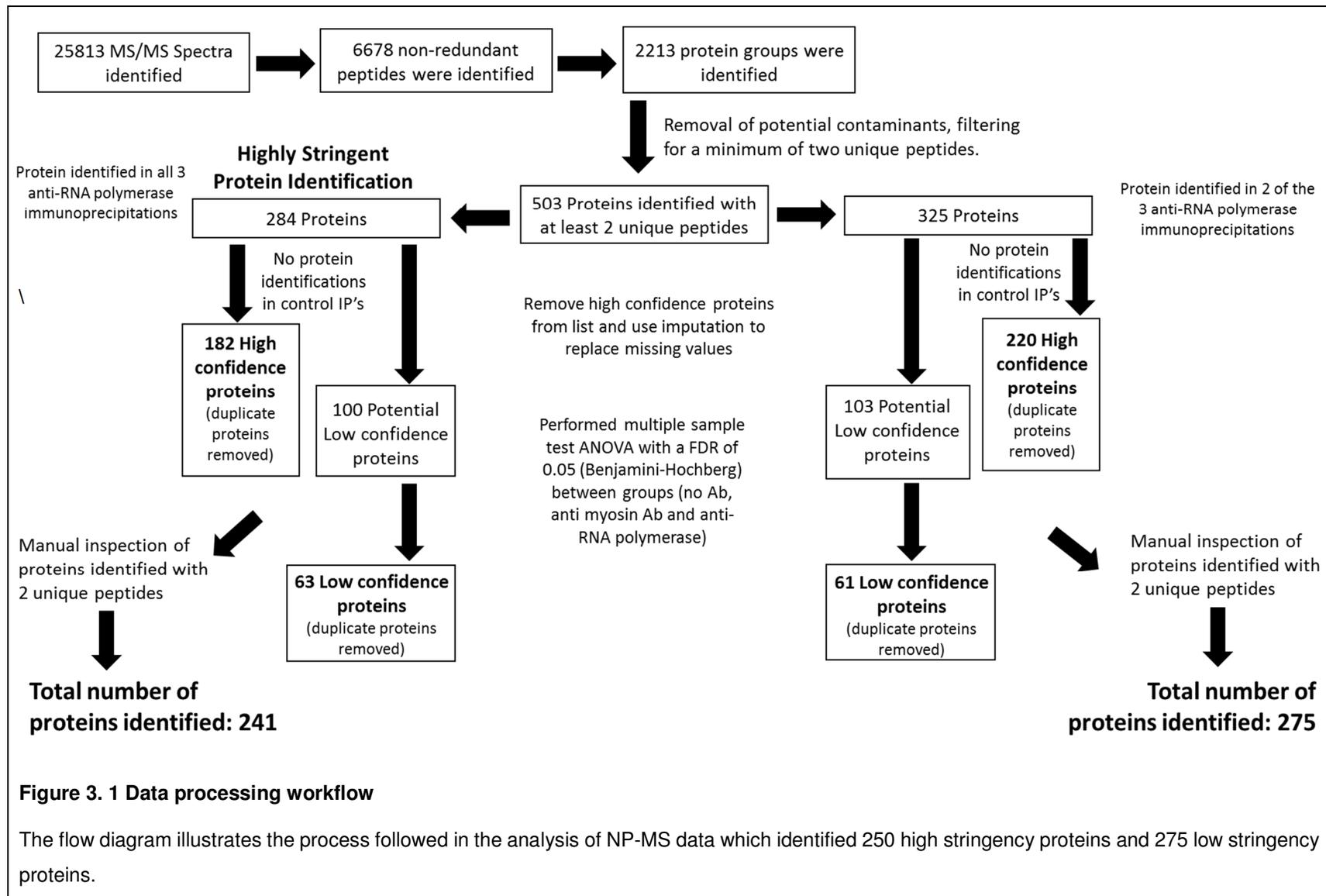
Data obtained from LC-MS/MS analysis was analysed using MaxQuant 1.5.0.25 to identify proteins via an automated database search of all tandem mass spectra against the *M. smegmatis* mc<sup>2</sup> 155 database (UP000000757) (124). Carbamidomethyl cysteine was set as a fixed modification and oxidized methionine, the possible addition of glycine to any lysine, serine and threonine (identified using Unimod: [http://www.unimod.org/modifications\\_list.php](http://www.unimod.org/modifications_list.php)), the possible di-methylation of lysine and arginine and the

addition of methylol and glycine on any histidine, asparagine, glutamine, tryptophan and tyrosine were used as variable modifications. Two missed tryptic cleavages were allowed and proteins were identified with a minimum of 1 unique peptide detected per protein. The protein and peptide false discovery rate (FDR) set at less than 0.01. A label-free approach was followed during MaxQuant data analysis and the “match between runs” algorithm was selected to detect sequenced peptides which were not selected for analysis in other experiments where after a Label Free Quantification (LFQ) calculation was applied. LFQ intensity values were used for statistical analysis using Perseus as part of the MaxQuant package. Intensity values were log 2 transformed and potential contaminants were removed prior to hierarchical clustering of the dataset. Two lists of proteins were identified using a very stringent and a less stringent approach and both these protein lists contain high and low confidence protein identifications. Data processing is summarised in Figure 3. 1.

Briefly, the very stringent protein identification list was compiled by selecting high confidence proteins when they were detected in all three biological replicates but not present in any of the negative controls. Following the removal of high confidence identified proteins from the dataset a low confidence list of proteins was generated. Similar to the high confidence protein identifications, low confidence protein candidates had to have been identified in all three biological replicate immunoprecipitations. The log transformed datasets were imputed using the Gaussian distribution relative to the standard deviation of measured values for each column separately in Perseus. Default settings were selected for imputation, where the width of the distribution was used for drawing random numbers from 30% of the standard deviation of the data and a down shift of 1.8 was used in relation to the standard deviation of the valid data (132). A multiple-sample test ANOVA between groups (group 1: anti-RNA polymerase antibody IP, group 2: anti-myosin antibody IP, group 3: protein G Dynabead® IP) with a FDR of 0.05 threshold was done. Proteins identified as significantly more present in the anti-RNA polymerase immunoprecipitation versus the control immunoprecipitations, with a fold change of at least 2 were included in the low confidence protein list.

The less stringent protein identification list was assembled by selecting high confidence proteins when they were identified in two of the three anti-RNA polymerase  $\beta$  subunit antibody immunoprecipitations but not present in any of the control immunoprecipitations. Low confidence proteins were identified as mentioned for the high stringency protein list.

Following identification of proteins for both the high and low stringency lists, manual spectral inspection was performed for proteins identified with only 2 unique peptides (133). Proteins were excluded from the high and low stringency identification lists when unique peptide spectra were found to have poor posterior error probability scores (PEP), major unexplained peaks, poor peptide coverage, majority y or b ions and low intensity peaks.



Proteins from both the high and low stringency proteins identification lists were described using data obtained from the Universal Protein Resource database (UniProt) (<http://www.UniProt.org/>). The Kyoto Encyclopaedia of Genes and Genomes (KEGG) search and colour pathway analysis tool ([http://www.genome.jp/kegg/tool/map\\_pathway2.html](http://www.genome.jp/kegg/tool/map_pathway2.html)) was used to map identified protein identifications in described metabolic pathways (134). Gene Ontology (GO) enrichment analyses was done using the Gene Ontology Enrichment Analysis Software Toolkit (GOEAST) (<http://omicslab.genetics.ac.cn/GOEAST/>) followed by the removal of redundant GO identifications using Reduce and Visualize Gene Ontology (REVIGO) (<http://revigo.irb.hr/>) (135, 136). Cytoscape 3.3.0 was used for the visualization of interactive graphs generated using REVIGO (137). The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) was used to map predicted protein-protein interaction data before interactive graph generation in Cytoscape (138).

### 3.2.5. Validation of NP-MS

#### 3.2.5.1. Competent Cells

##### *Chemical Competent E. coli*

Chemical competent *E. coli* was prepared by streaking out *E. coli* XL-1 Blue on LB agar plates containing tetracycline (50 µg/ml) followed by overnight incubation at 37°C. A single colony was picked and inoculated into 5 ml LB broth supplemented with tetracycline (50 µg/ml) and incubated at 37°C with shaking, overnight. *E. coli* was diluted 1:100 into double strength LB supplemented with 0.2% glucose and incubated at 37°C with shaking until the desired optical density of ~0.4. The *E. coli* culture was placed on ice and further steps were done at 4°C. Cells were centrifuged for 10 minutes at 3220xg at 4°C before being resuspended in 20 ml ice-cold 100 mM MgCl<sub>2</sub> followed by a 30 minute incubation on ice (4°C). Cells were centrifuged for 10 minutes at 4°C at 3220xg and resuspended in 2 ml ice-cold 100 mM CaCl<sub>2</sub> supplemented with 15% Glycerol before being aliquoted (170µl) into 1.5 ml Eppendorf® Safe-Lock tubes. Chemical competent *E. coli* was stored at -80°C.

##### *Electrocompetent M. smegmatis*

Electrocompetent *M. smegmatis* was prepared by inoculating a *M. smegmatis* starter culture into Difco™ 7H9 and incubating it at 37°C with shaking until the desired optical density of ~0.6. Cultures were placed on ice for 1 hour before cells were centrifuged at 3220xg for 10 minutes at 4°C. Cells were resuspended in 10% glycerol (4°C) of the same volume before repeating the centrifugation step. The wash step was repeated and cells were resuspended in 2 ml 10% glycerol per 100 ml culture. Electrocompetent *M. smegmatis* was stored at -80°C.

### 3.2.5.2. Primers

All primer sequences are described in Table 3. 2.

Sequencing information used for the design of primers was obtained from Smegmalist (<http://mycobrowser.epfl.ch/smegmalist.html>) and Tuberculist (<http://tuberculist.epfl.ch/>), respectively. For the selection of DNA-associated validation candidates the protein sequence analysis and classification tool, InterPro (<http://www.ebi.ac.uk/interpro/>), as well as the database of protein families, Pfam (<http://pfam.xfam.org/>) were used to identify predicted protein domains which might be nucleic acid associated. Primers used in this study were ordered from Integrated DNA Technologies (IDT).

Two sets of primers were used for the generation of the pNFLAG vector. The first set of primers was designed to incorporate the FLAG-tag sequence (GATTATAAAGATGACGACGATAAA) and an 8xGlycine linker chain when amplifying *MSMEG\_0615* from the vector pDMNI0615. The second set of primers were designed to facilitate In-Fusion® HD Cloning as stipulated by the manufacturer (Clontech, Takara). Briefly, primers were designed to have 15 bp complementary sequence to the linearized vector at the site of insertion and 15 bp complementary to the gene of interest. Restriction enzyme sequences between the complementary sequences were incorporated for the linearization of the vector. The PCR product generated using the first set of primers, served as the template for the second set of primers (Figure 3. 2).

Primers designed for the validation of NP-MS were designed to facilitate In-Fusion HD cloning as mentioned above. Primers were designed to PCR amplify *MSMEG\_0615*, *MSMEG\_1060*, *MSMEG\_2389*, *MSMEG\_2695*, *MSMEG\_3754*, *MSMEG\_4306* and *MSMEG\_5512*.

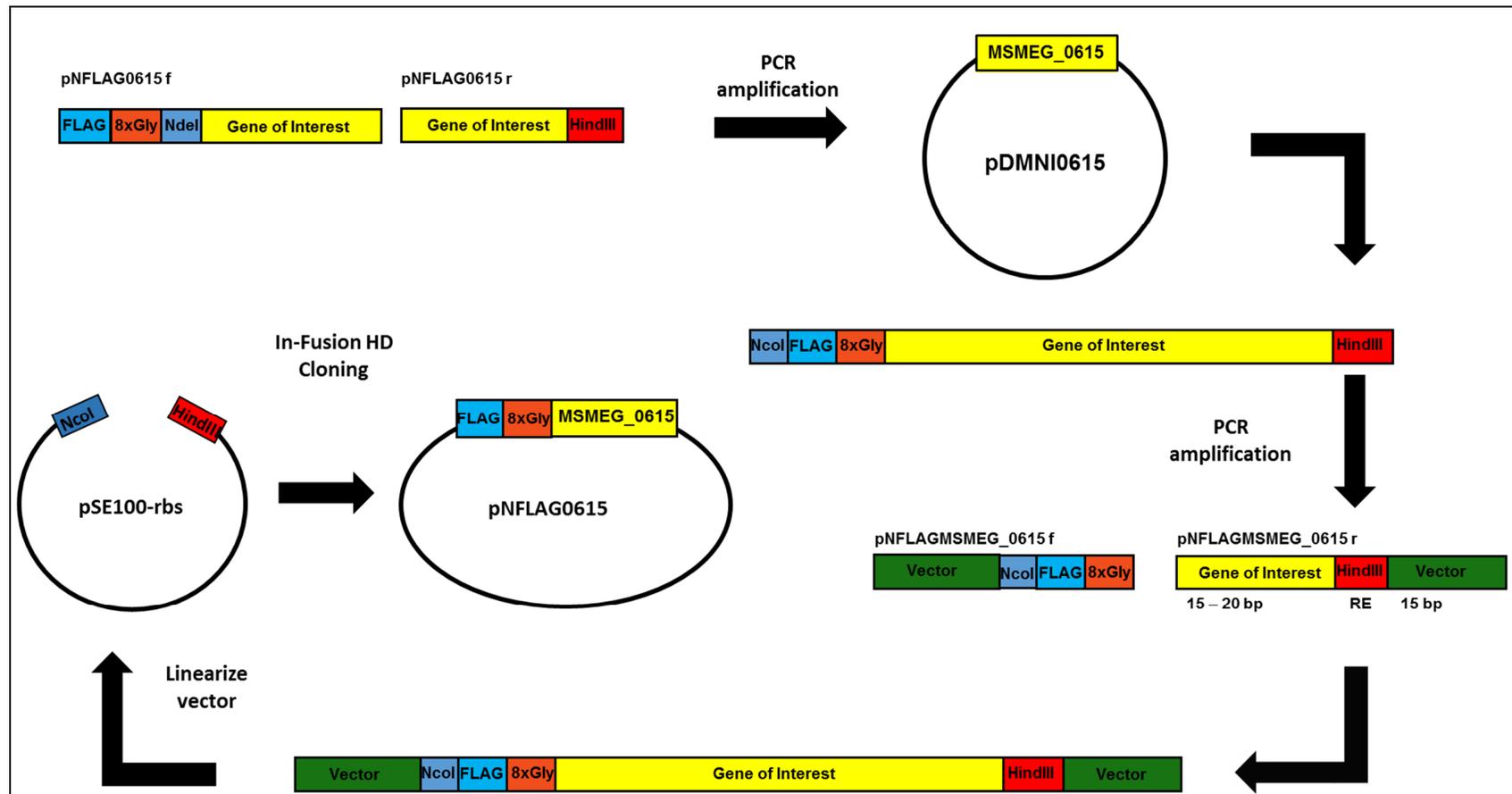
**Table 3. 2 Primer sequences**

Region	Primer name	Sequence (5' – 3')	Restriction site	Description
MSMEG_0615 insert in pDMNI	NFLAG0615 f	CCATGGGCAGCAGC <u>GATTATAAAGATGA</u> <u>CGACGATAAA</u> <u>CGTGTTGGTGGTGGTGGT</u> <u>GGTGGT</u> CATATGGGAAGTGACACGCTTG C	NcoI, NdeI	1898 bp product, including 56 bp for the incorporation of the N-terminal FLAG-tag and linker sequence, 9 bp for restriction enzyme cut sites and 1827 bp for MSMEG_0615 amplification
MSMEG_0615 insert in pDMNI	NFLAG0615 r	AAGCTTTCATGCCGGCACCGACAGTC	HindIII	1898bp product, including 6 bp for restriction enzyme cut sites and 1827 bp for the amplification of MSMEG_0615
FLAG-tagged MSMEG_0615	pNFLAGMSMEG_0615 f	<u>AGGAGATATACCATGGGCAGCAGCGATT</u> ATAAAGA	NdeI	1918 bp product, including 15 bp overhangs to facilitate In-Fusion HD® Cloning and the amplification of N-terminally FLAG-tagged MSMEG_0615
FLAG-tagged MSMEG_0615	pNFLAGMSMEG_0615 r	<u>TAATTAGCTAAAGCTTTCATGCCGGCACCC</u> GACAGT	HindIII	1918 bp product, including 15 bp overhangs to facilitate In-Fusion HD® Cloning and the amplification of N-terminally FLAG-tagged MSMEG_0615
MSMEG_1060, <i>M. smegmatis</i> genome	pNFLAGMSMEG_1060 f	<u>GTGGTGGTGGTCATATGGTGGGAAAAAT</u> TGTGACGAT	NdeI	438 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 405 bp for the amplification of MSMEG_1060
MSMEG_1060, <i>M. smegmatis</i> genome	pNFLAGMSMEG_1060 r	<u>TAATTAGCTAAAGCTTCTACACCCCTCAAC</u> TGCGAAG	HindIII	438 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 405 bp for the amplification of MSMEG_1060
MSMEG_2389, <i>M. smegmatis</i> genome	pNFLAGMSMEG_2389 f	<u>GTGGTGGTGGTCATATGAACAAAGCGGA</u> GCTCATCGA	NdeI	657 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 627 bp for the amplification of MSMEG_2389
MSMEG_2389, <i>M. smegmatis</i> genome	pNFLAGMSMEG_2389 r	<u>TAATTAGCTAAAGCTTTTACCTGCGGCC</u> TTCTTG	HindIII	657 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 627 bp for the amplification of MSMEG_2389
MSMEG_2695, <i>M. smegmatis</i> genome	pNFLAGMSMEG_2695 f	<u>GTGGTGGTGGTCATATGGCCAATCCGTT</u> CGTCAAGG	NdeI	870 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 840 bp for the amplification of MSMEG_2695
MSMEG_2695, <i>M. smegmatis</i> genome	pNFLAGMSMEG_2695 r	<u>TAATTAGCTAAAGCTTTTATTGGCCCAGC</u> GGGTTTT	HindIII	870 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 840 bp for the amplification of MSMEG_2695

MSMEG_3754, <i>M. smegmatis</i> genome	pNFLAGMSMEG_ 3754 f	<u>GTGGTGGTGGTCATATGGT</u> GATTCTCTTT TCAGCCGG	NdeI	1023 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 990 bp for the amplification of MSMEG_3754
MSMEG_3754, <i>M. smegmatis</i> genome	pNFLAGMSMEG_ 3754 r	TAATTAGCTAAAGCTTT <u>CAGGACAGCTCC</u> GTGATG	HindIII	1023 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 990 bp for the amplification of MSMEG_3754
MSMEG_4306, <i>M. smegmatis</i> genome	pNFLAGMSMEG_ 4306 f	<u>GTGGTGGTGGTCATATGAAAGCTGAAGT</u> AAGCCAACAAC	NdeI	759 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 729 bp for the amplification of MSMEG_4306
MSMEG_4306, <i>M. smegmatis</i> genome	pNFLAGMSMEG_ 4306 r	TAATTAGCTAAAGCTTT <u>CACTGCTTGACC</u> CGCAAC	HindIII	759 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 729 bp for the amplification of MSMEG_4306
MSMEG_5512, <i>M. smegmatis</i> genome	pNFLAGMSMEG_ 5512 f	<u>GTGGTGGTGGTCATATGGTGGT</u> GAGTCA ACCGAAGGA	NdeI	1419 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 1386 bp for the amplification of MSMEG_5512
MSMEG_5512, <i>M. smegmatis</i> genome	pNFLAGMSMEG_ 5512 r	TAATTAGCTAAAGCTTT <u>CAGCCATAGACG</u> GTCTCG	HindIII	1419 bp product, including 6 bp for restriction enzyme cut sites, 15 bp to facilitate In-Fusion HD® Cloning and 1386 bp for the amplification of MSMEG_5512
pNFLAG vectors	pNFLAG f	CCGAAATGAGCACGATCC	N/A	Sequencing final vectors
pNFLAG vectors	pNFLAG r	TAATTGGGGACCCTAGAGGTC	N/A	Sequencing final vectors

Nucleotides outlined in the primer NFLAG0615 f encodes the FLAG-tag and double underlined nucleotides within the same primer encodes the 8 x glycine linker chain between the FLAG-tag and the gene of interest.

Underlined: destination vector sequence incorporated into primer sequence to facilitate In-Fusion® HD cloning.



**Figure 3. 2 Construction of pNFLAG**

The gene MSMEG\_0615 was PCR amplified with the primer set NFLAG0615 from the plasmid pDMNI0615 to generate a PCR product that has an N-terminal FLAG-tag. The generated PCR amplicon was used to generate a second PCR product using the primer set pNFLAGMSMEG\_0615. The final PCR product was cloned into linear pSE100-rbs using In-Fusion<sup>®</sup> HD cloning.

### 3.2.5.3. PCR Amplification

Phusion High Fidelity DNA polymerase (NEB) and GC-rich buffer (Roche) was used for PCR amplification as described in Table 3. 3 and Table 3. 4. Melting temperatures were estimated by adding the multiplied amounts of cytosines and guanines (multiplied by four) and adenine and thymines (multiplied by two) together. Elongation times were calculated at 1 minute per 1 kb DNA.

**Table 3. 3 Master mix for Phusion High-Fidelity DNA Polymerase reaction**

Component	50 $\mu$ l reaction	Final Concentration
Water (nuclease free)	22.5 $\mu$ l	
5X Phusion HF Buffer	10 $\mu$ l	1X
GC rich solution (FastStart <sup>®</sup> (Roche))	10 $\mu$ l	1X
dNTPs	4 $\mu$ l	800 $\mu$ M
Phusion DNA Polymerase	0.5 $\mu$ l	1 unit
F primer	0.5 $\mu$ l	0.5 $\mu$ M
R primer	0.5 $\mu$ l	0.5 $\mu$ M
DNA	2 $\mu$ l	<250 ng

**Table 3. 4 PCR amplification conditions for Phusion High Fidelity DNA polymerase**

Number of Cycles	Temperature ( $^{\circ}$ C)	Time	Reaction
1	98	30 s	Activation Phusion Taq
35	98	10 s	Template denaturation
	Tm	30 s	Annealing of primers
	72	1 min/kb	Elongation
1	72	10 min	Elongation
1	4	$\infty$	Sample Storage

### 3.2.5.4. Plasmids

All plasmids used in this study are described in Table 3. 5.

**Table 3. 5 Plasmids**

Plasmid	Description	Size (bp)	Source/Reference
pSE100-csd-rbs	Mycobacterial expression vector containing <i>M. tuberculosis</i> Rv1464 ( <i>csd</i> ), a ribosome binding site (rbs), His-tag, thrombin cleave site, Pmyc1tetO promoter, Hyg <sup>R</sup>	6856	Jesmine Arries (MSc thesis 2015)
pDMNI0615	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_0615, SAGSAG linker, C-terminal <i>gfp</i> gene, psmyc promoter, Kan <sup>R</sup>	6914	Nastassja Steyn (MSc thesis 2012)
pNFLAG	Mycobacterial expression vector containing a N-terminal FLAG-tag, 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	5600	This study
pNFLAG0615	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_0615, N-terminal FLAG-tag with a 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	7427	This study
pNFLAG2389	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_2389, N-terminal FLAG-tag with a 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	6227	This study
pNFLAG1060	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_1060, N-terminal FLAG-tag with a 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	6008	This study
pNFLAG2695	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_2695, N-terminal FLAG-tag with a 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	6440	This study
pNFLAG3754	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_3754, N-terminal FLAG-tag with a 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	6593	This study
pNFLAG4306	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_4306, N-terminal FLAG-tag with a 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	6329	This study
pNFLAG5512	Mycobacterial expression vector containing <i>M. smegmatis</i> MSMEG_5512, N-terminal FLAG-tag with a 8 x Glycine linker, pmyc1tetO promoter, Hyg <sup>R</sup>	6989	This study
pTEK-4S-0X	<i>M. smegmatis</i> expression vector containing <i>revTetR</i> , psmyc promoter, Kan <sup>R</sup>	6466	Guo <i>et al.</i> 2007 (139)

### 3.2.5.5. Restriction Enzyme Digestion

Plasmid DNA was digested using restriction enzyme endonucleases as per manufacturer's instructions unless otherwise stated in text. Briefly, 1 µg plasmid DNA was digested using 10 U of restriction enzyme in a total volume of 10 µl at 37°C for 1-2 hours. Restriction enzyme nucleases were purchased from New England Biolabs (NEB) or Roche.

### 3.2.5.6. DNA Purification

For the purification of DNA embedded in agarose gel, PCR reactions or restriction enzyme digestions the Wizard® SV Gel and PCR Clean-up Kit (Promega) was used as per manufacturer's instructions.

### 3.2.5.7. DNA Quantification

The NanoDrop™ 2000c Spectrophotometer (Thermo Scientific) was used for the quantification of DNA samples.

### 3.2.5.8. DNA Sequencing

Sequencing was done by CAF at Stellenbosch University using an ABI 3730XL DNA Analyzer (Applied Biosystems) with appropriate DNA sequencing primers.

### 3.2.5.9. In-Fusion® HD Cloning

The In-Fusion® HD Cloning kit (Clontech) was used as per manufacturer's instructions for the generation of the pNFLAG vector as well as for the DNA association validation vectors for NP-MS. Isothermal/Gibson cloning makes use of three enzymes, an exonuclease, a polymerase and a ligase (140). The exonuclease digests either 3' or 5' overhangs (depending on the type of exonuclease) of both the linearized vector and the insert, generating large single stranded overhangs which are complementary. Subsequent to large overhangs annealing, the polymerase will fill all single stranded regions and the ligase will fuse the vector. Following this, generated vectors are propagated in *E. coli*.

Briefly, the vector, pSE100-rbs-csd (Figure 3. 3) (Jesmine Arries, MSc thesis 2015), was linearized using restriction enzymes NcoI and HindIII for the generation of pNFLAG. For the generation of validation vectors, pNFLAG was linearized using NdeI and HindIII. Following amplification and purification of PCR amplified inserts, the In-Fusion® HD cloning kit reaction was used with the appropriate controls (provided in the In-Fusion® HD Cloning kit) (Table 3. 6). Briefly, reaction mixture was prepared as depicted in Table 3. 6 with the amount of PCR insert and linearized vector added not exceeding half of the reaction volume. Reaction mixtures were incubated at 50°C for 15 minutes before being incubated on ice prior to *E. coli* transformation. The In-Fusion® HD cloning reaction was transformed with chemical competent *E. coli* (not TOP10 cells or their derivatives, these transformants result in a lower number of recombinant clones) with no more than 5

$\mu$ l of In-Fusion® cloning reaction being transformed with 50  $\mu$ l of chemical competent *E. coli*. The integrity of the genes cloned into pNFLAG were verified using sequencing.

**Table 3. 6 In-Fusion® HD cloning reaction**

Component	10 $\mu$ l reaction	Negative Control	Positive Control
Purified PCR Insert	10-200 ng*	0 ng	2 $\mu$ l (control insert)
Linearized Vector	50-200 ng**	1 $\mu$ l	1 $\mu$ l (pUC19 control vector)
5 X In-Fusion® HD Enzyme Premix	2 $\mu$ l	2 $\mu$ l	2 $\mu$ l
Water	to 10 $\mu$ l	to 10 $\mu$ l	to 10 $\mu$ l

\* Inserts size: <0.5 kb: 10-50 ng, 0.5 to 10 kb: 50-100 ng, >10 kb: 50-200 ng.

\*\* Vector size: <10kb: 50-100 ng, >10 kb: 50-200 ng.

### 3.2.5.10. Transformations

#### *E. coli* transformations

Plasmid transformations into chemical competent *E. coli* were done by adding 5  $\mu$ l of the In-Fusion® HD cloning mixture to 50  $\mu$ l of chemically competent *E. coli*. Transformation mixtures were incubated on ice (4°C) for 5 minutes followed by a 10 minute incubation at 37°C and lastly a 5 minute incubation on ice. Following the transformation, 1 ml of LB broth was added to the transformation mixture and a one generation recovery (20-30 minutes) time at 37°C with shaking was allowed before plating out on LB agar plates supplemented with the appropriate antibiotic.

#### *M. smegmatis* transformations

Plasmid transformations into *M. smegmatis* were done by adding 1  $\mu$ g of purified plasmid DNA to 200  $\mu$ l of electrocompetent *M. smegmatis*. Transformations were done using the Gene Pulser™ Electroporation system (Bio-Rad) and 0.2 cm electroporation cuvettes at 2.5 kV, 25  $\mu$ F, 125  $\mu$ Fd and 1000  $\Omega$ . Following electroporation, 1 ml 7H9 supplemented with AD and tween-80 was added to the reaction mixture and allowed to incubate at 37°C with shaking for 1 generation doubling time (3 hours) before plating out on 7H11 agar plates supplemented with the appropriate antibiotics.

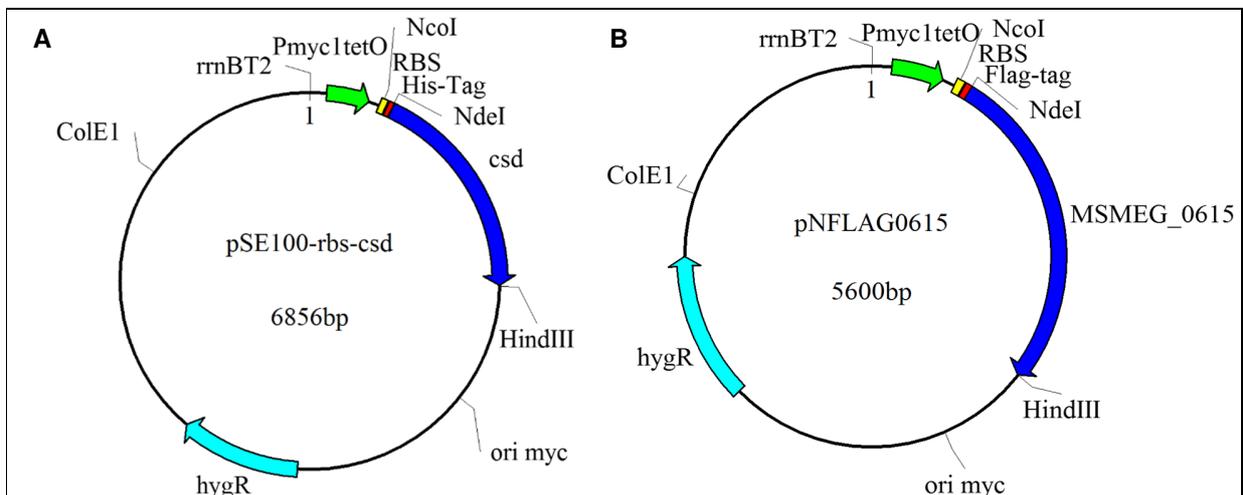
### 3.2.5.11. pNFLAG Plasmid Construction

The plasmid pNFLAG was created by modifying the generated vector pSE100-rbs-csd (Figure 3. 3). The before mentioned plasmid was constructed by excising the ribosome binding site (RBS), 6 x His-tag and thrombin cleavage site from pet-28a(+) and ligating the obtained fragment into pSE100 (Jasmine Arries, MSc thesis 2015) (Figure 3. 3). For the generation of the pNFLAG plasmid, the 6 x His-tag and thrombin cleavage site together with the *csd* insert was removed from the vector by partial digesting with NcoI (Roche) and HindIII (Roche). Briefly, restriction enzyme reactions were done in a final volume of 10 $\mu$ l,

containing 1 µg of vector DNA, 10 U of HindIII and 1 U of NcoI. Restriction enzyme digests were incubated at 37°C for 15 minutes. Enzymatic reactions were stopped with the addition of gel loading dye (Purple, NEB).

The N-terminal FLAG-tag and 8 x Glycine linker was amplified together with MSMEG\_0615 using the NFLAG0615 f and NFLAG0615 r primers described in Table 3. 2 from the plasmid pDMNI0615 (Nastassja Steyn, MSc thesis 2012). The amplified DNA fragment was purified and subsequently used as the template for the PCR amplification of the insert for the digested pSE100-rbs-csd plasmid. The insert was PCR amplified using pNFLAGMSMEG0615 f and pNFLAGMSMEG0615 r (Table 3. 2) which added the 15 bp overhangs required for In-Fusion® HD Cloning. Following purification, the generated inset was cloned into the linearized pSE100-csd-rbs plasmid using the In-Fusion® HD Cloning kit (Figure 3. 2). The integrity of the generated tag, linker and amplified insert was verified using sequencing.

To generate a pNFLAG construct without an insert, *MSMEG\_0615* was excised from pNFLAG0615 using restriction enzymes HindIII and NdeI and the DNA overhangs were blunted using the Klenow fragment (Roche). The DNA blunting reaction was incubated at 37°C for 2 hours. The digested and blunted vector was separated using agarose gel electrophoresis and excised before being purified using the Wizard® SV Gel and PCR Clean-up Kit. The purified linear vector was circularized using T<sub>4</sub> DNA ligase. Briefly, 100 ng vector was ligated using 3 U of T<sub>4</sub> DNA ligase and the reaction was incubated at room temperature for 2 hours. The integrity of the generated vector was verified using Sanger sequencing using the sequencing primers pNFLAG f and pNFLAG.



**Figure 3. 3 Vector maps**

pSE100-rbs-csd is a mycobacterial expression vector which allows for the conditional expression of N-terminally His-tagged proteins. This vector was modified to generate the vector pNFLAG0615 which allows for the conditional expression of N-terminally FLAG-tagged proteins in combination with pTEK-4S-0X.

### 3.2.5.12. DNA Association Assay

All buffers used for the DNA association assay are described in Table A. 1.

SDS-PAGE and Western blotting was done as described in section 3.2.3.7. and 3.2.3.8. using a mouse anti-FLAG antibody (1:4000 dilution) and a goat anti-mouse secondary antibody (1:10000 dilution) to confirm the expression of FLAG-tagged proteins in *M. smegmatis*.

*M. smegmatis* transformed with the plasmids pNFLAG, pNFLAG0615, pNFLAG2389, pNFLAG1060, pNFLAG2695, pNFLAG3754, pNFLAG4306 and pNFLAG5512 were cultured to an optical density of ~0.4 before cross-linking with formaldehyde (see section 3.2.3.1). Following quenching of the cross-linking reaction through the addition of glycine, cell pellets were frozen at -80°C overnight. Cell pellets were thawed on ice before being resuspended in 4 ml IP buffer I with protease inhibitors and sonicated as described in section 3.2.4.2. A mouse anti-FLAG primary antibody bound to protein G Dynabeads™ was used to enrich for FLAG-tagged proteins. Antibody bound beads and unbound beads (bead only control) were each incubated with 2 ml whole cell lysate for 2 hours at 4°C with rotation before being washed as described in section 3.2.4.3.

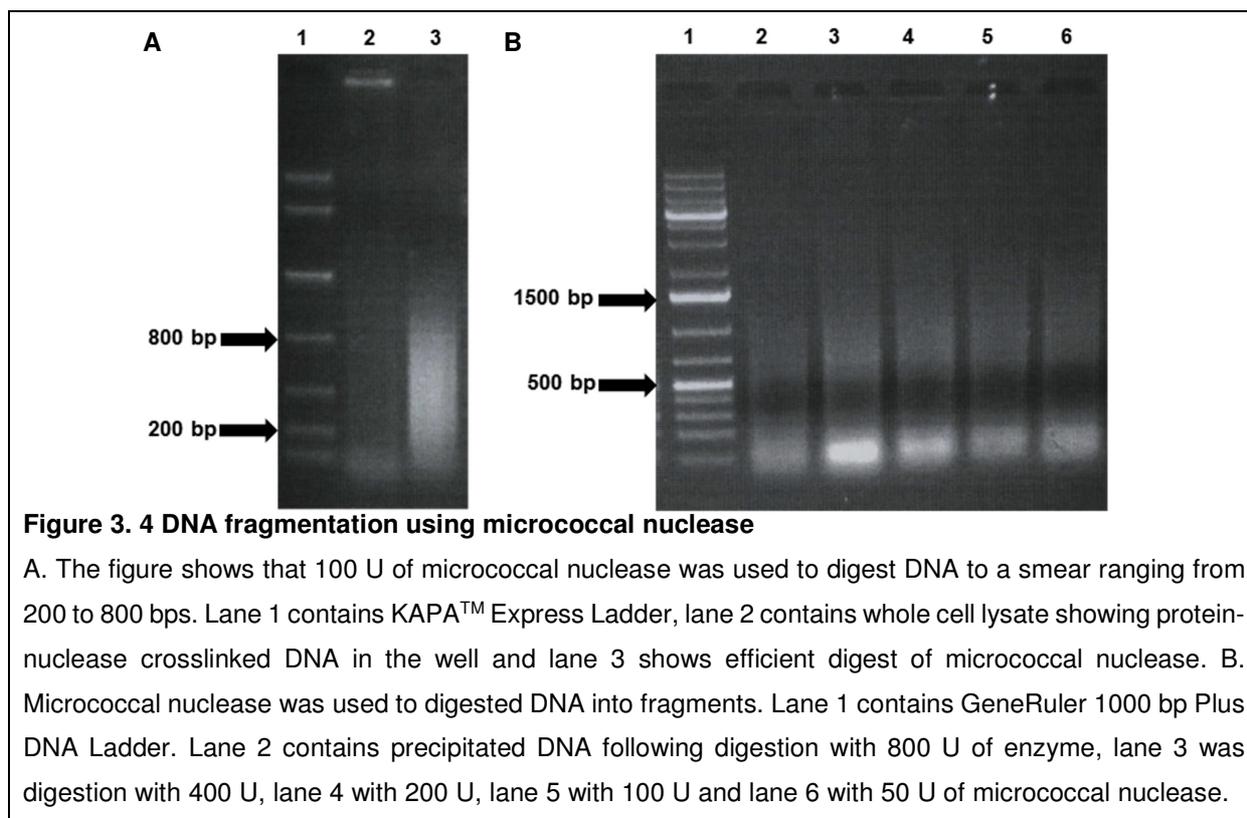
Prior to DNA precipitation, cross-linking of protein-DNA complexes were reversed by incubating magnetic beads in 100 µl elution buffer on a thermomixer for 1 hour at 65°C with shaking at 700 rpm. Proteinase K digestion was used to remove isolated proteins. Eluates were diluted with 100 µl nuclease free water before the addition of 50 µg proteinase K and an overnight incubation at 50°C. Sodium chloride – ethanol precipitation was used to precipitate isolated DNA. Briefly, a final concentration of 0.2 M NaCl and 420 µl 95% ethanol was added to each proteinase K digest to precipitate DNA. DNA precipitations were incubated on ice for 5 hours before centrifugation at 16 873xg for 30 minutes at room temperature. Tubes were inverted on tissue paper drained before washing with 500 µl 70% ethanol. Precipitations were washed by centrifuging at 16 873xg for 5 minutes before aspirating remaining ethanol. Pellets were allowed to dry at room temperature before being resuspended in 20 µl TE buffer. DNA was quantified using spectrophotometry as described in section 3.2.5.7 before separation using agarose gel electrophoresis as described in section 3.2.3.6.

### 3.3. Results

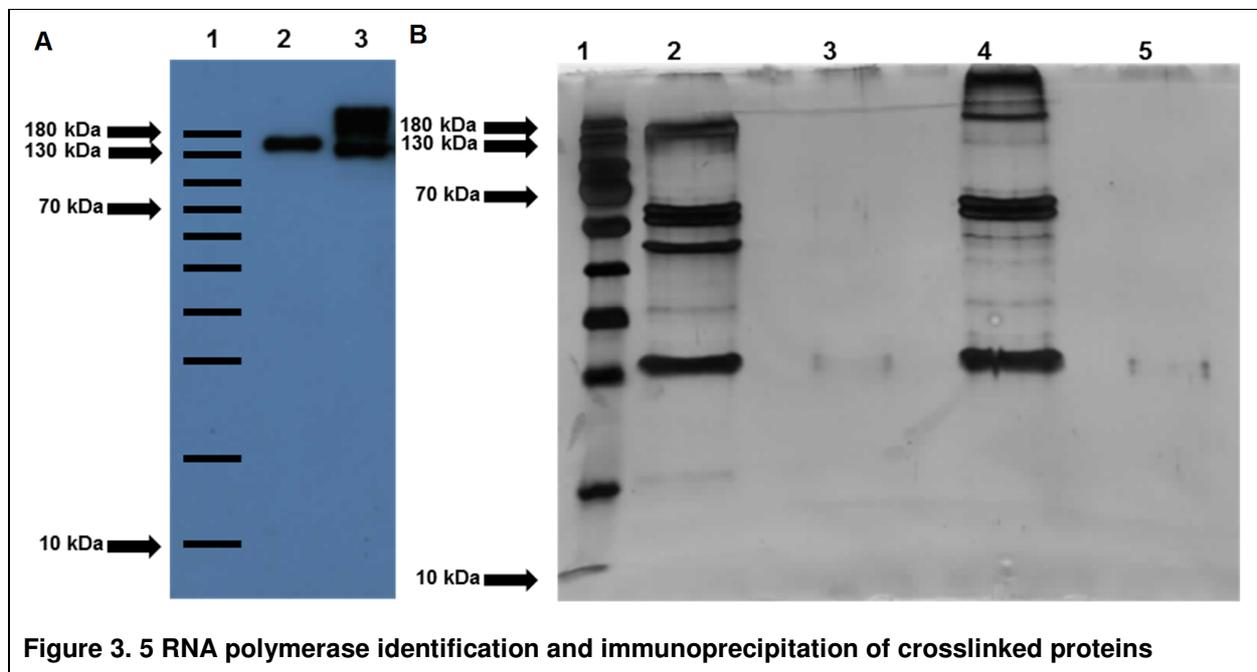
#### 3.3.1. Development of NP-MS

For the development for NP-MS, we made use of a ChIP-seq protocol previously described in a study investigating the *M. tuberculosis* transcriptome through the genome-wide mapping of RNA polymerase and NusA (62).

We first wished to demonstrate that formaldehyde was cross-linking DNA binding proteins to genomic DNA and that DNA was sufficiently fragmented to facilitate the immunoprecipitation of nucleoprotein complexes. Agarose gel electrophoresis was used to show that nucleoprotein complexes failed to migrate into the gel following formaldehyde cross-linking of *M. smegmatis*, whilst proteinase K treated and precipitated DNA from the same whole cell lysates demonstrated normal separation and a DNA smear (Figure 3. 4 A). A range of micrococcal nuclease concentrations was investigated to determine the optimum amount to yield genomic DNA fragments ranging from 200 - 800 bps (Figure 3. 4 B). One hundred units of enzyme was determined to be required for optimal fragmentation of genomic DNA in a total volume of 4 ml of crosslinked whole cell lysate (Figure 3. 4 A).



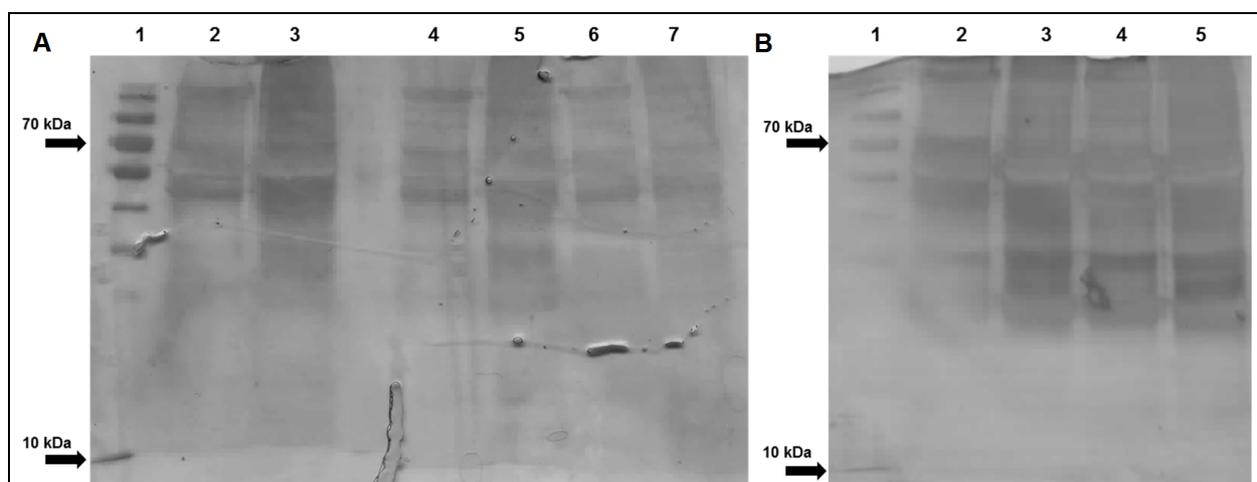
We next sought to demonstrate the ability of the anti-RNA polymerase antibody raised against the RNA polymerase  $\beta$ -subunit of *E. coli* to identify the RNA polymerase  $\beta$ -subunits in *M. smegmatis* (128.5 kDa) and *M. tuberculosis* (129.2 kDa). This was confirmed by Western blotting (Figure 3. 5). Following the identification of these homologue proteins in mycobacterial organisms using the anti-RNA polymerase  $\beta$ -subunit antibody, we wished to determine if nucleoproteins could be isolated using the anti-RNA polymerase antibody bound to protein G Dynabeads™ from formaldehyde cross-linked and untreated whole cell lysates. Using the ChIP-seq protocol as mentioned above and manufacturer's instructions, we used a total of 50  $\mu$ l of protein G Dynabeads™ per immunoprecipitation. As per manufacturer's instructions, we used 5  $\mu$ g of antibody bound to protein G Dynabeads™ per immunoprecipitation reaction as estimated by the requirement of 1  $\mu$ g of antibody per 500  $\mu$ g of whole cell lysate. We determined that approximately 2 - 2.5 mg of protein was present in 2ml of whole cell lysate using the RC-DC Assay (Bio-Rad).



**Figure 3. 5 RNA polymerase identification and immunoprecipitation of crosslinked proteins**

A. The western blot shows identification of the RNA polymerase  $\beta$ -subunit using the anti-RNA polymerase  $\beta$ -subunit antibody previously used in (62) for ChIP-seq. Lane 1 represents the sizes as present on the PVDF membrane for the PageRuler™ Prestained Protein Ladder. Lane 2 contains whole cell lysate from *M. tuberculosis* H37Rv and lane 3 contains whole cell lysate from *M. smegmatis*. B. The silver stained SDS-PAGE shows that proteins other than the RNA polymerase  $\beta$ -subunit is isolated from immunoprecipitations done on cross-linked *M. smegmatis* whole cell lysates. Lane 1 contains PageRuler™ Prestained Protein Ladder, lanes 2 and 3 were shows immunoprecipitation of untreated *M. smegmatis* whole cell lysates and lanes 4 and 5 shows immunoprecipitations for crosslinked *M. smegmatis* whole cell lysates. Lanes 2 and 4 shows proteins isolated using anti-RNA polymerase  $\beta$ -subunit antibody bound protein G Dynabeads™ and lanes 3 and 5 show protein G empty bead controls.

Following confirmation of the chosen approach to isolate proteins other than the RNA polymerase complex from formaldehyde crosslinked *M. smegmatis* cultures (Figure 3. 5), we sought to determine the optimal amount of time required for the immunoprecipitation of nucleoprotein complexes. To this end, antibody bound protein G Dynabeads™ were incubated with whole cell lysates obtained from uncross-linked and cross-linked *M. smegmatis* cultures for 2, 6, 8, 10 and 16 hours. Two hours was selected for the immunoprecipitation of nucleoprotein complexes as the amount of non-specific binding was seen to increase in non-cross-linked *M. smegmatis* immunoprecipitations after 10 and 16 hours when compared to 2 hours. Furthermore, no increase in the amount of immunoprecipitated protein was observed from cross-linked whole cell lysates when comparing immunoprecipitations for timepoints 2 to 8 hours (Figure 3. 6).

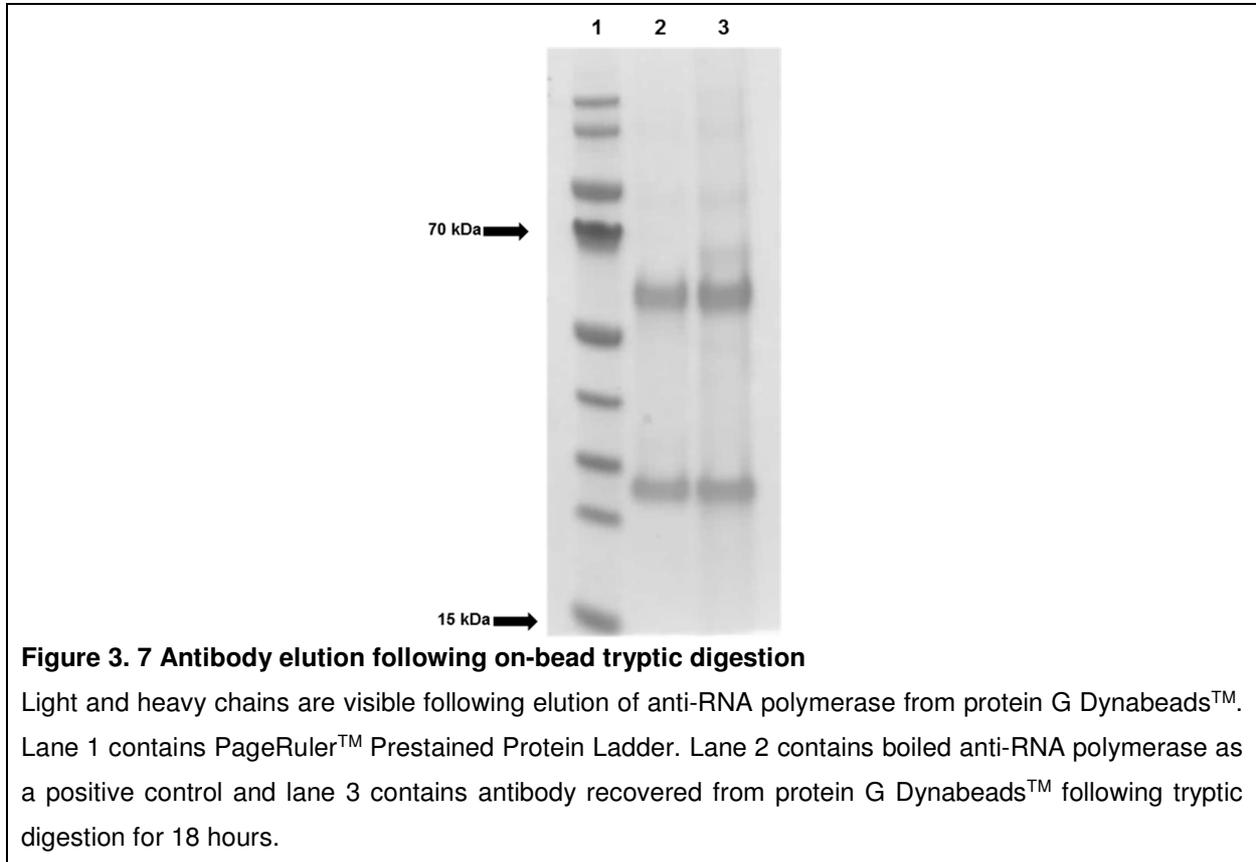


**Figure 3. 6 Immunoprecipitation time optimisation**

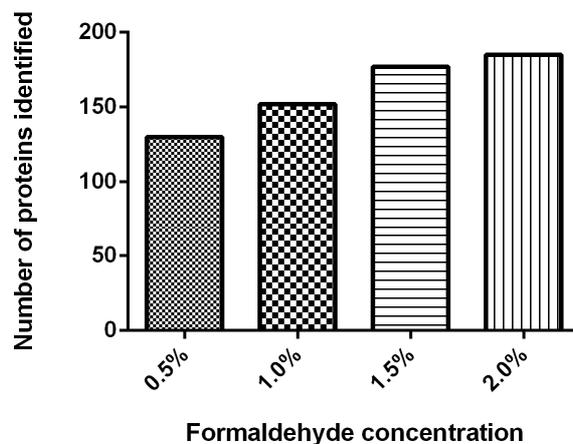
Silver stained SDS-PAGE gels show immunoprecipitations using anti-RNA polymerase bound protein G Dyabeads™. A. Lane 1 contains PageRuler™ Prestained Protein Ladder and lanes 2, 4, and 6 contains immunoprecipitations from uncross-linked *M. smegmatis* for hours 2, 6, and 8, respectively. Lanes 3, 5, and 7 contains immunoprecipitations from crosslinked *M. smegmatis* for hours 2, 6, and 8. B. Lane one contains PageRuler™ Prestained Protein Ladder and lanes 2 and 4 contains immunoprecipitations from uncross-linked *M. smegmatis* for hours 10 and 16. Lanes 3 and 5 contains immunoprecipitations from crosslinked *M. smegmatis* for hours 10 and 16.

Due to the overwhelming presence of the light and heavy chains of the anti-RNA polymerase  $\beta$ -subunit antibody when elution was performed by boiling in SDS reducing sample buffer (Figure 3. 5), on-matrix tryptic digestion was investigated for the elution of proteins as peptides. Mouse-derived IgG antibodies have previously been suggested to be resistant to tryptic digestion and to demonstrate that the same was true for our mouse-derived IgG antibody, beads recovered following tryptic digestion were boiled in SDS reducing sample buffer prior to separation using SDS-PAGE (Figure 3. 7) (141, 142). These results

demonstrated that on-bead tryptic digestion would not contaminate the resulting peptide mixture with the antibody used to isolate the RNA polymerase transcriptional complex and its associated proteins.



We next sought to determine the optimal formaldehyde concentration required for the isolation and identification of the RNA polymerase transcriptional complex and its associated proteins. *M. smegmatis* cultures were cross-linked for 10 minutes with 0.5%, 1.0%, 1.5%, and 2.0% formaldehyde (final concentration) at 37°C with gentle shaking as described in (62). Immunoprecipitations were done in biological duplicate and mass spectrometry analysis revealed that the number of protein identifications increased with higher concentrations of formaldehyde (Figure 3. 8). A search for nucleic acid associated proteins using the identifiers nucleo-, DNA, RNA, transcription, translation, ribosome, nucleic acid, chromosome, and ligase in each of these protein identification lists revealed that increased concentrations of formaldehyde resulted in the increased identification of proteins not associated with nucleic acids (Table 3. 7). These results suggested that increased formaldehyde concentrations may result in the crosslinking of proteins interacting with the targeted RNA polymerase transcriptional complex and its accompanying nucleic acid-associated proteins.



**Figure 3. 8 Formaldehyde crosslinking concentration**

Increasing concentrations of formaldehyde resulted in increased protein identifications following immunoprecipitation.

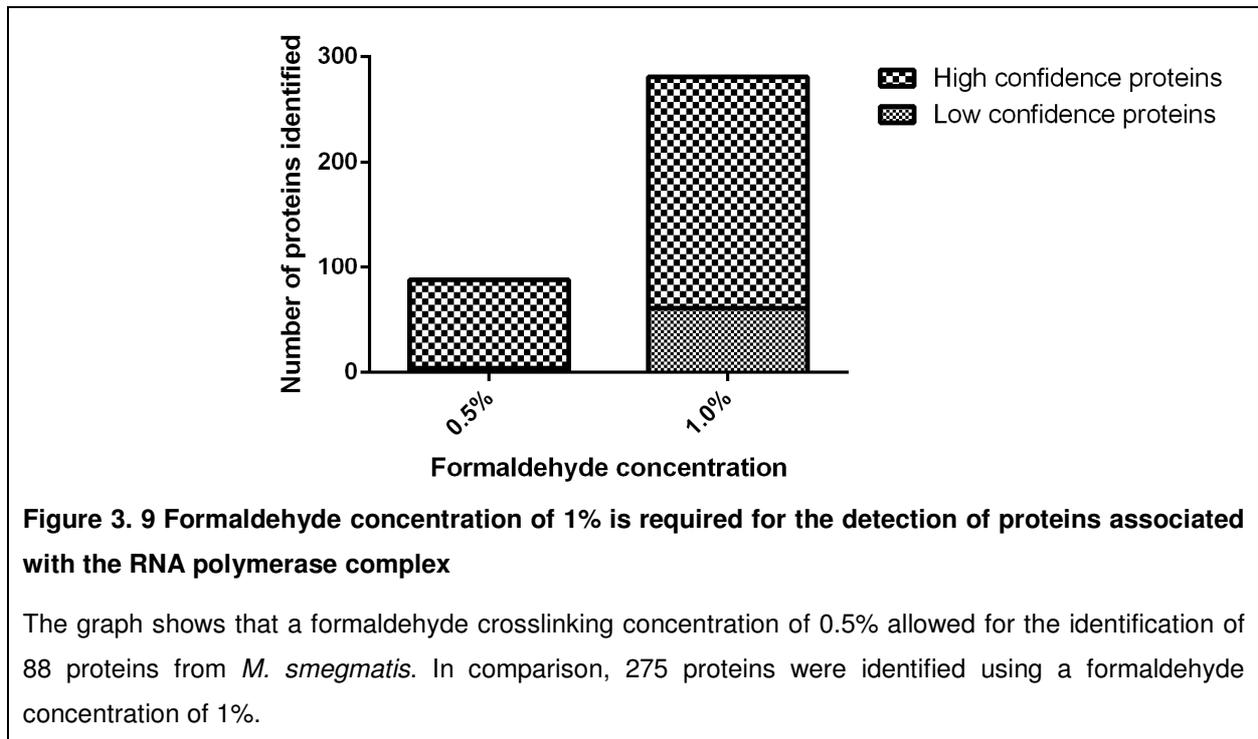
**Table 3. 7 Percentage nucleic acid-associated proteins identified with different formaldehyde concentrations**

Formaldehyde concentration	% Nucleic acid associated proteins
0.5%	58.0% (76/131)
1.0%	50.6% (77/152)
1.5%	50.2% (89/177)
2.0%	48.1% (89/185)

To avoid the identification of proteins not associated with the RNA polymerase complex and its associated proteins through increased formaldehyde concentrations, we decided to further investigate the identification of proteins using the formaldehyde crosslinking concentrations 0.5 and 1%. Furthermore, for these immunoprecipitations we included an empty bead control as well as an unrelated antibody control (anti human myosin antibody) which allowed us to identify non-specific protein identifications. Proteins from either experiment were identified as high or low confidence proteins depending on the identification of non-specific proteins in the negative controls. High confidence proteins were identified when a protein was present in at least two of the three anti-RNA polymerase immunoprecipitations but not present in any of the control immunoprecipitations. Following the removal of high confidence proteins from the list, missing abundance values for low confidence protein candidates were replaced using imputation. A subsequent ANOVA was performed with an FDR of 0.05 using the Benjamini-Hochberg correction to identify proteins which were significantly more abundant in the anti-RNA polymerase immunoprecipitations than in the

controls, suggesting that these proteins might also be associated with the RNA polymerase transcriptional complex.

A final concentration of 1% was selected for all future experiments given the low number of identified proteins in the 0.5% formaldehyde crosslinked experiment. Although the percentage of proteins associated with the searched identifiers mentioned above decreased from 48% to 41.9% when comparing identified proteins from the 0.5% to the 1% formaldehyde cross-linked immunoprecipitations, the number in proteins identified from the 1% cross-linked experiment were more than three times that of the 0.5% cross-linked experiment (Figure 3. 9).

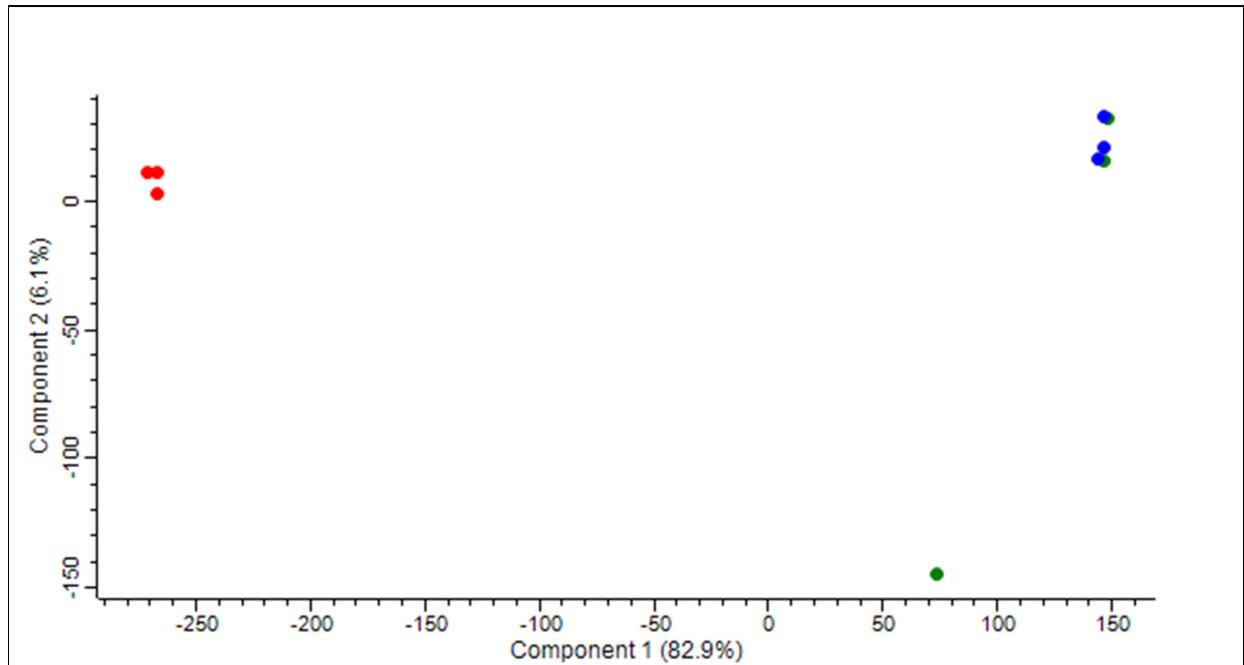


### 3.3.2. Identifying the Nucleic Acid Associated Proteins in *M. smegmatis*

To determine whether NP-MS was capable of the isolation and subsequent identification of the proteins which make up the RNA polymerase complex as well as its associated proteins, we prepared three biological replicate experiments using the developed approach. Two protein identification lists were assembled following a very stringent and a less stringent protein identification approach (Figure 3. 1). High stringency protein identification data can be found in Addendum B.3. and lower stringency data will be discussed in this thesis.

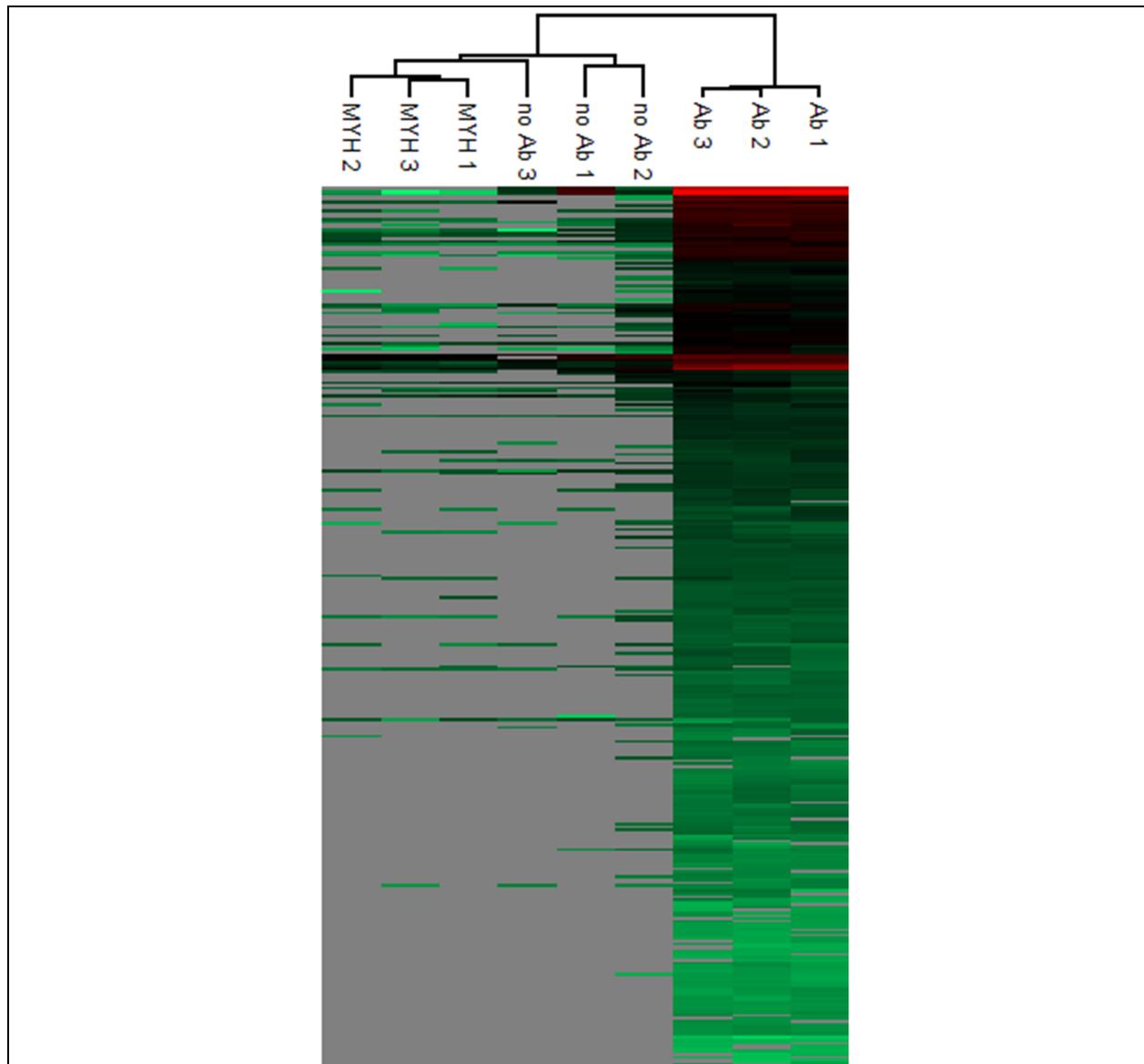
Proteins were deemed present when identified in at least two of the three biological replicate experiments with at least two unique peptides. Hierarchical clustering of columns and rows, displayed as a heatmap, as well as a two-dimensional principal component analysis (2D PCA) was done to show that anti-RNA polymerase immunoprecipitations were reproducible. The 2D PCA (Figure 3. 10) and the heatmap (Figure 3. 11) showed that biological replicate immunoprecipitations targeting the  $\beta$ -subunit of the RNA polymerase complex clustered together, separate from control immunoprecipitations (immunoprecipitations without antibody and an anti-myosin human antibody). The grey colouring in the heatmap indicates the absence of protein identifications within individual immunoprecipitations, suggesting that the wash buffers used in this study efficiently reduced the number of non-specific bound protein identifications.

We went on to identify proteins isolated using NP-MS. Proteins identified in protein G Dynabead™ and the anti-myosin human antibody control immunoprecipitations were treated as protein contaminants isolated during immunoprecipitation. Two hundred and twenty high confidence protein identifications were made when a protein was present in at least two of the three biological replicate experiments, but not in any of the immunoprecipitation controls (Figure 3. 1). Following the removal of high confidence proteins from the dataset we wished to identify low confidence proteins which were isolated in the anti-RNA polymerase immunoprecipitations but were also detected within the control immunoprecipitations. For the identification of low confidence proteins, the remaining data was imputed, separately for each immunoprecipitation using the Gaussian distribution, to replace the missing abundance values in control immunoprecipitations and the missing value from one of the three biological replicate experiments. A heatmap of imputed data was generated to show that anti-RNA polymerase immunoprecipitations clustered together and separately from control experiments (Figure 3. 12). A multiple sample test (ANOVA) was done between immunoprecipitation groups (with anti-RNA polymerase  $\beta$ -subunit antibody, without antibody and anti-myosin antibody) using the Benjamini-Hochberg correction with an FDR of 0.05 to identify proteins which were significantly more abundant within the anti-RNA polymerase immunoprecipitations than within the control immunoprecipitations. Following statistical testing, 61 low confidence proteins were identified with  $p$ -values less than 0.05. Manual inspection of spectra of proteins identified with 2 unique peptides resulted in the removal of six high confident protein identifications. In combination, a total of 275 proteins were identified using NP-MS (Table B. 1).



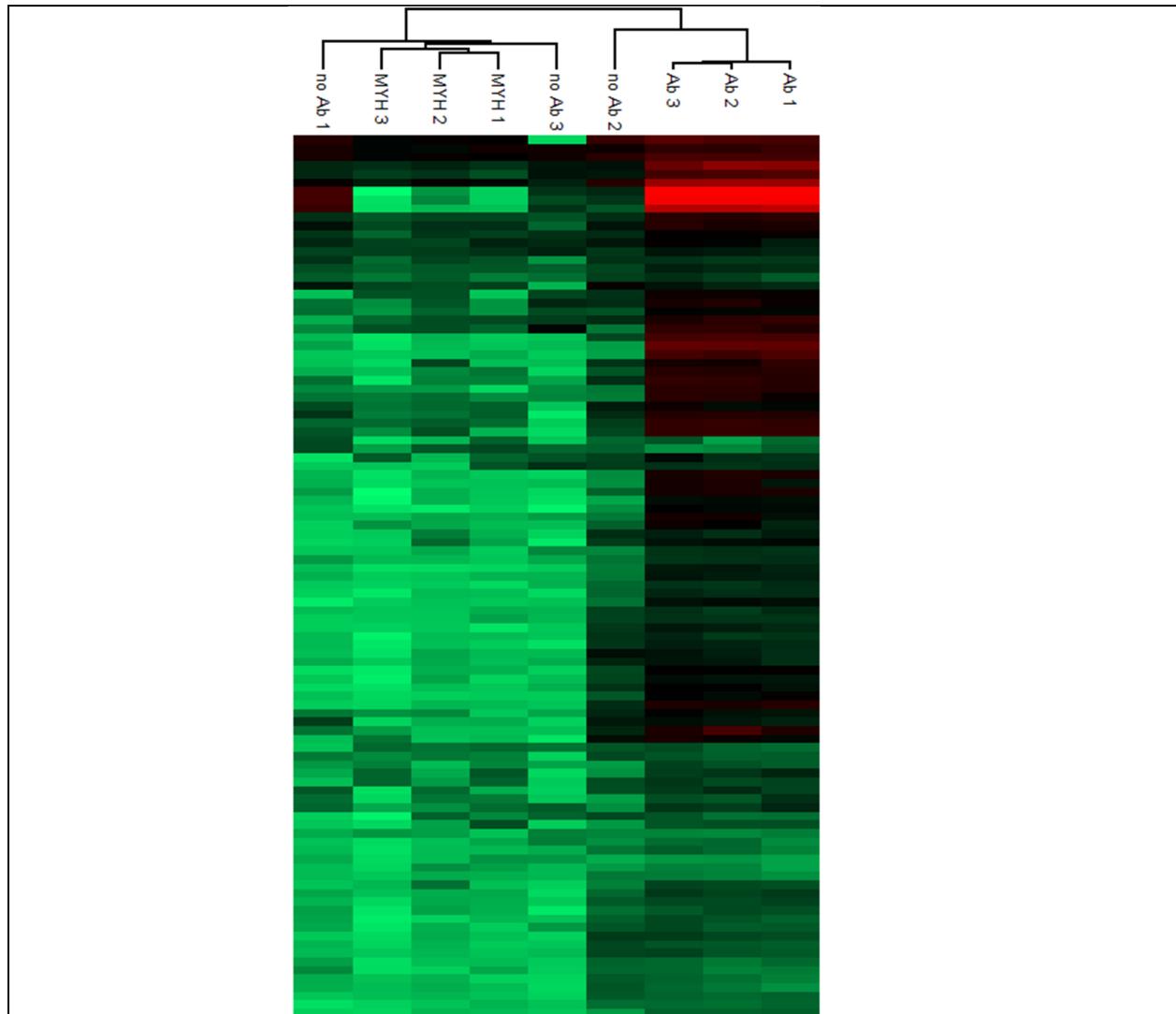
**Figure 3. 10 2D PCA of NP-MS LFQ Intensity data**

The 2D PCA shows that anti-RNA polymerase immunoprecipitations (red) clustered left, and separately from anti- human myosin (blue) and no antibody (green) control immunoprecipitations on the right.



**Figure 3. 11 Heatmaps representing less stringent NP-MS data**

The heatmap shows the clustering of the anti-RNA polymerase immunoprecipitation on the right and control immunoprecipitations left and center. The red colour is indicative of a higher abundance and the green of a lower abundance of a protein within an immunoprecipitated sample. The grey colouring within the heatmap suggests the absence of a protein within that immunoprecipitation.



**Figure 3. 12 Heatmap representing imputed low confidence protein data**

The heatmap shows imputed low confidence protein identification data following the removal of high confidence proteins from the dataset. The graph shows clustering of low confidence anti-RNA polymerase protein identifications on the right with control data clustering left and center.

To estimate the relative proportion of nucleic acid associated proteins isolated through the targeting of the RNA polymerase transcriptional complex, we made use of protein names and gene ontologies to describe identified proteins to search for the following identifiers: nucleo-, DNA, RNA, transcription, translation, ribosome, translation, nucleic acid, chromosome, and ligase. This approach revealed that 41.9% (115/281) proteins identified using NP-MS were nucleic acid associated proteins. These identified proteins included members of the RNA polymerase complex other than the RNA polymerase  $\beta$ -subunit such as rpoC, rpoD, and rpoZ. Several known DNA-associated proteins were identified which included DNA gyrase subunits A and B, DNA topoisomerase, DNA helicase, transcriptional regulators (e.g. MtrA, Crp, Rho, and WhiA) and NAPs (EspR, mIHF and HupB). RNA associated proteins identified included ribosomal proteins (50S and 30S subunits), translation initiation factors (InfA, infC), elongation factors (Efp, TypA, and TuF) and sigma factors (SigH and MysB).

To further investigate the possible processes in which NP-MS identified proteins are involved in, we made use of the pathway mapping tool KEGG. Eighty-two of the proteins identified using NP-MS mapped to metabolic pathways (Figure 3. 13). These pathway maps demonstrated that several proteins identified were associated with energy, lipid, carbohydrate, amino acid, and nucleic acid metabolism. These results demonstrated that during the enrichment of the RNA polymerase complex, various other energy metabolism proteins, possibly required by the RNA polymerase complex, are also isolated due to formaldehyde cross-linking and their close proximity to the complex.



### 3.3.2. Enrichment of Gene Ontologies Associated with Nucleic Acid Associated Proteins

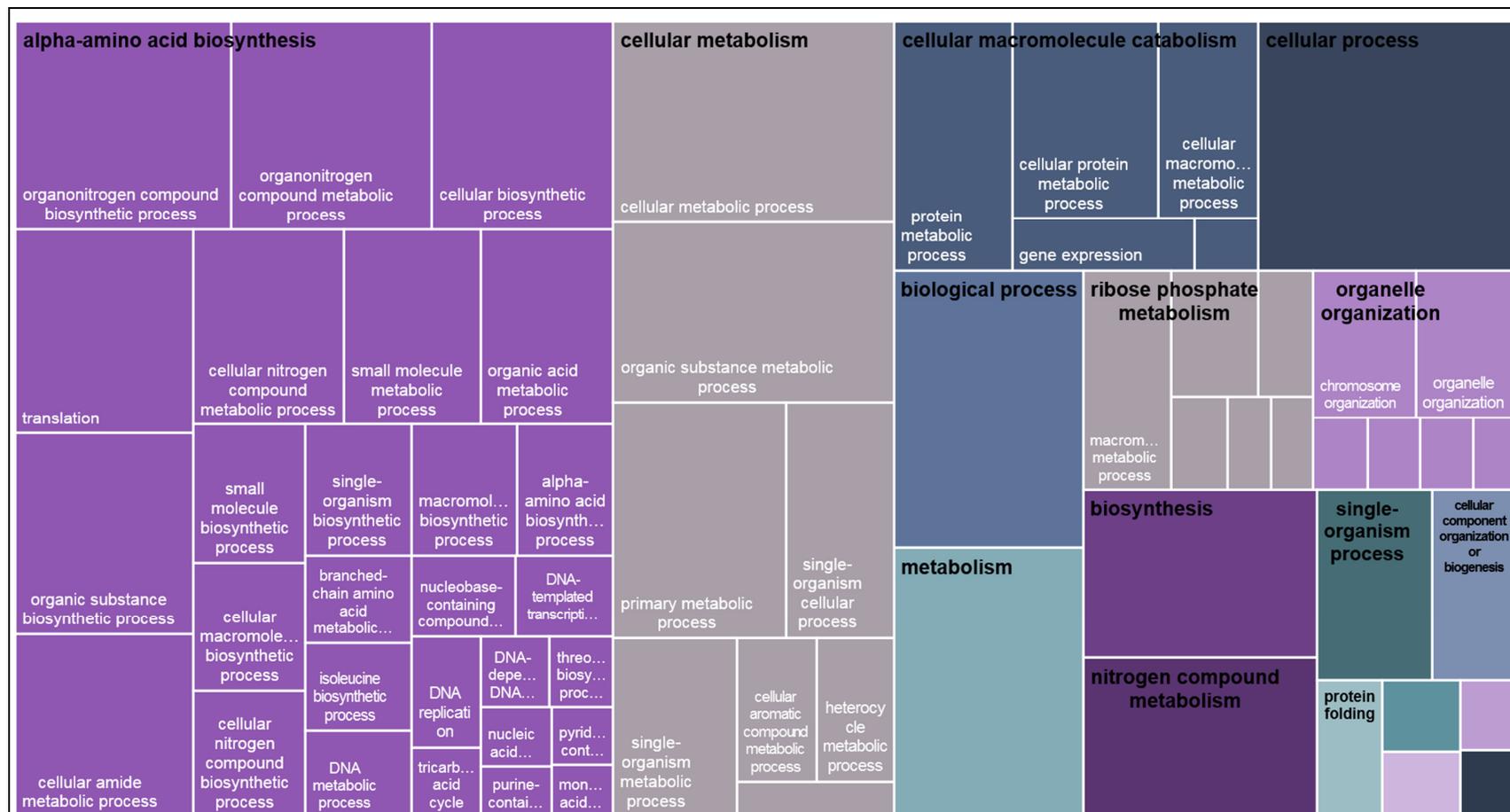
GO identities are used for the consistent description of genes and gene product attributes across three biological domains; molecular function, biological processes and cellular components, which are shared by all organisms (143, 144). GO identities used to describe the predicted functional attributes of proteins identified in this study was obtained from UniProt (Table B. 2) and we ascribed 343 unique GO identities to 275 identified proteins. Gene ontology enrichment analysis of identified proteins was done using GOEAST. GOEAST determines the significance of an input ID through hypergeometric testing, which is often used to identify sub-populations which are over- or under-represented within a sample (135). Gene ontology enrichment analysis revealed that 187 of the 343 uniquely identified GO terms were significantly enriched (Table B. 3). Enriched GO identities and their respective  $p$ -values were subjected to further analysis using REVIGO, which summarizes and visualizes long lists of GO terms by reducing the number of redundant GO identities (136). A similarity value of 0.7 was chosen for GO identities and the simRel similarity measure was selected for the analysis. SimRel is a functional similarity measure for comparing two GO identities with each other where a score of 0 is indicative of no similarity and 1 of a maximum similarity. REVIGO classified 104 of the enriched GO identities under the functional category biological processes, 60 as molecular function and the remaining 23 GO identities as cellular components.

Hierarchical clustering using treemaps was done to visualize the abundance of GO identities across three functional categories, biological processes (Figure 3. 14), molecular function (Figure 3. 15) and cellular components (Figure 3. 16). Treemaps display hierarchical data as a set of nested rectangles. Each rectangle represents a lower hierarchy GO term (a child GO term) and multiple rectangles of the same colour form a larger rectangle, representative of a higher hierarchy GO term (a parent GO term). Treemaps were assembled using the  $-\log_{10} p$ -values, which is indicative of the significance of each statistically relevant GO identity within its respective category.

Hierarchical clustering for the functional category biological processes revealed that the parent GO identity amide biosynthesis was most enriched (Figure 3. 14). This parent GO term, alpha-amino acid biosynthesis, umbrellas various child GO identities involved in DNA and RNA related processes such as translation, DNA replication, DNA templated transcription, DNA metabolic process, nucleobase containing compound metabolic process, branched chain amino acid metabolic process and translation. Other parent GO identities found to be enriched were cellular metabolism, cellular macromolecule catabolism, ribose phosphate metabolism, organelle organisation and nitrogen compound metabolism. Within the functional category molecular function, the parent GO identity, coenzyme binding, was identified as being the most enriched and ranked above the child GO terms nucleic acid binding, heterocyclic compound binding and organic cyclic compound binding which can all be associated with nucleic acids (Figure 3. 15). Other parent GO identities enriched within this functional category included manganese ion binding, rRNA binding, DNA-

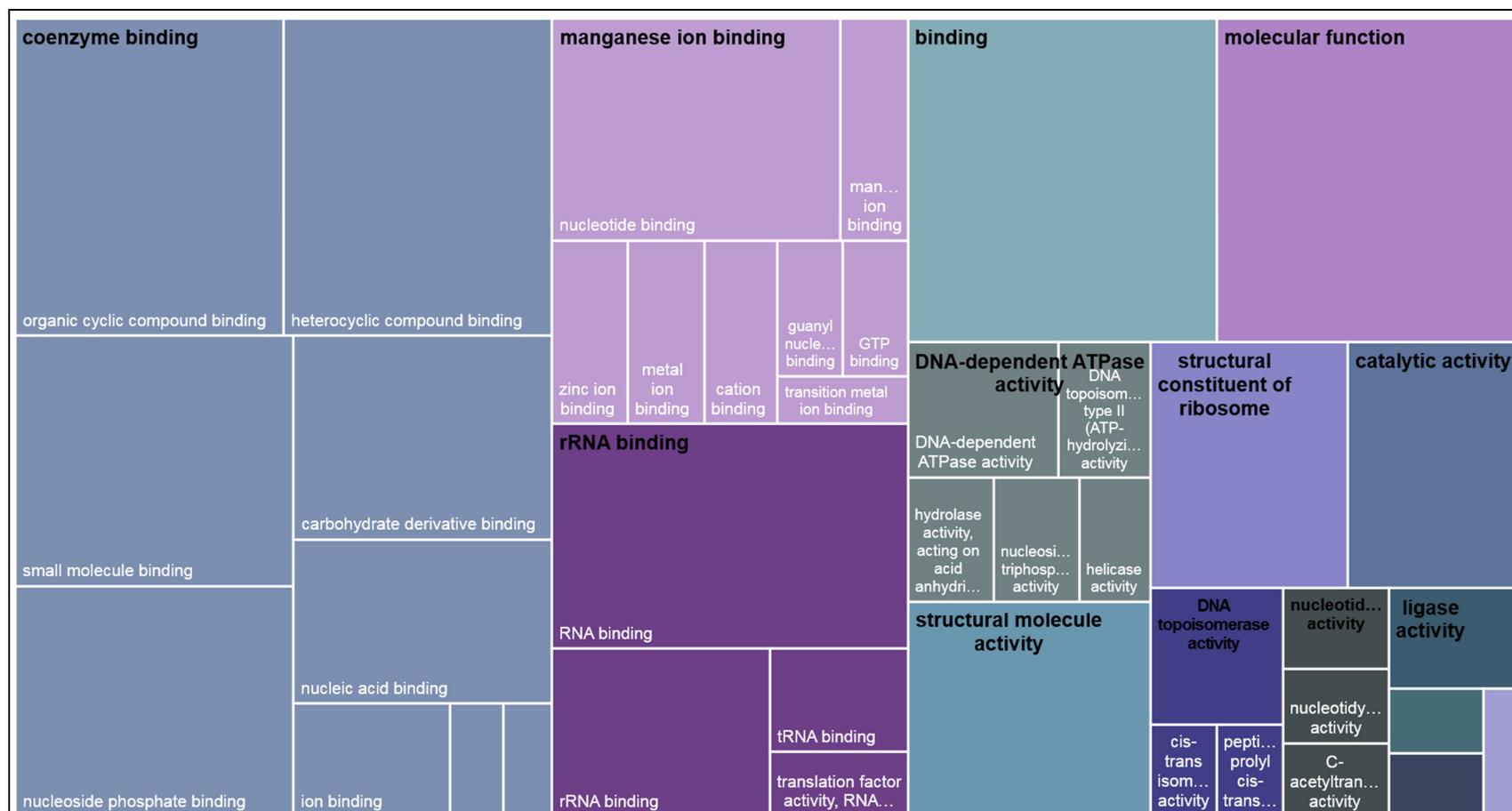
dependet ATPase activity, structural constituent of the ribosome, nucleotidyl transferase activity and DNA topoisomerase activity. Lastly, for the functional category cellular component, the higher hierarchy GO identity ribosome was shown to be the most abundant within this functional category and umbrellas the child GO identities ribonucleoprotein complex and chromosome which can all be associated with nucleic acids (Figure 3. 16). The absence of GO identities associated with the cell membrane such as integral component of membrane (GO:0016021) and anchored component of membrane (GO:0031225) implies that NP-MS is effective in the isolation of intracellular proteins.

These treemaps visually represent enriched GO identities and their predicted relationships. These graphs suggest that the proteins isolated using NP-MS have functional attributes which can be associated with the RNA polymerase transcriptional complex in *M. smegmatis* either through transcription or translation. Furthermore, the enrichment of GO identities associated with various metabolic processes (Figure 3. 14) correlates to what was shown with KEGG pathway mapping. These results are suggestive of proteins being isolated and identified which may be required for energy metabolism processes necessary by the RNA polymerase complex for transcription.



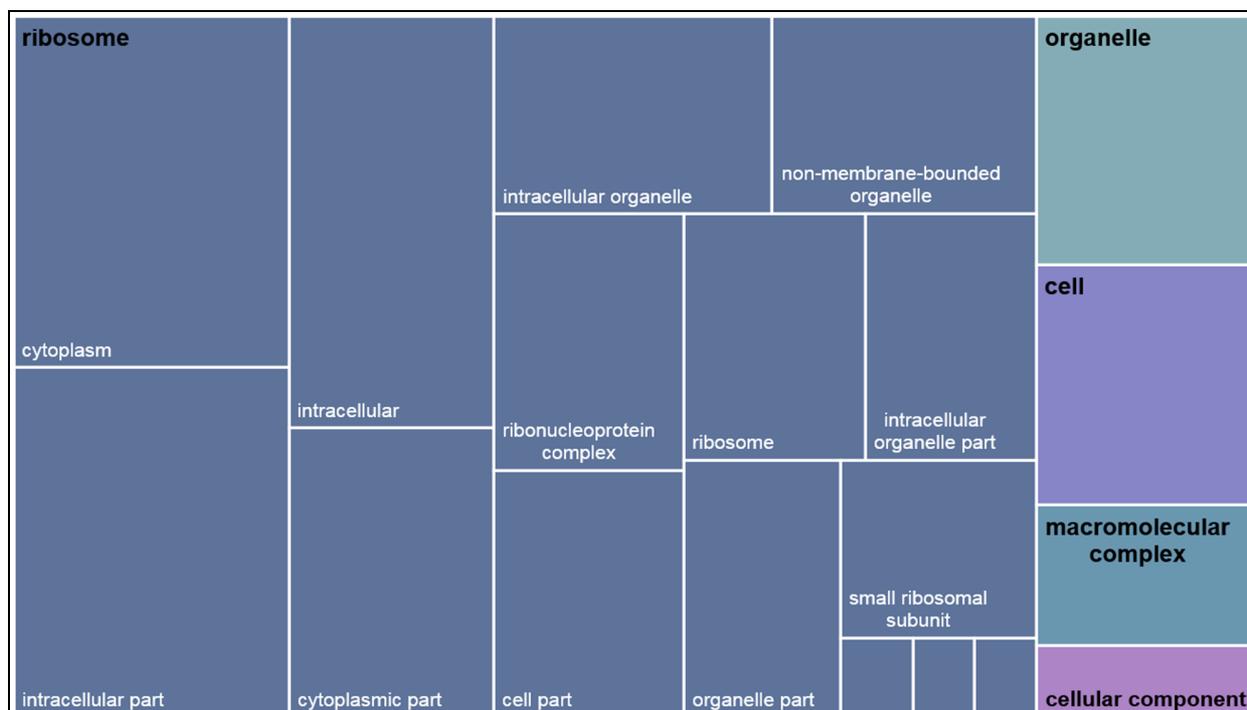
**Figure 3. 14 Hierarchical clustering of enriched biological processes GO identities**

Hierarchical clustering of biological processes GO identities shows higher hierarchy GO terms in black and lower hierarchy GO terms in white. The majority of the treemap is occupied by GO identities associated with alpha-amino acid biosynthesis. Other represented higher hierarchy GO identities include cellular metabolism, cellular macromolecule catabolism, ribose phosphate metabolism and organelle organization.



**Figure 3. 15 Hierarchical clustering of enriched molecular function GO identities**

The treemap of enriched molecular function GO identities shows higher hierarchy GO terms in black and lower hierarchy GO identities in white. The figure shows the enriched representation of GO identities involved in coenzyme binding. Other higher hierarchy enriched GO terms include manganese ion binding, rRNA binding, DNA-dependent ATPase activity, DNA topoisomerase activity and structural constituent of the ribosome.



**Figure 3. 16 Hierarchical clustering of enriched cellular component GO identities**

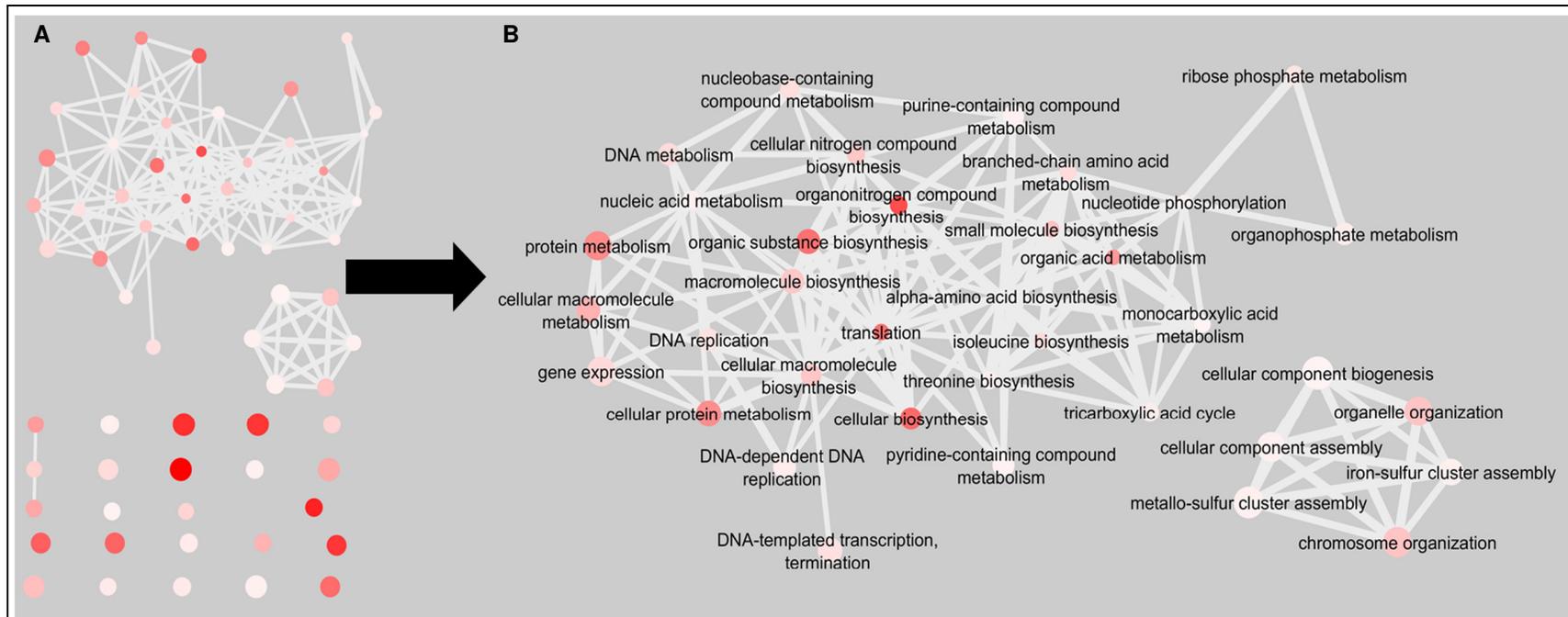
The figure shows enriched GO identities associated with the functional category cellular components. Higher hierarchy GO terms are displayed in black and lower hierarchy GO terms in white. The treemap shows the enrichment of GO terms associated with the higher order GO identity ribosomal subunit.

### 3.3.3. Gene Regulatory Networks of *M. smegmatis*

To better understand the relationship between the GO identities used to describe the proteins identified using NP-MS, we used REVIGO (136). Cytoscape was used to generate graphs for each of the functional categories, biological processes (Figure 3. 17), molecular function (Figure 3. 18) and cellular components (Figure 3. 19).

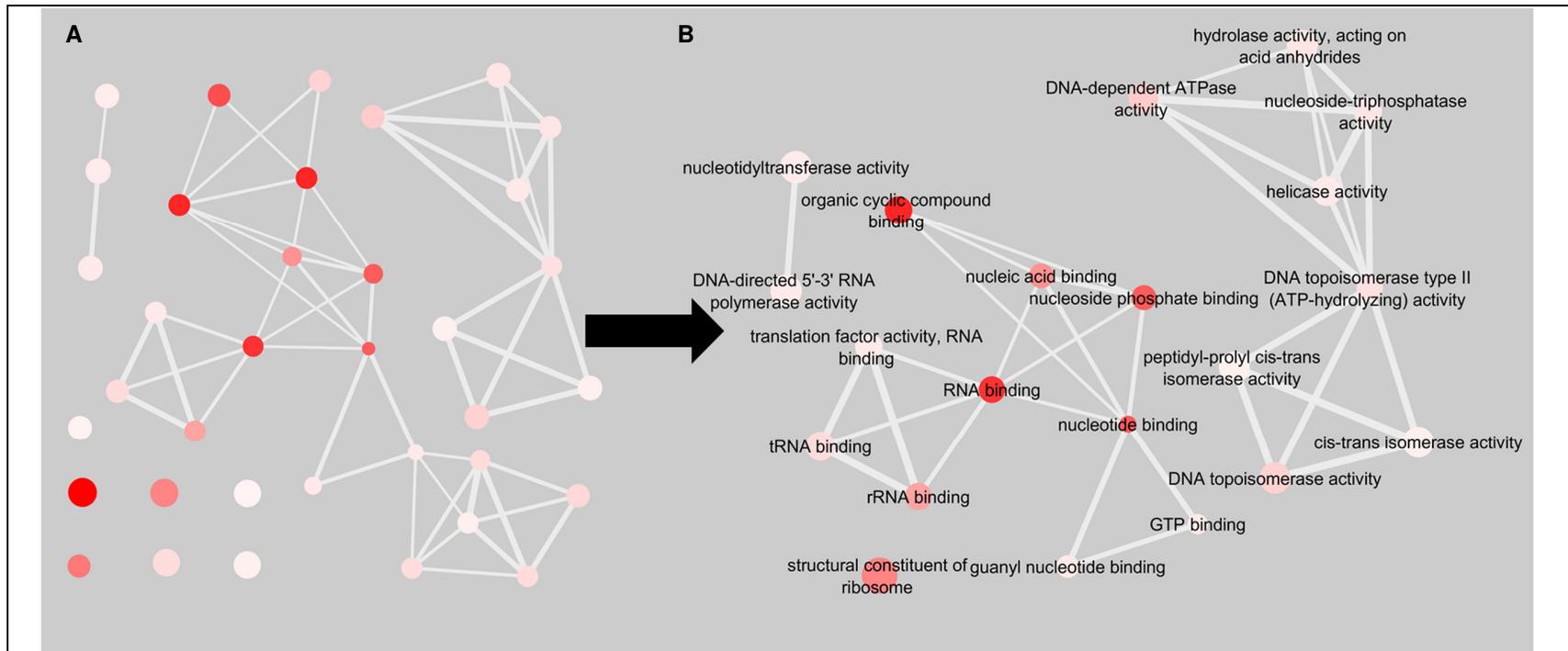
Interactive graphs generated by Cytoscape display enriched GO identities as nodes, with edges connecting interacting nodes suggesting a degree of similarity between the enriched GO terms. These interactive graphs demonstrate that some of the enriched GO identities associated with each of the functional categories share a degree of similarity, thereby suggesting that the proteins which they represent have similar functional attributes and may be representative of a protein network. Through the searching of identifiers such as DNA, chromosome, transcription, gene, RNA, ribosome, translation, and nucleotide, we demonstrate the GO identities associated with functional attributes expected to be used for the description of nucleic acids, create a subnetwork of interacting GO identities. These results, therefore, suggest that NP-MS enriched for a network of proteins which can be described with GO identities associated with nucleic acid associated proteins.

To further explore the possibility that NP-MS could be used to isolate a network of nucleic acid associated proteins, we used the STRING database to generate a predicted protein association network for the 275 proteins identified in this study (138). For the generation of the protein association network, we selected the experimental prediction method with a medium confidence score of 0.4. The data used by STRING for the construction of the predicted protein-protein interaction network was obtained from experimental data in *M. smegmatis*, *Saccharomyces cerevisiae*, and *Homo sapiens*. Network data generated by STRING was exported to Cytoscape for network visualization (Figure 3. 20). Four hundred and forty-seven possible protein-protein interactions were identified for 149 of the 275 identified proteins. The protein association network generated by STRING suggested that a network of nucleic acid associated proteins was being isolated using NP-MS.



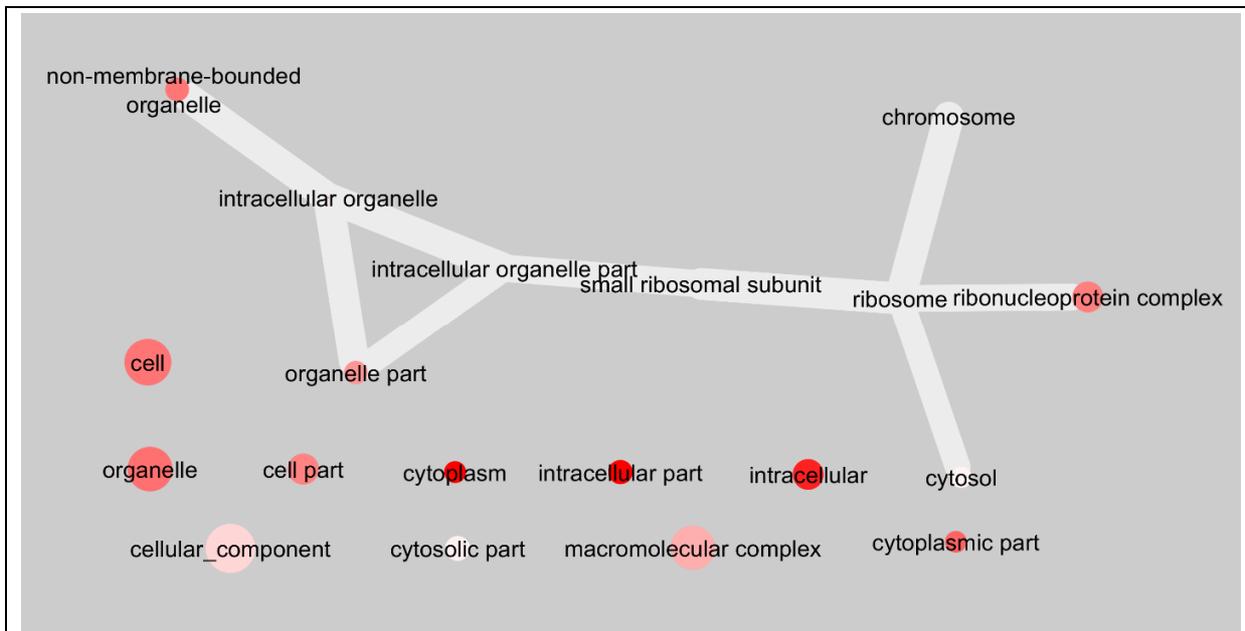
**Figure 3. 17 Network representation of biological processes GO identities**

The figure shows the GO identities associated with biological processes as a network. Each GO identity is represented as a node and interactions between nodes represent similarity between GO identities. The size of each node represents the frequency of GO identities within the data and the colour is indicative of the  $-\log_{10} p$ -value (the darker the more significant the enrichment). All GO terms associated with biological process are depicted in A. B shows a subnetwork generated by searching for GO IDs associated with DNA, chromosome, transcription, gene, RNA, ribosome, translation, nucleotide as well as their first connecting neighbours.



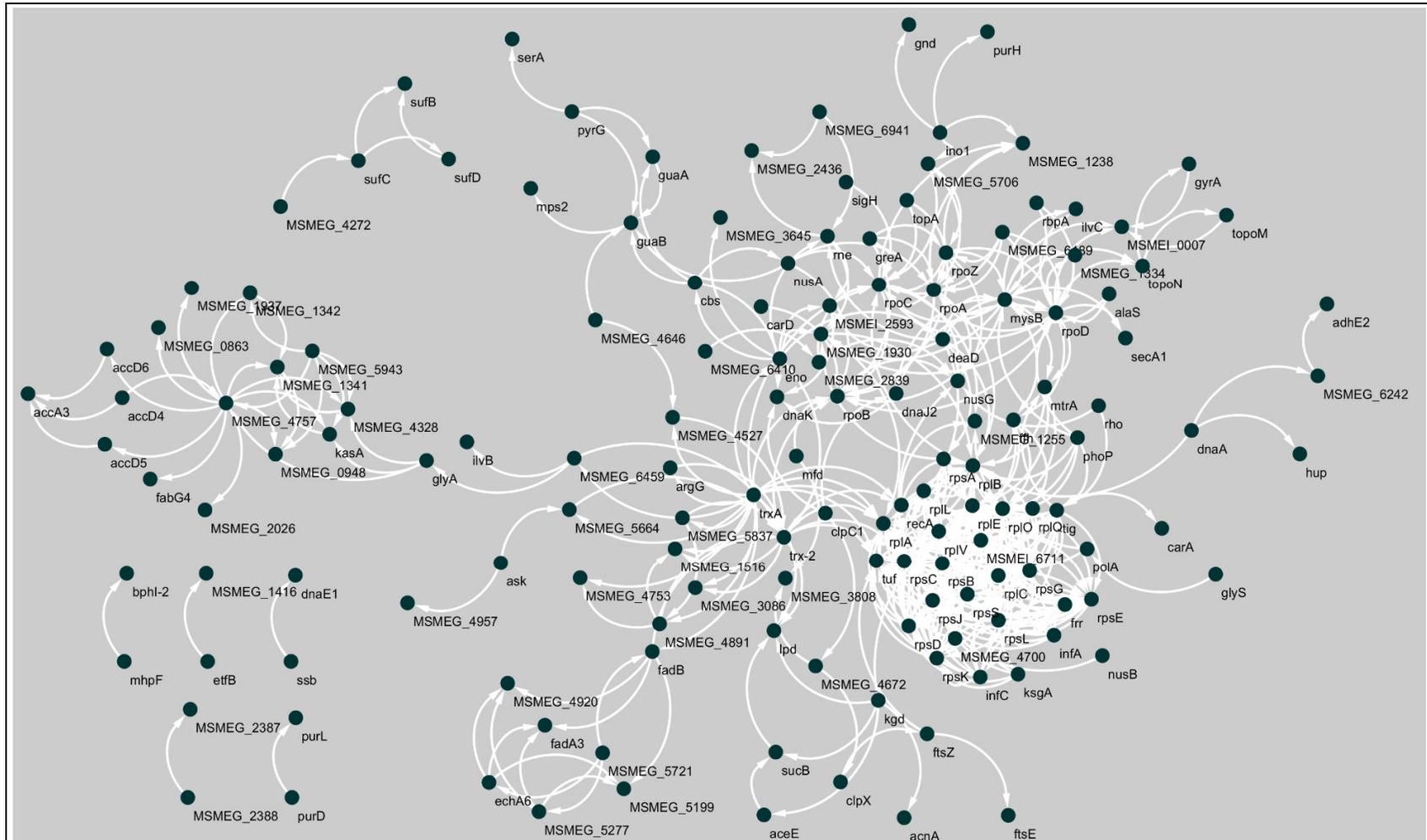
**Figure 3. 18 Network representation of molecular function GO identities**

The figure displays GO identities associated with molecular function as nodes with edges representing similarity between GO identities. Node size is indicative of GO identity frequency and the colour intensity is representative of the  $-\log_{10} p$ -value. All enriched GO identities are shown in A and B displays a subnetwork generated through searching for the identifiers DNA, chromosome, transcription, gene, RNA, ribosome, translation, nucleotide and their first connecting neighbours.



**Figure 3. 19 Network representation of cellular component GO identities**

The figure shows GO identities associated with cellular components as an interactive network. Enriched GO identities are represented by nodes, and similar GO identities are connected through edges. Node size shows the frequency of a GO identity within the sample and the colour intensity represents the  $-\log_{10} p$ -value.



**Figure 3. 20 Protein-protein interaction network of NP-MS identified proteins**

The figure shows a network of predicted protein-protein interactions of proteins identified using NP-MS. The STRING database was used to generate an interactive network of protein-protein interactions with arrows indicating the direction of the predicted interaction.

### 3.3.4. DNA Association of Isolated Proteins

To further validate the ability of NP-MS to accurately identify nucleic acid-associated proteins, we performed a DNA association assay. Selected proteins were cloned into an episomal expression vector, pNFLAG, from which these proteins would be expressed as N-terminal FLAG-tagged proteins. The NAP, HupB (MSMEG\_2389), was selected as a positive control for DNA binding and the cytosolic component of the ESX-3 secretion system, the *esx* conserved component A (*eccA*<sub>3</sub>, MSMEG\_0615) was selected as a negative control for nucleic acid associated proteins.

Five proteins were selected for DNA association validation studies, namely MSMEG\_1060, MSMEG\_2695, MSMEG\_3754, MSMEG\_4306, and MSMEG\_5512 through a comparison of conserved protein domains which have demonstrated affinity for nucleic acids in other well studied organisms (Table 3. 8).

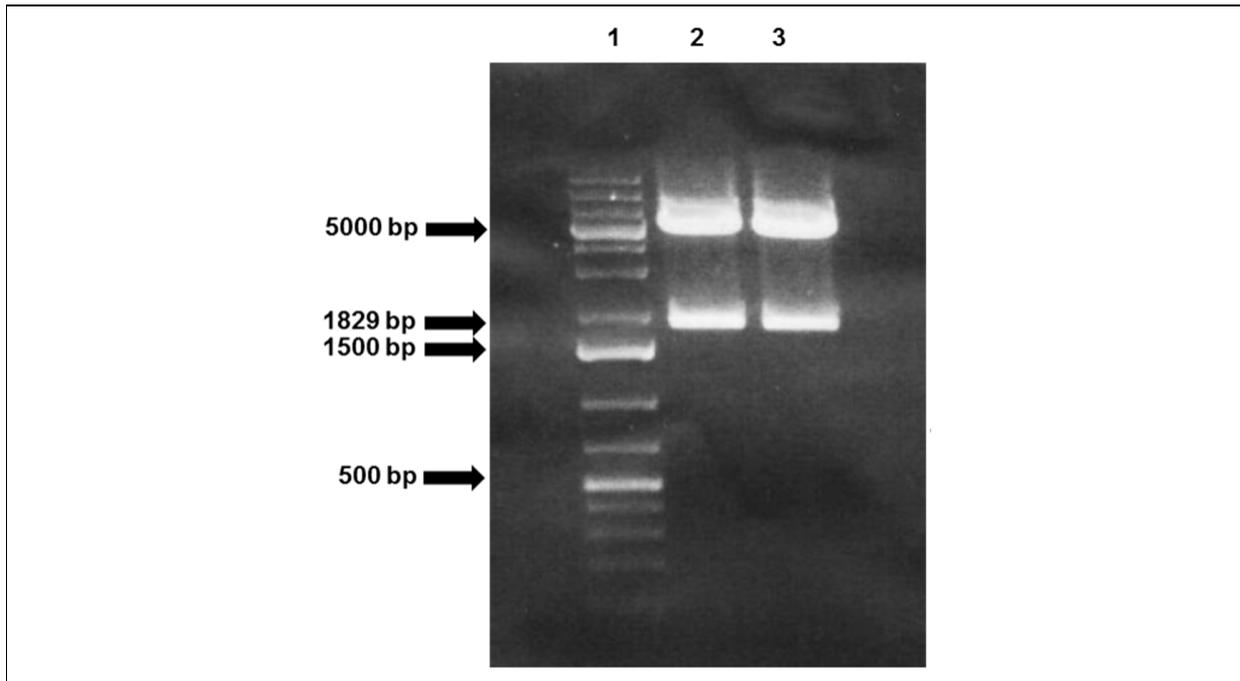
**Table 3. 8 DNA association validation candidate proteins**

Protein	Protein Domain	Description	Reference
MSMEG_1060	Lsr2	Lsr2 is a DNA-bridging protein in <i>Mycobacterium</i> .	(43, 145)
MSMEG_2695	PspA/IM30	PspA suppresses sigma54-dependent transcription, negative regulator of <i>E. coli</i> phage shock operon.	(145, 146)
MSMEG_3754	TPR	TPRs have shown involvement in cell cycle regulation, transcriptional control, and protein folding.	(145, 147)
MSMEG_4306	C4-type zinc ribbon	Structural modelling suggests that Zn-ribbon domain may bind nucleic acids.	(145, 148)
MSMEG_5512	RNA polymerase sigma factor 54 interaction domain	Interaction with sigma-54 factor and has ATPase activity. Half of the proteins identified with this domain might belong to signal transduction two-component systems.	(145, 149)

#### 3.3.4.1. Generation of pFLAG constructs

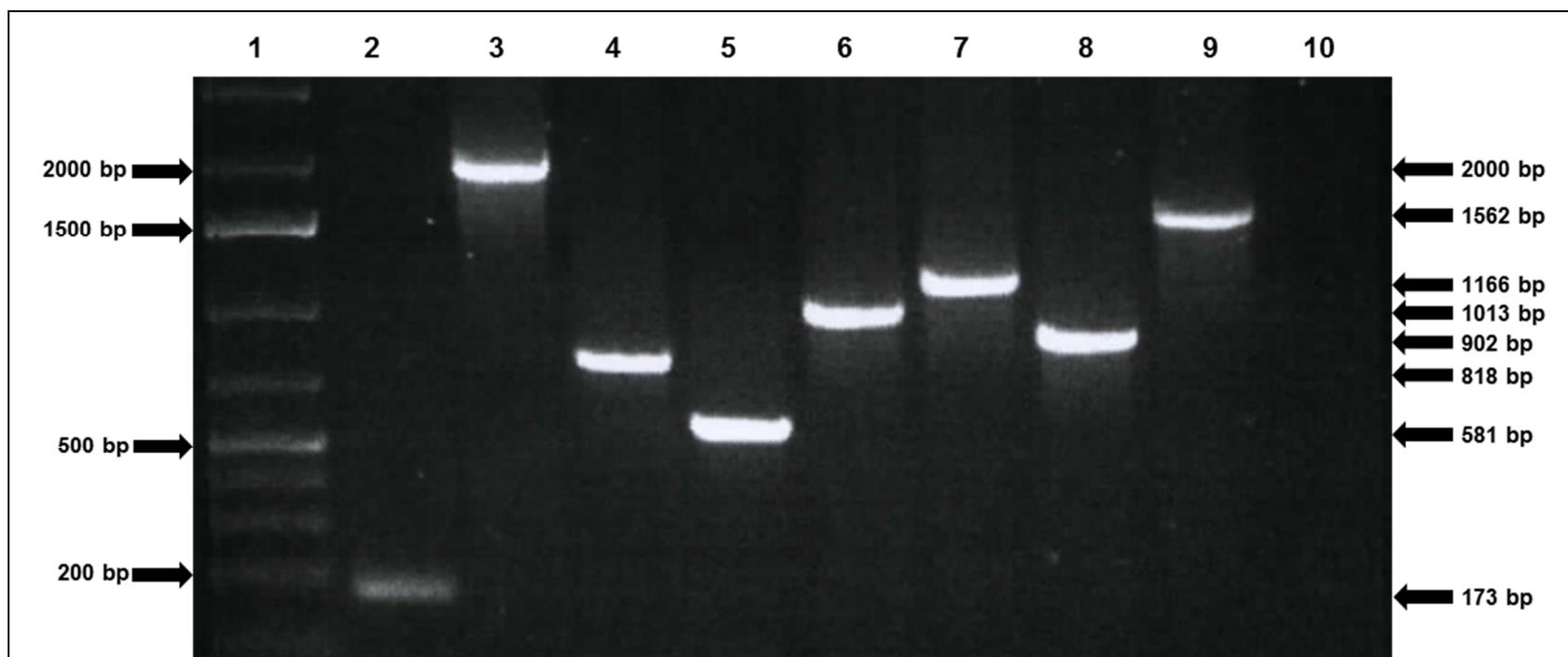
For the generation of the mycobacterial expression plasmid, pNFLAG0615, the N-terminally FLAG-tagged *MSMEG\_0615* gene was cloned into the linearized pSE100-rbs-csd vector (from which the 6 x histidine-tag, thrombin cleavage site and *csd* was excised) using In-Fusion® HD cloning. Following propagation of this expression plasmid in *E. coli*, colonies were picked and plasmid isolation as well as restriction enzyme digest was done to confirm the presence of the insert (Figure 3. 21). The integrity of the insert was confirmed using Sanger sequencing. We next sought to generate a gene insert free expression plasmid. To this end we excised the *MSMEG\_0615* gene before blunting the plasmid DNA using the Klenow fragment. Following propagation of the circularised plasmid, Sanger sequencing was done to verify that no insert was present and that no premature stop codon was generated during the blunting process. This plasmid was named pNFLAG.

The genes of interest *MSMEG\_2389*, *MSMEG\_1060*, *MSMEG\_2695*, *MSMEG\_3754*, *MSMEG\_4306*, and *MSMEG\_5512* were PCR amplified from *M. smegmatis* mc<sup>2</sup>155 using designed primer sets (Table 3. 2). The amplified PCR products were cloned into linearized pNFLAG using the In-Fusion® HD cloning kit prior to being propagated in *E. coli*. Presence of the inserts was confirmed using PCR (Figure 3. 22) and Sanger sequencing was done to verify the integrity of the inserts. These constructs were named pNFLAG2389, pNFLAG1060, pNFLAG2695, pNFLAG3754, pNFLAG4306, and pNFLAG5512.



**Figure 3. 21 Generation of pNFLAG0615 plasmid**

Agarose gel to separate DNA fragments resulting from digestion of FLAG-8xGly-*MSMEG\_0615* with NdeI and HindIII to confirm the presence of the FLAG-8xGly-*MSMEG\_0615* insert. Lane 1 contains GeneRuler 1000 bp Plus DNA Ladder and lanes 2 and 3 indicate plasmids screened from *E. coli* colonies isolated.

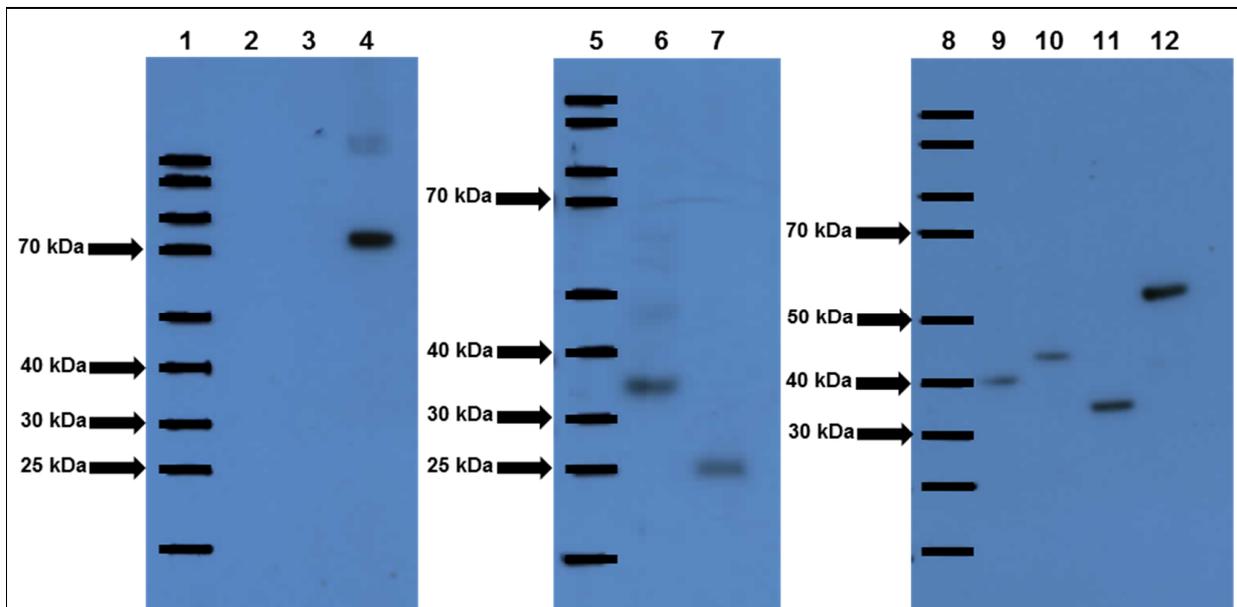


**Figure 3. 22 Generation of pNFLAG plasmid constructs**

The primers pNFLAG forward and reverse were used to PCR amplify the N-terminal FLAG-tag with the inserts cloned into the pNFLAG vector. Lane 1 contains GeneRuler 1kb Plus DNA Ladder and lane 10 contains the negative control for the PCR amplification reaction. Lanes 2 – 9 contains the PCR products generated for pNFLAG (empty vector control, 173 bp), *MSMEG\_0615* (2000 bp), *MSMEG\_2389* (818 bp), *MSMEG\_1060* (581 bp), *MSMEG\_2695* (1013 bp), *MSMEG\_3754* (1166 bp), *MSMEG\_4306* (902 bp), and *MSMEG\_5512* (1562 bp), respectively.

### 3.3.4.2. Expression of FLAG-tagged *M. smegmatis* proteins

Electrocompetent *M. smegmatis* mc<sup>2</sup>155 was transformed with the pNFLAG plasmids (pNFLAG, pNFLAG0615, pNFLAG2389, pNFLAG1060 and pNFLAG5512) and the pTEK-4S-0X plasmid, which encodes the reverse Tetracycline-binding repressor. Whole cell lysates of *M. smegmatis* transformants were prepared and proteins separated using SDS-PAGE. Western blotting was done with a mouse anti-FLAG primary antibody and a HRP-conjugated goat anti-mouse secondary antibody for the detection of FLAG-tagged proteins. Unfortunately, due to the small size of the FLAG-tag (4.7 kDa, including the linker) it could not be visualized on the western blot. Nevertheless, western blotting confirmed the expression of FLAG-tagged MSMEG\_0615, MSMEG\_2389, MSMEG\_1060, MSMEG\_2695, MSMEG\_3754, MSMEG\_4306, and MSMEG\_5512 (Figure 3. 23).



**Figure 3. 23 Detection of N-terminally FLAG-tagged protein expression in *M. smegmatis* using Western blotting**

The western blot confirms the expression of FLAG-tagged *M. smegmatis* proteins using the anti-FLAG antibody (1:4000) and a goat anti-mouse secondary antibody (1:10000). Lanes 1, 5, and 8 represents the sizes as present on the PVDF membrane for the PageRuler™ Prestained Protein Ladder. Lanes 2 to 3 shows no detection of FLAG-tagged proteins in *M. smegmatis* and *M. smegmatis* transformed with pNFLAG. Lane 4 shows expression of FLAG-tagged MSMEG\_0615 (69.1 kDa). Lanes 6 to 7 demonstrates the expression of FLAG-MSMEG\_2389 (25.9 kDa) and FLAG-MSMEG\_1060 (19 kDa), which are both proteins likely to be expressed as dimers). Lanes 9 to 12 show confirmation of FLAG-MSMEG\_2695 (35 kDa), FLAG-MSMEG\_3754 (40.2 kDa), FLAG-MSMEG\_4306 (31.8 kDa) and FLAG-MSMEG\_5512 (54.7 kDa).

### 3.3.4.3. DNA association assay

Whole cell lysates obtained from cross-linked *M. smegmatis* transformants were incubated with anti-FLAG antibody-bound protein G Dynabeads™. Following immunoprecipitation of FLAG-tagged proteins, cross-linking was reversed and all proteins isolated were digested using proteinase K. Sodium chloride and ethanol was used to precipitate any DNA-associated with the FLAG-tagged proteins. DNA precipitations were quantified spectrophotometrically (Table 3. 9) before being separated using agarose gel electrophoresis (Figure 3. 24).

*M. smegmatis* transformed with pNFLAG served as a FLAG-tag only negative control and was found to not be DNA-associated (Figure 3. 24, Figure 3. 25, Table 3. 9). The ESX conserved component EccA<sub>3</sub>, encoded by *MSMEG\_0615*, was selected as a non-DNA-associated negative control and the tagged protein FLAG-MSMEG\_0615 was found to not be DNA-associated (Figure 3. 24, Figure 3. 25, Table 3. 9).

The NAP, Hup, encoded by *MSMEG\_2389*, was selected as a positive DNA association control and FLAG-MSMEG\_2389 was found to be DNA-associated through spectrophotometric results (Table 3. 9, Figure 3. 25) and agarose gel electrophoresis (Figure 3. 24).

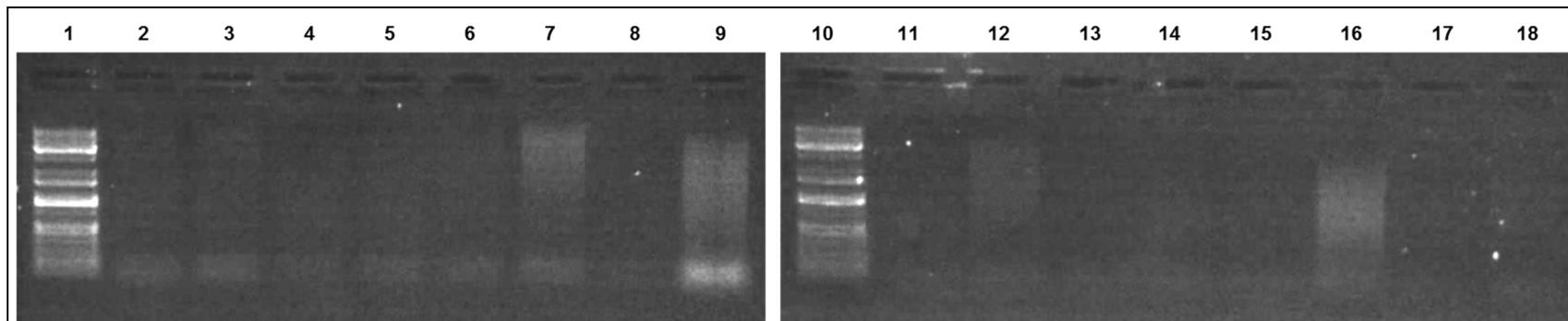
The gene, *MSMEG\_1060*, is predicted to encode a putative Lsr2 protein implying that MSMEG\_1060 might be a DNA binding protein. FLAG-MSMEG\_1060 was demonstrated to be DNA-associated both spectrophotometrically (Table 3. 9, Figure 3. 25) and through agarose gel electrophoresis (Figure 3. 24).

MSMEG\_2695 is predicted to contain a PspA domain. PspA has been shown to suppress sigma54-dependent transcription and is a negative regulator of the *E. coli* phase shock operon (146). FLAG-MSMEG\_2695 was determined to be associated with DNA both spectrophotometrically and through agarose gel electrophoresis (Figure 3. 24, Figure 3. 25, Table 3. 9).

The gene *MSMEG\_3754* is predicted to encode a TPR-domain of which some protein has shown association with transcriptional regulation (147). FLAG-MSMEG\_3754 was not found to be associated with DNA (Figure 3. 24, Figure 3. 25, Table 3. 9).

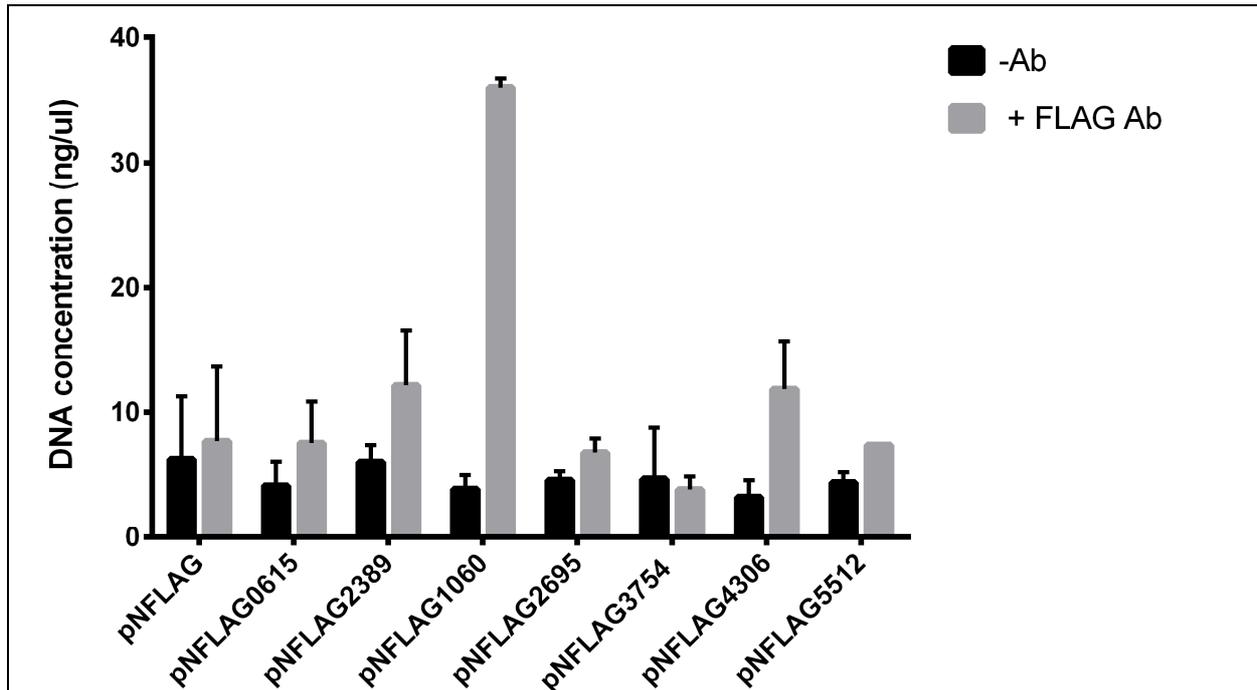
MSMEG\_4306 is predicted to contain a C4-type zinc ribbon domain which has been proposed to bind to nucleic acids (148). FLAG-MSMEG\_4306 was demonstrated to be associated with DNA both spectrophotometrically and through agarose gel electrophoresis (Figure 3. 24, Figure 3. 25, Table 3. 9).

The gene *MSMEG\_5512* is predicted to encode a magnesium chelatase with a predicted “regulation of transcription, DNA dependent” GO annotation. DNA precipitation of immunoprecipitated FLAG-MSMEG\_5512 was spectrophotometrically determined and implied DNA association. Additionally, a light smear could be observed following agarose gel electrophoresis (Figure 3. 24, Figure 3. 25, Table 3. 9).



**Figure 3. 24 FLAG-MSMEG\_1060, FLAG-MSMEG\_2695 and FLAG-MSMEG\_4306 are DNA-associated proteins**

Immunoprecipitation of N-terminally FLAG-tagged *M. smegmatis* proteins MSMEG\_1060, MSMEG\_2695, MSMEG\_3754 and MSMEG\_5512 was done to show DNA association. Lanes 1 and 10 contains the GeneRuler 1 kb Plus DNA Ladder. Even numbered lanes display no antibody control immunoprecipitations from *M. smegmatis* transformed with pNFLAG, pNFLAG0615, pNFLAG2389, pNFLAG1060, pNFLAG2695, pNFLAG3754, pNFLAG4306, and pNFLAG5512, respectively. Lanes 3 and 5 contains no DNA from the negative control anti-FLAG antibody immunoprecipitations of FLAG and FLAG-MSMEG\_0615. The positive control immunoprecipitation of FLAG-MSMEG\_2389 showed a smear following immunoprecipitation, suggesting DNA association. Likewise, lanes 9, 12, and 16 showed DNA association for anti-FLAG immunoprecipitations with FLAG-MSMEG\_1060, FLAG-MSMEG\_2695 and MSMEG\_4306. Lane 18 showed a light smear indicating DNA association for FLAG-MSMEG\_5512 however, no DNA smear was observed for FLAG-MSMEG\_3754.



**Figure 3. 25 Anti-FLAG Immunoprecipitation DNA concentrations**

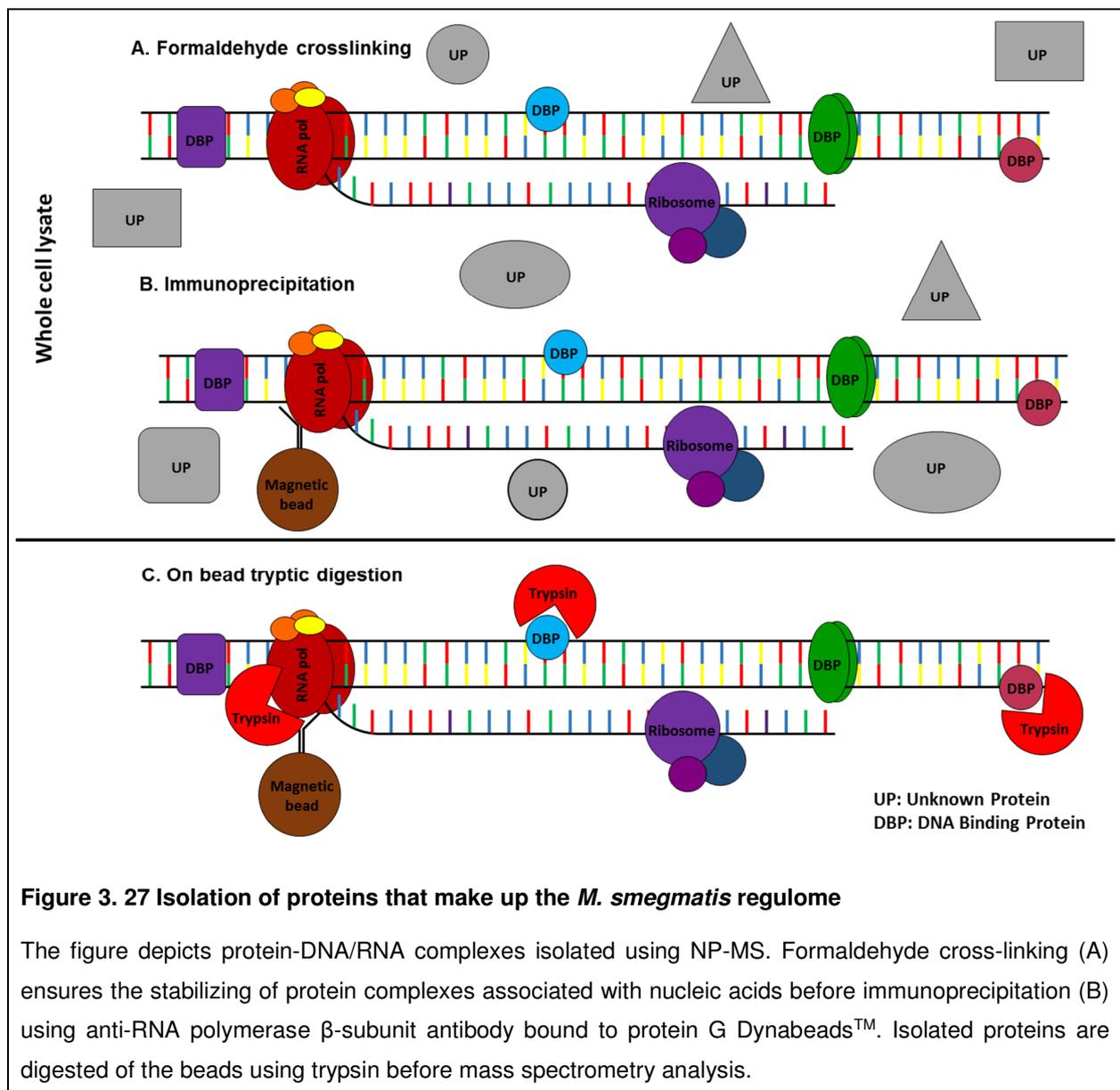
The figure displays DNA concentration values obtained following elution and precipitation of anti-FLAG immunoprecipitations. No antibody control immunoprecipitations for each of the *M. smegmatis* transformants are displayed in black and anti-FLAG immunoprecipitations in grey. The control immunoprecipitations pNFLAG and pNFLAG0615 provides a measurement of non-specific nucleic acid quantification. Our positive control for DNA association, pNFLAG2389 provides a signature for positive DNA association. These results suggested that following immunoprecipitation of N-terminally FLAG-tagged proteins, pNFLAG1060, pNFLAG2695, pNFLAG4306 and pNFLAG5512 expressed *M. smegmatis* proteins associated with DNA.

**Table 3. 9 Anti-FLAG immunoprecipitation DNA concentrations**

Sample ID	IP 1			IP 2		
	Nucleic Acid concentration ng/μl	260/280	260/230	Nucleic Acid concentration ng/μl	260/280	260/230
pNFLAG -Ab (negative control)	2.7	3.94	0.86	9.8	1.56	0.28
pNFLAG +Ab (negative control)	3.5	2.6	1.69	11.9	1.5	0.28
pNFLAG0615 -Ab (negative control)	2.8	2.63	1.16	5.5	1.66	0.23
pNFLAG0615 +Ab (negative control)	5.2	2.3	1.54	9.9	1.01	0.14
pNFLAG2389 -Ab (positive control)	5.1	2.56	2.35	7	1.65	0.28
pNFLAG2389 +Ab (positive control)	9	2.18	1.8	15.3	1.8	0.58
pNFLAG1060 -Ab	3.1	3.55	1.33	4.7	1.73	0.2
pNFLAG1060 +Ab	35.5	1.9	2.07	36.5	1.94	0.8
pNFLAG2695 -Ab	4.1	2.49	1.97	5.1	1.22	0.14
pNFLAG2695 +Ab	6	2.29	1.97	7.6	1.56	0.24
pNFLAG3754 -Ab	1.8	3.61	1.21	7.6	0.82	0.1
pNFLAG3754 +Ab	3.1	2.61	1.65	4.6	1.88	0.23
pNFLAG4306 -Ab	2.3	5.6	2.44	4.2	1.17	0.12
pNFLAG4306 +Ab	9.1	2.12	1.82	14.6	1.75	0.43
pNFLAG5512 -Ab	3.8	2.88	2.02	5	1.11	0.12
pNFLAG5512 +Ab	7.4	2.16	1.96	7.4	1.01	0.13



In this study, whole cell formaldehyde cross-linking was done followed by immunoprecipitation of DNA-protein and/or RNA-protein complexes using an anti-RNA polymerase  $\beta$ -subunit antibody bound to a magnetic bead (Figure 3. 27). RNA polymerase was selected as a “tag” for nucleic acids due to its association with both RNA and DNA, as well as its natural abundance. ChIP-seq demonstrated that the RNA polymerase  $\beta$ -subunit was present along the entire *M. tuberculosis* genome and similarly RNA-seq showed that RNA reads mapped to both coding and non-coding regions (62, 154). RNA polymerase was therefore deemed as a suitable tag for the isolation of DNA and RNA protein complexes.



Co-purification of “contaminant” proteins is a known limitation of AP-MS experiments (119, 129, 155). Therefore, to limit the purification of non-specifically bound proteins, protein-nucleic acid complexes were washed using several solutions consisting of various detergents and salts. The buffers used in this study to limit the identification of non-specific proteins were previously described in a ChIP-seq experiment which investigated the genome-wide distribution of RNA polymerase in *M. tuberculosis* (62). To further control for non-specific binding during immunoprecipitation, two affinity purification controls were selected. A no antibody (magnetic bead only) control was as well as an unrelated antibody control (anti-human myosin monoclonal antibody bound to a magnetic bead) was selected to limit the identification of non-specifically bound proteins as possible nucleic acid associated proteins. A heatmap (Figure 3. 11) of identified proteins following immunoprecipitation determined that the wash buffers used in this experiment were effective in reducing non-specific binding and thereby the identification of non-specific proteins as only 32 % of the total amount of proteins identified in at least two of three biological replicate experiments, were also detected in control immunoprecipitations (Figure 3. 1).

On-bead trypsin digestion of immunoprecipitated samples was done to elute peptides of isolated proteins before mass spectrometry analysis. Non-denatured tryptic shaving of magnetic beads was done to limit antibody contamination of the sample. Antibodies are known to exhibit resistance towards proteolytic digestion and specifically, tryptic digestion of mouse IgG antibodies has been shown to be inefficient (141, 142). Additionally, on-bead digestion facilitated single-shot in-solution based mass spectrometry. In-solution based mass spectrometry has many advantages to in-gel based mass spectrometry such as reduced run-time on the mass spectrometer (only one fraction vs. multiple fractions), reduced running costs due to shorter experimental running times and, shorter data analysis times due to the reduced number of fractions. In-gel based mass spectrometry does have the advantage of providing a higher resolution for peptide identification within a mixed protein sample however, this was not considered to be a limiting factor in this study due to protein sample set being less complex than that of a total proteome. Lastly, in-gel based mass spectrometry of the immunoprecipitated sample would have been limited by the presence of the light and heavy chains which make up the anti-RNA polymerase antibody used in NP-MS. The presence of these fragments on polyacrylamide gels would have resulted in the inability to identify proteins of similar sizes.

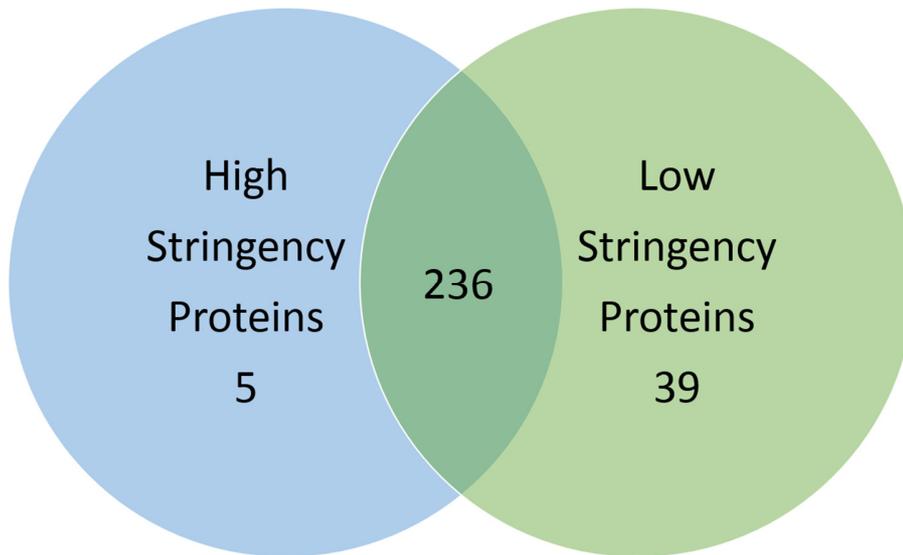
Data obtained from the LC-MS/MS analysis was analysed using MaxQuant to identify proteins through automated searching using the Andromeda search algorithm against the *M. smegmatis* mc<sup>2</sup>155 database. Formaldehyde treatment of proteins has been shown to result in various modifications on amino acids. The formation of these modifications is influenced by several factors such as the formaldehyde concentration, the rate of the particular cross-linking reaction, the position and local environment of the reactive amino acid, the pH and the components present within the cross-linking solution (156). Modifications on amino acids results in heavier peptides following tryptic digestion, which could ultimately be erroneously discarded during protein identification due to the inability of mapping to any protein sequences within the database.

Possible formaldehyde-induced modifications were identified through literature, the modifications of mass spectrometry database (Unimod), and a blind/unrestrictive search for post-translational modifications (157–159). Predicted modifications were searched against LC-MS/MS data to determine their respective frequencies within RNA polymerase immunoprecipitation experiments prepared using NP-MS (Table B. 1) (156, 160). For the identification of proteins through automated searching four variable modifications were selected, oxidized methionine, the addition of glycine on lysine, serine and threonine residues (hereafter referred to as “the addition of glycine”), the addition of methylol and glycine on any histidine, asparagine, glutamine, tryptophan and tyrosine (hereafter referred to as “the addition of methylol and glycine”) as well as the possible di-methylation of lysine and arginine residues. For the identification of proteins isolated using NP-MS we used a high stringency approach and a lower stringency approach. For the generation of a very stringent protein identification list, high confidence proteins were identified when a protein was present in all three of the anti-RNA polymerase immunoprecipitations and in none of the control immunoprecipitations. Low confidence proteins were identified following imputation to replace the missing values in control immunoprecipitations followed by a multi-factor ANOVA with an FDR of 0.05 using the Benjamini-Hochberg correction to identify proteins which were significantly more present in the anti-RNA polymerase immunoprecipitations vs. control immunoprecipitations. A total of 241 proteins were identified using the high stringency approach which consisted of 177 high confidence proteins and a further 73 low confidence proteins (Table B. 4, Figure 3. 1). The lower stringency identification list was assembled through the identification of high confidence proteins when a protein was present in at least two of the three biological replicate anti-RNA polymerase immunoprecipitation experiments and in none of the control immunoprecipitations. Low confidence proteins were identified as described for high confidence proteins. A total of 275 proteins were identified in the lower stringency protein identification list consisting of 214 high confidence proteins and 61 low confidence proteins (Table B. 2, Figure 3. 1).

Comparison of the high stringency and low stringency identification lists revealed that five of the 241 proteins identified in the high confidence list were not identified in the lower stringency list (Figure 3. 24). These five low confidence proteins are described in Table B. 6. Thirty-nine low confidence proteins were exclusively identified in the lower stringency list of which 18 (46.1%) were identified as nucleic acid associated proteins through searching for identifiers such as nucleo-, DNA, RNA, transcription, translation, ribosome, nucleic acid, chromosome, and ligase (Figure 3. 24) (Table B. 7). For this reason, in this thesis, we discussed the proteins identified in the lower stringency protein identification list.

Identification of nucleic acid associated proteins using the identifiers mentioned above revealed that 41.9% of the proteins identified in the lower stringency list could be considered nucleic acid associated proteins associated with the RNA polymerase transcriptional complex under the given conditions in this study. Proteins identified in this study which could be associated with the RNA polymerase transcriptional complex

included DNA polymerases, topoisomerases, helicases, NAPs, transcription factors, sigma factors, and ribosomes (Table B. 2).



**Figure 3. 28 High stringency vs lower stringency protein identification**

The Venn diagram shows that the majority of proteins identified in the higher stringency protein identification list was also identified in the lower stringency protein identification list (240 proteins). Four proteins were exclusively associated with the high stringency list and a further 41 proteins were only found in the lower stringency list.

Since the initial sequencing and annotation of the *M. smegmatis* genome was done, it became apparent that the sequence of this organism contained multiple errors which resulted in the incorrect translation to amino acid sequences for some proteins (161). Reasons for these sequencing errors have been attributed to the high GC content of the *M. smegmatis* genome as well as annotation shortfalls (162). To further complicate the analysis of the proteins identified using NP-MS we must bear in mind that a large proportion of the *M. smegmatis* genome contains genes which are predicted to encode hypothetical proteins. Hypothetical proteins are predicted proteins whose functions remain unknown and therefore in this study we made use of gene ontologies (GO) to describe the proteins identified using NP-MS. The online analysis tools GOEAST and REVIGO were used to identify enriched GO terms for the 275 proteins identified using NP-MS (Table B. 3). Enriched GO identities were classified into three functional categories biological processes, molecular function, and cellular components. Several GO terms which can be associated with nucleic acid associated proteins were identified in the functional categories biological processes (Figure 3.

14) and molecular function (Figure 3. 15) whereas for the functional category cellular components revealed only intracellular components such as ribosome were enriched (Figure 3. 16). These results demonstrated the ability of the chosen approach to identify nucleic acid-associated proteins in *M. smegmatis*.

Interactive graphs were generated using REVIGO and were used to visualise the similarities between enriched GO identities. Interactive graphs for the three functional categories biological processes (Figure 3. 17), molecular function (Figure 3. 18) and cellular components (Figure 3. 19) revealed that sub-networks of enriched GO terms used to describe nucleic acid-associated proteins could be isolated from enriched GO identity networks. These results suggested that the proteins identified using NP-MS shared a degree of similarity in predicted functional attributes and could possibly be part of a protein network associated with nucleic acids. To demonstrate that the predicted interactions between GO annotations could translate through on protein level, a predicted protein-protein interaction network was generated using STRING. The interactive network was assembled using experimental data gathered from protein-protein interaction databases. Four-hundred and forty-seven predicted protein-protein interactions were assigned to 149 of the 275 NP-MS identified proteins (Figure 3. 20). These results demonstrated that NP-MS was effective in the isolation and identification of a protein network associated with nucleic acids in *M. smegmatis*.

The data presented in this study showed that NP-MS was effective in the isolation and identification of proteins which are associated with both DNA and RNA. In a previous study investigating nucleoid-associated proteins in mitochondria, whole cell formaldehyde cross-linking combined with affinity purification of an over expressed FLAG-tagged DNA helicase followed by tandem mass spectrometry was used to identify proteins associated with the mitochondrial nucleoid. This approach was prone to protein contamination of proteins known to be involved in RNA metabolism and translation. The proteins identified in this study, which were associated with DNA and RNA, could therefore be considered the gene expression proteome of mitochondrial nucleoids (114). We consider NP-MS to be an improvement on this approach as the ectopic overexpression of a DNA helicase might influence gene expression and consequently the gene expression proteome. The targeting of intrinsically expressed RNA polymerase as a tag for DNA and RNA allows us to observe the gene expression proteome of *M. smegmatis* under native *in vitro* conditions. We anticipate that NP-MS could be used to study changes in the gene expression proteome when the organism in question is exposed to different environmental growth conditions.

#### **3.4.2 DNA Association of NP-MS Identified Proteins**

To validate the ability of NP-MS to isolate and identify nucleic acid associated proteins in *M. smegmatis*, we selected five proteins identified using NP-MS for DNA association analysis. MSMEG\_1060, MSMEG\_2695, MSMEG\_3754, MSMEG\_4306 and MSMEG\_5512 were selected through the identification of conserved protein domains which have been shown to associated with nucleic acids in other bacterial organisms (Table 3. 8).

Proteins of interest were expressed as FLAG-tagged proteins in *M. smegmatis* and anti-FLAG immunoprecipitations were done prior to proteinase K digestion and DNA precipitation. DNA quantification using NanoDrop™ spectrophotometric readings were done as well as DNA separation using agarose gel electrophoresis to determine DNA association with the proteins of interest. EccA<sub>3</sub> (encoded by *MSMEG\_0615*) was selected as a negative control for DNA association immunoprecipitations. EccA<sub>3</sub> is a cytoplasmic component of the ESX-3 secretion system and a predicted ATPase with no known DNA associations. The NAP, HupB (*MSMEG\_2389*) was selected as a positive DNA association control for DNA immunoprecipitation studies. DNA was recovered following immunoprecipitation for HupB, *MSMEG\_1060*, *MSMEG\_2695*, *MSMEG\_4306* and *MSMEG\_5512*. No DNA association could be determined for *MSMEG\_3754*. DNA association of the selected proteins identified using NP-MS showed that the developed methodology is effective in the isolation of proteins that are associated with DNA and therefore effective in the isolation of nucleic acid associated proteins.

DNA footprinting and ChIP-seq studies can be done to determine the DNA binding sequences of potential DNA binding proteins followed by electrophoretic mobility shift assays (EMSA) to assess the DNA binding affinity to the identified DNA sequences (163, 164). For the validation of predicted RNA binding proteins, RNA footprinting and EMSA's can be used to assess the RNA binding potential of identified proteins (165, 166). Predicted protein-protein associations (obtained from STRING) can further be explored using molecular techniques such as yeast two-hybrid screening or other co-immunoprecipitation techniques (167). Microscale thermophoresis (MST) is a technique in which the directed movement of particles in a microscopic temperature gradient are monitored to determine binding affinities of proteins, DNA, RNA and various other biomolecules (168). The use of MST for the validation of a previously unknown DNA, RNA and/or protein interactions is therefore an attractive approach for proteins identified within this study.

### 3.4.3 Limitations

NP-MS was effective in the isolation and identification of proteins that make up the gene expression proteome; however, the proposed methodology does have several limitations.

Prior to immunoprecipitation using an anti-RNA polymerase antibody, genomic DNA was digested into fragments of approximately 200 to 800bp. DNA fragmentation was done to facilitate the immunoprecipitation of DNA- and RNA-protein complexes. However, the fragmentation of genomic DNA restricts the identification of DNA-associated proteins to the immediate location of RNA polymerase. ChIP-seq has demonstrated the RNA polymerase complex to be present along the entire genome with a preference for putative promoter regions in both the exponential and stationary growth phases of *M. tuberculosis* (62). However, the possibility still remains that proteins associated with DNA which are not on fragments occupied by the RNA polymerase complex will not be identified using NP-MS.

To reduce the amount of non-specific binding of proteins which are not members of the gene expression proteome, various detergents (as SDS, NP-40, Triton X-100 and sodium deoxycholate) were used that are not recommended for mass spectrometry. These detergents inhibit the ionization of the peptides thereby limiting the detection abilities of peptides within the sample (169, 170). Additionally, the use of these detergents is potentially detrimental to the mass spectrometer as these polymers can cause a blockage of the liquid chromatography column. Ethyl acetate phase separation after tryptic digestion of in-solution proteomic sample preparation has been proposed as a possible step for peptide-bound detergent removal before mass spectrometry analysis. Ethyl acetate extraction has been reported to enhance peptide detection in peptide sample free of detergents but a loss of larger peptides associated with SDS has been reported (171). Methanol-chloroform phase separation was also considered for detergent purification of samples prior to mass spectrometry analysis. The purification of samples would be attractive for future experiments however, the potential loss of peptide could pose a problem as low confidence proteins are identified using LFQ intensity data. It would not be possible to determine what the loss would be per sample and whether the loss is uniform between biological replicates, therefore the use of the intensity data to identify proteins more abundant in the anti-RNA polymerase  $\beta$ -subunit immunoprecipitation versus control immunoprecipitation would be controversial.

Proteomic identification of nucleic acid associated proteins such as transcription factors is known to be problematic. The prevalence of more abundant proteins often masks the detection of less abundant proteins during mass spectrometric detection, therefore limiting the detection of low abundance proteins such as transcription factors (172, 173). During tandem mass spectrometry, high energy collisional dissociation of precursor ions allows for high-accuracy fragmentation. The resulting increased resolution in the detection of ionized peptides enables the improved identification of proteins (174). The fragmentation within the ion trap can however differ between biological experiments analysed and can subsequently limit the detection of low abundant proteins between biological replicate experiments. Therefore, for the identification of proteins we deemed a protein identified when detected in two of the three single shot proteomic experiments done with a minimum of two unique peptides.

Post-translational modifications made on proteins by formaldehyde cross-linking and glycine quenching potentially limit the detection of peptides using mass spectrometry. Unfortunately, not all potential variable modifications identified can be included in the protein identification search due to possibly increasing the false discovery rate. Therefore, no more than four variable modifications were included in the protein identification search. As discussed above the variable modifications selected for protein identification in this experiment were oxidized methionine, N-terminal acetylation, the addition of glycine and the possible dimethylation of lysine and arginine residues.

The KEGG pathway mapping tool was used to show to which metabolic processes NP-MS identified proteins mapped to, however, during this a large proportion of proteins submitted were unable to be mapped

to any pathways. Given the large number of proteins in mycobacterial organism which share very little sequence homology with proteins of other well studied organisms and have no described functions, this is not surprising. Our limited functional knowledge regarding the proteins encoded by mycobacteria does however limit our understanding of how these proteins influence bacterial physiology, and does limit the confidence to which online analysis tools such as KEGG can be used to describe proteins and the mycobacterial pathways in which they reside. Furthermore, gene ontology enrichment analysis was done using the customized microarray from the online analysis platform GOEAST to determine which GO identities were over-represented within our dataset through hypergeometric testing (135). Unfortunately, no dedicated gene ontology analysis tool exists for enrichment analysis in *M. smegmatis* genes or proteins, however, one is available for the analysis of *M. tuberculosis* genes/proteins (<http://www.geneontology.org/page/go-enrichment-analysis>). Enrichment data for the proteins identified using NP-MS was subjected to REVIGO to remove redundant GO identities and to visualise the enrichment data using treemaps. Cluster representative GO terms in treemaps were assembled using the *p*-values obtained from GOEAST and not through the prioritizing of lower- or higher-level GO identifications. Therefore, the development of a curated and dedicated gene ontology enrichment analysis tool will provide a higher level of confidence to the GO enrichment data used for downstream analyses.

This study serves as an initial screen to identify novel proteins which may be associated with the gene expression proteome of *M. smegmatis*. The information obtained from the *in silico* tools used to describe the predicted physical attributes of proteins is often derived from gene and protein homologies from unrelated species. Therefore, the possible protein, DNA and/or RNA associations of the proteins identified within this study remain to be validated using functional studies.

### 3.5. Conclusion

---

The targeting of nucleoprotein complexes using an anti-RNA polymerase  $\beta$ -subunit antibody immobilized on a magnetic bead following formaldehyde cross-linking allows for the identification of nucleic acid associated proteins by means of mass spectrometry analysis. NP-MS was shown to possess the necessary sensitivity required to detect low abundance proteins such as transcription factors, to be stringent in the isolation of proteins that are associated with DNA and/or RNA and to be reproducible between biological replicate experiments.

The majority of the proteins identified in this study have predicted functions which can be associated with nucleic acids. Several identified proteins are predicted to contain domains of unknown function (DUF), however, the accuracy with which proteins were identified using NP-MS suggests that these proteins might be novel DNA or RNA associated proteins. Further functional studies have to be performed to investigate the probable roles of these proteins within mycobacterial organisms.

Having developed, optimized, and validated NP-MS, we propose to use this new approach to study the proteins associated with the RNA polymerase transcriptional complex in *M. smegmatis* under various stress conditions. We anticipate that these studies could potentially shed new light on gene responses required for the adaption of mycobacterial organisms to stresses encountered in their environment. NP-MS also possesses the potential to easily be adapted for the investigation of nucleic acid associated proteins in organisms other than mycobacteria.

---

## Chapter 4: Nucleoproteins associated with exponential and stationary growth phase in *Mycobacterium smegmatis*

---

### 4.1. Introduction

---

Tuberculosis is classified as a granulomatous inflammatory disease, with macrophages, T and B lymphocytes, and fibroblasts aggregating to form granulomas (175). Infectious lesions are believed to consist of *Mycobacterium tuberculosis* subpopulations in different physiological states, with treatment resulting in the preliminary elimination of actively dividing mycobacteria, and resulting in the need for prolonged treatment to eliminate the remaining non-replicating bacteria (176–180). This non-replicating *M. tuberculosis* population is known to persist in the presence of functional immune systems for extended periods of time, which renders conventional chemotherapies which target cell division, DNA replication and protein expression, ineffective (181, 182). Several *in vivo* (e.g. the treated mouse model) and *in vitro* (e.g. Wayne hypoxia model, the 100-day static culture model, the Loebel nutrient starvation model and the controlled batch culture model) dormancy models have been used to investigate mycobacterial persistence (4, 183–185). Adaptation of mycobacteria to low oxygen is the most widely studied stress condition with the gradual decline in oxygen resulting in the formation of hypoxic conditions and a non-replicating population (5, 183, 186).

The physiological changes associated with stationary phase has been well documented in gram-negative bacteria such as *E. coli* and are known to include a smaller cell size, increased stress resistance, increased RNA stability and changes in protein synthesis (187, 188). Similarly, In *M. smegmatis*, stationary phase cells have been shown to have shorter cell lengths and a less domed appearance on agar plates, increased resistance to environmental stress, and increased mRNA stabilisation (189). Transcriptomic and proteomic analyses, employing some of the above mentioned *in vitro* dormancy models, have also been used to describe differential expression for stationary phase *M. smegmatis* and *M. tuberculosis* under hypoxic growth conditions and starvation (4, 6, 190–193). Furthermore, in *M. tuberculosis*, the number of coding RNA reads declined from 78.4% in exponential phase cultures to 11% in stationary phase (154). Despite the overall decline in transcription, mycobacteria have been shown to upregulate the expression of cell wall lipids, virulence associated proteins, sulphur metabolism proteins and the dormancy response regulon during stationary growth (5, 6, 91, 194, 195).

Following the development of Nucleoprotein – Mass Spectrometry (NP-MS) we wished to investigate the efficiency of NP-MS in the identification of nucleic acid associated proteins unique to environmental stress conditions experienced by the bacterium. To this end, *M. smegmatis* was cultured to exponential and early stationary phase before investigation using NP-MS.

## 4.2. Materials and Methods

---

All buffers used in this protocol are described in Table A. 1 within addendum A.

### 4.2.1. Bacterial strain and Culture Conditions

*M. smegmatis* mc<sup>2</sup>155 was grown in Difco™ Middlebrook 7H9 broth supplemented with AD and Tween-80 (Table A. 2) at 37°C with shaking. *M. smegmatis* growth curves were done in biological duplicates with three technical replicates per experiment. *M. smegmatis* cultures were grown for 27 hours with optical density readings taken at 3 hour intervals.

*M. smegmatis* cultures were screened for contamination using ZN staining as described in (131) and *M. smegmatis* stocks were stored at -80°C.

### 4.2.2. Chromatin Immunoprecipitation – Protein Mass Spectrometry

#### 4.2.2.1. Formaldehyde Cross-linking

Three biological replicate exponential cultures were collected at optical densities between 1.4 - 1.5 at a 600 nm wavelength (approximately 12 hours of growth) and likewise, three biological replicates of stationary growth phase cultures were collected at optical densities of ~3.0 at 600 nm after 24 hours of growth.

Exponential and stationary phase *M. smegmatis* mc<sup>2</sup>155 cultures were diluted to an optical density of 0.4 at 600 nm in a total volume of 50 ml using heated Erlenmeyer culture flasks and heated culture media prior to formaldehyde cross-linking. Diluted *M. smegmatis* exponential and stationary phase cultures were cross-linked using a 37% molecular grade formaldehyde solution at a final concentration of 1%. Cell cultures were incubated with formaldehyde for 10 minutes at 37°C with gentle shaking at 100 rpm. The formaldehyde cross-linking reaction was quenched with the addition of glycine to a final concentration of 125 mM and incubated at 37°C for 10 minutes. Cell cultures were pelleted at 3220xg for 10 minutes at 4°C. Cell cultures were washed with ice-cold TBS using centrifugation for 10 minutes at 3220xg at 4°C. This wash step was repeated before storing pellets at -80°C.

#### 4.2.2.2. Whole Cell Lysate Preparation

Exponential and Stationary phase whole cell lysates were prepared as described in Chapter 3, section 3.2.3.2. Following cell lysis and controlled DNA and RNA digestion, recovered supernatants from respective technical replicates for each biological replicate from the exponential and stationary phase cultures were pooled respectively.

#### 4.2.2.3. Protein Concentration Determination

Protein concentrations of exponential growth phase and stationary growth phase whole cell lysates were determined spectrophotometrically (Ultrospec 4051, LKB Biochrom) using the RC DC Protein Assay (Bio-Rad) according to manufacturer's instructions. Bovine Serum Albumin (BSA) was used to generate a standard curve ranging from 0 mg/ml to 2 mg/ml and the standard curve was plotted according to the BSA absorbance readings at 595 nm. Exponential and stationary phase whole cell lysate concentrations were determined from the standard curve.

#### 4.2.2.4. Immunoprecipitation

Protein G Dynabeads™ were prepared as stipulated by the manufacturer. Briefly, Dynabeads™ Protein G suspension was rotated for approximately 5 minutes prior to 50 µl of beads being aliquoted into 2 ml Eppendorf® Protein LoBind microcentrifuge tubes. To generate the resin which will allow for the immunoprecipitation of nucleoprotein complexes, each aliquot of Dynabeads™ was incubated with 5 µg anti-RNA polymerase β-subunit antibody (Biolegend) or 5 µg anti-muscle myosin (human) (MYH7) (Santa Cruz) antibody in 200 µl PBS-T, where appropriate. Antibodies were allowed to bind to beads for 10 minutes with rotation before being washed with PBS-T. Subsequent to antibody binding, 1 mg of cross-linked exponential and stationary phase whole cell lysates were added to antibody bound beads in a total volume of 2 ml (diluted with IP buffer I). Antibody coupled beads were incubated with cross-linked whole cell lysates for 2 hours at 4°C with rotation prior to extensive washing as described in Chapter 3, section 3.2.3.5., Table 3. 1.

#### 4.2.2.4. On-bead Tryptic Digestion

On-matrix tryptic digestion was done as described in Chapter 3, section 3.2.4.4. Peptides were recovered from magnetic beads using a DynaMag™ and stored in 1.5 ml Eppendorf® Protein LoBind microcentrifuge tubes at -80°C.

#### 4.2.2.5. Concentration and Desalting of Tryptic Digests

Tryptic digests were concentrated and desalted using stage-tips as described in Chapter 3, section 3.2.4.5. The desalted peptide solution was dried using the Concentrator*plus* (Eppendorf) before being resuspended in a 100 µl 2% acetonitrile and 0.1% formic acid solution.

#### 4.2.2.6. Methanol-Chloroform Detergent Removal

Methanol-Chloroform phase separation was used to remove detergents introduced to samples by various wash buffers used (Table A. 1). Briefly, 400 µl methanol and 100 µl of chloroform were added to 100 µl of desalted tryptic digest followed by vigorous mixing using a vortex mixer. Phase separation was achieved by the addition of 300 µl water and centrifugation at 14 000 rpm for 1 minute. Subsequently, the aqueous

phase was recovered and dried using the Concentrator<sup>plus</sup> (Eppendorf) prior to being resuspended in 20 µl solution A.

#### 4.2.2.7 Mass Spectrometry Analysis

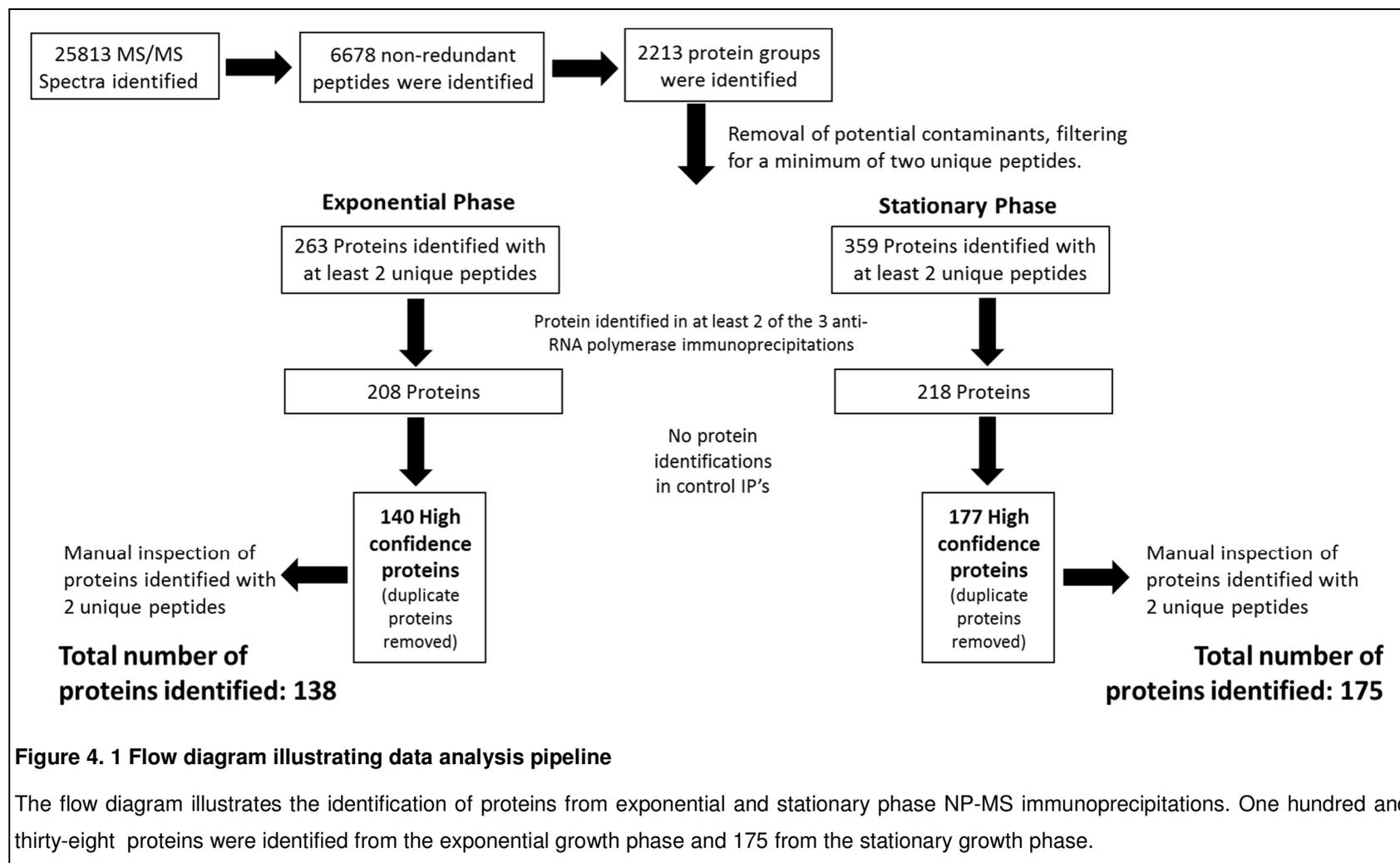
The Orbitrap Fusion<sup>TM</sup> Tribrid<sup>TM</sup> Mass Spectrometer (Thermo Fisher Scientific) was used for LC-MS/MS analysis in collaboration with Dr. M Vlok at the Mass Spectrometry Proteomics unit at the Central Analytical Facility (CAF) at Stellenbosch University. Liquid chromatography and mass spectrometry was done as described in Chapter 3, section 3.2.4.6.

#### 4.2.2.8 Data Analysis

LC-MS/MS data was analysed using MaxQuant 1.5.3.17 software to identify proteins through automated database searching of tandem spectra against the *M. smegmatis* mc<sup>2</sup> 155 database (UP000000757) as described in Chapter 3, section 3.2.4.7 (124, 144). Subsequent data analysis was performed using Perseus 1.5.8.5. LFQ Intensity values were log 2 transformed and potential contaminants were removed prior to hierarchical clustering of the dataset.

Since no proteins were identified in three of the six control immunoprecipitations for the exponential growth phase, missing values in the dataset could not be replaced by separate imputation for each column as done in Chapter 3: Development of Nucleoprotein - Mass Spectrometry. For the discussion of protein identification in this chapter we therefore focussed on high confidence protein identifications only. High confidence proteins were identified for the exponential and stationary phases when the identified protein was present in at least two of the three biological replicate immunoprecipitations but not present in any of the anti-myosin and protein G Dynabead<sup>®</sup> control immunoprecipitations (Figure 4. 1). For proteins identified from exponential and stationary phase immunoprecipitations with only 2 unique peptides, manual spectral inspection was performed (133). Proteins were excluded from the high and low stringency identification lists when unique peptide spectra were found to have poor posterior error probability scores (PEP), major unexplained peaks, poor peptide coverage, majority y or b ions and low intensity peaks.

Identified proteins were described using data obtained from UniProt (<http://www.UniProt.org/>) and gene ontology enrichment analysis was performed using GOEAST (<http://omicslab.genetics.ac.cn/GOEAST/>) (135, 144). Gene ontology enrichment data was subject to REVIGO (<http://revigo.irb.hr/>) to remove redundant GO identities (136). Comparison of exponential and stationary phase protein identifications were done using the Bioinformatics and Evolutionary Genomics Venn Diagram analysis tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). Gene ontology enrichment data of proteins found either exclusively in the exponential or stationary growth phases were subjected to the GO Terms Classification Counter (CateGORizer) to count and describe the predicted functional attributes of these proteins (196).



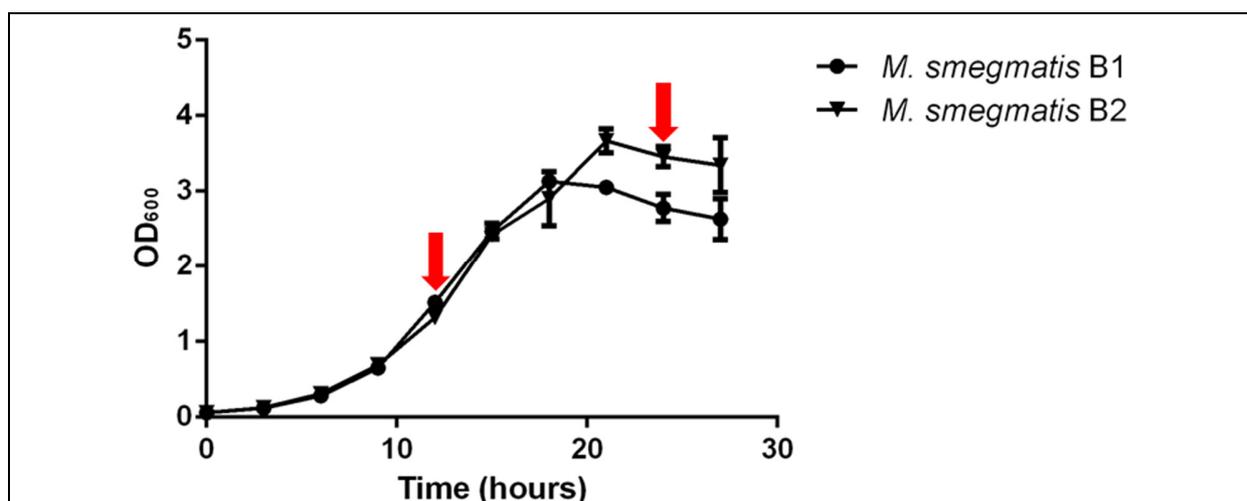
**Figure 4. 1 Flow diagram illustrating data analysis pipeline**

The flow diagram illustrates the identification of proteins from exponential and stationary phase NP-MS immunoprecipitations. One hundred and thirty-eight proteins were identified from the exponential growth phase and 175 from the stationary growth phase.

## 4.3 Results

### 4.3.1. Identifying Proteins Unique to the Exponential Phase and Stationary Phase

In this study, we sought to determine if NP-MS could be used to identify nucleic acid associated proteins that are unique to environmental conditions experienced by *M. smegmatis*. To this end, we investigated the nucleic acid associated proteins of the exponential and stationary growth phases of *M. smegmatis* using NP-MS (Figure 4. 2).



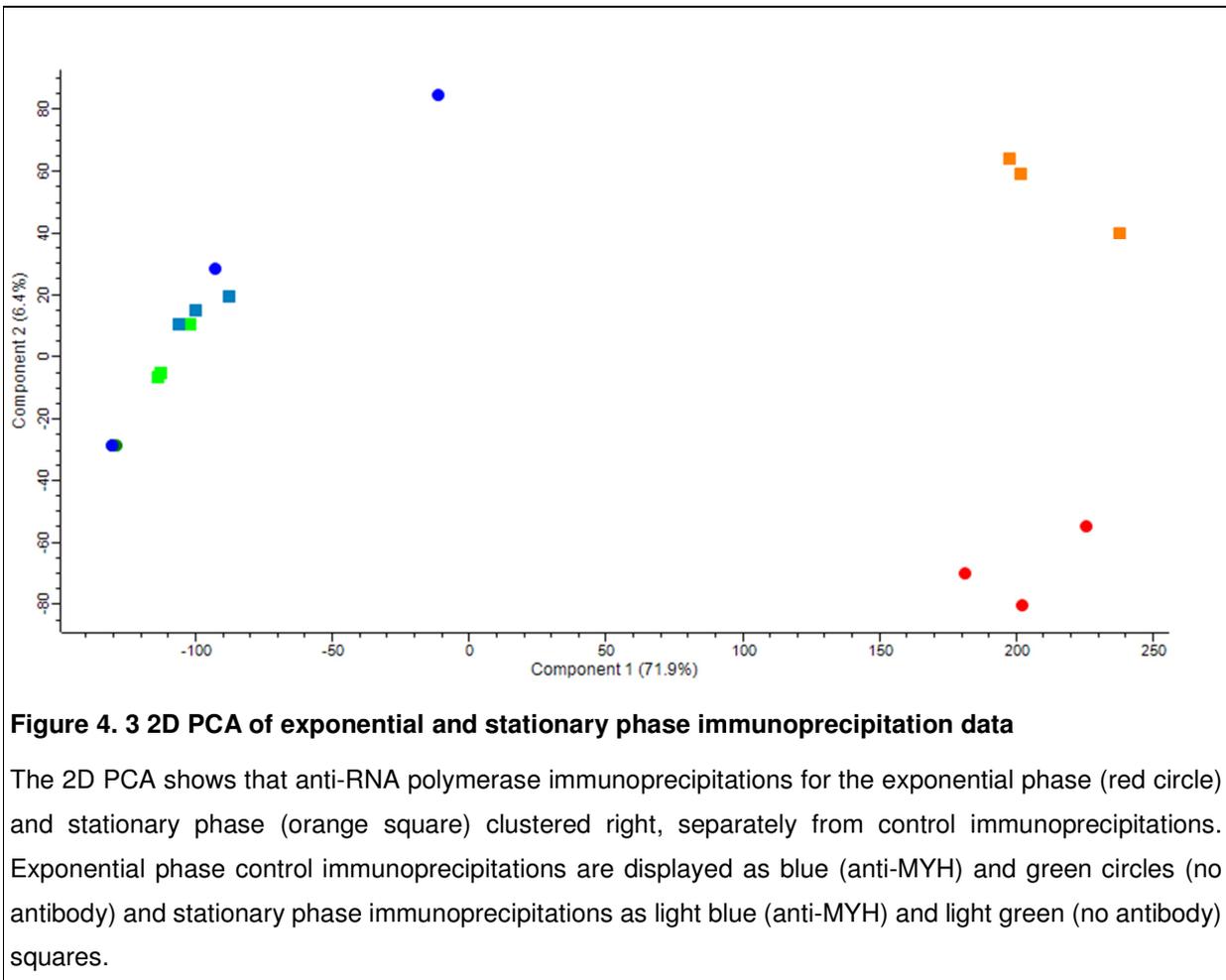
**Figure 4. 2** *M. smegmatis* growth curve

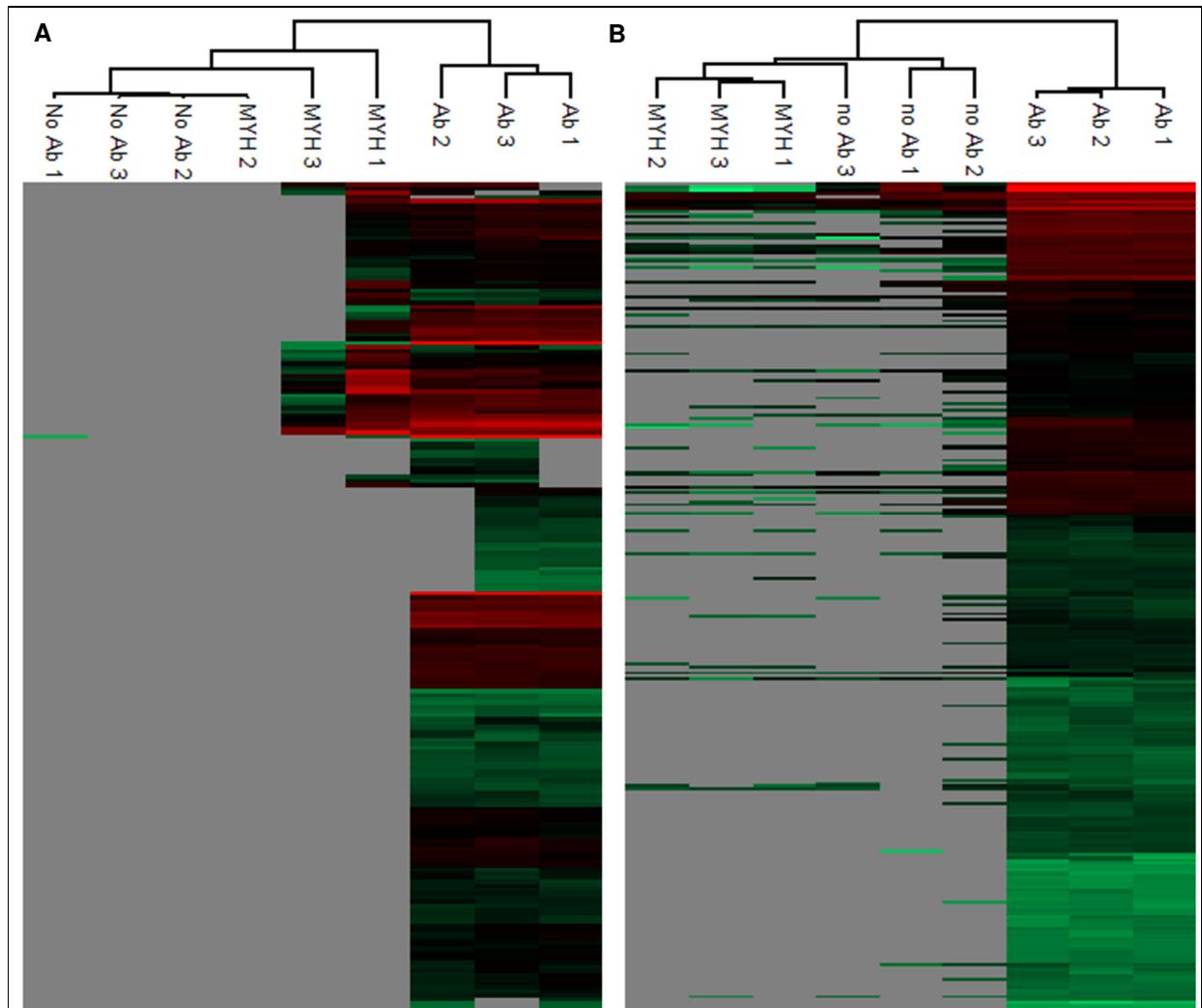
The *M. smegmatis* growth curve displays the optical densities measured at three-hour intervals for three technical replicates of two biological experiments. The red arrows indicate exponential and stationary growth phases at 12 (OD<sub>600</sub> = ~1.5) and 24 hours (OD<sub>600</sub> = ~3.0) respectively.

We first wished to identify NP-MS isolated proteins from both the exponential and stationary growth phases. Proteins were deemed identified when detected with at least two unique peptides through automated searching and when present in at least two of the three biological replicate anti-RNA polymerase  $\beta$ -subunit antibody immunoprecipitations. Hierarchical clustering represented as heatmaps, and a 2D principal component analysis demonstrated that anti-RNA polymerase immunoprecipitations for the exponential and stationary growth phases of *M. smegmatis* were reproducible. The 2D PCA (Figure 4. 3) and heatmaps for exponential and stationary phase immunoprecipitations (Figure 4. 4) show that anti-RNA polymerase antibody immunoprecipitations clustered separately from the anti-myosin human antibody and protein G Dynabeads™ immunoprecipitation controls. The grey colouring in heatmaps indicates the absence of a protein within individual immunoprecipitations. Since the majority of the protein identified were not present

in the control immunoprecipitations, this suggests that the washes used were efficient in reducing non-specific binding. Proteins identified in protein G Dynabead™ and the anti-myosin human antibody control immunoprecipitations are deemed to be protein contaminants identified during immunoprecipitation.

High confidence proteins for both the exponential and stationary phases were identified when a protein was found to be present in at least two of the three biological replicate anti-RNA polymerase immunoprecipitations, but not in any of the protein G Dynabead™ or anti-myosin human antibody control immunoprecipitations. Proteins identified with only unique spectra were subjected to manual spectral inspection to confirm confident identification using mass spectrometry. Following this approach, 138 (Table C. 1) and 175 (Table C. 3) proteins were identified from the exponential and stationary growth phases of *M. smegmatis*, respectively.



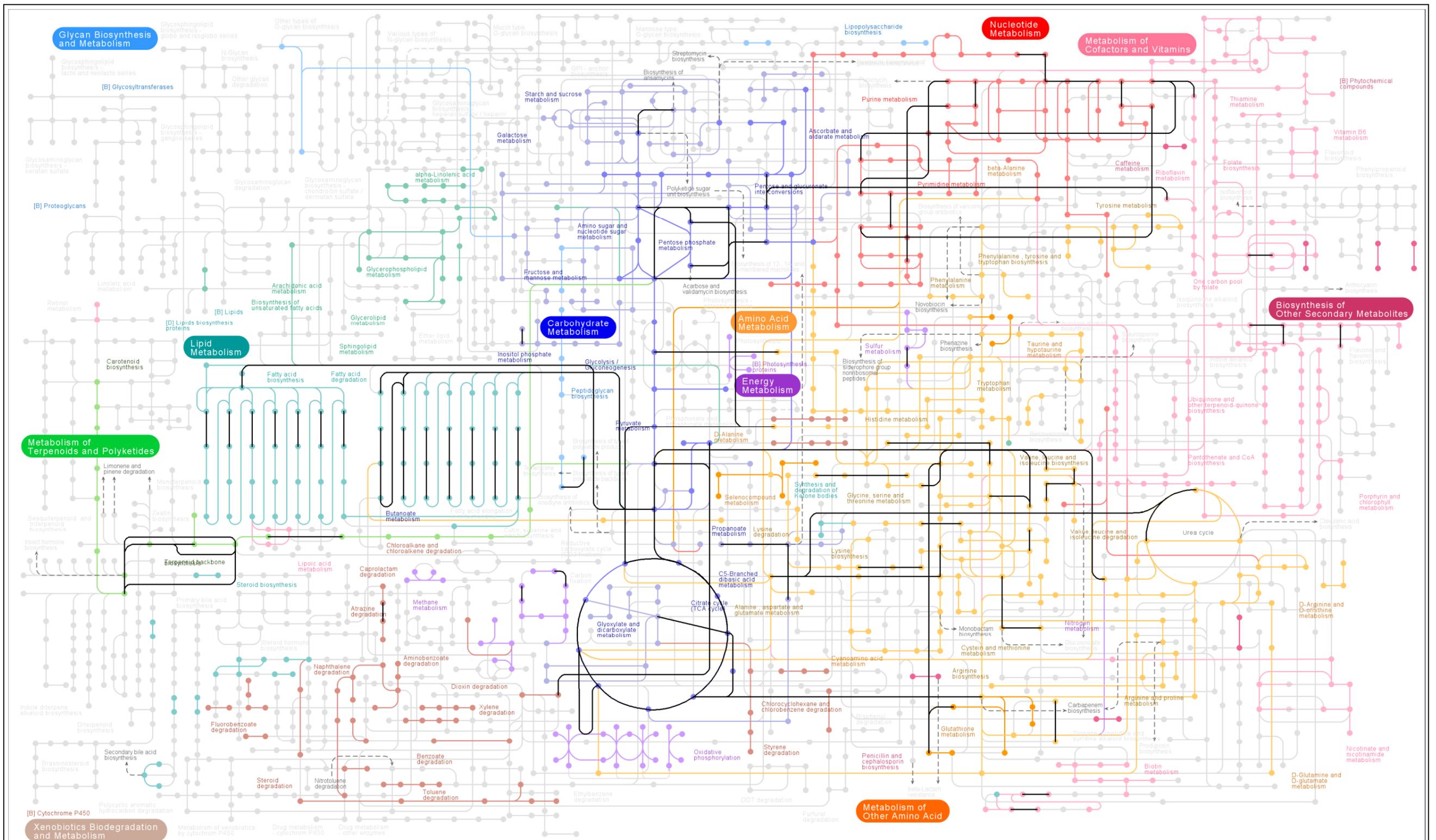


**Figure 4. 4 Heatmaps illustrating reproducibility of biological replicates for exponential and stationary phase immunoprecipitations.**

The heatmaps show the clustering of biological replicate experiments. Red is indicative of a higher intensity and green of a lower intensity of a protein within an immunoprecipitated sample. The grey colouring within the heatmap indicates the absence of a protein within that immunoprecipitation. The heatmaps of exponential (A) and stationary (B) growth phase immunoprecipitations show the clustering of the anti-RNA polymerase immunoprecipitation on the right and control immunoprecipitations (clustering left and center).

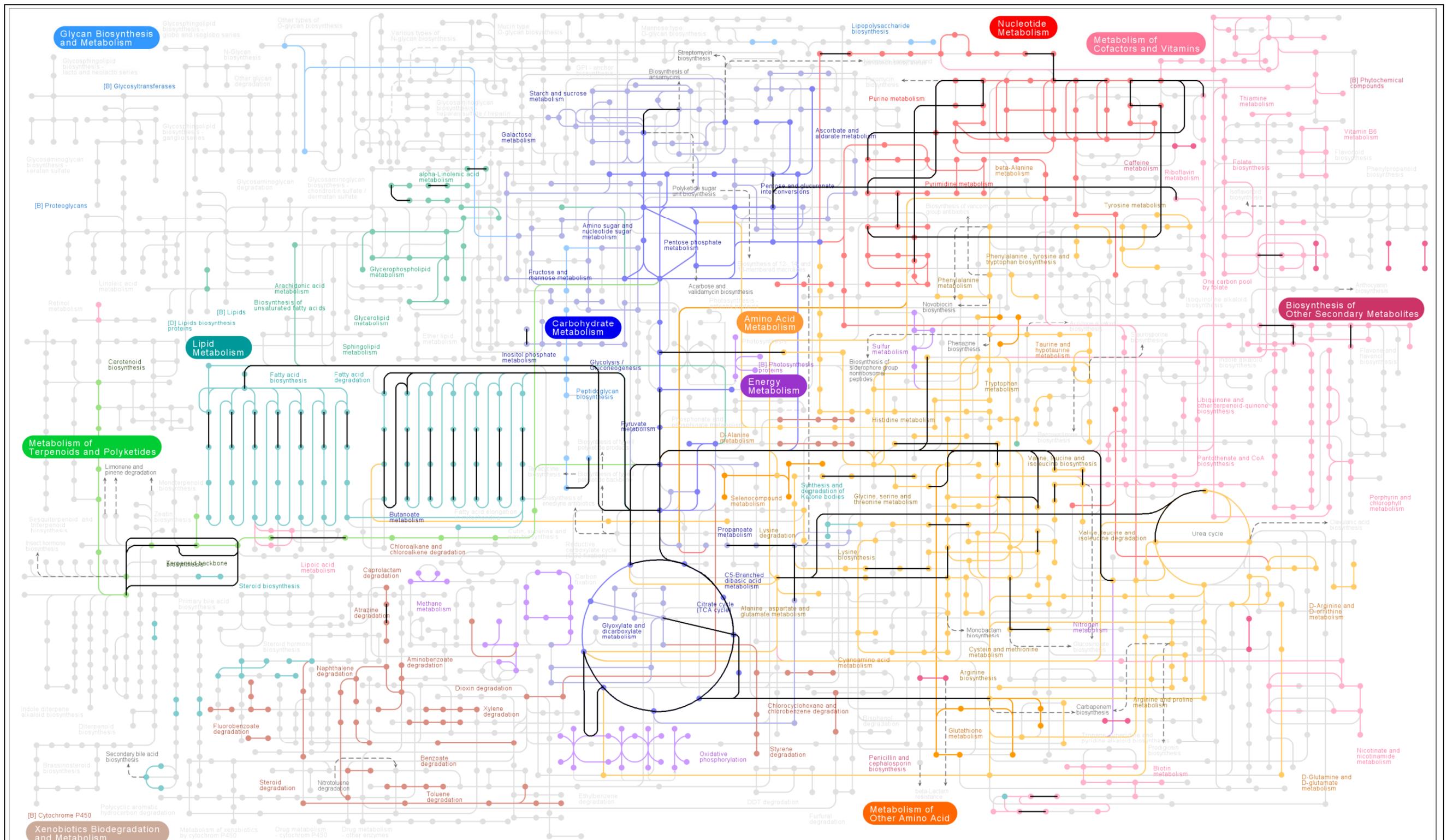
To estimate the relative proportion of nucleic acid associated proteins identified from the exponential and stationary growth phases, we made use of protein names and gene ontologies (obtained from UniProt) to search for the following identifiers: nucleo-, DNA, RNA, transcription, translation, ribosome, nucleic acid, chromosome, and ligase. This approach revealed that approximately 39.1% (54/138) of the proteins identified using NP-MS for the exponential growth phase were nucleic acid associated proteins. Similarly, approximately 42.8% (75/175) of the proteins identified from the stationary growth phase could be described using the identifiers searched. Despite the predicted proportion of nucleic acid associated proteins not being more than 50% of the total amount of proteins identified, other predicted functional attributes of proteins included GMP biosynthesis and “de novo” IMP biosynthetic process, which is related to purine biosynthesis for nucleic acids adenine and guanine.

Identified proteins for both the exponential and stationary phases were subjected to pathway mapping using KEGG. Fifty-six (Figure 4. 5) and 52 (Figure 4. 6) proteins from the exponential and stationary phase mapped to metabolic pathways, respectively. These pathway maps demonstrated that various proteins identified in either of these growth states using NP-MS were associated with not only nucleotide metabolism but also amino acid, carbon, lipid, and energy metabolism. These results demonstrate that during the affinity enrichment of the RNA polymerase complex as well as the DNA-associated proteins cross-linked to the bound DNA fragments, various other energy metabolism proteins are also isolated, possibly as a result of formaldehyde cross-linking of proteins in the vicinity.



**Figure 4. 5 Exponential phase protein metabolic pathway mapping**

High confidence proteins identified using NP-MS were subjected to pathway mapping using KEGG. Mapped pathways are represented in black and shows that NP-MS identified protein are involved in metabolic pathways associated with energy metabolism, lipid and carbohydrate metabolism, amino acid metabolism as well as nucleotide metabolism.



**Figure 4. 6 Stationary phase protein metabolic pathway mapping**

NP-MS stationary phase high confidence proteins were subjected to metabolic pathway mapping using KEGG. The figure shows that proteins isolated using NP-MS are predicted to be involved in energy metabolism, carbohydrate and lipid metabolism, amino acid metabolism and nucleotide metabolism.

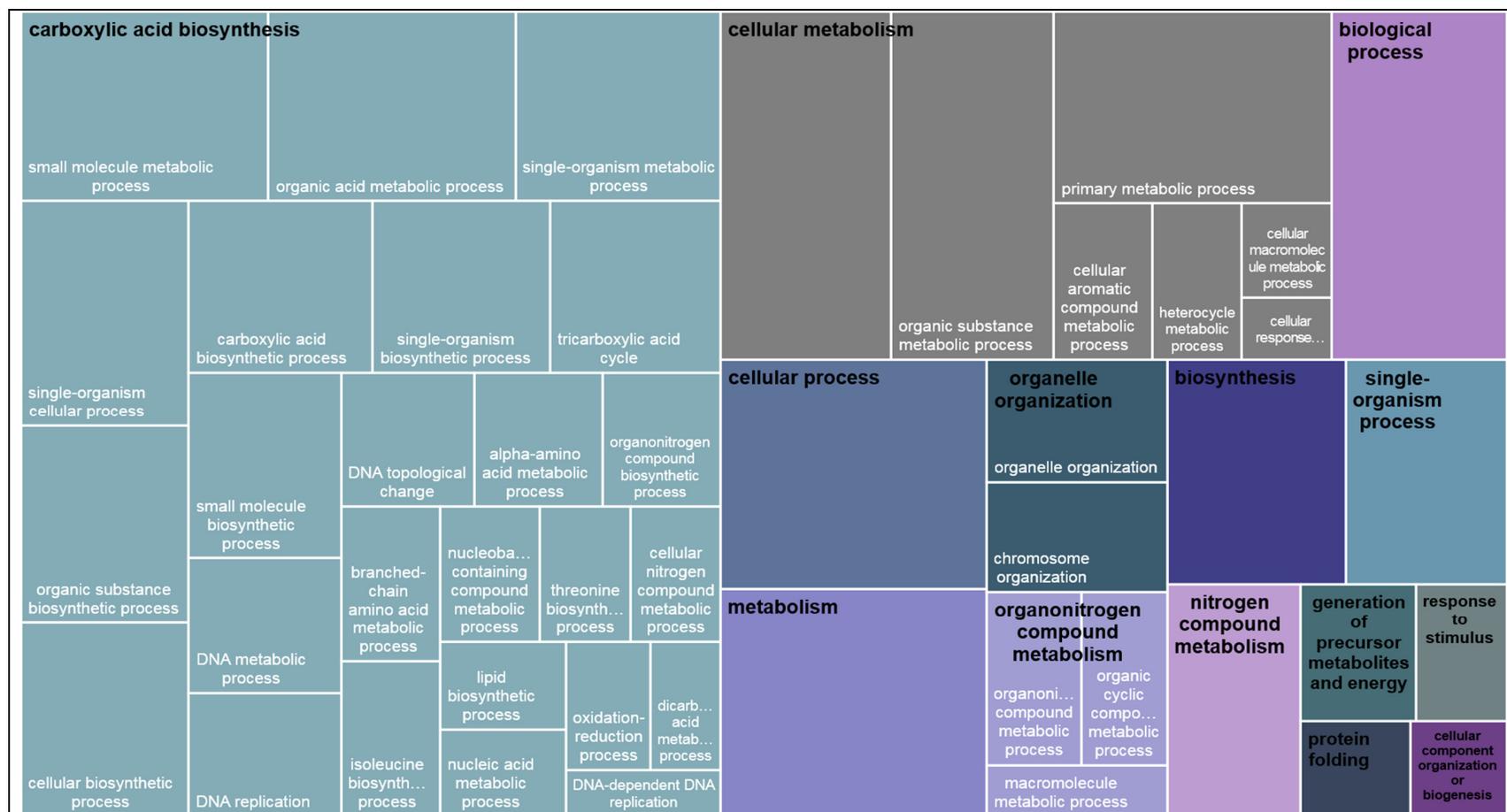
### 4.3.2. Gene Ontology Enrichment of Exponential and Stationary Phase Immunoprecipitations

#### *Gene Ontology Enrichment of Identified Proteins from the Exponential Growth Phase*

Exponential phase protein identifications were subjected to the online analysis tool, GOEAST to determine which GO identities were enriched within this dataset. Hypergeometric testing revealed that 114 GO identities were enriched and was also used to determine the frequencies of each of these GO terms within our dataset (Table C. 2). REVIGO classified 64 enriched GO identities under the functional category biological processes, another 46 under molecular function and the remaining 6 under the functional category cellular components. GO identity enrichment data was displayed as hierarchical treemaps for the three functional categories biological processes (Figure 4. 7), molecular function (Figure 4. 8), and cellular components (Figure 4. 9) based on the  $-\log_{10} p$ -values for each enriched GO identity.

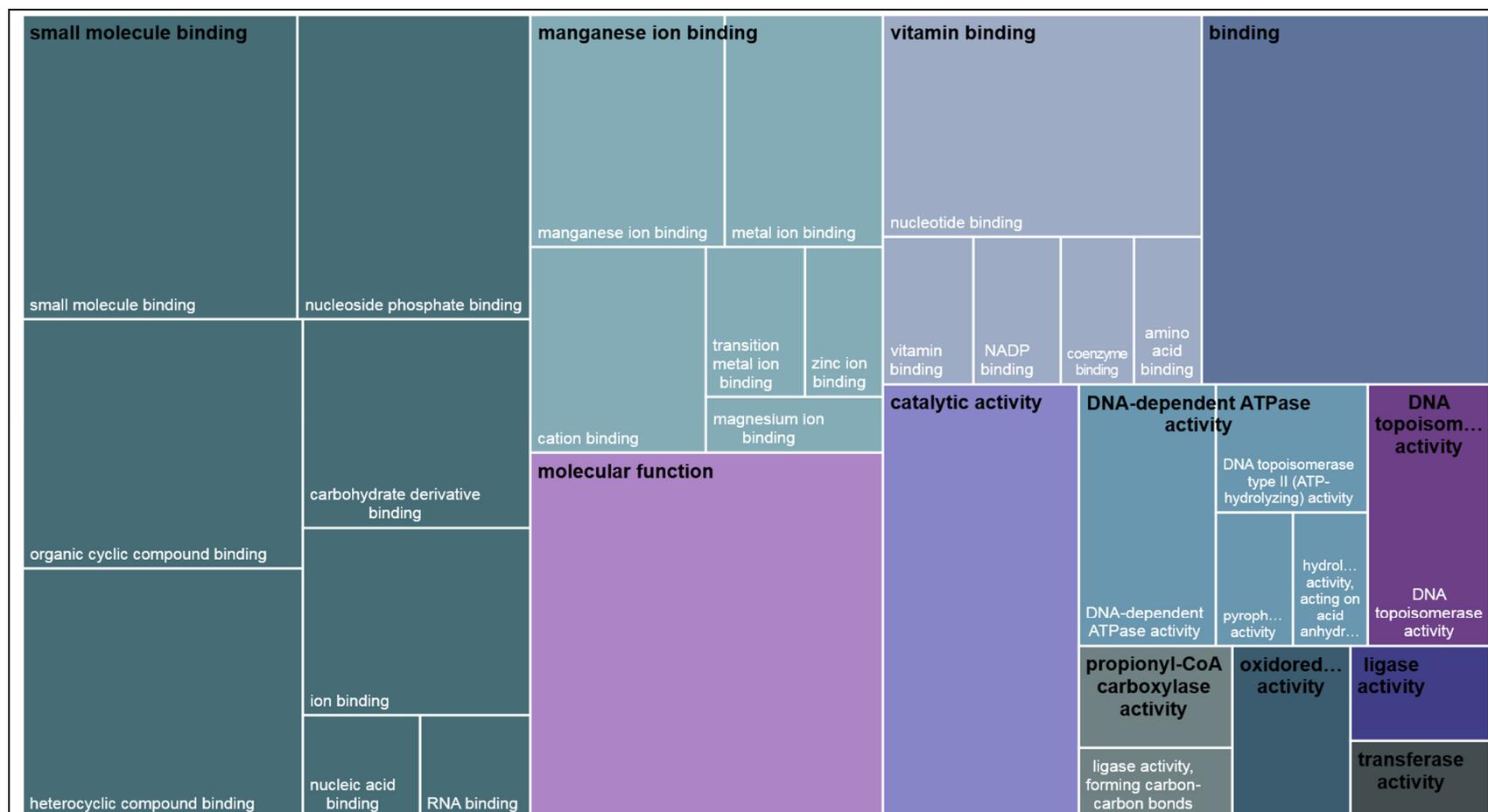
Hierarchical clustering within the functional category biological processes revealed that the higher hierarchy GO identity carboxylic acid biosynthesis was most enriched. This parent GO identity umbrellas GO terms associated with nucleic acids such as DNA metabolic process, DNA replication, DNA topological change and nucleic acid metabolic process. Other lower hierarchy GO terms associated with carboxylic acid biosynthesis included GO terms associated with amino acids such as alpha-amino acid metabolic process and branched chain-amino acid metabolic process as well as GO terms associated with energy metabolism such as tricarboxylic acid cycle and oxidation-reduction process. Other enriched higher hierarchy GO identities for this functional category included cellular metabolism, organonitrogen compound metabolism and organelle organisation. Within the functional category molecular function, the higher hierarchy GO identity small molecule binding was most abundant and envelopes lower hierarchy GO identities such as nucleoside binding, organic cyclic compound binding, carbohydrate derivative binding, nucleic acid binding and RNA binding. Other parent GO identities found to be enriched within this functional category included manganese ion binding, vitamin binding, DNA-dependent ATPase activity and DNA topoisomerase activity. Lastly, the most enriched parent GO identity within the functional category cellular components was chromosome, which enveloped the child GO terms cytoplasm and intracellular.

These treemaps provide a visual representation of the enriched GO identities and their predicted relationships which suggest that the proteins isolated from the exponential growth phase using NP-MS can be associated with the RNA polymerase transcription complex in *M. smegmatis*, either through its interaction with DNA or RNA or through energy metabolism which is required by the complex for translation. Furthermore, the cross-linking of proteins to nucleic acids allows for the identification of DNA-associated proteins such as DNA polymerases which are required for DNA replication as well as transcription factors.



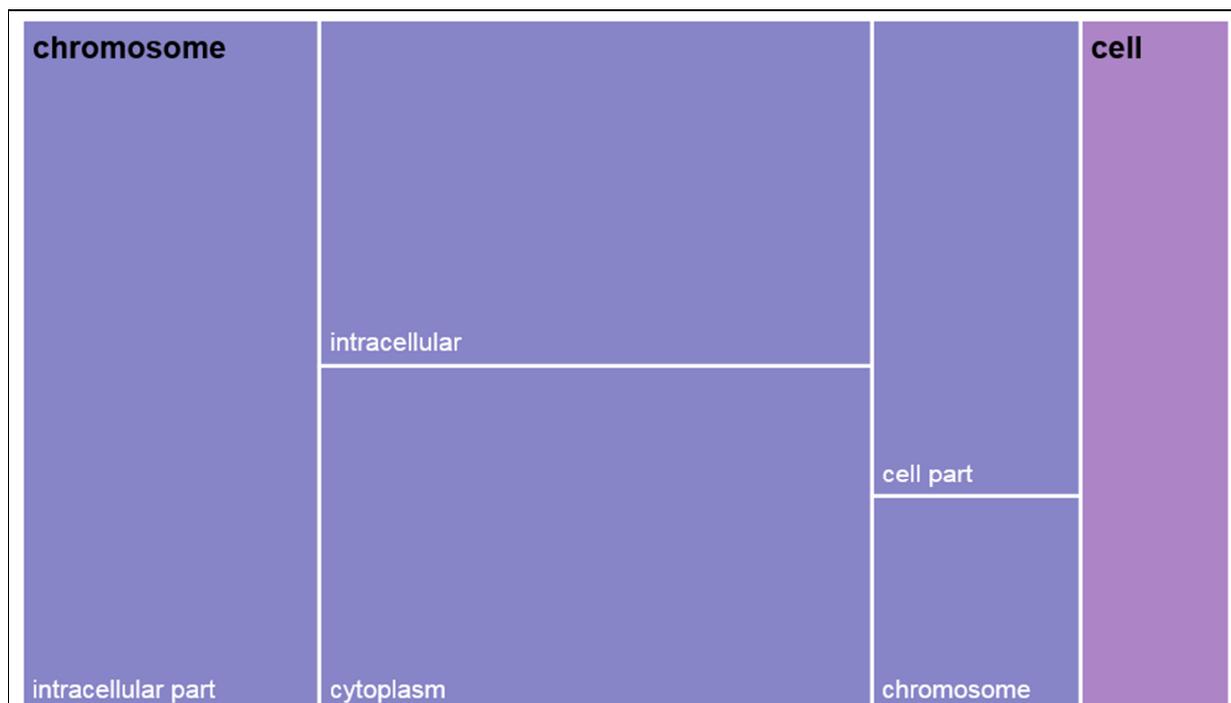
**Figure 4. 7 Hierarchical clustering of enriched biological processes GO identities for exponential phase NP-MS identified proteins**

Hierarchical clustering of enriched GO identities associated with biological processes revealed the parent GO term carboxylic acid biosynthesis to be the most enriched within this functional category. Other enriched higher hierarchy GO terms included cellular metabolism, organonitrogen compound metabolism, organelle organization, generation of precursor metabolites and energy and protein folding.



**Figure 4. 8 Hierarchical cluster of enriched molecular function enriched GO terms for exponential phase proteins**

The treemap of enriched GO identities associated with the functional category molecular function shows that the parent GO term small molecule binding was most enriched. This parent GO identity envelopes various child GO identities which can be associated with nucleic acid associated proteins. Other enriched parent GO identities included manganese ion binding, vitamin binding and DNA-dependent ATPase activity.

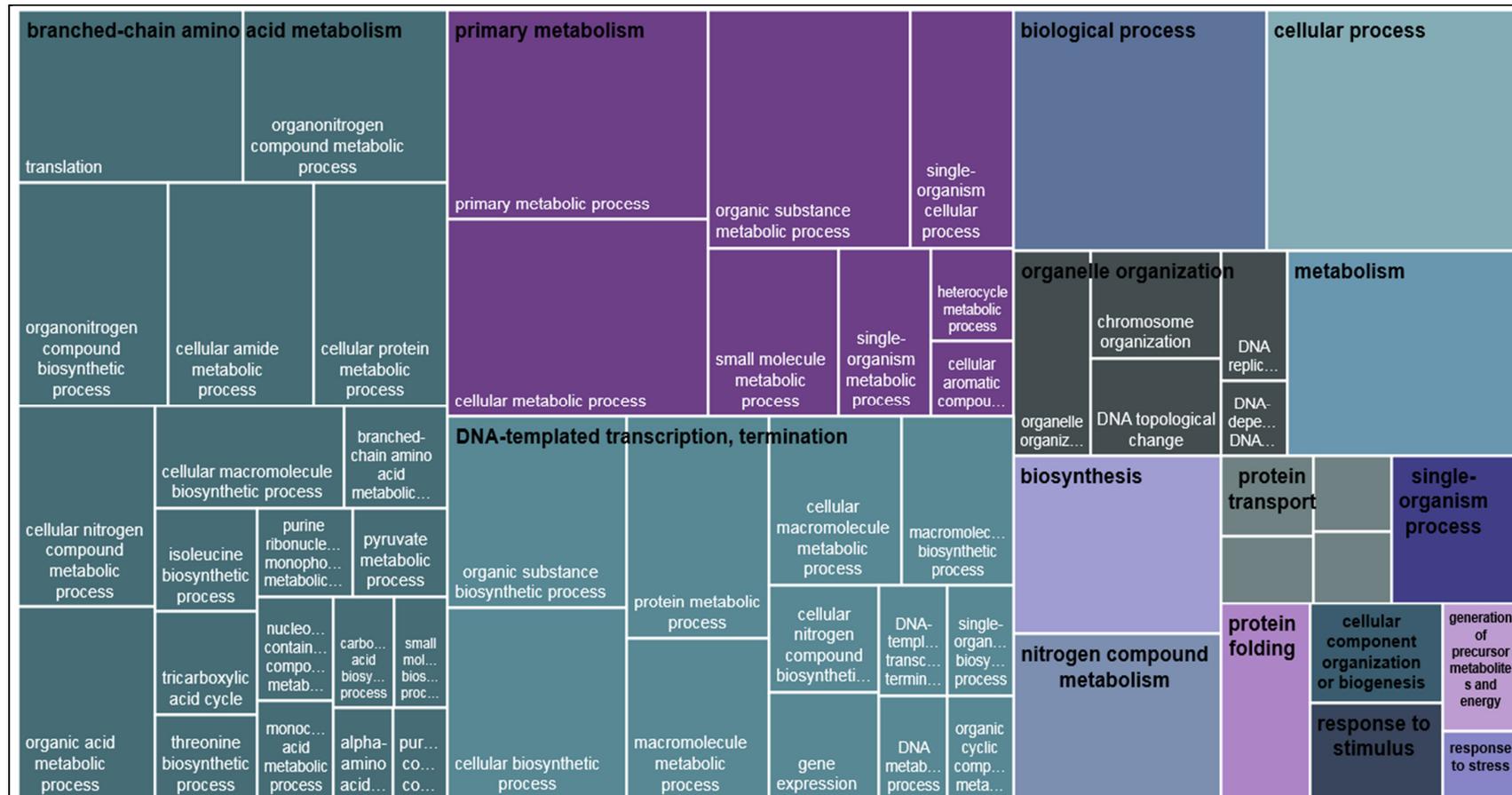


**Figure 4. 9 Hierarchical clustering of enriched cellular component GO terms for the exponential growth phase**

The treemap shows that chromosome was the most enriched GO term associated with the functional category cellular components.

#### *Gene Ontology Enrichment of Identified Proteins from the Stationary Growth Phase*

We went on to make use of hypergeometric testing of stationary phase protein identifications using GOEAST to identify 154 enriched GO identities and their respective frequencies within this dataset (Table C. 4). REVIGO categorized 92 of these enriched GO terms under the functional category biological processes, 45 as molecular function and the remaining 17 as cellular components. Categorized enriched GO identities were displayed as treemaps for the three functional categories biological processes (Figure 4. 10), molecular function (Figure 4. 11) and cellular components (Figure 4. 12).



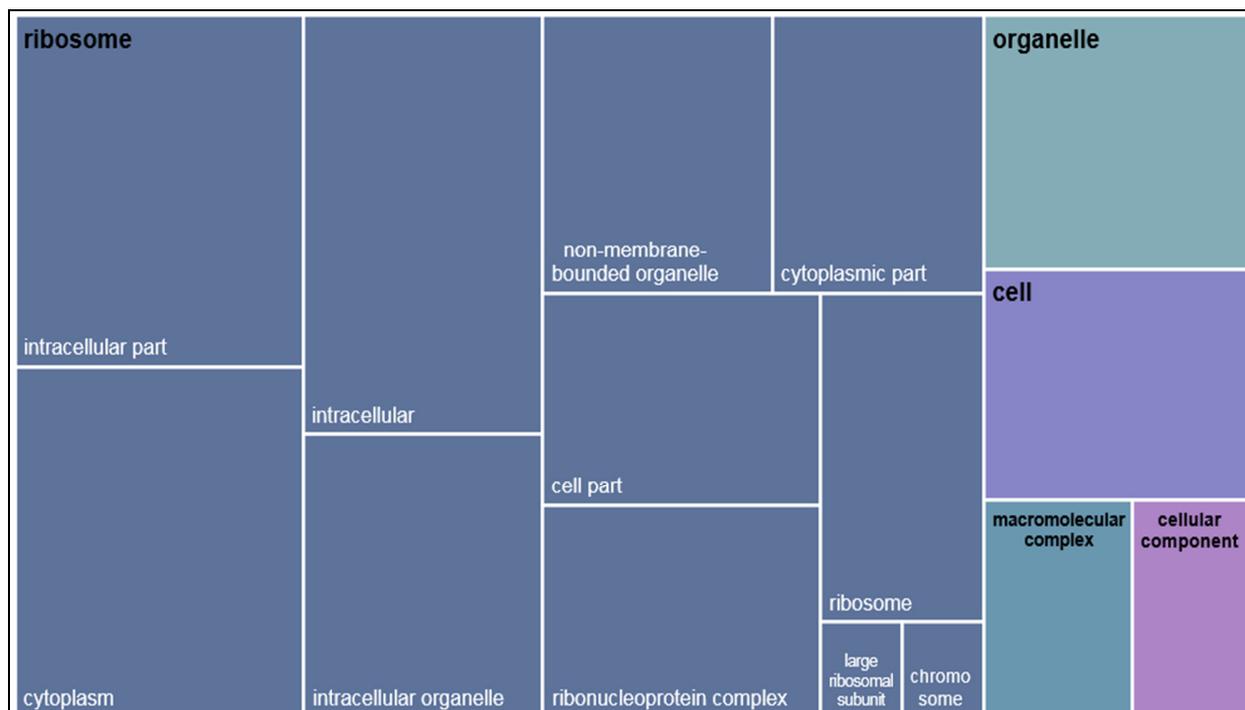
**Figure 4. 10 Hierarchical clustering of biological processes GO terms of stationary phase proteins**

Hierarchical clustering of enriched GO identities associated with biological processes revealed the parent GO term branched-chain amino acid metabolism was most abundant within this functional category and umbrellas the lower hierarchy GO terms translation, organonitrogen compound metabolic process and cellular protein metabolic process. Other higher hierarchy GO terms included primary metabolism and DNA templated transcription, termination and organelle organisation.



**Figure 4. 11 Hierarchical clustering of molecular function GO identities of stationary phase proteins**

The treemaps show that the higher hierarchy magnesium ion binding was most abundant within this functional category which envelopes child GO terms such as nucleoside phosphate binding, nucleotide binding and organic cyclic compound binding. Other parent GO terms found to be enriched within this functional category included rRNA binding, structural constituent of the ribosome and DNA-dependent ATPase activity.



**Figure 4. 12 Hierarchical clustering of cellular component GO terms of stationary phase proteins**

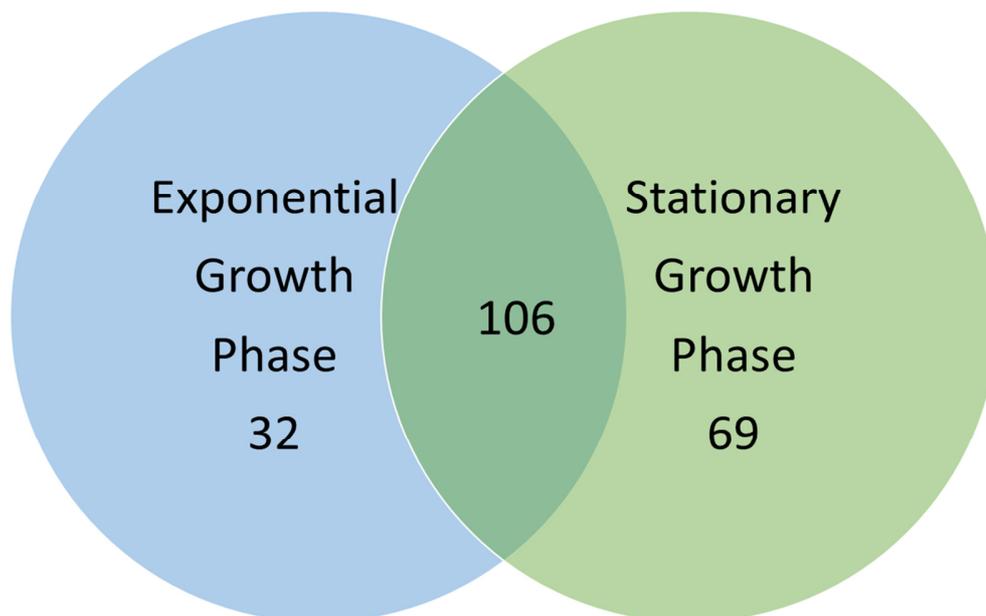
The higher hierarchy GO term ribosome were most abundant within this functional category and envelopes lower hierarchy GO terms ribosome, ribonucleoprotein complex, and chromosome.

Treemaps for the functional category biological processes demonstrated that the higher hierarchy GO identity branched-chain amino acid metabolism was most abundant within this functional category. Associated GO identities included translation, cellular protein metabolic process, and organonitrogen compound metabolic process which can all be associate with the RNA polymerase transcriptional complex. Furthermore, the energy metabolism GO term tricarboxylic acid cycle was also found to be enriched. Other parent GO terms associated with biological processes included primary metabolism and DNA-templated transcription, termination. Within the functional category molecular function, the higher hierarchy GO identity magnesium ion binding was most found to be most enriched and umbrellas lower hierarchy GO terms such as nucleoside phosphate binding, nucleotide binding, organic cyclic compound binding and nucleic acid binding. Other enriched parent GO identities within this functional category included rRNA binding, structural constituent of ribosome and DNA-dependent ATPase activity. Lastly, the most enriched parent GO identity within the functional category cellular components was ribosome, which enveloped the child GO terms cytoplasm, intracellular, ribonucleoprotein complex and non-membrane bound organelle.

Similar to what was observed for the proteins from the exponential phase, these treemaps showed that various predicted functional attributes of the proteins identified for the stationary phase can be related to the RNA polymerase transcriptional complex. These treemaps highlighted the enrichment of not only DNA and RNA associated functions but also energy metabolism processes required by the organism to allow for transcription and translation.

#### 4.3.3. Nucleoid-associated Proteins unique to each of the Growth Phases

Following the identification of proteins isolated using NP-MS for both the exponential and stationary growth phases, we sought to determine which of these proteins were unique to each growth phase. A two-dimensional venn diagram shows (Figure 4. 13) that 33 proteins were uniquely identified within the exponential growth phase (Table 4. 1) and a further 70 within the stationary growth phase (Table 4. 2). Comparison of the exponential and stationary growth phase proteins revealed that 107 proteins were identified in both growth phases.



**Figure 4. 13 Venn diagram of NP-MS identified exponential and stationary growth phase proteins**

The Venn diagram shows that the majority of the proteins identified by NP-MS are present in both the exponential and stationary growth phase, while 33 proteins were unique to the exponential growth phase and 70 were unique to the stationary growth phase.

**Table 4. 1 Unique proteins of the exponential growth phase**

UniProt	MSMEG Annotation	RV homologue	Gene Name	Protein Names	Gene Ontology
A0QNJ7	MSMEG_0067	Rv3876		Uncharacterized protein	
A0QP32	MSMEG_0255	Rv0211	<i>pckG</i>	Phosphoenolpyruvate carboxykinase	cytoplasm; gluconeogenesis; GTP binding; manganese ion binding; phosphoenolpyruvate carboxykinase (GTP) activity
Q3L885	MSMEG_0408		<i>pks</i>	Type I polyketide synthase	erythronolide synthase activity; oxidoreductase activity; phosphopantetheine binding; zinc ion binding
A0QQC8	MSMEG_0709	Rv0350	<i>dnaK</i>	Chaperone protein DnaK	ATP binding; protein folding
A0QQF9	MSMEG_0741			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0QQH2	MSMEG_0754	Rv0360c		Uncharacterized protein	
A0QQU2	MSMEG_0876			Short chain dehydrogenase	oxidoreductase activity
A0QR89	MSMEG_1028 MSMEG_2308			Geranylgeranyl reductase	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor
A0QR94	MSMEG_1033 MSMEG_2313	Rv3048c	<i>nrdF2</i>	Ribonucleoside-diphosphate reductase subunit beta	deoxyribonucleoside diphosphate metabolic process; deoxyribonucleotide biosynthetic process; DNA replication; metal ion binding; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor; ribonucleoside-diphosphate reductase complex
A0QS45	MSMEG_1346	Rv0640	<i>rplK</i>	50S ribosomal protein L11	large ribosomal subunit rRNA binding; ribosome; structural constituent of ribosome; translation
A0QS49	MSMEG_1350		<i>cmA1</i>	Cyclopropane-fatty-acyl-phospholipid synthase 1	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
A0QSL8	MSMEG_1524	Rv3457c	<i>rpoA</i>	DNA-directed RNA polymerase subunit alpha	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QUX1	MSMEG_2367	Rv3009c	<i>gatB</i>	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit B	ATP binding; carbon-nitrogen ligase activity, with glutamine as amido-N-donor; translation
A0QUZ0	MSMEG_2388	Rv2987c	<i>leuD</i>	3-isopropylmalate dehydratase small subunit	3-isopropylmalate dehydratase activity; 3-isopropylmalate dehydratase complex; leucine biosynthetic process
A0QWY0	MSMEG_3103	Rv1449c	<i>tkt</i>	Transketolase	metal ion binding; transketolase activity
A0QWY3	MSMEG_3106	Rv1454c		Quinone oxidoreductase	NADPH:quinone reductase activity; zinc ion binding
P71534	MSMEG_3150	Rv1483	<i>fabG</i>	3-oxoacyl-[acyl-carrier-protein] reductase	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity; fatty acid elongation; NADP binding

A0R033	MSMEG_4244	Rv2178c	<i>aroG</i>	Phospho-2-dehydro-3-deoxyheptonate aldolase	3-deoxy-7-phosphoheptulonate synthase activity; aromatic amino acid family biosynthetic process; chorismate biosynthetic process; manganese ion binding; plasma membrane
A0R061	MSMEG_4272	Rv2204c		Iron-sulfur cluster assembly accessory protein	iron-sulfur cluster assembly; iron-sulfur cluster binding; structural molecule activity
A0R069	MSMEG_4281	Rv2213	<i>pepA</i>	Probable cytosol aminopeptidase	aminopeptidase activity; cytoplasm; manganese ion binding; metalloexopeptidase activity
A0R095	MSMEG_4306	Rv2229c		Uncharacterized protein	
A0R0W1	MSMEG_4527	Rv2391		Ferredoxin sulfite reductase	4 iron, 4 sulfur cluster binding; heme binding; metal ion binding; sulfite reductase (ferredoxin) activity
A0R2B1	MSMEG_5049	Rv1248c	<i>kgd</i>	Multifunctional 2-oxoglutarate metabolism enzyme	2-hydroxy-3-oxoadipate synthase activity; 2-oxoglutarate decarboxylase activity; dihydrolipoyllysine-residue succinyltransferase activity; metal ion binding; oxoglutarate dehydrogenase (succinyl-transferring) activity; thiamine pyrophosphate binding; tricarboxylic acid cycle
A0R2U8	MSMEG_5240	Rv1098c	<i>fumC</i>	Fumarate hydratase class II	fumarate hydratase activity; fumarate metabolic process; tricarboxylic acid cycle; tricarboxylic acid cycle enzyme complex
A0R425	MSMEG_5680	Rv0887c		Glyoxalase family protein	dioxygenase activity
A0R451	MSMEG_5706	Rv0861c		DNA or RNA helicase of superfamily protein II	ATP binding; ATP-dependent DNA helicase activity; DNA binding; nucleotide-excision repair
A0R4S6	MSMEG_5937	Rv3534c	<i>bphI-2</i>	4-hydroxy-2-oxovalerate aldolase 2	4-hydroxy-2-oxovalerate aldolase activity; aromatic compound catabolic process; manganese ion binding
A0R561	MSMEG_6077	Rv3583c	<i>carD</i>	RNA polymerase-binding transcription factor	
A0R5Q2	MSMEG_6271	Rv3710	<i>leuA</i>	2-isopropylmalate synthase	2-isopropylmalate synthase activity; leucine biosynthetic process
A0R656	MSMEG_6431	Rv3849		Uncharacterized protein	sequence-specific DNA binding
A0R683	MSMEG_6458	Rv3858c	<i>gltD</i>	Glutamate synthase, small subunit	flavin adenine dinucleotide binding; glutamate biosynthetic process; iron-sulfur cluster binding; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor

A0R6Q9	MSMEG_6638	Rv1133c	<i>metE</i>	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity; methionine biosynthetic process; zinc ion binding
--------	------------	---------	-------------	--	--

**Table 4. 2 Unique proteins of the stationary growth phase**

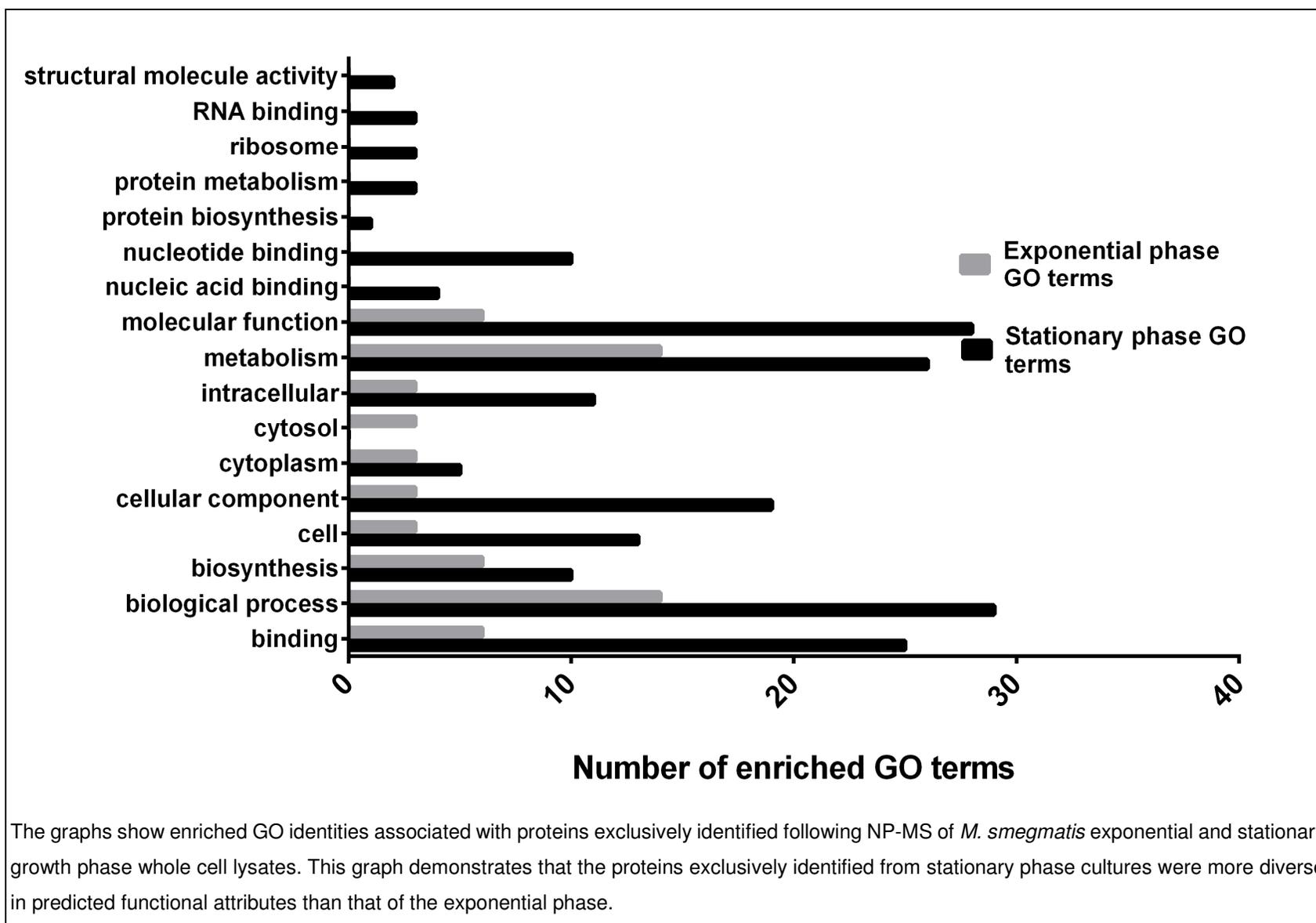
UniProt	MSMEG Annotation	RV homologue	Gene Name	Protein Names	Gene Ontology
A0QNF6	MSMEG_0024	Rv0009	<i>ppiA</i>	Peptidyl-prolyl cis-trans isomerase	peptidyl-prolyl cis-trans isomerase activity; protein folding
A0QNZ3	MSMEG_0216			3-hydroxyacyl-CoA dehydrogenase	oxidoreductase activity
A0QQB0	MSMEG_0690	Rv0338c		Putative iron-sulfur-binding reductase	iron-sulfur cluster binding
A0QRA5	MSMEG_1046 MSMEG_2326	Rv3041c		ABC transporter ATP-binding protein	ATPase activity; ATP binding
A0QRB8	MSMEG_1060			Putative Lsr2 protein (Uncharacterized protein)	
A0QS46	MSMEG_1347	Rv0641	<i>rplA</i>	50S ribosomal protein L1	large ribosomal subunit; regulation of translation; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QS62	MSMEG_1364	Rv0651	<i>rplJ</i>	50S ribosomal protein L10	large ribosomal subunit rRNA binding; ribosome; ribosome biogenesis; structural constituent of ribosome; translation
A0QSB1	MSMEG_1416	Rv0688		Pyridine nucleotide-disulphide oxidoreductase	cell redox homeostasis; ferredoxin-NAD <sup>+</sup> reductase activity; flavin adenine dinucleotide binding
A0QSD3	MSMEG_1438	Rv0703	<i>rplW</i>	50S ribosomal protein L23	nucleotide binding; ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSD8	MSMEG_1443	Rv0708	<i>rplP</i>	50S ribosomal protein L16	ribosome; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QSE0	MSMEG_1445	Rv0710	<i>rpsQ</i>	30S ribosomal protein S17	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSF9	MSMEG_1465	Rv0714	<i>rplN</i>	50S ribosomal protein L14	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSG1	MSMEG_1467	Rv0716	<i>rplE</i>	50S ribosomal protein L5	ribosome; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QSG3	MSMEG_1469	Rv0718	<i>rpsH</i>	30S ribosomal protein S8	ribosome; rRNA binding; structural constituent of ribosome; translation

A0QSG8	MSMEG_1474	Rv0723	<i>rpO</i>	50S ribosomal protein L15	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSL6	MSMEG_1522	Rv3459c	<i>rpsK</i>	30S ribosomal protein S11	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSP8	MSMEG_1556	Rv3443c	<i>rpM</i>	50S ribosomal protein L13	ribosome; structural constituent of ribosome; translation
A0QSS3	MSMEG_1582	Rv3418c	<i>groS</i>	10 kDa chaperonin (GroES protein)	ATP binding; cytoplasm; protein folding
A0QSS4	MSMEG_1583	Rv3417c	<i>groL2</i>	60 kDa chaperonin 2 (GroEL protein 2)	ATP binding; cytoplasm; protein refolding
A0QSU4	MSMEG_1603	Rv3410c		Uncharacterized oxidoreductase	oxidoreductase activity
A0QT20	MSMEG_1682			Flavin-containing monooxygenase FMO	monooxygenase activity
A0QTK6	MSMEG_1878	Rv3241c		Sigma 54 modulation protein/ribosomal protein S30EA	primary metabolic process; ribosome
A0QU00	MSMEG_2026			Short-chain dehydrogenase/reductase SDR	oxidoreductase activity
A0QUV6	MSMEG_2351	Rv3029c	<i>effB</i>	Electron transfer flavoprotein beta subunit	electron carrier activity
A0QUV7	MSMEG_2352	Rv3028c	<i>effA</i>	Electron transfer flavoprotein, alpha subunit	electron carrier activity; flavin adenine dinucleotide binding
A0QUY9	MSMEG_2387	Rv2988c	<i>leuC</i>	Isopropylmalate isomerase	3-isopropylmalate dehydratase activity; 4 iron, 4 sulfur cluster binding; leucine biosynthetic process; metal ion binding
A0QV14	MSMEG_2412	Rv2967c	<i>pyc</i>	Pyruvate carboxylase	ATP binding; biotin carboxylase activity; gluconeogenesis; metal ion binding; pyruvate carboxylase activity
A0QV37	MSMEG_2435	Rv2909c	<i>rpsP</i>	30S ribosomal protein S16	ribosome; structural constituent of ribosome; translation
A0QV42	MSMEG_2440	Rv2904c	<i>rplS</i>	50S ribosomal protein L19	ribosome; structural constituent of ribosome; translation
A0QVB9	MSMEG_2520	Rv2889c	<i>tsf</i>	Elongation factor Ts (EF-Ts)	cytoplasm; translation elongation factor activity
A0QVM7	MSMEG_2628	Rv2839c	<i>infB</i>	Translation initiation factor IF-2	cytoplasm; GTPase activity; GTP binding; translation initiation factor activity
A0QW25	MSMEG_2782	Rv2676c		Uncharacterized protein	
A0QWS8	MSMEG_3050	Rv1388	<i>mihF</i>	Integration host factor	nucleic acid binding
A0QXA3	MSMEG_3227	Rv1617	<i>pyk</i>	Pyruvate kinase	glycolytic process; magnesium ion binding; potassium ion binding; pyruvate kinase activity
A0QYB0	MSMEG_3595			Uncharacterized protein	
A0QYB1	MSMEG_3596			AAA ATPase	
A0QYG9	MSMEG_3654	Rv1821	<i>secA2</i>	Protein translocase subunit SecA 2	ATP binding; cytoplasm; intracellular protein transmembrane transport; plasma membrane; protein import; protein targeting

A0QYP9	MSMEG_3738	Rv1713	<i>engA</i>	GTPase Der (GTP-binding protein EngA)	GTP binding; ribosome biogenesis; transferase activity
A0QYW6	MSMEG_3811	Rv1636		Universal stress protein family protein, putative	cytoplasm; response to stress
A0QZ83	MSMEG_3932	Rv2031c	<i>hspX</i>	Heat shock protein hspX	
A0QZ96	MSMEG_3945	Rv1996 Rv2005c R2623		Universal stress protein family protein	response to stress
A0QZA1	MSMEG_3950	Rv2026c		Universal stress protein	response to stress
A0R012	MSMEG_4222	Rv2150c	<i>ftsZ</i>	Cell division protein FtsZ	barrier septum assembly; cell division site; cytoplasm; FtsZ-dependent cytokinesis; GTPase activity; GTP binding; protein complex; protein polymerization
A0R0B3	MSMEG_4326	Rv2244	<i>acpM</i>	Meromycolate extension acyl carrier protein	ACP phosphopantetheine attachment site binding involved in fatty acid biosynthetic process; cytoplasm
A0R0B4	MSMEG_4327	Rv2245	<i>kasA</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 1	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R0C7	MSMEG_4340	Rv2259	<i>adhE2</i>	Zinc-dependent alcohol dehydrogenase	oxidoreductase activity; zinc ion binding
A0R0I3	MSMEG_4396			Isochorismatase hydrolase	hydrolase activity
A0R0I8	MSMEG_4401			Phosphonoacetaldehyde hydrolase	hydrolase activity
A0R0Q9	MSMEG_4474			Acyl-CoA oxidase	acyl-CoA dehydrogenase activity; acyl-CoA oxidase activity; fatty acid beta-oxidation; flavin adenine dinucleotide binding; peroxisome
A0R102	MSMEG_4571	Rv2412	<i>rpsT</i>	30S ribosomal protein S20	ribosome; rRNA binding; structural constituent of ribosome; translation
A0R171	MSMEG_4646	Rv2455c		Pyruvate flavodoxin/ferredoxin oxidoreductase-like protein	2-oxoglutarate synthase activity
A0R1V9	MSMEG_4891	Rv2428		Alkyl hydroperoxide reductase subunit C	peroxidase activity; peroxiredoxin activity
A0R218	MSMEG_4954	Rv1297	<i>rho</i>	Transcription termination factor Rho	ATP binding; DNA-templated transcription, termination; helicase activity; regulation of transcription, DNA-templated; RNA binding; RNA-dependent ATPase activity
A0R248	MSMEG_4985	Rv1284		Carbonic anhydrase	carbonate dehydratase activity; zinc ion binding
A0R2T0	MSMEG_5222	Rv1112	<i>ychF</i>	Ribosome-binding ATPase	ATPase activity; ATP binding; GTP binding; ribosomal large subunit binding; ribosome binding

A0R2V2	MSMEG_5244	Rv3133c	<i>devR</i>	LuxR family two-component response regulator	DNA binding; phosphorelay signal transduction system; sequence-specific DNA binding transcription factor activity; transcription, DNA-templated
A0R2V4	MSMEG_5246	Rv2032	<i>acg</i>	Uncharacterized protein	oxidoreductase activity
A0R2V6	MSMEG_5248	Rv1094	<i>desA2</i>	Acyl-acyl-carrier protein desaturase	acyl-[acyl-carrier-protein] desaturase activity; fatty acid metabolic process
A0R3B8	MSMEG_5415	Rv1023	<i>eno</i>	Enolase	cell surface; extracellular region; glycolytic process; magnesium ion binding; phosphopyruvate hydratase activity; phosphopyruvate hydratase complex
A0R3D2	MSMEG_5431	Rv1015c	<i>rplY</i>	50S ribosomal protein L25 (General stress protein CTC)	5S rRNA binding; ribosome; structural constituent of ribosome; translation
A0R3M3	MSMEG_5524	Rv0952	<i>sucD</i>	Succinyl-CoA ligase	ATP binding; ATP citrate synthase activity; cofactor binding; succinate-CoA ligase (ADP-forming) activity
A0R4C9	MSMEG_5789	Rv3117 Rv0815c		Putative thiosulfate sulfurtransferase	thiosulfate sulfurtransferase activity
A0R4J1	MSMEG_5852	Rv0772	<i>purD</i>	Phosphoribosylglycinamide synthetase	de novo IMP biosynthetic process; ATP binding; magnesium ion binding; manganese ion binding; phosphoribosylamine-glycine ligase activity; purine nucleobase biosynthetic process
A0R577	MSMEG_6094	Rv3598c	<i>lysS</i>	Lysyl-tRNA synthetase	ATP binding; cytoplasm; lysine-tRNA ligase activity; lysyl-tRNA aminoacylation; magnesium ion binding; nucleic acid binding
A0R5H1	MSMEG_6189	Rv3676		Crp/Fnr family transcriptional regulator	DNA binding; intracellular; sequence-specific DNA binding transcription factor activity; transcription, DNA-templated
A0R5H5	MSMEG_6193	Rv3679		Anion-transporting ATPase	arsenite-transmembrane transporting ATPase activity; ATP binding
A0R628	MSMEG_6403	Rv3808c	<i>glfT2</i>	Galactofuranosyl transferase	capsule polysaccharide biosynthetic process; cell wall macromolecule biosynthetic process; cell wall organization; membrane; metal ion binding; transferase activity; transferase activity, transferring glycosyl groups
A0R7F9	MSMEG_6897	Rv0053	<i>rpsF</i>	30S ribosomal protein S6	ribosome; rRNA binding; structural constituent of ribosome; translation
A0R7G8	MSMEG_6907	Rv0044c		MmcI protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen

We next sought to describe the proteins exclusively identified in either of the different growth states. To this end, gene ontology enrichment analysis was done using GOEAST for each group of uniquely identified proteins. Enrichment revealed that 22 and 76 associated GO terms were enriched for the uniquely identified proteins of the exponential (Table C. 6) and stationary (Table C. 7) growth phases, respectively. The GO class counting tool, CateGORizer, was used to count and classify enriched GO terms whilst filtering out redundant GO identities for exponential and stationary phase (Figure 4. 14) enrichment data. These results demonstrated that the proteins uniquely identified from the stationary phase were more diverse in predicted functional attributes than those of the exponential phase. These results might be representative of more proteins being identified uniquely from the stationary phase than the exponential growth phase, thereby giving rise to more enriched GO terms. Interestingly, the GO terms uniquely enriched with the stationary growth phase included various GO identities associated with nucleic acids, such as RNA binding, ribosome, nucleotide binding, and nucleic acid binding.



#### 4.4. Discussion

---

Following the development of NP-MS to study nucleic acid associated proteins, we wished to investigate whether NP-MS could be used to identify nucleic acid associated proteins that are unique to specific environmental stimuli experienced by *M. smegmatis*. In this study, we explored the ability of NP-MS to identify proteins which may be required by *M. smegmatis* to adapt to the stationary growth phase. NP-MS was successfully used to identify 138 proteins from the exponential growth phase (Table C. 1) and a further 175 proteins from the stationary growth phase (Table C. 3). Identification of nucleic acid associated proteins with identifiers such as DNA, RNA, nucleic acid etc., revealed that approximately 39.1% and 42.8% of the proteins identified using NP-MS for the exponential and stationary growth phases, respectively, were associated with GO terms and protein names likely to be identifiers of nucleic acid associated proteins. Although the predicted number of nucleic acid associated proteins were found to be lower than 50% of the total number of NP-MS identified proteins for both growth states, metabolic pathway analysis revealed that other proteins isolated were predicted to be involved in energy, amino acid, carbon and lipid metabolic processes (Figure 4. 5, Figure 4. 6). The isolation and identification of these proteins are likely as a result of formaldehyde cross-linking fixing these associated proteins to the translational complex prior to immunoprecipitation.

Gene ontology enrichment analysis was done to describe the enriched predicted functional attributes of the proteins identified using NP-MS. These analyses were displayed as treemaps to visually demonstrate the enrichment of each of these GO identities, but to also demonstrate their grouping with other GO terms, suggesting an involvement in similar processes. Treemaps for exponential phase and stationary NP-MS identified proteins revealed that various enriched GO identities could be associated with nucleic acids as well as with processes such as energy metabolism and protein metabolic processes. These treemaps together with the pathway mapping demonstrate the ability of NP-MS to isolate and identify members of the RNA polymerase transcriptional complex as well as members of various metabolic pathways.

Comparison of the total number of protein identified between the two growth phases (207 proteins) revealed that 32 proteins were exclusively identified in the exponential growth phase (Table 4. 1) and a further 69 within the stationary growth phase immunoprecipitations (Table 4. 2). Gene ontology enrichment of the proteins exclusively identified in either of the growth states found that the proteins exclusively identified from stationary phase immunoprecipitations were more diverse in predicted functional attributes than those of the exponential growth phase (Figure 4. 14). Given the differences in the number of proteins uniquely identified in either of these growth states, it was unclear whether the diversity in predicted functional attributes were as a result of this discrepancy. However, upon closer inspection, we found that the enriched GO identities that are unique to the proteins associated with the stationary growth phase included GO terms such as RNA binding, nucleic acid binding, nucleotide binding, and ribosome. Enrichment of these GO

identities might be suggestive of transcriptional changes occurring within the organism which could allow it to adapt to the stresses experienced during the stationary growth phase.

#### 4.4.1. Uniquely identified proteins of the exponential phase

Transcriptomic analysis of aerated exponential and stationary phase cultures have revealed that the transcription of predicted open reading frames decline from 78.4% to 11% during the transition from exponential to stationary growth (154). Comparison of proteins uniquely identified from the exponential growth phase of *M. smegmatis* to transcriptomic of from *M. tuberculosis* exponential and stationary growth phases revealed that all *M. tuberculosis* homologues of proteins exclusively identified from the *M. smegmatis* exponential growth phase had more exponential phase sense reads than for the stationary phase (Table 3. 3). Correlation of these datasets demonstrated that NP-MS was effectively isolating proteins which were being transcribed during the exponential growth phase.

**Table 3. 3 Transcriptomic data of uniquely identified exponential phase *M. smegmatis* proteins and their *M. tuberculosis* homologues**

MSMEG annotation	Rv homologue	Exponential sense reads(154)	Stationary sense reads(154)	Ratio of reads
MSMEG_0255	Rv0211	5067	1109.5	4.57
MSMEG_0709	Rv0350	7297	291	25.08
MSMEG_0754	Rv0360c	224.6	15.5	14.49
MSMEG_1346	Rv0640	1769.8	197	8.98
MSMEG_5706	Rv0861c	485.6	57	8.52
MSMEG_5680	Rv0887c	62.4	12	5.20
MSMEG_5240	Rv1098c	2941	100	29.41
MSMEG_6638	Rv1133c	26381.4	484	54.51
MSMEG_5049	Rv1248c	4845.8	298.5	16.23
MSMEG_3103	Rv1449c	4970.2	358	13.88
MSMEG_3106	Rv1454c	579.4	93	6.23
MSMEG_3150	Rv1483	1001.8	49	20.44
MSMEG_4244	Rv2178c	1998.6	153.5	13.02
MSMEG_4272	Rv2204c	1050.6	117.5	8.94
MSMEG_4281	Rv2213	1664.8	130	12.81
MSMEG_4306	Rv2229c	229.6	6	38.27
MSMEG_4527	Rv2391	8007.8	89	89.98
MSMEG_2388	Rv2987c	142.8	5	28.56
MSMEG_2367	Rv3009c	796.6	83.5	9.54
MSMEG_1033/2313	Rv3048c	682.4	38.5	17.72
MSMEG_1524	Rv3457c	4103	82	50.04
MSMEG_5937	Rv3534c	133.8	43.5	3.08
MSMEG_6077	Rv3583c	2577.2	327.5	7.87
MSMEG_6271	Rv3710	1852	239	7.75

<i>MSMEG_6431</i>	<i>Rv3849</i>	718.6	40	17.97
<i>MSMEG_6458</i>	<i>Rv3858c</i>	914	18.5	49.41
<i>MSMEG_0067</i>	<i>Rv3876</i>	7823.8	659.5	11.86

A further comparison of proteins exclusively identified in the exponential growth phase of *M. smegmatis* to transcriptomic data of *M. tuberculosis* exponential growth revealed that MetE, encoded by *MSMEG\_6638* and *Rv1133c*, was the 16<sup>th</sup> most abundant transcript in exponential phase *M. tuberculosis* cultures (154). MetE is a predicted methyltransferase which has been demonstrated to be required for *in vitro* *M. tuberculosis* growth (48). MetE is characterised by a cobalamin-independent methionine synthase domain which catalysis the synthesis of the amino acid methionine through the transfer of a methyl group to the  $\alpha$ -amino acid homocysteine which is a homologue of the amino acid cysteine (197).

#### 4.4.2. NP-MS Identifies Universal Stress Response Proteins in the Stationary Phase

Stationary phase bacteria have been shown to be more resilient to environmental stresses such as heat shock, oxidative stress and osmotic challenge (198–200). We therefore aimed to identify stress proteins exclusively associated with the stationary growth phase. Six possible stress proteins were identified using GO terms and protein names obtained from UniProt (Table 4. 4). *M. tuberculosis* encodes for eight USPs (*rv1636*, *rv2028c*, *rv2624c*, *rv3134c*, *rv1996*, *rv2005c*, *rv2026c* and *rv2623*) and NP-MS successfully identified three of the five *M. smegmatis* homologues in this study namely *MSMEG\_3811*, *MSMEG\_3945* and *MSMEG\_3950* (201).

**Table 4. 4 NP-MS stress proteins identified from the stationary growth phase**

UniProt ID	MSMEG ID	Rv Homologue	Gene Name	Description	Gene Ontology
A0QTK6	MSMEG_1878	Rv3241c	<i>hpf</i>	Ribosome hibernation promoting factor	primary metabolic process; ribosome
A0QYW6	MSMEG_3811	Rv1636		Universal stress protein family protein, putative	cytoplasm; response to stress
*A0QZ83	MSMEG_3932	Rv2031c	<i>hspX</i>	Heat shock protein HspX	
*A0QZ96	MSMEG_3945	Rv1996, Rv2005c, Rv2623		Universal stress protein family protein	response to stress
A0QZA1	MSMEG_3950	Rv2026c		universal stress protein family protein	response to stress
*A0R2V2	MSMEG_5244	Rv3133c	<i>devr</i>	Two component transcriptional regulatory protein devr	DNA binding; phosphorelay signal transduction system; sequence-specific DNA binding transcription factor activity; transcription, DNA-templated

\* Genes of the devR regulon

Most notably, the well-studied DevR, which forms part of the DevR-DevS two-component regulatory system, was identified using NP-MS (202). In *M. tuberculosis* DevR, also known as the dormancy response regulator DosR, has been shown to be under the control of two kinases, DevS (Rv3132c) and DosT (Rv2027c) (203). Phosphorylation of DevR has been demonstrated to result in the transcriptional regulation of the DosR regulon, comprising of 48 genes under hypoxic, nitric oxide, carbon monoxide, and redox stress conditions in *M. tuberculosis* (204–211). DevR was reported to be absent from aerated stationary phase cultures in the past however, in this study and a transcriptomic analysis study, DevR as well as members of the DosR regulon have been shown to be present in lysates of *M. smegmatis* and *M. tuberculosis*, respectively (154, 205). Upon further inspection, we discovered that NP-MS identified five members of the DosR regulon of which four (MSMEG\_3932, MSMEG\_3945, MSMEG\_5244 and MSMEG\_5246) were only found in the stationary phase of NP-MS identified proteins (Table 4. 4). The *ahpC* homologue, MSMEG\_4891, was identified in both the exponential and stationary growth phase NP-MS immunoprecipitations, despite its predicted role in oxidative stress response (Table C. 1 and Table C. 3) (212).

Two DevR regulated proteins were also identified as stressed proteins in Table 4. 4, heat shock protein (HspX) (MSMEG\_3932) and MSMEG\_3945. Despite its name, *hspX* was demonstrated to not be upregulated by heat shock and was found to not be required for *M. tuberculosis* growth *in vitro*, however a *M. tuberculosis hspX* replacement mutant displayed impaired growth in bone-marrow-derived macrophages and THP-1 cells (48–50, 95, 213–216). Transcriptional analysis has revealed that *hspX* is the most abundant coding transcript in aerated stationary phase cultures and that expression of this stress protein

increases during the transition from exponential to stationary growth (154, 217, 218) Furthermore, HspX has been shown to be induced under several stress conditions, including oxygen and nutrient starvation (4, 186). These results suggest a role for HspX in the adaptation to stress conditions similar to what is predicted to be experienced by *M. tuberculosis* during latent mycobacterial infections or dormancy. The cellular localisation of HspX to the cell membrane has led researchers to speculate that the increased expression of this stress protein results in a thickened cell wall which may help mycobacteria adapt to long-term survival (186). The second DevR regulated protein detected using NP-MS, MSMEG\_3945, has three *M. tuberculosis* homologues, Rv1996, Rv2005c and Rv2623 and all are annotated as universal stress proteins. Similar to HspX, *rv2623* has been demonstrated to be upregulated under oxygen limiting conditions and nitrosative stress (204, 206, 208, 219, 220). Infection models have also been used to show upregulation of *rv2623* in human and mouse derived macrophages as well as in mice with persistent *M. tuberculosis* infection (105, 221, 222). Furthermore, transcriptomic analysis has revealed *rv2623* to be the fifth and *rv1996* the 17<sup>th</sup> most abundant transcripts within aerated *M. tuberculosis* stationary phase cultures (154). Functional studies have revealed that deletion of *rv2623* resulted in a hypervirulent phenotype whilst overexpression resulted in attenuation of mycobacterial growth (223). These results, together with the confirmed ATP-binding activity of Rv2623, has suggested that Rv2623 possibly regulates mycobacterial growth in an ATP-dependent manner (223). Interestingly, ectopic overexpression of HspX has also resulted in an attenuated growth phenotype, further suggesting that these two proteins have similar growth-regulatory functions in mycobacterial dormancy (217, 223). The third DevR regulated protein identified from the stationary phase in *M. smegmatis*, MSMEG\_5246, is a conserved hypothetical protein containing a nitroreductase-like domain with a predicted oxidoreductase activity (Table 4. 2). The *M. tuberculosis* homologue, *rv2032* also known as *acg*, has been shown to be upregulated under hypoxic growth conditions and during macrophage infection (211, 224). Deletion of *Acg* from *M. tuberculosis* has been shown to result in loss of viability in macrophages as well as in both immunocompetent and incompetent murine infection models (225). Investigation into the possible nitroreductase activity of *Acg* has revealed the *acg* deletion mutant to be uncompromised under nitrosative and oxidative stress conditions, whilst showing that *acg* deletion mutant was shown to be more sensitive than the wild type strain to prodrugs nitrofurantoin and nitrofurantoin (225). These data suggested that *rv2032* does not function as nitroreductase but is an essential virulence factor required for mycobacterial viability during infection.

Further comparison of aerated *M. tuberculosis* stationary phase culture transcripts with *M. smegmatis* homologue proteins isolated using NP-MS revealed Rv3241c (MSMEG\_1878) to also be one of the 50 most abundant stationary phase transcripts (154). Interestingly, MSMEG\_1878, also known as the ribosome hibernation promoting factor (Hpf), was also identified as a stress related protein in Table 4. 4. This sigma 54/S30AE modulation protein has been demonstrated to be induced under carbon starvation and comparison of exponential and stationary growth phase *M. smegmatis* cultures revealed Hpf to be a stationary phase protein (4, 226, 227). Sigma 54/S30AE modulation proteins are known to promote

dimerization of 70S ribosomes in *E. coli*, leading to the formation of 100S ribosomes during the stationary phase. The formation of these 100S ribosome units has been termed “ribosomal hibernation”, which has been shown to result in the inhibition of translation and the stabilisation of mRNA (228, 229). Unlike in *E. coli*, Hpf acts to stabilise the 70S ribosomal units (consisting of 50S and 30S subunits) in *M. smegmatis*. Interestingly, the amount of detected Hpf was shown to be lower than the amount of ribosomal proteins with which it associates, resulting destabilisation of the 70S ribosomal units and an increase in the detectable levels of 50S and 30S ribosomal subunits in the stationary growth phase (227). Even though Hpf does not promote dimerization of 70S ribosomal units like seen in *E. coli*, mycobacterial mRNA has been demonstrated to stabilise with transition into the stationary growth phase (189). Interestingly, the 50S ribosomal protein L25 (MSMEG\_5431), also known as RplY, was also identified in the stationary growth phase using NP-MS (Table 4. 2). This structural component of the ribosome has been designated a general stress protein due to similarities to Ctc, a *Bacillus subtilis* stress protein (230).

The last possible stress protein identified in Table 4. 4, MSMEG\_3811, is predicted to be a nucleotide binding protein which contain a universal stress protein A domain (231). *E. coli uspA* has been shown to induced under conditions of growth arrest and *uspA* deficient cells have demonstrated poor stationary phase survival (231). UspA has been shown to be an auto-phosphorylating serine/threonine phosphoprotein which physically protects DNA from UV radiation and mitomycin C through physically interacting with the DNA molecule and acting as a shield (231, 232). UspA domain containing proteins have been suggested to have similar functions in dormant mycobacteria as reduced mitomycin c effects have been demonstrated in hypoxic dormant *M. bovis* BCG (233). Protein homology has also identified UspA domains in two other stress proteins identified using NP-MS, MSMEG\_3945 and MSMEG\_3950 (Table 4. 4). Functional studies have revealed that MSMEG\_3811 and its *M. tuberculosis* homologue Rv1636, binds cyclic AMP (cAMP), a known universal stress protein ligand, with a higher affinity than ATP (234). Furthermore, the affinity of Rv1636 and MSMEG\_3811 for cAMP is higher than that of CRP and that of the cAMP-regulated protein acyltransferase Rv0998 (235, 236). These results, together with the knowledge that Rv1636 is the twentieth most abundant protein in the *M. tuberculosis* proteome and the previous identification of MSMEG\_3811 in the early exponential growth phase of *M. smegmatis* (Chapter 3, Table B. 2) suggests that this universal stress protein (USP) is not required for adaptation to the stationary growth phase but may rather be required for the binding of cAMP (58, 234). Inspection of the exponential phase immunoprecipitation data revealed MSMEG\_3811 to have been isolated in all three biological replicate experiments for the exponential growth phase however, due to detection in one of the six control immunoprecipitations, MSMEG\_3811 was not identified as an NP-MS exponential phase protein (data not shown). MSMEG\_3811 and Rv1636 have been proposed to bind to cAMP, which mycobacterial species have been shown to express in high volumes, to regulate downstream cAMP actions prior to secretion (234, 237–239). The secretion of cAMP has been proposed to modulate the host macrophage environment during infection (234).

Comparison of proteins uniquely identified from the stationary growth phase of *M. smegmatis* to transcriptomic data from *M. tuberculosis* exponential and stationary growth phases revealed that all homologues of stress proteins identified in Table 4. 4 except for Rv1636, had more sense reads in stationary phase cultures compared to exponential phase *M. tuberculosis* cultures (154). Many of the proteins exclusively identified from stationary phase cultures were not demonstrated to have more sense reads in the stationary phase when compared to the exponential growth phase cultures, indicating that these proteins are no longer being actively transcribed. The identification of these proteins may be as a result of stationary phase cells inheriting these proteins which were produced in large quantities during the exponential growth phase. Furthermore, it is noteworthy to mention that in our study we investigated the proteins associated with the early stationary phase in *M. smegmatis* however, the transcriptomics study investigated the transcriptome of late stationary phase *M. tuberculosis*. Discordances between transcriptome and proteome studies are well known and is possibly the result of posttranscriptional/posttranslational changes (240, 241). Therefore, the comparison of these two datasets emphasises the need for multiple investigative strategies being used to study the physiological changes associated with growth.

Notably, no starvation-induced proteins, such as described by Betts *et al.*, were identified in proteins unique to the stationary phase (4). This is not surprising as models used to investigate starvation in *M. smegmatis* are known to culture this organism in PBS for 14 days (242). Likewise, studies where starvation has been investigated for *M. tuberculosis* has reported culturing in PBS for a minimum of 6 weeks (4, 243). Investigation of mycobacterial starvation is made difficult by the reported abilities of *M. tuberculosis* to store both glycogen and lipids (244, 245). Furthermore, *M. tuberculosis* is known to encode two iron storage proteins, BfrA and BfrB, which have also been shown to provide protection against oxidative stress and hypoxia (246).

#### 4.4.3. Limitations

In this study, we showed that NP-MS could be used to identify proteins unique to environmental stresses experienced by *M. smegmatis*. Specifically, in this study we made use of NP-MS to identify stress proteins associated the mycobacterial stationary phase. Unfortunately, there is no way of quantifying which of the proteins identified using NP-MS were being actively translated and which of the identified proteins were associated with the targeted RNA polymerase transcriptional complex.

As discussed before, various detergents such as SDS, Triton X-100, and Nonidet P-40 as well as salts (NaCl, sodium deoxycholate and LiCl) were used in wash buffers to reduce non-specific binding during immunoprecipitation of nucleoprotein complexes. These detergents are known to influence the ability of peptide ionization, effectively limiting their detection and subsequent identification through mass spectrometry (169, 170). Furthermore, these detergents are detrimental to optimal functioning of the mass

spectrometer as these compounds could potentially result in the blockage of the liquid chromatography column. As such, peptide samples prepared using NP-MS were subjected to chloroform-methanol phase separation to remove detergents introduced to the samples (247). Detergent removal using methanol-chloroform phase separation does potentially result in the loss of peptides from the sample, thereby limiting our use of the LFQ intensity values of identified proteins for the identification of low confidence proteins. Furthermore, the controls used in this study are crucial for the identification of low confidence proteins following stringent filtering of high confidence protein identifications. The inability to identify any proteins in three of the control immunoprecipitations for the exponential phase resulted in the failure to replace missing values through separate imputation for each column as done in Chapter 3, section 3.2.4.15. It is unclear whether the loss of protein identification in these control immunoprecipitations were solely as a result of the detergent removal step included, as there were proteins identified in all six control immunoprecipitations for stationary phase NP-MP prepared samples.

Overall the number of high confidence proteins identified was lower than for Chapter 3. This discrepancy could be due to several factors. Firstly, due to unforeseen circumstances, a different supplier of the anti-RNA polymerase antibody was used for experiments in Chapter 4 than was used for Chapter 3. The replacement antibody, supplied by Biolegend, was also raised against the 8RB13 epitope of the *E. coli* RNA polymerase  $\beta$ -subunit, however differences in sensitivity cannot be excluded. Secondly, a blank sample was run in between every NP-MS fraction during mass spectrometry analysis in Chapter 4. This was done to reduce build-up of detergents on the liquid chromatography column which could have potentially resulted in a blockage. For analysis in Chapter 3, blank runs were only included between batches of biological replicates i.e. no-antibody control replicates, non-specific antibody replicates and anti-RNA polymerase  $\beta$ -subunit replicates. It is unclear how the detergents used in this study and the condition of the column used for liquid chromatography along with the increased mass spectrometry analysis time influenced the retention time of peptides on the column, thereby potentially limiting the identification of proteins following mass spectrometry analysis.

#### 4.5. Conclusion

---

In this study, we successfully used NP-MS to identify three USPs in *M. smegmatis* of which two (MSMEG\_3945 and MSMEG\_3950) were supported by literature to be required for adaptation to stress conditions in the stationary growth phase. In *M. tuberculosis*, USPs have been proposed to be required for the adaptation to stresses such as hypoxia, nitric oxide, carbon monoxide as well as for survival within the host (208–211, 215, 248). Many of these USPs have been shown to not be required for *in vitro* growth despite their upregulation under adverse growth conditions, this is possibly due to these proteins being functionally redundant (249). Furthermore, NP-MS also identified HspX, which was demonstrated to be the most abundant stationary phase transcript from aerated stationary phase *M. tuberculosis* cultures as well as Hpf, another abundant transcript from the *M. tuberculosis* stationary phase (154). Most notably, NP-MS successfully isolated and identified the DevR transcriptional regulator from stationary phase cultures which regulate the expression of approximately 48 genes under an array of stress conditions (204–207, 209–211). These results, therefore, demonstrate the ability of NP-MS to effectively identify proteins required for adaptation to environmental stress conditions.

---

## Chapter 5: Summary

---

### 5.1. Conclusion

---

Understanding how complex regulatory changes are mediated in mycobacteria in response to changing growth conditions, is crucial in understanding how pathogenic organisms like *M. tuberculosis* adapt to host environments. Growth limiting conditions such as oxygen deprivation, nutrient starvation, nitrosative stress, and pH shock have been shown to result in complex transcriptional changes (4–7). These intricate transcriptional responses are likely mediated through several regulatory components, however, the methodologies employed to investigate these components are limited in that they focus on the study of single regulatory elements. In this study, we proposed that affinity chromatography together with mass spectrometry could be used to identify and investigate members of the RNA polymerase transcriptional complex as well as its associated proteins which will include DNA and RNA associated proteins. We hypothesized that by targeting the proteins which facilitate transcriptional changes, we can identify the effector proteins which mediate the required responses for mycobacterial survival.

Formaldehyde cross-linking followed by affinity purification of nucleoprotein complexes through the targeting of RNA polymerase facilitated the identification of various DNA- and RNA-associated proteins using mass spectrometry. Furthermore, the developed method, Nucleoprotein – Mass Spectrometry (NP-MS), also identified various proteins required for energy, carbon, lipid and amino acid metabolism. These results demonstrated the ability of the chosen approach to not only isolate the RNA transcriptional complex but also its associated translational complex and energy metabolism proteins required for these biological processes.

To validate the ability of NP-MS to specifically isolate and identify nucleoproteins and particularly DNA-associated proteins, we selected several proteins for further analysis. Through a comparison of conserved protein domains, which have demonstrated some affinity for nucleic acids in other well studied organisms, five proteins were selected for DNA association validation studies. Genes encoding these proteins were cloned into the episomal expression vector pNFLAG, which facilitated N-terminal FLAG-tagging of possible DNA-associated proteins. We showed that NP-MS identified proteins MSMEG\_1060, MSMEG\_2695, MSMEG\_4306 and MSMEG\_5512 as possible DNA-associated proteins following affinity purification of FLAG-tagged proteins. Although direct DNA interaction by these proteins remains to be determined, these results provide original data. GO identities used to describe the predicted functional attributes of proteins failed to identify DNA association for the proteins MSMEG\_1060, MSMEG\_2695 and MSMEG\_4306.

Confirmation of NP-MS' ability to identify proteins associated with the RNA polymerase transcriptional complex allowed us to speculate that we could use the developed approach to investigate environmental

stress in *M. smegmatis*. In support of this, NP-MS was successfully used to identify proteins associated with the exponential and stationary growth phases of *M. smegmatis*. Proteins found to be exclusively identified from the exponential growth phase were in accordance with data from *M. tuberculosis* transcriptional analysis studies (154). In addition, NP-MS successfully identified three (MSMEG\_3811, MSMEG\_3945 and MSMEG\_3950) of the five predicted universal stress proteins in *M. smegmatis* as well as the ribosome hibernation promoting factor (hpf), heat shock protein X (hspX) and the two-component transcriptional regulator DevR. All but one of these proteins, MSMEG\_3811, have been demonstrated to have some response in relation to stress experienced in various mycobacterial organisms. The results demonstrated the ability of NP-MS to identify proteins associated with the RNA polymerase transcriptional complex from *M. smegmatis* cultures experiencing environmental stress.

## 5.2. Future directions

---

NP-MS, as reported here offers opportunities for novel insight into the RNA polymerase transcriptional complex and its associated proteins. However, it must be acknowledged that the technical limitations associated with NP-MS remains a considerable constraint for the identification of RNA polymerase associated proteins involved in cellular processes such as transcription and translation. The use of an assortment of detergents to reduce non-specific binding during immunoprecipitation reactions may be effective in increasing the specificity of the anti-RNA polymerase immunoprecipitation, however, the technical difficulties these present for mass spectrometry remains an obstacle. In future, replacement of the wash buffers used in this study with detergent free wash buffers will eliminate the need to make use of methanol-chloroform phase separations for detergent removal. Although this may result in an increase of non-specifically bound proteins, low confidence protein identification can still be made using the LFQ intensity data subsequent to mass spectrometry analysis (119). Following improvement of proteomic sample preparation, we believe that NP-MS can effectively be used to investigate the proteins in the cell which mediate change under stress condition such as hypoxia, low pH, nutrient starvation, temperature shock and exposure to antibiotics. Investigation into the proteins associated with the RNA polymerase transcriptional complex under these stresses, which are known to occur within the host environment, could broaden our understanding of *M. tuberculosis* gene regulation during host infection. Furthermore, we believe that by improving the wash buffers used, NP-MS can not only be used to investigate transcriptional responses to stress in mycobacteria, but in other bacterial organisms as well.

Results from this study demonstrated the ability of NP-MS to identify known proteins associated with nucleic acids. Not surprisingly, given the number of predicted open reading frames in *M. smegmatis* that are believed to encode proteins of unknown function, several proteins with ambiguous or no predicted functional attributes were identified. The possible DNA-, RNA- and/or protein associations of these proteins remain to be verified through the use of functional studies. Possible DNA interactions of identified proteins could

possibly be investigated using techniques such as DNA footprinting and ChIP-seq followed by electrophoretic mobility shift assays (EMSA) to determine the affinity of investigated proteins for the identified target DNA sequences (163, 164). Likewise, RNA footprinting and EMSA can potentially be employed to investigate the functional properties of predicted RNA associated proteins (250). Furthermore, the use of highly sensitive techniques such as microscale thermophoresis is appealing for the investigation into the DNA, RNA and possible protein binding capabilities of NP-MS identified proteins (168).

Several proteins identified in this study (MSMEG\_1060, MSMEG\_2695, MSMEG\_4306 and MSMEG\_5512) were expressed as N-terminal FLAG-tagged proteins and were shown to be associated with DNA. In future, ChIP-seq can be used to not only reveal the targeted DNA binding sequence of the proteins in question, but also hint towards the regulatory roles that these proteins might have on gene expression, if any.

Results obtained from these studies will not only provide more functional information regarding the proteins expressed by mycobacteria but will aid in our understanding of how transcriptional responses are mediated in mycobacterial organisms. The valuable information gained from these studies have the potential to identify novel drug development targets and may aid in the development of new and more effective anti-tuberculosis treatment regimens.

---

## References

---

1. **WHO**. 2016. Global Tuberculosis Report 2016. Cdc 2016 214.
2. **WHO**. 2015. Global Tuberculosis Report 2015.
3. **WHO**. 2010. Multidrug and extensively drug-resistant TB (M/XDR-TB).
4. **Betts JC, Lukey PT, Robb LC, McAdam R a., Duncan K**. 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol Microbiol* **43**:717–731.
5. **Voskuil MI, Visconti KC, Schoolnik GK**. 2004. *Mycobacterium tuberculosis* gene expression during adaptation to stationary phase and low-oxygen dormancy. *Tuberculosis* **84**:218–227.
6. **Hampshire T, Soneji S, Bacon J, James BW, Hinds J, Laing K, Stabler RA, Marsh PD, Butcher PD**. 2004. Stationary phase gene expression of *Mycobacterium tuberculosis* following a progressive nutrient depletion: A model for persistent organisms? *Tuberculosis* **84**:228–238.
7. **Deb C, Lee CM, Dubey VS, Daniel J, Abomoelak B, Sirakova TD, Pawar S, Rogers L, Kolattukudy PE**. 2009. A novel in vitro multiple-stress dormancy model for *Mycobacterium tuberculosis* generates a lipid-loaded, drug-tolerant, dormant pathogen. *PLoS One* **4**.
8. **Parish T, Smith D a, Kendall S, Casali N, Bancroft GJ, Stoker NG**. 2003. Deletion of Two-Component Regulatory Systems Increases the Virulence of *Mycobacterium tuberculosis* Deletion of Two-Component Regulatory Systems Increases the Virulence of *Mycobacterium tuberculosis*. *Society* **71**:1134–1140.
9. **Minch KJ, Rustad TR, Peterson EJR, Winkler J, Reiss DJ, Ma S, Hickey M, Brabant W, Morrison B, Turkarslan S, Mawhinney C, Galagan JE, Price ND, Baliga NS, Sherman DR**. 2015. The DNA-binding network of *Mycobacterium tuberculosis*. *Nat Commun* **6**:5829.
10. **Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh a, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail M a, Rajandream M a, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG**. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537–544.

11. **Boshoff HIM, Barry CE.** 2005. Tuberculosis — metabolism and respiration in the absence of growth. *Nat Rev Microbiol* **3**:70–80.
12. **Boshoff HIM, Myers TG, Copp BR, McNeil MR, Wilson MA, Barry CE.** 2004. The Transcriptional Responses of *Mycobacterium tuberculosis* to Inhibitors of Metabolism: NOVEL INSIGHTS INTO DRUG MECHANISMS OF ACTION. *J Biol Chem* **279**:40174–40184.
13. **Dreyfuss G, Kim VN, Kataoka N.** 2002. Messenger-Rna-Binding Proteins and the Messages They Carry. *Nat Rev Mol Cell Biol* **3**:195–205.
14. **Burd CG, Dreyfuss G.** 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**:615–621.
15. **Lunde BM, Moore C, Varani G.** 2007. {RNA}-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* **8**:479–490.
16. **Van Assche E, Van Puyvelde S, Vanderleyden J, Steenackers HP.** 2015. RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol* **6**:1–16.
17. **Ohniwa RL, Ushijima Y, Saito S, Morikawa K.** 2011. Proteomic analyses of nucleoid-associated proteins in *Escherichia coli*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, and *Staphylococcus aureus*. *PLoS One* **6**:e19172.
18. **Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, Gomes A, Rustad T, Dolganov G, Glotova I, Abeel T, Mahwinney C, Kennedy AD, Allard R, Brabant W, Krueger A, Jaini S, Honda B, Yu W-H, Hickey MJ, Zucker J, Garay C, Weiner B, Sisk P, Stolte C, Winkler JK, Van de Peer Y, Iazzetti P, Camacho D, Dreyfuss J, Liu Y, Dorhoi A, Mollenkopf H-J, Drogaris P, Lamontagne J, Zhou Y, Piquenot J, Park ST, Raman S, Kaufmann SHE, Mohney RP, Chelsky D, Moody DB, Sherman DR, Schoolnik GK.** 2013. The *Mycobacterium tuberculosis* regulatory network and hypoxia. *Nature* **499**:178–83.
19. **Dillon SC, Dorman CJ.** 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* **8**:185–195.
20. **Dorman CJ.** 2014. Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation. *J Mol Microbiol Biotechnol* **24**:316–331.
21. **Luijsterburg MS, Noom MC, Wuite GJLL, Dame RT.** 2006. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J Struct Biol* **156**:262–272.

22. **Rice PA, Yang SW, Mizuuchi K, Nash HA.** 1996. Crystal structure of an IHF-DNA complex: A protein-induced DNA U-turn. *Cell* **87**:1295–1306.
23. **Mouw KW, Rice PA.** 2007. Shaping the *Borrelia burgdorferi* genome: Crystal structure and binding properties of the DNA-bending protein Hbb. *Mol Microbiol* **63**:1319–1330.
24. **Vis H, Mariani M, Vorgias CE, Wilson KS, Kaptein R, Boelens R.** 1995. Solution structure of the HU protein from *Bacillus stearothermophilus*. *J Mol Biol* **254**:692–703.
25. **Swinger KK, Lemberg KM, Zhang Y, Rice PA.** 2003. Flexible DNA bending in HU-DNA cocrystal structures. *EMBO J* **22**:3749–3760.
26. **Browning DF, Grainger DC, Busby SJ.** 2010. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol* **13**:773–780.
27. **Azam T, Ishihama A.** 1999. Twelve species of the nucleoid-associated protein 481 from *Escherichia coli*. *Seq Recognit Specif DNA Bind ...* **274**:33105–33113.
28. **Kolter R.** 1997. Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps . Protection of DNA during Oxidative Stress by the Nonspecific DNA-Binding Protein Dps **179**:5188–5194.
29. **Keatch SA, Leonard PG, Ladbury JE, Dryden DTF.** 2005. StpA protein from *Escherichia coli* condenses supercoiled DNA in preference to linear DNA and protects it from digestion by DNase I and EcoKI. *Nucleic Acids Res* **33**:6540–6546.
30. **Colangeli R, Haq a, Arcus VL, Summers E, Magliozzo RS, McBride a, Mitra a K, Radjainia M, Khajo a, Jacobs WR, Salgame P, Alland D.** 2009. The multifunctional histone-like protein Lsr2 protects mycobacteria against reactive oxygen intermediates. *Proc Natl Acad Sci U S A* **106**:4414–4418.
31. **Basu D, Khare G, Singh S, Tyagi A, Khosla S, Mande SC.** 2009. A novel nucleoid-associated protein of *Mycobacterium tuberculosis* is a sequence homolog of GroEL. *Nucleic Acids Res* **37**:4944–4954.
32. **Grainger DC, Hurd D, Goldberg MD, Busby SJW.** 2006. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res* **34**:4642–4652.
33. **Blasco B, Chen JM, Hartkoorn R, Sala C, Uplekar S, Rougemont J, Pojer F, Cole ST.** 2012. Virulence Regulator EspR of *Mycobacterium tuberculosis* Is a Nucleoid-Associated Protein. *PLoS Pathog* **8**:e1002621.

34. **Pandey SD, Choudhury M, Yousuf S, Wheeler PR, Gordon S V., Ranjan A, Sritharan M.** 2014. Iron-Regulated Protein HupB of *Mycobacterium tuberculosis* Positively Regulates Siderophore Biosynthesis and Is Essential for Growth in Macrophages. *J Bacteriol* **196**:1853–1865.
35. **Castaing B, Zelwer C, Laval J, Boiteux S.** 1995. HU protein of *Escherichia coli* binds specifically to DNA that contains single-strand breaks or gaps. *J Biol Chem*.
36. **Kamashev D, Rouviere-Yaniv J.** 2000. The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J* **19**:6527–35.
37. **Ghosh S, Grove A.** 2004. Histone-like protein HU from *Deinococcus radiodurans* binds preferentially to four-way DNA junctions. *J Mol Biol* **337**:561–71.
38. **Mukherjee A, Bhattacharyya G, Grove A.** 2008. The C-terminal domain of HU-related histone-like protein Hlp from *Mycobacterium smegmatis* mediates DNA end-joining. *Biochemistry* **47**:8744–53.
39. **Azam TA, Iwata A, Nishimura A, Azam T a LI, Ueda S.** 1999. Growth Phase-Dependent Variation in Protein Composition of the *Escherichia coli* Nucleoid Growth Phase-Dependent Variation in Protein Composition of the *Escherichia coli* Nucleoid. *J Bac* **181**:6361.
40. **Dame RT.** 2005. The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol* **56**:858–870.
41. **Zhang A, Rimsky S, Reaban ME, Buc H, Belfort M.** 1996. *Escherichia coli* protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics. *EMBO J* **15**:1340–9.
42. **Free A, Dorman CJ.** 1995. Coupling of *Escherichia coli* hns mRNA levels to DNA synthesis by autoregulation: implications for growth phase control. *Mol Microbiol* **18**:101–113.
43. **Gordon BRG, Imperial R, Wang L, Navarre WW, Liu J.** 2008. Lsr2 of *Mycobacterium* represents a novel class of H-NS-like proteins. *J Bacteriol* **190**:7052–7059.
44. **Werlang ICR, Schneider CZ, Mendonça JD, Palma MS, Basso L a., Santos DS.** 2009. Identification of Rv3852 as a nucleoid-associated protein in *Mycobacterium tuberculosis*. *Microbiology* **155**:2652–2663.
45. **Prabhakar S, Annapurna PS, Jain NK, Dey a B, Tyagi JS, Prasad HK.** 1998. Identification of an immunogenic histone-like protein (HLPMT) of *Mycobacterium tuberculosis*. *Tuber Lung Dis* **79**:43–53.

46. **Pedulla ML, Lee MH, Lever DC, Hatfull GF.** 1996. A novel host factor for integration of mycobacteriophage L5. *Proc Natl Acad Sci U S A* **93**:15411–6.
47. **Liu Y, Wang H, Cui T, Zhou X, Jia Y, Zhang H, He Z-G.** 2016. NapM, a new nucleoid-associated protein, broadly regulates gene expression and affects mycobacterial resistance to anti-tuberculosis drugs. *Mol Microbiol* **0**:n/a-n/a.
48. **Sassetti CM, Boyd DH, Rubin EJ.** 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**:77–84.
49. **Sassetti CM, Rubin EJ.** 2003. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**:12989–94.
50. **Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM.** 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog* **7**:1–9.
51. **Chen JM, Ren H, Shaw JE, Wang YJ, Li M, Leung AS, Tran V, Berbenetz NM, Kocincova D, Yip CM, Reyrat JM, Liu J.** 2008. Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. *Nucleic Acids Res* **36**:2123–2135.
52. **Gordon BRG, Li Y, Cote A, Weirauch MT, Ding P, Hughes TR, Navarre WW, Xia B, Liu J.** 2011. Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc Natl Acad Sci U S A* **108**:10690–5.
53. **Blasco B, Stenta M, Alonso-Sarduy L, Dietler G, Peraro MD, Cole ST, Pojer F.** 2011. Atypical DNA recognition mechanism used by the EspR virulence regulator of *Mycobacterium tuberculosis*. *Mol Microbiol* **82**:251–264.
54. **Blasco B, Japaridze A, Stenta M, Wicky BIM, Dietler G, Peraro MD, Pojer F, Cole ST.** 2014. Functional dissection of intersubunit interactions in the espr virulence regulator of *Mycobacterium tuberculosis*. *J Bacteriol* **196**:1889–1900.
55. **Mishra A, Vij M, Kumar D, Taneja V, Mondal AK, Bothra A, Rao V, Ganguli M, Taneja B.** 2013. Integration host factor of *Mycobacterium tuberculosis*, mIHF, compacts DNA by a bending mechanism. *PLoS One* **8**:e69985.
56. **Sharadamma N, Harshavardhana Y, Ravishankar A, Anand P, Chandra NR, Muniyappa K.** 2014. Molecular dissection of *Mycobacterium tuberculosis* Integration Host Factor Reveals Novel Insights into the Mode of DNA binding and Nucleoid Compaction. *J Biol Chem* **0**–37.

57. **Bhowmick T, Ghosh S, Dixit K, Ganesan V, Ramagopal U a, Dey D, Sarma SP, Ramakumar S, Nagaraja V.** 2014. Targeting *Mycobacterium tuberculosis* nucleoid-associated protein HU with structure-based inhibitors. *Nat Commun* **5**:4124.
58. **Schubert OT, Mouritsen J, Ludwig C, R??st HL, Rosenberger G, Arthur PK, Claassen M, Campbell DS, Sun Z, Farrah T, Gengenbacher M, Maiolica A, Kaufmann SHE, Moritz RL, Aebersold R.** 2013. The Mtb proteome library: A resource of assays to quantify the complete proteome of *Mycobacterium tuberculosis*. *Cell Host Microbe* **13**:602–612.
59. **Pedulla ML, Hatfull GF.** 1998. Characterization of the mlHF gene of *Mycobacterium smegmatis*. *J Bacteriol* **180**:5473–5477.
60. **Wolfe LM, Mahaffey SB, Kruh NA, Dobos KM.** 2010. Proteomic definition of the cell wall of *Mycobacterium tuberculosis*. *J Proteome Res* **9**:5816–5826.
61. **de Souza G a, Arntzen MØ, Fortuin S, Schürch AC, Målen H, McEvoy CRE, van Soolingen D, Thiede B, Warren RM, Wiker HG.** 2011. Proteogenomic analysis of polymorphisms and gene annotation divergences in prokaryotes using a clustered mass spectrometry-friendly database. *Mol Cell Proteomics* **10**:M110.002527.
62. **Uplekar S, Rougemont J, Cole ST, Sala C.** 2013. High-resolution transcriptome and genome-wide dynamics of RNA polymerase and NusA in *Mycobacterium tuberculosis*. *Nucleic Acids Res* **41**:961–77.
63. **Sharadamma N, Khan K, Kumar S, Patil KN, Hasnain SE, Muniyappa K.** 2011. Synergy between the N-terminal and C-terminal domains of *Mycobacterium tuberculosis* HupB is essential for high-affinity binding, DNA supercoiling and inhibition of RecA-promoted strand exchange. *FEBS J* **278**:3447–62.
64. **Gupta M, Sajid A, Sharma K, Ghosh S, Arora G, Singh R, Nagaraja V, Tandon V, Singh Y.** 2014. HupB, a nucleoid-associated protein of *Mycobacterium tuberculosis*, is modified by serine/threonine protein kinases in vivo. *J Bacteriol* **196**:2646–2657.
65. **Kumar S, Sardesai AA, Basu D, Muniyappa K, Hasnain SE.** 2010. DNA clasping by mycobacterial HU: The c-terminal region of HupB mediates increased specificity of DNA binding. *PLoS One* **5**:1–10.
66. **Tyagi JS, Prasad HK, Prabhakar S, Mishra a, Singhal a, Katoch VM, Thakral SS.** 2004. Use of the hupB Gene Encoding a Histone-Like Protein of *Mycobacterium tuberculosis* as a Target for Detection and Use of the hupB Gene Encoding a Histone-Like Protein of *Mycobacterium*

- tuberculosis* as a Target for Detection and Differentiation of *M. tuberc.* Society **42**:2724–2732.
67. **Sharadamma N, Harshavardhana Y, Singh P, Muniyappa K.** 2010. *Mycobacterium tuberculosis* nucleoid-associated DNA-binding protein H-NS binds with high-affinity to the holliday junction and inhibits strand exchange promoted by RecA protein. *Nucleic Acids Res* **38**:3555–3569.
  68. **Ghosh S, Padmanabhan B, Anand C, Nagaraja V.** 2016. Lysine acetylation of the *Mycobacterium tuberculosis* HU protein modulates its DNA binding and genome organization. *Mol Microbiol* **0**:1–12.
  69. **Pandey SD, Choudhury M, Sritharan M.** 2014. Transcriptional regulation of *Mycobacterium tuberculosis* hupB gene expression. *Microbiology* **160**:1637–1647.
  70. **Zhao N, Sun M, Burns-Huang K, Jiang X, Ling Y, Darby C, Ehrt S, Liu G, Nathan C.** 2015. Identification of Rv3852 as an Agrimophol-Binding Protein in *Mycobacterium tuberculosis*. *PLoS One* **10**:e0126211.
  71. **Ghosh S, Indi SS, Nagaraja V.** 2013. Regulation of lipid biosynthesis, sliding motility, and biofilm formation by a membrane-anchored nucleoid-associated protein of *Mycobacterium tuberculosis*. *J Bacteriol* **195**:1769–78.
  72. **Odermatt NT, Sala C, Benjak A, Kolly GS, Vocat A, Lupien A, Cole ST.** 2017. Rv3852 (H-NS) of *Mycobacterium tuberculosis* is not involved in nucleoid compaction and virulence regulation. *J Bacteriol* JB.00129-17.
  73. **Fayet O, Ziegelhoffer T, Georgopoulos C.** 1989. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J Bacteriol* **171**:1379–85.
  74. **Xu Z, Horwich a L, Sigler PB.** 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388**:741–750.
  75. **Kong TH, Coates a R, Butcher PD, Hickman CJ, Shinnick TM.** 1993. *Mycobacterium tuberculosis* expresses two chaperonin-60 homologs. *Proc Natl Acad Sci U S A* **90**:2608–2612.
  76. **Aravindhavan V, Christy AJ, Roy S, Ajitkumar P, Narayanan PR, Narayanan S.** 2009. *Mycobacterium tuberculosis* groE promoter controls the expression of the bicistronic groESL1 operon and shows differential regulation under stress conditions. *FEMS Microbiol Lett* **292**:42–9.
  77. **Sielaff B, Lee KS, Tsai FTF.** 2011. Structural and functional conservation of *Mycobacterium tuberculosis* GroEL paralogs suggests that GroEL1 is a chaperonin. *J Mol Biol* **405**:831–9.

78. **Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Hatfull GF.** 2005. GroEL1: A dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. *Cell* **123**:861–873.
79. **Qamra R, Srinivas V, Mande SC.** 2004. *Mycobacterium tuberculosis* GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins. *J Mol Biol* **342**:605–617.
80. **Kumar CMS, Khare G, Srikanth C V., Tyagi AK, Sardesai AA, Mande SC.** 2009. Facilitated Oligomerization of Mycobacterial GroEL: Evidence for Phosphorylation-Mediated Oligomerization. *J Bacteriol* **191**:6525–6538.
81. **Hu Y, Henderson B, Lund PA, Tormay P, Ahmed MT, Gurcha SS, Besra GS, Coates ARM.** 2008. A *Mycobacterium tuberculosis* mutant lacking the groEL homologue cpn60.1 is viable but fails to induce an inflammatory response in animal models of infection. *Infect Immun* **76**:1535–1546.
82. **Qamra R, Mande SC, Coates ARM, Henderson B.** 2005. The unusual chaperonins of *Mycobacterium tuberculosis*. *Tuberculosis* **85**:385–394.
83. **Laal S, Sharma YD, Prasad HK, Murtaza A, Singh S, Tangri S, Misra RS, Nath I.** 1991. Recombinant fusion protein identified by lepromatous sera mimics native *Mycobacterium leprae* in T-cell responses across the leprosy spectrum. *Proc Natl Acad Sci U S A* **88**:1054–1058.
84. **Summers EL, Meindl K, Usón I, Mitra AK, Radjainia M, Colangeli R, Alland D, Arcus VL.** 2012. The structure of the Oligomerization Domain of Lsr2 from *Mycobacterium tuberculosis* reveals a mechanism for chromosome organization and protection. *PLoS One* **7**:1–12.
85. **Qu Y, Lim CJ, Whang YR, Liu J, Yan J.** 2013. Mechanism of DNA organization by *Mycobacterium tuberculosis* protein Lsr2. *Nucleic Acids Res* **41**:5263–5272.
86. **Gordon BRG, Li Y, Wang L, Sintsova A, van Bakel H, Tian S, Navarre WW, Xia B, Liu J.** 2010. Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci* **107**:5154–5159.
87. **Colangeli R, Helb D, Vilchèze C, Hazbón MH, Lee C-G, Safi H, Sayers B, Sardone I, Jones MB, Fleischmann RD, Peterson SN, Jacobs WR, Alland D.** 2007. Transcriptional Regulation of Multi-Drug Tolerance and Antibiotic-Induced Responses by the Histone-Like Protein Lsr2 in *M. tuberculosis*. *PLoS Pathog* **3**:e87.
88. **Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, Deschavanne P.** 2007. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli.

- Mol Biol Evol **24**:1861–1871.
89. **Jang J, Becq J, Gicquel B, Deschavanne P, Neyrolles O.** 2008. Horizontally acquired genomic islands in the tubercle bacilli. Trends Microbiol **16**:303–308.
  90. **Raghavan S, Manzanillo P, Chan K, Dovey C, Cox JS.** 2008. Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. Nature **454**:717–721.
  91. **Cox JS, Chen B, McNeil M, Jacobs WR.** 1999. Complex lipid determines tissue-specific replication of *Mycobacterium tuberculosis* in mice. Nature **402**:79–83.
  92. **Camacho LR, Ensergueix D, Perez E, Gicquel B, Guilhot C.** 1999. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol Microbiol **34**:257–267.
  93. **Yu J, Tran V, Li M, Huang X, Niu C, Wang D, Zhu J, Wang J, Gao Q, Liu J.** 2012. Both phthiocerol dimycocerosates and phenolic glycolipids are required for virulence of *Mycobacterium marinum*. Infect Immun **80**:1381–1389.
  94. **Stewart GR, Wernisch L, Stabler R, Mangan JA, Hinds J, Laing KG, Young DB, Butcher PD.** 2002. Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. Microbiology **148**:3129–3138.
  95. **Rustad TR, Harrell MI, Liao R, Sherman DR.** 2008. The enduring hypoxic response of *Mycobacterium tuberculosis*. PLoS One **3**:1–8.
  96. **Wong DK, Lee B, Horwitz M a, Gibson W.** 1999. Identification of Fur , Aconitase , and Other Proteins Expressed by *Mycobacterium tuberculosis* under Conditions of Low and High Concentrations of Iron by Combined Two-Dimensional Gel Electrophoresis and Mass Spectrometry Identification of Fur , Aconitase **67**:327–336.
  97. **Bartek IL, Woolhiser LK, Baughn a. D, Basaraba RJ, Jacobs WR, Lenaerts a. J, Voskuil MI.** 2014. *Mycobacterium tuberculosis* Lsr2 Is a Global Transcriptional Regulator Required for Adaptation to Changing Oxygen Levels and Virulence. MBio **5**:e01106-14.
  98. **Du Y, Zhang H, He Y, Huang F, He ZG.** 2012. *Mycobacterium smegmatis* Lsr2 physically and functionally interacts with a new flavoprotein involved in bacterial resistance to oxidative stress. J Biochem **152**:479–486.
  99. **Dahl JL, Kraus CN, Boshoff HIM, Doan B, Foley K, Avarbock D, Kaplan G, Mizrahi V, Rubin H, Barry CE.** 2003. The role of RelMtb-mediated adaptation to stationary phase in long-term

- persistence of *Mycobacterium tuberculosis* in mice. Proc Natl Acad Sci U S A **100**:10026–31.
100. **Yeruva VC, Duggirala S, Lakshmi V, Kolarich D, Altmann F, Sritharan M.** 2006. Identification and characterization of a major cell wall-associated iron-regulated envelope protein (Irep-28) in *Mycobacterium tuberculosis*. Clin Vaccine Immunol **13**:1137–1142.
  101. **Lee BH, Murugasu-Oei B, Dick T.** 1998. Upregulation of a histone-like protein in dormant *Mycobacterium smegmatis*. Mol Gen Genet **260**:475–479.
  102. **Shires K, Steyn L.** 2001. The cold-shock stress response in *Mycobacterium smegmatis* induces the expression of a histone-like protein. Mol Microbiol **39**:994–1009.
  103. **Whiteford DC, Klingelhoets JJ, Bambenek MH, Dahl JL.** 2011. Deletion of the histone-like protein (Hlp) from *Mycobacterium smegmatis* results in increased sensitivity to UV exposure, freezing and isoniazid. Microbiology **157**:327–35.
  104. **Dosanjh NS, Rawat M, Chung JH, Av-Gay Y.** 2005. Thiol specific oxidative stress response in Mycobacteria. FEMS Microbiol Lett **249**:87–94.
  105. **Monahan IM, Betts J, Banerjee DK, Butcher PD.** 2001. Differential expression of mycobacterial proteins following phagocytosis by macrophages. Microbiology **147**:459–471.
  106. **Takatsuka M, Osada-Oka M, Satoh EF, Kitadokoro K, Nishiuchi Y, Niki M, Inoue M, Iwai K, Arakawa T, Shimoji Y, Ogura H, Kobayashi K, Rambukkana A, Matsumoto S.** 2011. A histone-like protein of mycobacteria possesses ferritin superfamily protein-like activity and protects against DNA damage by Fenton reaction. PLoS One **6**:e20985.
  107. **Av-gay Y, Everett M.** 2000. The eukaryotic-like Ser / Thr protein kinases of *Mycobacterium tuberculosis*. Trends Microbiol Rev **8**:238–244.
  108. **Junger MA, Aebersold R, Junger MA, Aebersold R, Junger MA, Aebersold R.** 2014. Mass spectrometry-driven phosphoproteomics: Patterning the systems biology mosaic. Wiley Interdiscip Rev Dev Biol **3**:83–112.
  109. **Greenstein AE, Grundner C, Echols N, Gay LM, Lombana TN, Mietskowski CA, Pullen KE, Sung PY, Alber T.** 2006. Structure/function studies of Ser/Thr and Tyr protein phosphorylation in *Mycobacterium tuberculosis*. J Mol Microbiol Biotechnol **9**:167–181.
  110. **Canova MJ, Kremer L, Molle V.** 2009. The *Mycobacterium tuberculosis* GroEL1 chaperone is a substrate of Ser/Thr protein kinases. J Bacteriol **191**:2876–2883.

111. **Deol P, Vohra R, Saini AK, Singh A, Chandra H, Chopra P, Das TK, Tyagi AK, Singh Y.** 2005. Role of *Mycobacterium tuberculosis* Ser / Thr Kinase PknF : Implications in Glucose Transport and Cell Division. *J Bacteriol* **187**:3415–3420.
112. **Gopalaswamy R, Narayanan S, Jacobs WR, Av-Gay Y.** 2008. *Mycobacterium smegmatis* biofilm formation and sliding motility are affected by the serine/threonine protein kinase PknF. *FEMS Microbiol Lett* **278**:121–127.
113. **Pinault L, Han JS, Kang CM, Franco J, Ronning DR.** 2013. Zafirlukast inhibits complexation of Lsr2 with DNA and growth of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **57**:2134–2140.
114. **Rajala N, Hensen F, Wessels HJCT, Ives D, Gloerich J, Spelbrink JN.** 2015. Whole cell formaldehyde cross-linking simplifies purification of mitochondrial nucleoids and associated proteins involved in mitochondrial gene expression. *PLoS One* **10**:1–20.
115. **Wang W, Li G-W, Chen C, Xie XS, Zhuang X.** 2011. Chromosome Organization by a Nucleoid-Associated Protein in Live Bacteria. *Science (80- )* **333**:1445–1449.
116. **Gavin A-C, Bösch M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon A-M, Cruciat C-M, Remor M, Höfert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier M-A, Copley RR, Edlmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G.** 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**:141–147.
117. **Butland G, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A.** 2005. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**:531–7.
118. **Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman Y V, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert J-P, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams S-L, Moran MF, Morin GB, Topaloglou T, Figeys D.** 2007. Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3**:89.
119. **Keilhauer EC, Hein MY, Mann M.** 2015. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol Cell Proteomics* **14**:120–35.

120. **Guruharsha KG, Rual J, Zhai B, Mintseris J, Vaidya N, Beekman C, Wong C, Rhee D, Cenaj O, McKillig E, Shah S, Spatleton M, Wan K, Yu C, Artavanis-Tsakonas S.** 2011. A Protein Complex Network of *Drosophila melanogaster*. *Cell* **147**:690–703.
121. **Giambruno R, Grebien F, Stukalov A, Knoll C, Planyavsky M, Rudashevskaya EL, Colinge J, Superti-Furga G, Bennett KL.** 2013. Affinity purification strategies for proteomic analysis of transcription factor complexes. *J Proteome Res* **12**:4018–4027.
122. **Roque ACA, Lowe CR.** 2008. Affinity chromatography: history, perspectives, limitations and prospects. *Methods Mol Biol* **421**:1–21.
123. **Dunham WH, Mullin M, Gingras AC.** 2012. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *Proteomics* **12**:1576–1590.
124. **Cox J, Mann M.** 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**:1367–72.
125. **Eng J.** 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**:976,989.
126. **Craig R, Beavis RC.** 2004. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **20**:1466–1467.
127. **Kim S, Gupta N, Pevzner PA.** 2008. [journal of proteome] pevnzer kim Spectral Probabilities and Generating Functions of Tandem Mass Spectra a Strike Against the Decoy Database-supplement- Supplementary\_Info **7**:1–6.
128. **Gibson TJ, Seiler M, Veitia R a.** 2013. The transience of transient overexpression. *Nat Methods* **10**:715–21.
129. **Mellacheruvu D, Wright Z, Couzens AL, Lambert J, St-denis N, Li T, Miteva Y V, Hauri S, Sardu ME, Yew T, Halim VA, Bagshaw RD, Hubner NC, Bouchard A, Faubert D, Fermin D, Dunham WH, Heck AJR, Choi H, Gstaiger M.** 2013. The CRAPome: a Contaminant Repository for Affinity Purification Mass Spectrometry Data. *Nat Methods* **10**:730–736.
130. **Snapper SB, Melton RE, Mustafa S, Kieser T, Jr WRJ.** 1990. Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* **4**:1911–1919.
131. **Ellis RC, Zabrowarny LA.** 1993. Safer staining method for acid fast bacilli. *J Clin Pathol* **46**:559–560.

132. **Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J.** 2016. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**:731–40.
133. **Tabb DL, Friedman DB, Ham A-JL.** 2006. Verification of automated peptide identifications from proteomic tandem mass spectra. *Nat Protoc* **1**:2213–2222.
134. **Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M.** 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **27**:29–34.
135. **Zheng QQQ, Wang XJ.** 2008. GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res* **36**:358–363.
136. **Supek F, Bosnjak M, Skunca N, Smuc T.** 2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**.
137. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski BB, Ideker T.** 2003. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* **13**:2498–2504.
138. **Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Von Mering C.** 2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**:D447–D452.
139. **Guo X V, Monteleone M, Klotzsche M, Kamionka A, Hillen W, Braunstein M, Ehrh S, Schnappinger D.** 2007. Silencing *Mycobacterium smegmatis* by using tetracycline repressors. *J Bacteriol* **189**:4614–23.
140. **Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO.** 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**:343–5.
141. **Parham P.** 1983. On the fragmentation of monoclonal IgG1, IgG2a, and IgG2b from BALB/c mice. *J Immunol* **131**:2895–2902.
142. **Van de Water J, Deininger SO, Macht M, Przybylski M, Gershwin ME.** 1997. Detection of molecular determinants and epitope mapping using MALDI-TOF mass spectrometry. *Clin Immunol Immunopathol* **85**:229–235.
143. **Gene T, Consortium O.** 2008. The Gene ontology project in 2008. *Nucleic Acids Res* **36**:440–444.

144. **Bateman A, Martin MJ, O'Donovan C, Magrane M, Apweiler R, Alpi E, Antunes R, Arganiska J, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Chavali G, Cibrian-Uhalte E, Da Silva A, De Giorgi M, Dogan T, Fazzini F, Gane P, Castro LG, Garmiri P, Hatton-Ellis E, Hieta R, Huntley R, Legge D, Liu W, Luo J, Macdougall A, Mutowo P, Nightingale A, Orchard S, Pichler K, Poggioli D, Pundir S, Pureza L, Qi G, Rosanoff S, Saidi R, Sawford T, Shypitsyna A, Turner E, Volynkin V, Wardell T, Watkins X, Zellner H, Cowley A, Figueira L, Li W, McWilliam H, Lopez R, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Boeckmann B, Bolleman J, Boutet E, Breuza L, Casal-Casas C, De Castro E, Coudert E, Cuche B, Doche M, Dornevil D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G, Lara V, Lemercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nospikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey AL, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Suzek BE, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Yerramalla MS, Zhang J.** 2015. UniProt: A hub for protein information. *Nucleic Acids Res* **43**:D204–D212.
145. **Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, Mcanulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C.** 2009. InterPro: The integrative protein signature database. *Nucleic Acids Res* **37**:211–215.
146. **Elderkin S, Jones S, Schumacher J, Studholme D, Buck M.** 2002. Mechanism of action of the *Escherichia coli* phage shock protein PspA in repression of the AAA family transcription factor PspF. *J Mol Biol* **320**:23–37.
147. **Goebel M, Yanagida M.** 1991. The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem Sci* **16**:173–177.
148. **Rigden DJ.** 2011. *Ab Initio* Modeling Led Annotation Suggests Nucleic Acid Binding Function for Many DUFs. *Omi A J Integr Biol* **15**:431–438.
149. **Albright LM, Huala E, Ausubel FM.** 1989. Prokaryotic signal transduction mediated by sensor and regulator protein pairs. *Annu Rev Genet* **23**:311–336.
150. **Valerio Orlando, Helen Strutt RP.** 1997. Analysis of chromatin structure in vivo. *Methods* **12**:105–

- 114.
151. **Feldman MY.** 1973. Reactions of nucleic acids and nucleoproteins with formaldehyde. *Prog Nucleic Acid Res Mol Biol* **13**:1–49.
152. **Hoffman EA, Frey BL, Smith LM, Auble DT.** 2015. Formaldehyde crosslinking: A tool for the study of chromatin complexes. *J Biol Chem* **290**:26404–26411.
153. **Wu CH, Chen S, Shortreed MR, Kreitinger GM, Yuan Y, Frey BL, Zhang Y, Mirza S, Cirillo LA, Olivier M, Smith LM.** 2011. Sequence-specific capture of protein-DNA complexes for mass spectrometric protein identification. *PLoS One* **6**.
154. **Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, Rose G, Perkins TT, Parkhill J, Dougan G, Young DB.** 2011. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. *PLoS Pathog* **7**.
155. **Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimermancic P, Greninger AL, Pico AR.** 2014. Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. *Nat Protoc* **9**:2539–54.
156. **Metz, B., Kersten, G., Hoogerhout, P., Brugghe, H., Timmermans, H., Jong, A., Meiring, H., Hove, J., Hennink, W., Crommelin, D., Jiskoo W, Metz B, Metz, B., Kersten, G., Hoogerhout, P., Brugghe, H., Timmermans, H., Jong, A., Meiring, H., Hove, J., Hennink, W., Crommelin, D., Jiskoo W, Metz B, Metz, B., Kersten, G., Hoogerhout, P., Brugghe, H., Timmermans, H., Jong, A., Meiring, H., Hove, J., Hennink, W., Crommelin, D., Jiskoo W, Metz B.** 2003. Identification of Formaldehyde-induced Modifications in Proteins: REACTIONS WITH MODEL PEPTIDES. *J Biol Chem* **279**:6235–6243.
157. **Tsur D, Tanner S, Zandi E, Bafna V, Pevzner PA.** 2005. Identification of post-translational modifications via blind search of mass-spectra. *Proc - 2005 IEEE Comput Syst Bioinforma Conf CSB 2005* **2005**:157–166.
158. **Tanner S, Pevzner PA, Bafna V.** 2006. Unrestrictive identification of post-translational modifications through peptide mass spectrometry. *Nat Protoc* **1**:67–72.
159. **Dasari S, Chambers MC, Codreanu SG, Liebler DC, Collins BC, Pennington SR, Gallagher WM, Tabb DL.** 2011. Sequence tagging reveals unexpected modifications in toxicoproteomics. *Chem Res Toxicol* **24**:204–216.
160. **Toews J, Rogalski JC, Clark TJ, Kast J.** 2008. Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions. *Anal Chim Acta* **136**

- 618:168–183.
161. **Deshayes C, Perrodou E, Gallien S, Euphrasie D, Schaeffer C, Van-Dorselaer A, Poch O, Lecompte O, Reyrat J-M.** 2007. Interrupted coding sequences in *Mycobacterium smegmatis*: authentic mutations or sequencing errors? *Genome Biol* **8**:R20.
162. **Potgieter MG, Nakedi KC, Ambler JM, Nel AJM, Garnett S, Soares NC, Mulder N, Blackburn JM.** 2016. Proteogenomic analysis of *Mycobacterium smegmatis* using high resolution mass spectrometry. *Front Microbiol* **7**.
163. **Galas DJ, Schmitz a.** 1978. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res* **5**:3157–3170.
164. **Anderson JN, Limited IRLP.** 1986. *Nucleic Acids Research* **9**:8513–8533.
165. **Bendak K, Loughlin FE, Cheung V, O'Connell MR, Crossley M, MacKay JP.** 2012. A rapid method for assessing the RNA-binding potential of a protein. *Nucleic Acids Res* **40**:1–11.
166. **Gagnon KT, Maxwell ES.** 2011. *Rna* **703**:275–291.
167. **Young KH.** 1998. Yeast two-hybrid: so many interactions, (in) so little time... *Biol Reprod* **58**:302–311.
168. **Duhr S, Braun D.** 2006. Why molecules move along a temperature gradient. *Proc Natl Acad Sci U S A* **103**:19678–82.
169. **Katayama H, Nagasu T, Oda Y.** 2001. Improvement of in-gel digestion protocol for peptide mass fingerprinting by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **15**:1416–1421.
170. **Zhang N, Li L.** 2004. Effects of common surfactants on protein digestion and matrix-assisted laser desorption/ionization mass spectrometric analysis of the digested peptides using two-layer sample preparation. *Rapid Commun Mass Spectrom* **18**:889–896.
171. **Yeung YG, Stanley ER.** 2010. Rapid detergent removal from peptide samples with ethyl acetate for mass spectrometry analysis. *Curr Protoc Protein Sci* 1–5.
172. **Jiang D, Jarrett HW, Haskins WE.** 2009. Methods for proteomic analysis of transcription factors. *J Chromatogr A* **1216**:6881–6889.
173. **Nagore L, Nadeau R, Guo Q, Jadhav Y, Jarrett H, Haskins W.** 2013. Purification & Characterization of Transcription Factors. *Mass Spectrom Rev* **32**:386–398.

174. **Olsen J V, Macek B, Lange O, Makarov A, Horning S, Mann M.** 2007. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* **4**:709–12.
175. **de Brito T, Franco MF.** 1994. Granulomatous inflammation. *Rev Inst Med Trop Sao Paulo*.
176. **Dickinson BJM, Mitchison DA.** 1966. *Tubercle, Lond.*, (1966). **47**, 370 370–380.
177. **Dickinson JM, Mitchison DA.** 1970. Observations in vitro on the suitability of pyrazinamide for intermittent chemotherapy of tuberculosis. *Tubercle* **51**:389–396.
178. **Grosset J.** 1980. Bacteriologic basis of short-course chemotherapy for tuberculosis. *Clin Chest Med* **1**:231–41.
179. **Hu Y, Coates ARM, Mitchison DA.** 2003. Sterilizing activities of fluoroquinolones against rifampin-tolerant populations of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **47**:653–657.
180. **Alnimr AM.** 2015. Dormancy models for *Mycobacterium tuberculosis*: A minireview **647**:641–647.
181. **Baena A, Porcelli SA.** 2009. Evasion and subversion of antigen presentation by *Mycobacterium tuberculosis*: REVIEW ARTICLE. *Tissue Antigens* **74**:189–204.
182. **Connolly LE, Edelstein PH, Ramakrishnan L.** 2007. Why is long-term therapy required to cure tuberculosis? *PLoS Med* **4**:435–442.
183. **Wayne LG, Sohaskey CD.** 2001. NONREPLICATING PERSISTENCE OF MYCOBACTERIUM TUBERCULOSIS \*.
184. **James BW, Bacon J, Hampshire T, Morley K, Marsh PD.** 2002. In vitro gene expression dissected: Chemostat surgery for *Mycobacterium tuberculosis*. *Comp Funct Genomics* **3**:345–347.
185. **Hu Y, Coates AR, Mitchison DA.** 2006. Sterilising action of pyrazinamide in models of dormant and rifampicin-tolerant *Mycobacterium tuberculosis*. *Int J Tuberc Lung Dis* **10**:317–322.
186. **Cunningham AF, Spreadbury CL.** 1998. Mycobacterial stationary phase induced by low oxygen tension: cell wall thickening and localization of the 16-kilodalton alpha-crystallin homolog. *J Bacteriol* **180**:801–8.
187. **Matin A.** 1990. Molecular analysis of the starvation stress in *Escherichia coli*. *FEMS Microbiol Ecol* **74**:185–195.
188. **Siegele DA, Kolter R.** 1992. Life after log. *J Bacteriol* **174**:345–348.
189. **Smeulders MJ, Keer J, Speight R a, Williams HD.** 1999. Adaptation of *Mycobacterium smegmatis*

- to Stationary Phase Adaptation of *Mycobacterium smegmatis* to Stationary Phase **181**:270–283.
190. **Blokpoel MCJ, Blokpoel MCJ, Keer J, Keer J, Williams HD, Williams HD.** 2005. Global Analysis of Proteins Synthesized by *Mycobacterium smegmatis* Provides Direct Evidence for Physiological Heterogeneity in Stationary-Phase Cultures. *Society* **187**:6691–6700.
  191. **Toole RO, Smeulders MJ, Blokpoel MC, Kay EJ, Lougheed K, Williams HD.** 2003. A Two-Component Regulator of Universal Stress Protein Expression and Adaptation to Oxygen Starvation in *Mycobacterium smegmatis* **185**:1543–1554.
  192. **Ntolosi B a, Betts J, Zappe H, Powles R, Steyn LM.** 2001. Growth phase-associated changes in protein expression in *Mycobacterium smegmatis* identify a new low molecular weight heat shock protein. *Tuberculosis (Edinb)* **81**:279–289.
  193. **Wang R, Prince JT, Marcotte EM.** 2005. Mass spectrometry of the M-smegmatis proteome: Protein expression levels correlate with function, operons, and codon bias. *Genome Res* **15**:1118–1126.
  194. **Fazio TG, Roth JR.** 1996. Evidence that the CysG protein catalyzes the first reaction specific to B12 synthesis in *Salmonella typhimurium*, insertion of cobalt. *J Bacteriol* **178**:6952–6959.
  195. **Wooff E, Michell SL, Gordon S V., Chambers MA, Bardarov S, Jacobs WR, Hewinson RG, Wheeler PR.** 2002. Functional genomics reveals the sole sulphate transporter of the *Mycobacterium tuberculosis* complex and its relevance to the acquisition of sulphur in vivo. *Mol Microbiol* **43**:653–663.
  196. **Hu Z-L, Bao J, Reecy J.** 2008. CateGORizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. *Online J Bioinforma* **9**:108–112.
  197. **Ferrer JL, Ravanel S, Robert M, Dumas R.** 2004. Crystal Structures of cobalamin-independent methionine synthase complexed with zinc, homocysteine, and methyltetrahydrofolate. *J Biol Chem* **279**:44235–44238.
  198. **Jenkins DE, Schultz JE, Matin a.** 1988. Starvation-induced cross protection against heat or H<sub>2</sub>O<sub>2</sub> challenge in *Escherichia coli*. *J Bacteriol* **170**:3910–3914.
  199. **Jenkins DE, Chaisson SA, Matin A.** 1990. Starvation-induced cross-protection against osmotic challenge in *Escherichia coli*. *J Bacteriol* **172**:2779–2781.
  200. **Lange R, Hengge-Aronis R.** 1991. Identification of a central regulator of stationary phase gene expression in *Escherichia coli*. *Mol Microbiol* **5**:49–59.

201. **Toole RO, Williams HD.** 2003. Universal stress proteins and *Mycobacterium tuberculosis* **154**:387–392.
202. **Saini DK, Malhotra V, Dey D, Pant N, Das TK, Tyagi JS.** 2004. DevR-DevS is a bona fide two-component system of *Mycobacterium tuberculosis* that is hypoxia-responsive in the absence of the DNA-binding domain of DevR. *Microbiology* **150**:865–875.
203. **Roberts DM, Liao RP, Wisedchaisri G, Hol WGJ, Sherman DR.** 2004. Two Sensor Kinases Contribute to the Hypoxic Response of *Mycobacterium tuberculosis* \* **279**:23082–23087.
204. **Boon C, Li R, Qi R, Dick T.** 2001. Proteins of *Mycobacterium bovis* BCG Induced in the Wayne Dormancy Model Proteins of *Mycobacterium bovis* BCG Induced in the Wayne Dormancy Model. *J Bacteriol* **183**:2672–2676.
205. **Boon C, Dick T.** 2002. *Mycobacterium bovis* BCG Response Regulator Essential for Hypoxic Dormancy *Mycobacterium bovis* BCG Response Regulator Essential for Hypoxic Dormancy. *J Bacteriol* **184**:6760–6767.
206. **Park H-D, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, Schoolnik GK, Sherman DR.** 2003. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Mol Microbiol* **48**:833–43.
207. **Chao MC, Rubin EJ.** 2010. Letting sleeping dosLie: does dormancy play a role in tuberculosis? *Annu Rev Microbiol* **64**:293–311.
208. **Voskuil MI, Schnappinger D, Visconti KC, Harrell MI, Dolganov GM, Sherman DR, Schoolnik GK.** 2003. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. *J Exp Med* **198**:705–713.
209. **Kumar A, Toledo JC, Patel RP, Lancaster JR, Steyn AJC.** 2007. *Mycobacterium tuberculosis* DosS is a redox sensor and DosT is a hypoxia sensor. *Proc Natl Acad Sci* **104**:11568–11573.
210. **Shiloh MU, Manzanillo P, Cox JS.** 2008. *Mycobacterium tuberculosis* Senses Host-Derived Carbon Monoxide during Macrophage Infection. *Cell Host Microbe* **3**:323–330.
211. **Sherman DR, Voskuil M, Schnappinger D, Liao R, Harrell MI, Schoolnik GK.** 2001. Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha -crystallin. *Proc Natl Acad Sci U S A* **98**:7534–7539.
212. **Sherman DR, Mdluli K, Hickey MJ, Barry CE, Stover CK.** 1999. AhpC, oxidative stress and drug resistance in *Mycobacterium tuberculosis*. *Biofactors* **10**:211–217.

213. **Kendall SL, Movahedzadeh F, Rison SCG, Wernisch L, Parish T, Duncan K, Betts JC, Stoker NG.** 2004. The *Mycobacterium tuberculosis* dosRS two-component system is induced by multiple stresses. *Tuberculosis* **84**:247–255.
214. **Converse PJ, Karakousis PC, Klinkenberg LG, Kesavan AK, Ly LH, Allen SS, Grosset JH, Jain SK, Lamichhane G, Manabe YC, McMurray DN, Nuermberger EL, Bishai WR.** 2009. Role of the dosR-dosS two-component regulatory system in *Mycobacterium tuberculosis* virulence in three animal models. *Infect Immun* **77**:1230–1237.
215. **Yuan Y, Crane DD, Simpson RM, Zhu YQ, Hickey MJ, Sherman DR, Barry CE.** 1998. The 16-kDa alpha-crystallin (Acr) protein of *Mycobacterium tuberculosis* is required for growth in macrophages. *Proc Natl Acad Sci U S A* **95**:9578–9583.
216. **Stewart JN, Rivera HN, Karls R, Quinn FD, Roman J, Rivera-Marrero CA.** 2006. Increased pathology in lungs of mice after infection with an  $\alpha$ -crystallin mutant of *Mycobacterium tuberculosis*: Changes in cathepsin proteases and certain cytokines. *Microbiology* **152**:233–244.
217. **Yuan Y, Crane DD, Barry CE.** 1996. Stationary phase-associated protein expression in *Mycobacterium tuberculosis*: Function of the mycobacterial alpha-crystallin homolog. *J Bacteriol* **178**:4484–4492.
218. **Hu YM, Butcher PD, Sole K, Mitchison DA, Coates ARM.** 1998. Protein synthesis is shutdown in dormant *Mycobacterium tuberculosis* and is reversed by oxygen or heat shock **158**:139–145.
219. **Ohno H, Zhu G, Mohan VP, Chu D, Kohno S, Jacobs WR, Chan J.** 2003. The effects of reactive nitrogen intermediates on gene expression in *Mycobacterium tuberculosis*. *Cell Microbiol* **5**:637–648.
220. **Rosenkrands I, Slayden RA, Crawford J, Aagaard C, Clifton EB, Andersen P.** 2002. Hypoxic Response of *Mycobacterium tuberculosis* Studied by Metabolic Labeling and Proteome Analysis of Cellular and Extracellular Proteins. *Society* **184**:3485–3491.
221. **Florczyk M a, Mccue LA, Purkayastha A, Currenti E, Wolin MJ, Mcdonough K a, Kathleen a.** 2003. A Family of *acr* -Coregulated *Mycobacterium tuberculosis* Genes Shares a Common DNA Motif and Requires Rv3133c ( dos R or dev R ) for Expression A Family of *acr*-Coregulated *Mycobacterium tuberculosis* Genes Shares a Common DNA Motif and Requires Rv3133c ( do. *Infect Immun* **71**:5332.
222. **Shi L, Jung Y-J, Tyagi S, Gennaro ML, North RJ.** 2003. Expression of Th1-mediated immunity in mouse lungs induces a *Mycobacterium tuberculosis* transcription pattern characteristic of

- nonreplicating persistence. *Proc Natl Acad Sci* **100**:241–246.
223. **Drumm JE, Mi K, Bilder P, Sun M, Lim J, Bielefeldt-ohmann H, Basaraba R, So M, Zhu G, Tufariello JM, Izzo AA, Orme IM, Almo SC, Leyh TS, Chan J.** 2009. *Mycobacterium tuberculosis* Universal Stress Protein Rv2623 Regulates Bacillary Growth by ATP-Binding: Requirement for Establishing Chronic Persistent Infection **5**.
224. **Purkayastha A, McCue LA, McDonough KA.** 2002. Identification of a *Mycobacterium tuberculosis* putative classical nitroreductase gene whose expression is coregulated with that of the *acr* gene within macrophages, in standing versus shaking cultures, and under low oxygen conditions. *Infect Immun* **70**:1518–1529.
225. **Hu Y, Coates ARM.** 2011. *Mycobacterium tuberculosis* *acg* gene is required for growth and virulence in vivo. *PLoS One* **6**.
226. **Berney M, Cook GM.** 2010. Unique flexibility in energy metabolism allows mycobacteria to combat starvation and hypoxia. *PLoS One* **5**.
227. **Trauner A, Loughheed KEA, Bennett MH, Hingley-Wilson SM, Williams HD.** 2012. The dormancy regulator DosR controls ribosome stability in hypoxic mycobacteria. *J Biol Chem* **287**:24053–24063.
228. **Ueta M, Ohniwa RL, Yoshida H, Maki Y, Wada C, Wada A.** 2008. Role of HPF (hibernation promoting factor) in translational activity in *Escherichia coli*. *J Biochem* **143**:425–433.
229. **Aiso T, Yoshida H, Wada A, Ohki R.** 2005. Modulation of mRNA stability participates in stationary-phase-specific expression of ribosome modulation factor. *J Bacteriol* **187**:1951–1958.
230. **Schmalisch M, Langbein I, Stulke J, Stülke J.** 2002. The general stress protein Ctc of *Bacillus subtilis* is a ribosomal protein. *J Mol Microbiol Biotechnol* **4**:495–501.
231. **Freestone P, Nyström T, Trinei M, Norris V.** 1997. The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *J Mol Biol* **274**:318–24.
232. **Diez A, Gustavsson N, Nystrom T.** 2000. The universal stress protein A of *Escherichia coli* is required for resistance to DNA damaging agents and is regulated by a RecA/FtsK-dependent regulatory pathway. *Mol Microbiol* **36**:1494–1503.
233. **Peh HL, Toh A, Murugasu-Oei B, Dick T.** 2001. In vitro activities of mitomycin C against growing and hypoxic dormant tubercle bacilli. *Antimicrob Agents Chemother* **45**:2403–2404.
234. **Banerjee A, Adolph RS, Gopalakrishnapai J, Kleinboelting S, Emmerich C, Steegborn C,**

- Visweswariah SS.** 2015. A Universal Stress Protein ( USP ) in Mycobacteria Binds cAMP  
**290:**12731–12743.
235. **Stapleton M, Haq I, Hunt DM, Arnvig KB, Artymiuk PJ, Buxton RS, Green J.** 2010. *Mycobacterium tuberculosis* cAMP receptor protein (Rv3676) differs from the *Escherichia coli* paradigm in its cAMP binding and DNA binding properties and transcription activation properties. *J Biol Chem* **285:**7016–7027.
236. **Nambi S, Basu N, Visweswariah SS.** 2010. cAMP-regulated protein lysine acetylases in mycobacteria. *J Biol Chem* **285:**24313–24323.
237. **Lowrie D, Jackett P, Ratcliffe N.** 1975. Mycobacterium microti may protect itself from intracellular destruction by releasing cyclic AMP into phagosomes. *Nature* **254:**83–86.
238. **Dass BKM, Sharma R, Shenoy AR, Mattoo R, Visweswariah SS.** 2008. Cyclic AMP in mycobacteria: Characterization and functional role of the Rv1647 ortholog in *Mycobacterium smegmatis*. *J Bacteriol* **190:**3824–3834.
239. **Bai G, Schaak D.** 2009. cAMP levels within *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG increase upon infection of macrophages. *FEMS Immunol &*
240. **Seliger B, Dressler SP, Wang E, Kellner R, Recktenwald C, Lottspeich F, Marincola FM, Baumgärtner M, Atkins D, Lichtenfels R.** 2012. Combined analysis of transcriptome and proteome data as a tool for the identification of candidate biomarkers in renal cell carcinoma. *Proteomics* **9:**1567–1581.
241. **Fernie a. R, Stitt M.** 2012. On the Discordance of Metabolomics with Proteomics and Transcriptomics: Coping with Increasing Complexity in Logic, Chemistry, and Network Interactions *Scientific Correspondence. Plant Physiol* **158:**1139–1145.
242. **Wu ML, Gengenbacher M, Dick T.** 2016. Mild nutrient starvation triggers the development of a small-cell survival morphotype in mycobacteria. *Front Microbiol* **7:**1–11.
243. **Dietrich J, Roy S, Rosenkrands I, Lindenstrøm T, Filskov J, Rasmussen EM, Cassidy J, Andersen P.** 2015. Differential influence of nutrient starved *Mycobacterium tuberculosis* on adaptive immunity results in progressive TB disease and pathology. *Infect Immun* **83:**IAI.01055-15.
244. **Antoine AD, Tepper BS.** 1969. Environmental control of glycogen and lipid content of *Mycobacterium tuberculosis*. *J Bacteriol* **100:**538–539.
245. **Reed MB, Gagneux S, DeRiemer K, Small PM, Barry CE.** 2007. The W-Beijing lineage of

- Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. *J Bacteriol* **189**:2583–2589.
246. **Khare G, Nangpal P, Tyagi AK.** 2017. Differential roles of iron storage proteins in maintaining the iron homeostasis in *Mycobacterium tuberculosis*. *PLoS One* **12**:1–18.
247. **Wessel D, Flügge UI.** 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **138**:141–143.
248. **Kumar A, Deshane JS, Crossman DK, Bolisetty S, Yan BS, Kramnik I, Agarwal A, Steyn AJC.** 2008. Heme oxygenase-1-derived carbon monoxide induces the *Mycobacterium tuberculosis* dormancy regulon. *J Biol Chem* **283**:18032–18039.
249. **Hingley-Wilson SM, Loughheed KEA, Ferguson K, Leiva S, Williams HD.** 2010. Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable in vitro. *Tuberculosis* **90**:236–244.
250. **Nilsen TW.** 2014. RNase footprinting to map sites of RNA-protein interactions. *Cold Spring Harb Protoc* **2014**:677–682.
251. **Sala C, Haouz A, Saul FA, Miras I, Rosenkrands I, Alzari PM, Cole ST.** 2009. Genome-wide regulon and crystal structure of Blal (Rv1846c) from *Mycobacterium tuberculosis*. *Mol Microbiol* **71**:1102–1116.
252. **Davis SE, Mooney R a., Kanin EI, Grass J, Landick R, Ansari AZ.** 2011. Mapping *E. coli* RNA polymerase and associated transcription factors and identifying promoters genome-wide *Methods in Enzymology*, 1sted. Elsevier Inc.

## Addendum A: Buffers

Table A. 1 NP-MS Buffers

Buffer	Component	Source	Volume/Weight/ Concentration	Diluent	pH	Storage	Reference
Glycine	Glycine	Merck	2.5 M	Water		Room Temperature	(251)
Tris	Tris	Merck	1M		7.5,	Room	
	HCl	Merck	To pH		8.0	Temperature	
Micrococcal nuclease resuspension buffer	Tris pH 8.0		10 mM	Water		-20°C	(252)
	Glycerol	Merck	50%				
TAE	1 MTris	Merck	40 mM	Water		Room Temperature	
	Acetic acid	Merck	20 mM				
	EDTA	Sigma-Aldrich	1 mM				
CaCl <sub>2</sub> mix	Tris-HCl pH 8.0		0.5 M	Water		-20°C	(252)
	CaCl <sub>2</sub>	Sigma-Aldrich	50 mM				
PBS-T	PBS	WhiteSci	1X	Water		Room temperature	
	Tween-20	Sigma-Aldrich	0.02%				
Immunoprecipitation buffer I	Tris-HCl pH 8.0		100 mM	Water		Room Temperature	(252)
	NaCl	Sigma-Aldrich	300 mM				
	Triton X-100	Merck	2%				
Immunoprecipitation buffer II	Hepes-KOH pH 7.5	Sigma-Aldrich	50 mM	Water		Room Temperature	(33)
	NaCl	Sigma-Aldrich	150 mM				
	EDTA	Sigma-Aldrich	1 mM				
	Triton X-100	Merck	1%				
	Sodium deoxycholate	Sigma-Aldrich	0.1%				
	SDS	Sigma-Aldrich	0.1%				
IP Buffer II + 500 mM NaCl	Immunoprecipitation buffer II		90%			Room Temperature	(251)
	NaCl	Sigma-Aldrich	10%				
IP Buffer II + 750 mM NaCl	Immunoprecipitation buffer II		85%			Room Temperature	
	NaCl	Sigma-Aldrich	15%				
LiCl Wash Buffer	Tris-HCl pH 8.0	Merck	10 mM	Water		Room Temperature	(251)
	LiCl	Sigma-Aldrich	250 mM				

	EDTA	Sigma-Aldrich	1 mM			
	Igepal ® CA-630	Sigma-Aldrich	0.5%			
	Sodium deoxycholate	Sigma-Aldrich	0.5%			
Ammonium Bicarbonate	NH <sub>4</sub> CHO <sub>3</sub>	Sigma-Aldrich	50 mM	Water		Freshly prepared
MgCl <sub>2</sub>	MgCl <sub>2</sub>	Sigma-Aldrich	100 mM	Water		4°C (Autoclaved)
CaCl <sub>2</sub>	CaCl <sub>2</sub>	Sigma-Aldrich	100 mM	Water		4°C (Autoclaved)
	Glycerol	Merck	15%			
Elution Buffer	Tris-HCl pH 7.5		50 mM	Water		Room Temperature (251)
	EDTA	Sigma-Aldrich	10 mM			
	SDS	Sigma-Aldrich	1%			
TBS Buffer	Tris	Merck	50 mM	Water	7.5	4°C
	NaCl	Sigma-Aldrich	150 mM			
TE Buffer	Tris-HCl pH 7.5	Merck	10 mM	Water		Room Temperature
	EDTA	Sigma-Aldrich	1 mM			

**Table A. 2 Culture Media**

Culture Medium	Composition	Source	Sterilisation
Luria-Bertani broth (LB)	0.5% Sodium chloride 1% Tryptone 0.5% Yeast extract	Sigma-Aldrich Merck Merck	Autoclave
Luria-Bertani agar (LB agar)	1.2% Bacteriological agar 0.5% Sodium chloride 1% Tryptone 0.5% Yeast extract	Merck Sigma-Aldrich Merck Merck	Autoclave
Difco™ Middlebrook 7H9	Middlebrook 7H9 (0.47%), supplemented with Bovine serum albumin (0.5%), Glucose (0.2%), Glycerol (0.2%) and Tween-80 (0.05%)	Becton Dickinson (BD)	Filter Sterilisation
BBL™ Seven H11 Base	BBL™ Seven H11 Base (1.9%), supplemented with Bovine serum albumin (0.5%), Glucose (0.2%), Glycerol (0.2%) and Tween-80 (0.05%)	Becton Dickinson (BD)	Filter Sterilisation

**Table A. 3 SDS-PAGE and Silver Staining Buffers**

Buffer	Component	Source	Volume/Weight/ Concentration	Diluent	Storage	pH	
Separating buffer	Tris	Merck	1.5 M	Water	Room temperature	pH 8.8	
Stacking buffer	Tris	Merck	1.0 M	Water	Room temperature	pH 6.8	
2X Reducing sample buffer	SDS	Merck	3.4 ml		Room temperature, dark container		
	Glycerol	Merck	2 ml				
	SDS (20%)	Sigma-Aldrich	3 ml				
	Bromophenol blue	Sigma-Aldrich	500 $\mu$ l				
	EDTA	Sigma-Aldrich	200 $\mu$ l				
	B-Mercaptoethanol	Sigma-Aldrich	1 ml				
Running buffer (2L)	Tris	Merck	6 g	Water	Room temperature		
	Glycine	Merck	28.8 g				
	SDS (10%)	Sigma-Aldrich	20 ml				
Fixing solution (1L)	Methanol	Sigma-Aldrich	500 ml	Water	Room temperature		
	Acetic acid	Sigma-Aldrich	100 ml				
	Ammonium Acetate	Sigma-Aldrich	7.708 g				
Developing solution (1L)	Formaldehyde	Merck	973 $\mu$ l	Water	Room temperature		
	Sodium Carbonate	Sigma-Aldrich	20 g				

**Table A. 4 Separating gel composition**

<b>Component</b>	<b>Source</b>	<b>12% Separating Gel (2 x 0.75 mm mini-slab gels)</b>
Acrylamide/bisacrylamide	Sigma-Aldrich	4.0 ml
1.5 M Tris-HCl pH 8.8	Merck	2.5 ml
10 % SDS	Sigma-Aldrich	0.1 ml
Water		3.35 ml
TEMED	Sigma-Aldrich	10 $\mu$ l
20% Ammonium persulfate	Sigma-Aldrich	50 $\mu$ l

**Table A. 5 Stacking gel composition**

<b>Component</b>	<b>Source</b>	<b>3% Separating Gel (4 x 0.75 mm mini-slab gels)</b>
Acrylamide/bisacrylamide	Sigma-Aldrich	1.3 ml
1.0 M Tris-HCl pH 6.8	Merck	1.25 ml
10 % SDS	Sigma-Aldrich	0.1 ml
Water		7.4 ml
TEMED	Sigma-Aldrich	20 $\mu$ l
20% Ammonium persulfate	Sigma-Aldrich	50 $\mu$ l

**Table A. 6 Western Blotting buffers**

<b>Buffer</b>	<b>Composition</b>	<b>Diluent</b>	<b>Storage</b>
Blotting/Transfer buffer	25 mM Tris 192 mM Glycine 20% Methanol	Water	4°C
TBS-T Wash Buffer	20 mM Tris 137 mM NaCl Tween-20	Water	Room Temperature
Blocking buffer	5% BSA	TBS-T Wash Buffer	Make fresh

---

## Addendum B: NP-MS results

---

### B.1. Formaldehyde induced modification frequency determination

---

**Table B. 1 Formaldehyde induced modification frequency determination**

Modifications	Weight added	Amino acid/ terminus modified	Number of proteins identified*	MS/MS count peptides for rpoB	Number of Unique peptides for rpoB	Number of proteins with Variable modification
Oxidation (M),	16	M	532	1275	90	33
Acetyl (Protein N-term)	42	N-terminus				
Addition of Glycine	57	K, S, T	544	1082	90	29
Thiazolidine (C)	12	N-terminus	529	1089	91	0
Methylol and Glycine	87	H, N, Q, W, Y	532	992	90	30
Modification of Arginine	99	R	511	1051	90	6
Modification of Arginine 2	198	R	528	1085	90	0
Methylol	30	C, K, H	530	1086	91	4
Modification of Tryptophan	12	W	523	1105	91	4
Modification of Tyrosine	174	Y	525	1090	91	3
Dimethylation	28	K	531	1077	91	17

\* proteins identified for two unique peptides

## B.2. Low confidence protein identification

Table B. 2 Less stringent protein identification list in *M. smegmatis*

UniProt Annotation	MSMEG Annotation	RV homologue	Gene Name	Protein Names	Gene Ontology
<b>High Confidence Proteins</b>					
A0QNE0	MSMEG_0005	Rv0005	<i>gyrB</i>	DNA gyrase subunit B	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; magnesium ion binding
P48354	MSMEG_0006	Rv0006	<i>gyrA</i>	DNA gyrase subunit A	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; response to antibiotic
A0QNF5	MSMEG_0023	Rv0008c	<i>cwsA</i>	Cell wall synthesis protein	cell cycle; cell division; cell wall biogenesis; integral component of plasma membrane; protein localization; regulation of cell shape
A0QNI9	MSMEG_0059	Rv3868	<i>eccA1</i>	ATPase, AAA family protein	ATP binding
A0QNJ7	MSMEG_0067	Rv3876	<i>espI</i>	Uncharacterized protein	
A0QNM0	MSMEG_0092	Rv0144		Probable transcriptional regulatory protein	DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0QNN0	MSMEG_0102	Rv0154c	<i>fadE2</i>	Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QNY0	MSMEG_0203			IS1096, tnpR protein	
A0QNZ3	MSMEG_0216	Rv0216		3-hydroxyacyl-CoA dehydrogenase	oxidoreductase activity
A0QP06	MSMEG_0229	Rv0189c	<i>ilvD</i>	Dihydroxy-acid dehydratase	4 iron, 4 sulfur cluster binding; dihydroxy-acid dehydratase activity; isoleucine biosynthetic process; metal ion binding; valine biosynthetic process
A0QP11	MSMEG_0234	Rv0198c	<i>zmp1</i>	Metallopeptidase	metalloendopeptidase activity
A0QP15	MSMEG_0238			CoA binding domain protein	cofactor binding
A0QP16	MSMEG_0239	Rv3340	<i>metC</i>	O-acetylserine sulfhydrylase	O-acetylhomoserine aminocarboxypropyltransferase activity; pyridoxal phosphate binding
A0QP20	MSMEG_0243			Uncharacterized protein	
A0QP32	MSMEG_0255	Rv0211	<i>pckG</i>	Phosphoenolpyruvate carboxykinase	cytoplasm; gluconeogenesis; GTP binding; manganese ion binding; phosphoenolpyruvate carboxykinase (GTP) activity
A0QPE7	MSMEG_0372	Rv0242c	<i>fabG4</i>	3-oxoacyl-acyl-carrier protein reductase	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity
A0QPG3	MSMEG_0388		<i>tylF</i>	Macrocin-O-methyltransferase	macrocin O-methyltransferase activity

A0QPH5	MSMEG_0400			Peptide synthetase	catalytic activity
Q3L891	MSMEG_0402		<i>mps2</i>	Linear gramicidin synthetase subunit D	isomerase activity; oxidoreductase activity
Q3L887	MSMEG_0406	Rv0244c	<i>fadE5</i>	Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; butyryl-CoA dehydrogenase activity
Q3L885	MSMEG_0408		<i>pkS</i>	Type I modular polyketide synthase	erythronolide synthase activity; oxidoreductase activity; phosphopantetheine binding; zinc ion binding
A0QPI4	MSMEG_0409	Rv1182		Condensation domain protein	
A0QPJ0	MSMEG_0415	Rv0245		NADH-fmn oxidoreductase	FMN binding; riboflavin reductase (NADPH) activity
A0QPN1	MSMEG_0456			DNA topoisomerase	ATP binding; chromosome; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change
A0QPN2	MSMEG_0457			DNA gyrase subunit B-like protein	ATP binding; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change
A0QPV4	MSMEG_0531			Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QQB0	MSMEG_0690	Rv0338c		Iron-sulfur cluster-binding protein	iron-sulfur cluster binding
A0QQF9	MSMEG_0741			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0QQH2	MSMEG_0754	Rv0360c		Uncharacterized protein	
A0QQP0	MSMEG_0824	Rv0421c		Uncharacterized protein	
A0QQP1	MSMEG_0825	Rv0422c	<i>thiD</i>	Phosphomethylpyrimidine kinase	ATP binding; phosphomethylpyrimidine kinase activity; thiamine biosynthetic process
A0QQP2	MSMEG_0826	Rv0423c	<i>thiC</i>	Thiamine biosynthesis protein ThiC	4 iron, 4 sulfur cluster binding; lyase activity; thiamine biosynthetic process; thiamine diphosphate biosynthetic process; zinc ion binding
A0QQS3	MSMEG_0859			Transcriptional regulator, TetR family	DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0QQS8	MSMEG_0863	Rv0439c		Short chain dehydrogenase	oxidoreductase activity
A0QQW8	MSMEG_0903	Rv0462	<i>lpdA</i>	Dihydrolipoyl dehydrogenase	cell redox homeostasis; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
Q3I5Q7	MSMEG_0919	Rv0475	<i>hbhA</i>	Uncharacterized protein	cell adhesion; cell surface; heparin binding; pathogenesis
A0QR13	MSMEG_0948	Rv0504c		Uncharacterized protein	
A0QR19	MSMEG_0954	Rv0511	<i>hemD</i>	Uroporphyrinogen-III synthase	methyltransferase activity; tetrapyrrole biosynthetic process; uroporphyrinogen-III synthase activity
A0QR89	MSMEG_1028			Geranylgeranyl reductase	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor
A0QRB8	MSMEG_1060			Uncharacterized protein	
A0QRM0	MSMEG_1165	Rv0566c		Uncharacterized protein	
A0QRU1	MSMEG_1238			Type III restriction enzyme, res subunit	ATP binding; DNA binding; Type I site-specific deoxyribonuclease activity

A0QRV8	MSMEG_1255			DNA helicase	ATP binding; ATP-dependent DNA helicase activity; DNA binding
A0QS36	MSMEG_1334	Rv0634c		Glyoxalase II, GloB	hydroxyacylglutathione hydrolase activity
A0QS41	MSMEG_1341	Rv0636		MaoC-like dehydratase	
A0QS42	MSMEG_1342	Rv0637		Uncharacterized protein	
A0QS49	MSMEG_1350		<i>cmaA1</i>	Mycolic acid synthase UmaA	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
A0QS81	MSMEG_1383	Rv0670	<i>nfo</i>	Endonuclease IV	deoxyribonuclease IV (phage-T4-induced) activity; DNA binding; DNA repair; zinc ion binding
A0QSB1	MSMEG_1416	Rv0688		Pyridine nucleotide-disulphide oxidoreductase	cell redox homeostasis; ferredoxin-NAD <sup>+</sup> reductase activity; flavin adenine dinucleotide binding
A0QSL0	MSMEG_1516			Thioredoxin reductase	oxidoreductase activity; phosphorelay signal transduction system
A0QSN7	MSMEG_1543	Rv0458		Eptc-inducible aldehyde dehydrogenase	aldehyde dehydrogenase (NAD) activity
A0QSU3	MSMEG_1602	Rv3411c	<i>guaB</i>	IMP dehydrogenase	adenyl nucleotide binding; GMP biosynthetic process; IMP dehydrogenase activity; metal ion binding
A0QSV0	MSMEG_1610	Rv3396c	<i>guaA</i>	GMP synthase	ATP binding; glutamine metabolic process; GMP biosynthetic process; GMP synthase (glutamine-hydrolyzing) activity; pyrophosphatase activity
A0QSX4	MSMEG_1635	Rv3368c		Nitroreductase family protein	oxidoreductase activity
A0QSZ6	MSMEG_1657	Rv3336c	<i>trpS</i>	Tryptophan-tRNA ligase	ATP binding; cytoplasm; tryptophan-tRNA ligase activity; tryptophanyl-tRNA aminoacylation
A0QT08	MSMEG_1670	Rv3318	<i>sdhA</i>	Succinate dehydrogenase flavoprotein subunit	electron transport chain; flavin adenine dinucleotide binding; succinate dehydrogenase activity; tricarboxylic acid cycle
A0QT17	MSMEG_1679		<i>amiB1</i>	Amidohydrolase AmiB1	hydrolase activity
A0QT18	MSMEG_1680			Uncharacterized protein	
A0QT19	MSMEG_1681			Endoribonuclease L-PSP superfamily protein	
A0QTE7	MSMEG_1813	Rv3280	<i>accD5</i>	Propionyl-CoA carboxylase beta chain	propionyl-CoA carboxylase activity
A0QTF4	MSMEG_1821	Rv3274c	<i>fadE25</i>	Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; butyryl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QTK2	MSMEG_1874	Rv3246c	<i>mtrA</i>	DNA-binding response regulator	cytoplasm; DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
P71533	MSMEG_1881	Rv3240c	<i>secA1</i>	Protein translocase subunit SecA 1	ATP binding; cytoplasm; intracellular protein transmembrane transport; plasma membrane; protein import; protein targeting
A0QTP2	MSMEG_1914	Rv3223c	<i>sigH</i>	RNA polymerase sigma-H factor	DNA binding; DNA-templated transcription, initiation; sequence-specific DNA binding transcription factor activity; sigma factor activity
A0QEQ8	MSMEG_1930	Rv3211		DEAD/DEAH box helicase	ATP binding; ATP-dependent helicase activity; nucleic acid binding
A0QTR5	MSMEG_1937	Rv3206c	<i>MoeB</i>	Molybdopterin biosynthesis protein	catalytic activity
A0QU00	MSMEG_2026			Short chain dehydrogenase	oxidoreductase activity

A0QU53	MSMEG_2080	Rv3140	<i>fadE23</i>	Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity
A0QU61	MSMEG_2089	Rv3102c	<i>ftsE</i>	Cell division ATP-binding protein	ATPase activity; ATP binding; cell division
A0QUE0	MSMEG_2174			DNA helicase	ATP binding; ATP-dependent DNA helicase activity; DNA binding
P0CH00	MSMEG_2299	Rv3051c	<i>nrdE2</i>	Ribonucleotide reductase R1 subunit 2	ATP binding; DNA replication; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
P0CH37	MSMEG_2317	Rv3045	<i>adhC2</i>	NADP-dependent alcohol dehydrogenase C 2	alcohol dehydrogenase (NADP+) activity; zinc ion binding
A0QRA8	MSMEG_2329	Rv3038c		Methyltransferase type 11	methyltransferase activity
A0QUX1	MSMEG_2367	Rv3009c	<i>gatB</i>	Aspartylglutamyl-tRNA amidotransferase subunit B	ATP binding; carbon-nitrogen ligase activity, with glutamine as amido-N-donor; translation
A0QUX6	MSMEG_2372	Rv3003c	<i>ilvB</i>	Acetolactate synthase	acetolactate synthase activity; flavin adenine dinucleotide binding; isoleucine biosynthetic process; magnesium ion binding; thiamine pyrophosphate binding; valine biosynthetic process
A0QUY6	MSMEG_2382	Rv2993c		5-carboxymethyl-2-hydroxyisovalerate delta-isomerase	isomerase activity
A0QUY9	MSMEG_2387	Rv2988c	<i>leuC</i>	3-isopropylmalate dehydratase, large subunit	3-isopropylmalate dehydratase activity; 4 iron, 4 sulfur cluster binding; leucine biosynthetic process; metal ion binding
A0QUZ0	MSMEG_2388	Rv2987c	<i>leuD</i>	3-isopropylmalate dehydratase small subunit	3-isopropylmalate dehydratase activity; 3-isopropylmalate dehydratase complex; leucine biosynthetic process
A0QV32	MSMEG_2430	Rv2916c	<i>ffh</i>	Signal recognition particle protein	7S RNA binding; GTPase activity; GTP binding; signal recognition particle; SRP-dependent cotranslational protein targeting to membrane
A0QV38	MSMEG_2436	Rv2908c		Uncharacterized protein	RNA binding
A0QVE0	MSMEG_2541	Rv2882c	<i>frr</i>	Ribosome-recycling factor	cytoplasm; translational termination
A0QVM4	MSMEG_2625	Rv2841c	<i>nusA</i>	Transcription termination/antitermination protein	regulation of DNA-templated transcription, termination; RNA binding; sequence-specific DNA binding transcription factor activity
A0QVU2	MSMEG_2695	Rv2744c		35 kDa protein	
Q59560	MSMEG_2723	Rv2737c	<i>recA</i>	Recombinase A	ATP binding; cytoplasm; damaged DNA binding; DNA-dependent ATPase activity; DNA recombination; DNA repair; single-stranded DNA binding; SOS response
A0QVZ3	MSMEG_2750	Rv2711	<i>ideR</i>	Iron-dependent repressor	DNA binding; sequence-specific DNA binding transcription factor activity; transition metal ion binding
A0QVZ5	MSMEG_2752	Rv2710	<i>mysB</i>	RNA polymerase sigma factor	DNA binding; DNA-templated transcription, initiation; sequence-specific DNA binding transcription factor activity; sigma factor activity
A0QW02	MSMEG_2758	Rv2703	<i>rpoD</i>	RNA polymerase sigma factor SigA	cytoplasm; DNA binding; sequence-specific DNA binding transcription factor activity; sigma factor activity; transcription initiation from bacterial-type RNA polymerase promoter
A0QW25	MSMEG_2782	Rv2676c		Uncharacterized protein	

A0QW71	MSMEG_2839	Rv2986c	<i>hupB</i>	Transcriptional accessory protein	DNA binding; DNA repair; hydrolase activity, acting on ester bonds; RNA binding
A0QWG8	MSMEG_2937	Rv2606c	<i>pdxS</i>	Pyridoxal 5'-phosphate synthase subunit	lyase activity; pyridoxal phosphate biosynthetic process
A0QWQ0	MSMEG_3020	Rv2033c		Uncharacterized protein	
A0QWQ4	MSMEG_3025	Rv2555c	<i>alaS</i>	Alanine-tRNA ligase	alanine-tRNA ligase activity; alanyl-tRNA aminoacylation; ATP binding; cytoplasm; tRNA binding; zinc ion binding
A0QWR4	MSMEG_3035	Rv2534c	<i>efp</i>	Elongation factor P	cytoplasm; peptide biosynthetic process; translation elongation factor activity
A0QWR5	MSMEG_3036	Rv2533c	<i>nusB</i>	N utilization substance protein B homolog	DNA-templated transcription, termination; regulation of transcription, DNA-templated; RNA binding
A0QWS4	MSMEG_3046	Rv1383	<i>carA</i>	Carbamoyl-phosphate synthetase glutamine chain	'de novo' UMP biosynthetic process; arginine biosynthetic process; ATP binding; carbamoyl phosphate biosynthetic process; carbamoyl-phosphate synthase (glutamine-hydrolyzing) activity; glutamine catabolic process
A0QWT1	MSMEG_3053	Rv1390	<i>rpoZ</i>	DNA-directed RNA polymerase subunit omega	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QWT7	MSMEG_3059	Rv1400c		Esterase	hydrolase activity
A0QWV0	MSMEG_3072	Rv1415	<i>ribAB</i>	Riboflavin biosynthesis protein	3,4-dihydroxy-2-butanone-4-phosphate synthase activity; GTP binding; GTP cyclohydrolase II activity; magnesium ion binding; manganese ion binding; riboflavin biosynthetic process; zinc ion binding
A0QWV9	MSMEG_3081	Rv1423	<i>whiA</i>	Putative sporulation transcription regulator	DNA binding; regulation of sporulation; regulation of transcription, DNA-templated; transcription, DNA-templated
A0QWW4	MSMEG_3086	Rv1438	<i>tpiA</i>	Triosephosphate isomerase	cytoplasm; gluconeogenesis; glycolytic process; pentose-phosphate shunt; triose-phosphate isomerase activity
A0QWY3	MSMEG_3106	Rv1454c		Quinone oxidoreductase	NADPH:quinone reductase activity; zinc ion binding
A0QWZ9	MSMEG_3122	Rv1461	<i>sufB</i>	FeS assembly protein SufB	iron-sulfur cluster assembly
A0QX00	MSMEG_3123	Rv1462	<i>sufD</i>	FeS assembly protein SufD	iron-sulfur cluster assembly
A0QX01	MSMEG_3124	Rv1463	<i>sufC</i>	FeS assembly ATPase SufC	ATPase activity; ATP binding; transport
A0QX15	MSMEG_3138	Rv1471	<i>trx</i>	Thioredoxin	cell redox homeostasis; glycerol ether metabolic process; protein disulfide oxidoreductase activity
P71534	MSMEG_3150	Rv1483	<i>fabG</i>	3-oxoacyl-[acyl-carrier-protein] reductase FabG	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity; fatty acid elongation; NADP binding
A0QX55	MSMEG_3178	Rv1547	<i>dnaE1</i>	DNA-directed DNA polymerase	3'-5' exonuclease activity; DNA binding; DNA-directed DNA polymerase activity; DNA replication
A0QX95	MSMEG_3219	Rv1611	<i>trpC</i>	Indole-3-glycerol phosphate synthase	indole-3-glycerol-phosphate synthase activity; tryptophan biosynthetic process
A0QYB0	MSMEG_3595			Uncharacterized protein	
A0QYB1	MSMEG_3596			AAA ATPase	

A0QYD4	MSMEG_3619	Rv1856c		Short chain dehydrogenase	oxidoreductase activity
A0QYD6	MSMEG_3621	Rv1854c	<i>ndh</i>	NADH dehydrogenase	flavin adenine dinucleotide binding; NADH dehydrogenase activity
A0QYE7	MSMEG_3632	Rv1844c	<i>gnd</i>	6-phosphogluconate dehydrogenase, decarboxylating	NADP binding; pentose-phosphate shunt; phosphogluconate dehydrogenase (decarboxylating) activity
A0QYF7	MSMEG_3642	Rv1832	<i>gcvP</i>	Glycine	glycine decarboxylation via glycine cleavage system; glycine dehydrogenase (decarboxylating) activity; pyridoxal phosphate binding
A0QYG0	MSMEG_3645	Rv1829		Uncharacterized protein	nuclease activity
A0QYQ7	MSMEG_3746	Rv1699	<i>pyrG</i>	CTP synthase	'de novo' CTP biosynthetic process; ATP binding; CTP synthase activity; glutamine metabolic process
A0QYR3	MSMEG_3754	Rv1691		TPR-repeat-containing protein	
A0QYS6	MSMEG_3770	Rv1658	<i>argG</i>	Argininosuccinate synthase	arginine biosynthetic process; argininosuccinate synthase activity; ATP binding; cytoplasm
A0QYU8	MSMEG_3793	Rv1641	<i>infC</i>	Translation initiation factor IF-3	cytoplasm; translation initiation factor activity
A0QYW4	MSMEG_3808	Rv1638	<i>uvrA</i>	UvrABC system protein A	ATPase activity; ATP binding; cytoplasm; DNA binding; excinuclease ABC activity; excinuclease repair complex; nucleotide-excision repair; SOS response; zinc ion binding
A0QYZ2	MSMEG_3839	Rv1629	<i>polA</i>	DNA polymerase I	3'-5' exonuclease activity; DNA binding; DNA-dependent DNA replication; DNA-directed DNA polymerase activity
A0QZ11	MSMEG_3858	Rv2050	<i>rbpA</i>	RNA polymerase-binding protein RbpA	bacterial-type RNA polymerase core enzyme binding; positive regulation of transcription, DNA-templated; response to antibiotic; transcription, DNA-templated
A0QZ49	MSMEG_3897	Rv2112c	<i>dop</i>	Pup deamidase/depupylase	ATP binding; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides; metal ion binding; modification-dependent protein catabolic process; peptidase activity; proteasomal protein catabolic process; protein pupylation
A0QZ54	MSMEG_3902	Rv2115c	<i>mpa</i>	Mycobacterial proteasome ATPase	ATPase activity; ATP binding; modification-dependent protein catabolic process; proteasomal protein catabolic process; proteasome-activating nucleotidase complex; protein unfolding
A0QZX6	MSMEG_4185	Rv2124c	<i>methH</i>	Methionine synthase	cobalamin binding; intracellular; methionine synthase activity; pteridine-containing compound metabolic process; S-adenosylmethionine-homocysteine S-methyltransferase activity; zinc ion binding
A0QZY0	MSMEG_4189	Rv2130c	<i>mshC</i>	Mycothiol ligase	ATP binding; cysteine-glucosaminylinositol ligase activity; mycothiol biosynthetic process; zinc ion binding
A0R006	MSMEG_4217	Rv2145c	<i>wag31</i>	Cell wall synthesis protein	cell cycle; cell division; cytoplasm; regulation of cell shape
A0R012	MSMEG_4222	Rv2150c	<i>ftsZ</i>	Cell division protein	barrier septum assembly; cell division site; cytoplasm; FtsZ-dependent cytokinesis; GTPase activity; GTP binding; protein complex; protein polymerization
A0R029	MSMEG_4240	Rv2173		Polyprenyl synthetase	isoprenoid biosynthetic process; transferase activity

A0R033	MSMEG_4244	Rv2178c	<i>aroG</i>	3-deoxy-D-arabino-heptulosonate phosphate synthase	7-	3-deoxy-7-phosphoheptulonate synthase activity; aromatic amino acid family biosynthetic process; chorismate biosynthetic process; manganese ion binding; plasma membrane
A0R061	MSMEG_4272	Rv2204c		Iron-sulfur cluster assembly accessory protein		iron-sulfur cluster assembly; iron-sulfur cluster binding; structural molecule activity
A0R069	MSMEG_4281	Rv2213	<i>pepA</i>	Probable cytosol aminopeptidase		aminopeptidase activity; cytoplasm; manganese ion binding; metalloexopeptidase activity
A0R072	MSMEG_4283	Rv2215	<i>sucB</i>	2-oxoglutarate dehydrogenase, component, succinyltransferase	E2 dihydrolipoamide	dihydrolipoyllysine-residue succinyltransferase activity
A0R090	MSMEG_4301			Acyl-CoA synthase		catalytic activity
A0R095	MSMEG_4306	Rv2229c		Uncharacterized protein		
A0R0A1	MSMEG_4313			Glyoxalase/bleomycin protein/dioxygenase	resistance	dioxygenase activity
A0R0B0	MSMEG_4323	Rv2241	<i>aceE</i>	Pyruvate dehydrogenase E1 component		glycolytic process; pyruvate dehydrogenase (acetyl-transferring) activity
A0R0B6	MSMEG_4329	Rv2247	<i>accD6</i>	Propionyl-CoA carboxylase beta chain		propionyl-CoA carboxylase activity
A0R0C7	MSMEG_4340	Rv2259	<i>adhE2</i>	Zinc-dependent alcohol dehydrogenase		oxidoreductase activity; zinc ion binding
A0R0Q9	MSMEG_4474			Acyl-CoA oxidase		acyl-CoA dehydrogenase activity; acyl-CoA oxidase activity; fatty acid beta-oxidation; flavin adenine dinucleotide binding; peroxisome
A0R0R9	MSMEG_4485	Rv2357c	<i>glyS</i>	Glycine-tRNA ligase		ATP binding; cytoplasm; glycine-tRNA ligase activity; glycyl-tRNA aminoacylation
A0R0T8	MSMEG_4504	Rv2373c	<i>dnaJ</i>	Chaperone protein		ATP binding; cytoplasm; DNA replication; protein folding; response to heat; zinc ion binding
A0R0W1	MSMEG_4527	Rv2391		Ferredoxin sulfite reductase		4 iron, 4 sulfur cluster binding; heme binding; metal ion binding; sulfite reductase (ferredoxin) activity
A0R152	MSMEG_4626	Rv2444c	<i>rne</i>	Ribonuclease E		cytoplasm; metal ion binding; mRNA processing; ribonuclease activity; RNA binding; rRNA processing
A0R171	MSMEG_4646	Rv2455c		Pyruvate synthase		2-oxoglutarate synthase activity
A0R196	MSMEG_4671	Rv2457c	<i>clpX</i>	ATP-dependent Clp protease subunit	ATP-binding	ATP binding; protein folding; zinc ion binding
A0R197	MSMEG_4672	Rv2460c	<i>clpP</i>	Endopeptidase Clp		cytoplasm; serine-type endopeptidase activity
A0R199	MSMEG_4674	Rv2462c	<i>tig</i>	Trigger factor		cell cycle; cell division; cytoplasm; peptidyl-prolyl cis-trans isomerase activity; protein folding; protein transport
A0R1C3	MSMEG_4700	Rv2477c		ABC-transporter protein, component	ATP binding	ATPase activity; ATP binding
A0R1G3	MSMEG_4742			Clavaldehyde dehydrogenase		oxidoreductase activity
A0R1H3	MSMEG_4753	Rv2521		Antioxidant, AhpC/TSA family protein		peroxidase activity; peroxiredoxin activity
A0R1Y7	MSMEG_4920	Rv1323		Probable acetyl-CoA acetyltransferase		acetyl-CoA C-acetyltransferase activity

A0R218	MSMEG_4954	Rv1297	<i>rho</i>	Transcription termination factor	ATP binding; DNA-templated transcription, termination; helicase activity; regulation of transcription, DNA-templated; RNA binding; RNA-dependent ATPase activity
A0R221	MSMEG_4957	Rv1294		Homoserine dehydrogenase	amino acid binding; homoserine dehydrogenase activity; isoleucine biosynthetic process; methionine biosynthetic process; NADP binding; threonine biosynthetic process
A0R239	MSMEG_4976			Isochorismatase hydrolase	hydrolase activity
A0R242	MSMEG_4979	Rv1285	<i>cysD</i>	Sulfate adenylyltransferase subunit 2	ATP binding; hydrogen sulfide biosynthetic process; sulfate adenylyltransferase (ATP) activity; sulfate assimilation; sulfate reduction
A0R248	MSMEG_4985	Rv1284		Carbonic anhydrase	carbonate dehydratase activity; zinc ion binding
A0R2A4	MSMEG_5042	Rv1253	<i>deaD</i>	ATP-dependent RNA helicase DeaD	ATP binding; ATP-dependent RNA helicase activity; cellular response to cold; cytoplasm; ribosomal large subunit assembly; RNA binding; RNA catabolic process
A0R2D0	MSMEG_5068	Rv1229c	<i>mrp</i>	ATP-binding Mrp protein	ATP binding
A0R2J0	MSMEG_5132	Rv1165	<i>typA</i>	GTP-binding protein TypA/BipA	GTPase activity; GTP binding; translation elongation factor activity
A0R2Q5	MSMEG_5197	Rv2724c		Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0R2Q7	MSMEG_5199			Acetyl-CoA acetyltransferase	acetyl-CoA C-acyltransferase activity
A0R2T0	MSMEG_5222	Rv1112	<i>ychF</i>	Ribosome-binding ATPase YchF	ATPase activity; ATP binding; GTP binding; ribosomal large subunit binding; ribosome binding
A0R2V6	MSMEG_5248	Rv1094	<i>desA2</i>	Acyl-[ACP] desaturase	acyl-[acyl-carrier-protein] desaturase activity; fatty acid metabolic process
A0R2V7	MSMEG_5249	Rv1093	<i>glyA</i>	Serine hydroxymethyltransferase	cytoplasm; glycine biosynthetic process from serine; glycine hydroxymethyltransferase activity; pyridoxal phosphate binding; tetrahydrofolate interconversion
A0R2W9	MSMEG_5261	Rv1082	<i>mca</i>	Mycothioliol S-conjugate amidase	hydrolase activity; mycothiol-dependent detoxification; mycothiol metabolic process; zinc ion binding
A0R2X8	MSMEG_5270	Rv1077	<i>cbs</i>	Cystathionine beta-synthase	adenyl nucleotide binding; cystathionine beta-synthase activity; cysteine biosynthetic process via cystathionine; cytoplasm
A0R2Y5	MSMEG_5277	Rv1070c	<i>echA8</i>	Enoyl-CoA hydratase/isomerase	enoyl-CoA hydratase activity; isomerase activity
A0R3B5	MSMEG_5412	Rv1926c		Uncharacterized protein	extracellular space
A0R3C5	MSMEG_5423	Rv1020	<i>mfd</i>	Transcription-repair-coupling factor	ATP binding; cytoplasm; damaged DNA binding; helicase activity; regulation of transcription, DNA-templated; transcription-coupled nucleotide-excision repair, DNA damage recognition
A0R3D6	MSMEG_5435	Rv1013		Putative ligase	ligase activity
A0R3D9	MSMEG_5438	Rv1010	<i>ksgA</i>	Ribosomal RNA small subunit methyltransferase A	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase activity; cytoplasm; RNA binding; rRNA (adenine-N6,N6-)-dimethyltransferase activity

A0R3L1	MSMEG_5512	Rv0958		Magnesium chelatase	ATP binding; regulation of transcription, DNA-templated
A0R3L4	MSMEG_5515	Rv0957	<i>purH</i>	Bifunctional purine biosynthesis protein	'de novo' IMP biosynthetic process; IMP cyclohydrolase activity; phosphoribosylaminoimidazolecarboxamide formyltransferase activity
A0R3Y5	MSMEG_5639	Rv0905	<i>echA6</i>	Enoyl-CoA hydratase	enoyl-CoA hydratase activity; isomerase activity
A0R409	MSMEG_5664			Peptidyl-prolyl cis-trans isomerase	peptidyl-prolyl cis-trans isomerase activity; protein folding
A0R425	MSMEG_5680	Rv0887c		Glyoxalase family protein	dioxygenase activity
A0R441	MSMEG_5696	Rv0871		'Cold-shock' DNA-binding domain protein	cytoplasm; DNA binding; regulation of transcription, DNA-templated
A0R451	MSMEG_5706	Rv0861c		DNA or RNA helicase of superfamily protein II	ATP binding; ATP-dependent DNA helicase activity; DNA binding; nucleotide-excision repair
A0R461	MSMEG_5715			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0R465	MSMEG_5720	Rv0860	<i>fadB</i>	Enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase	3-hydroxyacyl-CoA dehydrogenase activity; coenzyme binding; fatty acid metabolic process
A0R466	MSMEG_5721	Rv0859		Acetyl-CoA acetyltransferase	acetyl-CoA C-acyltransferase activity
A0R4B3	MSMEG_5773	Rv0824c	<i>des</i>	Acyl-acyl-carrier protein desaturase DesA1	acyl-[acyl-carrier-protein] desaturase activity; fatty acid metabolic process
A0R4G4	MSMEG_5824	Rv0803	<i>purL</i>	Phosphoribosylformylglycinamide synthase subunit II	'de novo' IMP biosynthetic process; ATP binding; cytoplasm; magnesium ion binding; phosphoribosylformylglycinamide synthase activity
A0R4H6	MSMEG_5837			Glutathione peroxidase	glutathione peroxidase activity; response to oxidative stress
A0R4J1	MSMEG_5852	Rv0772	<i>purD</i>	Phosphoribosylamine-glycine ligase	de novo' IMP biosynthetic process; ATP binding; magnesium ion binding; manganese ion binding; phosphoribosylamine-glycine ligase activity; purine nucleobase biosynthetic process
A0R4L1	MSMEG_5872	Rv0757	<i>phoP</i>	DNA-binding response regulator	DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R4S6	MSMEG_5937	Rv3534c	<i>bphI-2</i>	4-hydroxy-2-oxovalerate aldolase 2	4-hydroxy-2-oxovalerate aldolase activity; aromatic compound catabolic process; manganese ion binding
A0R4S7	MSMEG_5939	Rv3535c	<i>mhpF</i>	Acetaldehyde dehydrogenase 2	acetaldehyde dehydrogenase (acetylating) activity; aromatic compound catabolic process; NAD binding
A0R4T0	MSMEG_5943	Rv3538		Putative peroxisomal multifunctional enzyme type 2	3alpha,7alpha,12alpha-trihydroxy-5beta-cholest-24-enoyl-CoA hydratase activity; 3-hydroxyacyl-CoA dehydrogenase activity
A0R561	MSMEG_6077	Rv3583c	<i>carD</i>	RNA polymerase-binding transcription factor	
A0R566	MSMEG_6082	Rv3588c	<i>cynT</i>	Carbonic anhydrase	carbonate dehydratase activity; carbon utilization; zinc ion binding
A0R577	MSMEG_6094	Rv3598c	<i>lysS</i>	Lysine-tRNA ligase	ATP binding; cytoplasm; lysine-tRNA ligase activity; lysyl-tRNA aminoacylation; magnesium ion binding; nucleic acid binding
A0R5D9	MSMEG_6157	Rv3646c	<i>topA</i>	DNA topoisomerase I	DNA binding; DNA topoisomerase type I activity; DNA topological change; magnesium ion binding

A0R5M2	MSMEG_6241			ATPase associated with various cellular activities, AAA-5	ATPase activity; ATP binding
A0R5N7	MSMEG_6256	Rv3708c	<i>asd</i>	Aspartate-semialdehyde dehydrogenase	'de novo' L-methionine biosynthetic process; aspartate-semialdehyde dehydrogenase activity; cytoplasm; diaminopimelate biosynthetic process; isoleucine biosynthetic process; lysine biosynthetic process via diaminopimelate; N-acetyl-gamma-glutamyl-phosphate reductase activity; NAD binding; NADP binding; threonine biosynthetic process
A0R5N8	MSMEG_6257	Rv3709c	<i>ask</i>	Aspartokinase	amino acid binding; aspartate kinase activity; lysine biosynthetic process via diaminopimelate; threonine biosynthetic process
A0R616	MSMEG_6391	Rv3799c	<i>accD4</i>	Propionyl-CoA carboxylase beta chain	propionyl-CoA carboxylase activity
A0R617	MSMEG_6392	Rv3800c		Polyketide synthase	biosynthetic process; hydrolase activity, acting on ester bonds; phosphopantetheine binding; transferase activity
A0R628	MSMEG_6403	Rv3808c	<i>glfT2</i>	Galactofuranosyl transferase	capsule polysaccharide biosynthetic process; cell wall macromolecule biosynthetic process; cell wall organization; membrane; metal ion binding; transferase activity; transferase activity, transferring glycosyl groups
A0R635	MSMEG_6410	Rv3818		Putative Rieske 2Fe-2S iron-sulfur protein	2 iron, 2 sulfur cluster binding; metal ion binding; oxidoreductase activity
A0R656	MSMEG_6431	Rv3849	<i>espR</i>	Uncharacterized protein	sequence-specific DNA binding
A0R683	MSMEG_6458	Rv3858c	<i>glfD</i>	Glutamate synthase, NADH/NADPH, small subunit	flavin adenine dinucleotide binding; glutamate biosynthetic process; iron-sulfur cluster binding; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor
A0R684	MSMEG_6459	Rv3859c		Ferredoxin-dependent glutamate synthase 1	glutamate biosynthetic process; glutamate synthase (ferredoxin) activity
A0R696	MSMEG_6471			Glycine/D-amino acid oxidase	oxidoreductase activity
A0R6D6	MSMEG_6511			Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0R6D7	MSMEG_6512			Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0R6Q9	MSMEG_6638	Rv1133c	<i>metE</i>	Cobalamin-independent methionine synthase	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity; methionine biosynthetic process; zinc ion binding
A0R7F7	MSMEG_6895	Rv0055	<i>rpsR</i>	30S ribosomal protein S18	ribosome; rRNA binding; structural constituent of ribosome; translation
A0R7J0	MSMEG_6934	Rv3914	<i>trx</i>	Thioredoxin	cell redox homeostasis; glycerol ether metabolic process; protein disulfide oxidoreductase activity
A0R7J3	MSMEG_6938	Rv3917c	<i>parB</i>	Putative chromosome-partitioning protein parB	DNA binding
A0R7J4	MSMEG_6939	Rv3918c		Chromosome partitioning protein ParA	

A0R7J6	MSMEG_6941	Rv3920c		R3H domain-containing protein	nucleic acid binding
A0R7K1	MSMEG_6947	Rv0001	<i>dnaA</i>	Chromosomal replication initiator protein	ATP binding; cytoplasm; DNA replication initiation; DNA replication origin binding; regulation of DNA replication
<b>Low Confidence Proteins</b>					
A0QNF6	MSMEG_0024	Rv0009	<i>ppiA</i>	Peptidyl-prolyl cis-trans isomerase	peptidyl-prolyl cis-trans isomerase activity; protein folding
A0QQC8	MSMEG_0709	Rv0350	<i>dnaK</i>	Chaperone protein	ATP binding; protein folding
A0QQX6	MSMEG_0911	Rv0467	<i>aceA</i>	Isocitrate lyase	carboxylic acid metabolic process; isocitrate lyase activity
A0QQX8	MSMEG_0913	Rv0469	<i>umaA</i>	Methoxy mycolic acid synthase 1	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
A0QS44	MSMEG_1345	Rv0639	<i>nusG</i>	Transcription termination/antitermination protein	DNA-templated transcription, elongation; DNA-templated transcription, termination; regulation of DNA-templated transcription, elongation; transcription antitermination
A0QS46	MSMEG_1347	Rv0641	<i>rplA</i>	50S ribosomal protein L1	large ribosomal subunit; regulation of translation; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QS63	MSMEG_1365	Rv0652	<i>rplL</i>	50S ribosomal protein L7/L12	ribosome; structural constituent of ribosome; translation
P60281	MSMEG_1367	Rv0667	<i>rpoB</i>	DNA-directed RNA polymerase subunit beta	DNA binding; DNA-directed RNA polymerase activity; response to antibiotic; ribonucleoside binding; transcription, DNA-templated
A0QS66	MSMEG_1368	Rv0668	<i>rpoC</i>	DNA-directed RNA polymerase subunit beta'	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QS96	MSMEG_1398	Rv0682	<i>rpsL</i>	30S ribosomal protein S12	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation; tRNA binding
A0QS97	MSMEG_1399	Rv0683	<i>rpsG</i>	30S ribosomal protein S7	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation; tRNA binding
A0QS98	MSMEG_1401	Rv0685	<i>tuf</i>	Elongation factor	cytoplasm; GTPase activity; GTP binding; translation elongation factor activity
A0QSD0	MSMEG_1435	Rv0700	<i>rpsJ</i>	30S ribosomal protein S10	ribosome; structural constituent of ribosome; translation; tRNA binding
A0QSD1	MSMEG_1436	Rv0701	<i>rplC</i>	50S ribosomal protein L3	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSD4	MSMEG_1439	Rv0704	<i>rplB</i>	50S ribosomal protein L2	large ribosomal subunit; rRNA binding; structural constituent of ribosome; transferase activity; translation
A0QSD5	MSMEG_1440	Rv0705	<i>rpsS</i>	30S ribosomal protein S19	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSD6	MSMEG_1441	Rv0706	<i>rplV</i>	50S ribosomal protein L22	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSD7	MSMEG_1442	Rv0707	<i>rpsC</i>	30S ribosomal protein S3	mRNA binding; rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSG1	MSMEG_1467	Rv0716	<i>rplE</i>	50S ribosomal protein L5	ribosome; rRNA binding; structural constituent of ribosome; translation; tRNA binding

A0QSG6	MSMEG_1472	Rv0721	<i>rpsE</i>	30S ribosomal protein S5	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSG8	MSMEG_1474	Rv0723	<i>rplO</i>	50S ribosomal protein L15	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSL3	MSMEG_1519	Rv3462c	<i>infA</i>	Translation initiation factor	cytoplasm; translation initiation factor activity
A0QSL6	MSMEG_1522	Rv3459c	<i>rpsK</i>	30S ribosomal protein S11	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSL7	MSMEG_1523	Rv3458c	<i>rpsD</i>	30S ribosomal protein S4	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSL8	MSMEG_1524	Rv3457c	<i>rpoA</i>	DNA-directed RNA polymerase subunit alpha	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QSL9	MSMEG_1525	Rv3456c	<i>rplQ</i>	50S ribosomal protein L17	ribosome; structural constituent of ribosome; translation
A0QSZ1	MSMEG_1652		<i>metC</i>	O-acetylhomoserine sulfhydrylase	O-acetylhomoserine aminocarboxypropyltransferase activity; pyridoxal phosphate binding
A0QSZ3	MSMEG_1654	Rv0066c	<i>icd2</i>	Isocitrate dehydrogenase, NADP-dependent	isocitrate dehydrogenase (NADP+) activity; metal ion binding; tricarboxylic acid cycle
A0QTE1	MSMEG_1807	Rv3285	<i>accA3</i>	Acetyl-/propionyl-coenzyme A carboxylase alpha chain	ATP binding; biotin carboxylase activity; metal ion binding
A0QU54	MSMEG_2081	Rv3139		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QUV6	MSMEG_2351	Rv3029c	<i>etfB</i>	Electron transfer flavoprotein beta subunit	electron carrier activity
A0QUX8	MSMEG_2374	Rv3001c	<i>ilvC</i>	Ketol-acid reductoisomerase	coenzyme binding; isoleucine biosynthetic process; ketol-acid reductoisomerase activity; valine biosynthetic process
A0QUY2	MSMEG_2378		<i>serA</i>	D-3-phosphoglycerate dehydrogenase	amino acid binding; L-serine biosynthetic process; NAD binding; phosphoglycerate dehydrogenase activity
Q9ZHC5	MSMEG_2389	Rv2986c	<i>hup</i>	DNA-binding protein HU homolog (Hlp)	cell wall; chromosome condensation; DNA binding; extracellular region
A0QVB8	MSMEG_2519	Rv2890c	<i>rpsB</i>	30S ribosomal protein S2	small ribosomal subunit; structural constituent of ribosome; translation
A0QVQ5	MSMEG_2656	Rv2783c	<i>pnp</i>	Polyribonucleotide nucleotidyltransferase	3'-5'-exoribonuclease activity; cytoplasm; magnesium ion binding; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity; RNA binding; RNA processing
A0QWS8	MSMEG_3050	Rv1388	<i>mihF</i>	Integration host factor	nucleic acid binding
A0QWY0	MSMEG_3103	Rv1449c	<i>tkt</i>	Transketolase	metal ion binding; transketolase activity
A0QX20	MSMEG_3143	Rv1475c	<i>acnA</i>	Aconitate hydratase A	aconitate hydratase activity; iron-sulfur cluster binding; metal ion binding; tricarboxylic acid cycle
A0QXA3	MSMEG_3227	Rv1617	<i>pyk</i>	Pyruvate kinase	glycolytic process; magnesium ion binding; potassium ion binding; pyruvate kinase activity
A0QYW6	MSMEG_3811	Rv1636		Universal stress protein family protein, putative	cytoplasm; response to stress

A0QYY6	MSMEG_3833	Rv1630	<i>rpsA</i>	30S ribosomal protein S1	nucleotidyltransferase activity; ribosome; RNA binding; structural constituent of ribosome; translation
A0R0B4	MSMEG_4327	Rv2245	<i>kasA</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 1	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R0B5	MSMEG_4328	Rv2246	<i>kasB</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 2	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R1H7	MSMEG_4757	Rv2524c		Fatty acid synthase	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process; fatty acid synthase complex; nucleotide binding
A0R1V9	MSMEG_4891	Rv2428	<i>aphC</i>	Alkyl hydroperoxide reductase subunit C	peroxidase activity; peroxiredoxin activity
A0R2B1	MSMEG_5049	Rv1248c	<i>kgd</i>	Multifunctional 2-oxoglutarate metabolism enzyme	2-hydroxy-3-oxoadipate synthase activity; 2-oxoglutarate decarboxylase activity; dihydrolipoyllysine-residue succinyltransferase activity; metal ion binding; oxoglutarate dehydrogenase (succinyl-transferring) activity; thiamine pyrophosphate binding; tricarboxylic acid cycle
A0R2X1	MSMEG_5263	Rv1080c	<i>greA</i>	Transcription elongation factor	DNA binding; regulation of DNA-templated transcription, elongation; transcription, DNA-templated
A0R2Y1	MSMEG_5273	Rv1074c	<i>fadA3</i>	Acetyl-CoA acetyltransferase	transferase activity, transferring acyl groups other than amino-acyl groups
A0R3B8	MSMEG_5415	Rv1023	<i>eno</i>	Enolase	cell surface; extracellular region; glycolytic process; magnesium ion binding; phosphopyruvate hydratase activity; phosphopyruvate hydratase complex
A0R3M3	MSMEG_5524	Rv0952	<i>sucD</i>	Succinyl-CoA ligase	ATP binding; ATP citrate synthase activity; cofactor binding; succinate-CoA ligase (ADP-forming) activity
A0R417	MSMEG_5672	Rv0896	<i>gltA</i>	Citrate synthase	cellular carbohydrate metabolic process; citrate (Si)-synthase activity; cytoplasm; tricarboxylic acid cycle
A0R574	MSMEG_6091		<i>clpC1</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein metabolic process
A0R5E1	MSMEG_6159		<i>cspA</i>	Probable cold shock protein A	cytoplasm; DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R5H1	MSMEG_6189	Rv3676	<i>crp</i>	Crp/Fnr family transcriptional regulator	DNA binding; intracellular; sequence-specific DNA binding transcription factor activity; transcription, DNA-templated
A0R5M3	MSMEG_6242			Iron-containing alcohol dehydrogenase	1,3-propanediol dehydrogenase activity; alcohol dehydrogenase (NAD) activity; metal ion binding
A0R618	MSMEG_6393		<i>fadD32</i>	Acyl-CoA synthase	ligase activity
A0R729	MSMEG_6759		<i>glpK</i>	Glycerol kinase	ATP binding; glycerol-3-phosphate metabolic process; glycerol catabolic process; glycerol kinase activity
A0R730	MSMEG_6761			Glycerol-3-phosphate dehydrogenase	glycerol-3-phosphate dehydrogenase complex; glycerol-3-phosphate metabolic process; sn-glycerol-3-phosphate:ubiquinone-8 oxidoreductase activity
Q9AFI5	MSMEG_6896	Rv0054	<i>ssb</i>	Single-stranded DNA-binding protein (SSB)	DNA replication; single-stranded DNA binding

---

A0R7G6	MSMEG_6904	Rv0046c	<i>ino1</i>	Inositol-3-phosphate synthase	inositol-3-phosphate synthase activity; inositol biosynthetic process; phospholipid biosynthetic process
--------	------------	---------	-------------	-------------------------------	---

---

**Table B. 3 Lower stringency protein identification gene ontology enrichment data**

<b>GOID</b>	<b>Ontology</b>	<b>Term</b>	<b>Number of identifications</b>	<b>p-value</b>
GO:0000166	molecular function	nucleotide binding	76	7.34E-14
GO:0001882	molecular function	nucleoside binding	55	1.82E-12
GO:0001883	molecular function	purine nucleoside binding	54	5.14E-12
GO:0003674	molecular function	molecular function	245	5.70E-21
GO:0005488	molecular function	binding	165	3.31E-21
GO:0005524	molecular function	ATP binding	48	7.12E-10
GO:0017076	molecular function	purine nucleotide binding	54	5.54E-12
GO:0030554	molecular function	adenyl nucleotide binding	48	7.69E-10
GO:0032549	molecular function	ribonucleoside binding	55	1.82E-12
GO:0032550	molecular function	purine ribonucleoside binding	54	5.14E-12
GO:0032553	molecular function	ribonucleotide binding	55	1.34E-10
GO:0032555	molecular function	purine ribonucleotide binding	54	5.14E-12
GO:0032559	molecular function	adenyl ribonucleotide binding	48	7.12E-10
GO:0035639	molecular function	purine ribonucleoside triphosphate binding	54	5.14E-12
GO:0036094	molecular function	small molecule binding	81	5.34E-15
GO:0097159	molecular function	organic cyclic compound binding	133	3.53E-18
GO:0097367	molecular function	carbohydrate derivative binding	57	1.56E-11
GO:1901265	molecular function	nucleoside phosphate binding	76	7.34E-14
GO:1901363	molecular function	heterocyclic compound binding	133	3.53E-18
GO:0003676	molecular function	nucleic acid binding	74	2.07E-09
GO:0003824	molecular function	catalytic activity	187	1.62E-09
GO:0003916	molecular function	DNA topoisomerase activity	5	2.04E-04
GO:0003918	molecular function	DNA topoisomerase type II (ATP-hydrolyzing) activity	4	2.73E-03
GO:0008094	molecular function	DNA-dependent ATPase activity	8	6.07E-05
GO:0016462	molecular function	pyrophosphatase activity	24	5.85E-03
GO:0016817	molecular function	hydrolase activity, acting on acid anhydrides	24	6.89E-03
GO:0016818	molecular function	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	24	6.13E-03
GO:0017111	molecular function	nucleoside-triphosphatase activity	23	6.97E-03
GO:0061505	molecular function	DNA topoisomerase II activity	4	2.73E-03
GO:0006139	biological process	nucleobase-containing compound metabolic process	56	1.76E-03
GO:0006259	biological process	DNA metabolic process	17	1.05E-03
GO:0006265	biological process	DNA topological change	5	2.04E-04
GO:0006725	biological process	cellular aromatic compound metabolic process	65	1.81E-04
GO:0006807	biological process	nitrogen compound metabolic process	102	8.52E-13
GO:0006996	biological process	organelle organization	8	1.39E-05
GO:0008150	biological process	biological process	158	9.66E-18
GO:0008152	biological process	metabolic process	136	2.88E-17

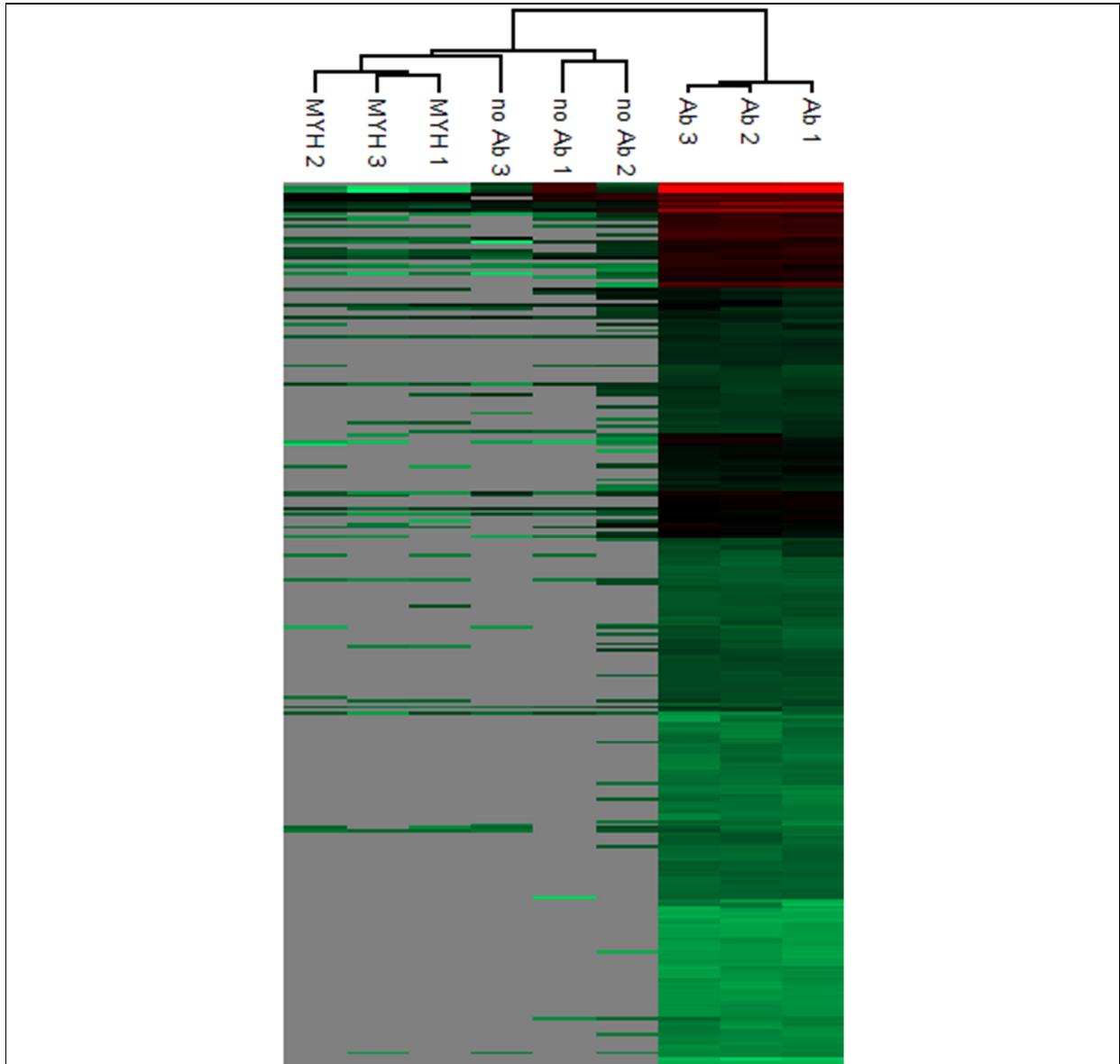
GO:0009987	biological process	cellular process	146	1.16E-21
GO:0016043	biological process	cellular component organization	15	3.50E-06
GO:0034641	biological process	cellular nitrogen compound metabolic process	84	2.93E-10
GO:0043170	biological process	macromolecule metabolic process	69	5.32E-07
GO:0044237	biological process	cellular metabolic process	129	6.41E-19
GO:0044238	biological process	primary metabolic process	111	6.33E-14
GO:0044260	biological process	cellular macromolecule metabolic process	68	4.55E-07
GO:0046483	biological process	heterocycle metabolic process	64	2.53E-04
GO:0051276	biological process	chromosome organization	7	1.31E-05
GO:0071103	biological process	DNA conformation change	6	1.70E-04
GO:0071704	biological process	organic substance metabolic process	129	3.73E-17
GO:0071840	biological process	cellular component organization or biogenesis	17	5.23E-06
GO:0090304	biological process	nucleic acid metabolic process	44	4.31E-02
GO:1901360	biological process	organic cyclic compound metabolic process	65	2.68E-04
GO:0006260	biological process	DNA replication	8	3.00E-03
GO:0006261	biological process	DNA-dependent DNA replication	4	2.67E-02
GO:0009058	biological process	biosynthetic process	98	1.82E-13
GO:0009059	biological process	macromolecule biosynthetic process	54	3.25E-05
GO:0034645	biological process	cellular macromolecule biosynthetic process	54	2.13E-05
GO:0044249	biological process	cellular biosynthetic process	93	6.47E-13
GO:1901576	biological process	organic substance biosynthetic process	93	2.31E-12
GO:0005575	cellular component	cellular component	80	4.07E-04
GO:0005622	cellular component	intracellular	70	5.93E-19
GO:0005623	cellular component	cell	78	4.21E-12
GO:0005694	cellular component	chromosome	3	8.63E-02
GO:0043226	cellular component	organelle	23	1.68E-12
GO:0043228	cellular component	non-membrane-bounded organelle	22	5.14E-12
GO:0043229	cellular component	intracellular organelle	23	1.29E-12
GO:0043232	cellular component	intracellular non-membrane-bounded organelle	22	5.14E-12
GO:0044424	cellular component	intracellular part	66	1.16E-21
GO:0044464	cellular component	cell part	74	5.80E-11
GO:0005737	cellular component	cytoplasm	64	1.16E-21
GO:0043167	molecular function	ion binding	52	2.54E-04
GO:0043169	molecular function	cation binding	45	1.77E-03
GO:0046872	molecular function	metal ion binding	45	1.40E-03
GO:0050896	biological process	response to stimulus	17	5.16E-02
GO:0044699	biological process	single-organism process	78	6.78E-08
GO:0044763	biological process	single-organism cellular process	74	5.42E-09
GO:0044085	biological process	cellular component biogenesis	8	8.86E-02
GO:0010467	biological process	gene expression	47	7.98E-04
GO:0044271	biological process	cellular nitrogen compound biosynthetic process	59	2.56E-05

GO:0003995	molecular function	acyl-CoA dehydrogenase activity	10	5.94E-02
GO:0016491	molecular function	oxidoreductase activity	65	9.67E-02
GO:0048037	molecular function	cofactor binding	28	7.65E-02
GO:0050660	molecular function	flavin adenine dinucleotide binding	14	9.21E-02
GO:0050662	molecular function	coenzyme binding	21	5.72E-02
GO:0016836	molecular function	hydro-lyase activity	12	5.73E-02
GO:0006082	biological process	organic acid metabolic process	41	4.09E-09
GO:0006549	biological process	isoleucine metabolic process	5	9.82E-04
GO:0009081	biological process	branched-chain amino acid metabolic process	5	9.82E-04
GO:0009082	biological process	branched-chain amino acid biosynthetic process	5	9.82E-04
GO:0009097	biological process	isoleucine biosynthetic process	5	9.82E-04
GO:0016053	biological process	organic acid biosynthetic process	20	2.09E-04
GO:0044281	biological process	small molecule metabolic process	55	2.34E-09
GO:0044283	biological process	small molecule biosynthetic process	27	8.70E-06
GO:0044710	biological process	single-organism metabolic process	62	6.64E-08
GO:0044711	biological process	single-organism biosynthetic process	38	3.03E-05
GO:1901564	biological process	organonitrogen compound metabolic process	62	2.93E-14
GO:1901566	biological process	organonitrogen compound biosynthetic process	55	2.83E-15
GO:1901605	biological process	alpha-amino acid metabolic process	18	2.32E-04
GO:1901607	biological process	alpha-amino acid biosynthetic process	15	8.29E-05
GO:0006520	biological process	cellular amino acid metabolic process	20	3.71E-04
GO:0006573	biological process	valine metabolic process	3	8.63E-02
GO:0008652	biological process	cellular amino acid biosynthetic process	13	2.56E-03
GO:0009099	biological process	valine biosynthetic process	3	5.72E-02
GO:0019752	biological process	carboxylic acid metabolic process	34	3.74E-07
GO:0043436	biological process	oxoacid metabolic process	34	9.36E-07
GO:0046394	biological process	carboxylic acid biosynthetic process	18	4.35E-04
GO:0005525	molecular function	GTP binding	7	1.58E-02
GO:0019001	molecular function	guanyl nucleotide binding	7	1.58E-02
GO:0032561	molecular function	guanyl ribonucleotide binding	7	1.58E-02
GO:0030145	molecular function	manganese ion binding	6	8.62E-04
GO:0046914	molecular function	transition metal ion binding	23	5.16E-02
GO:0008270	molecular function	zinc ion binding	17	1.40E-03
GO:0006790	biological process	sulfur compound metabolic process	8	1.59E-02
GO:0006793	biological process	phosphorus metabolic process	18	1.58E-02
GO:0006796	biological process	phosphate-containing compound metabolic process	17	3.04E-02
GO:0019637	biological process	organophosphate metabolic process	17	1.97E-02
GO:0051186	biological process	cofactor metabolic process	13	5.72E-02
GO:0004386	molecular function	helicase activity	7	1.58E-02
GO:0006163	biological process	purine nucleotide metabolic process	9	4.31E-02

GO:0006753	biological process	nucleoside phosphate metabolic process	12	6.97E-02
GO:0009117	biological process	nucleotide metabolic process	12	6.69E-02
GO:0009119	biological process	ribonucleoside metabolic process	8	8.86E-02
GO:0009123	biological process	nucleoside monophosphate metabolic process	10	1.26E-02
GO:0009126	biological process	purine nucleoside monophosphate metabolic process	9	6.91E-03
GO:0009150	biological process	purine ribonucleotide metabolic process	9	2.43E-02
GO:0009161	biological process	ribonucleoside monophosphate metabolic process	10	7.13E-03
GO:0009167	biological process	purine ribonucleoside monophosphate metabolic process	9	6.91E-03
GO:0009259	biological process	ribonucleotide metabolic process	11	6.95E-03
GO:0009260	biological process	ribonucleotide biosynthetic process	7	9.03E-02
GO:0019693	biological process	ribose phosphate metabolic process	12	5.85E-03
GO:0072521	biological process	purine-containing compound metabolic process	9	6.97E-02
GO:1901135	biological process	carbohydrate derivative metabolic process	16	4.83E-02
GO:1901657	biological process	glycosyl compound metabolic process	10	2.69E-02
GO:0016874	molecular function	ligase activity	20	1.74E-03
GO:0006412	biological process	translation	25	1.68E-12
GO:0006518	biological process	peptide metabolic process	26	1.92E-12
GO:0019538	biological process	protein metabolic process	28	2.93E-10
GO:0043043	biological process	peptide biosynthetic process	26	6.47E-13
GO:0043603	biological process	cellular amide metabolic process	26	2.78E-11
GO:0043604	biological process	amide biosynthetic process	26	1.04E-12
GO:0044267	biological process	cellular protein metabolic process	27	4.93E-10
GO:0006091	biological process	generation of precursor metabolites and energy	9	2.14E-02
GO:0006099	biological process	tricarboxylic acid cycle	5	2.66E-02
GO:0009060	biological process	aerobic respiration	5	5.16E-02
GO:0005829	cellular component	cytosol	3	8.63E-02
GO:0032991	cellular component	macromolecular complex	27	2.08E-07
GO:0044444	cellular component	cytoplasmic part	25	1.31E-13
GO:0044445	cellular component	cytosolic part	3	5.72E-02
GO:0003723	molecular function	RNA binding	32	4.06E-17
GO:0030529	cellular component	intracellular ribonucleoprotein complex	20	2.78E-11
GO:1990904	cellular component	ribonucleoprotein complex	20	2.78E-11
GO:0006353	biological process	DNA-templated transcription, termination	4	2.73E-03
GO:0072524	biological process	pyridine-containing compound metabolic process	6	7.20E-02
GO:0000049	molecular function	tRNA binding	6	1.29E-03
GO:0008135	molecular function	translation factor activity, RNA binding	5	1.53E-02
GO:0003899	molecular function	DNA-directed RNA polymerase activity	4	1.75E-02
GO:0016779	molecular function	nucleotidyltransferase activity	9	2.43E-02
GO:0034062	molecular function	RNA polymerase activity	4	1.75E-02

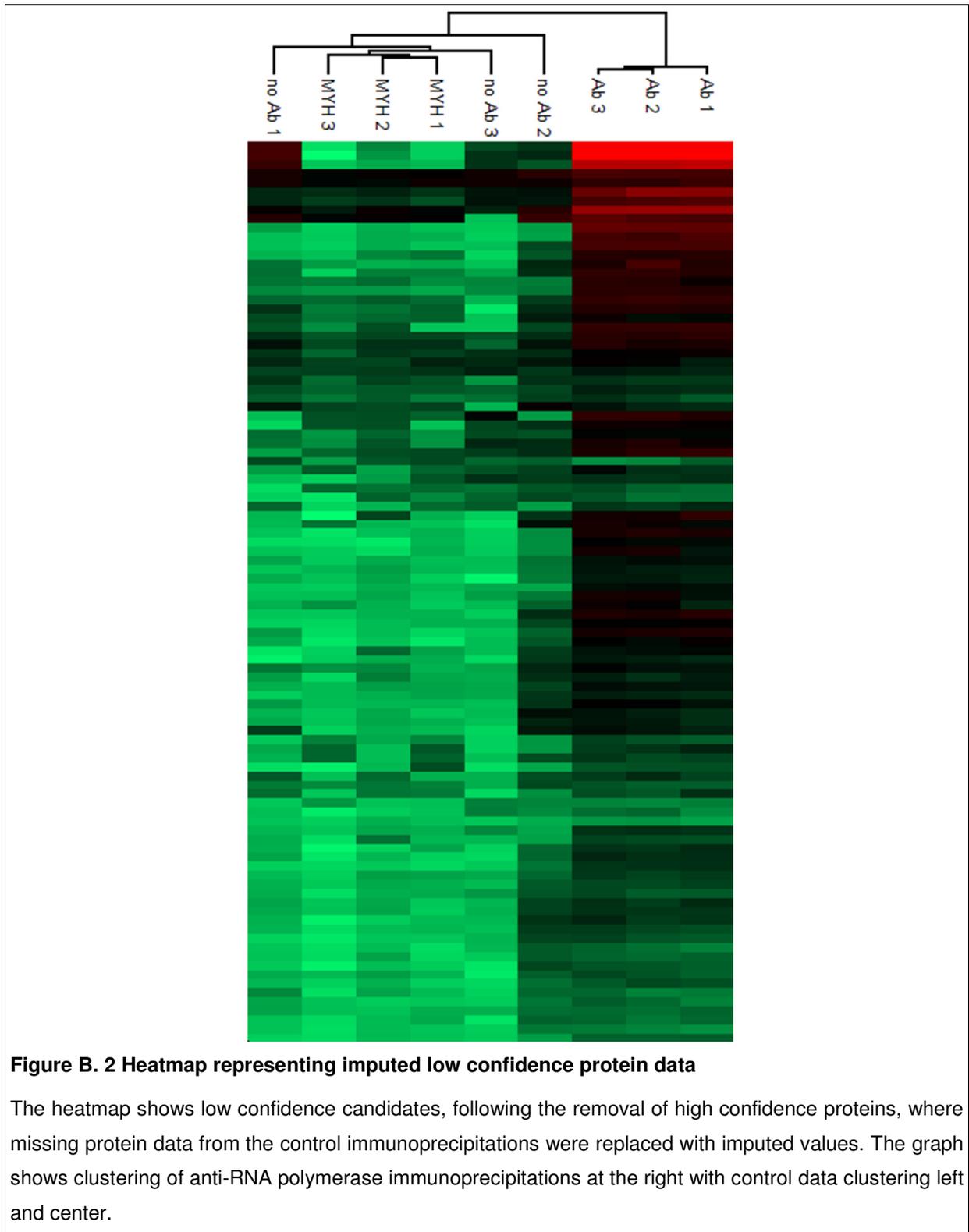
GO:0006096	biological process	glycolytic process	4	5.16E-02
GO:0006165	biological process	nucleoside diphosphate phosphorylation	4	5.16E-02
GO:0006757	biological process	ATP generation from ADP	4	5.16E-02
GO:0009135	biological process	purine nucleoside diphosphate metabolic process	4	5.16E-02
GO:0009179	biological process	purine ribonucleoside diphosphate metabolic process	4	5.16E-02
GO:0009185	biological process	ribonucleoside diphosphate metabolic process	4	5.16E-02
GO:0032787	biological process	monocarboxylic acid metabolic process	10	9.21E-02
GO:0046031	biological process	ADP metabolic process	4	5.16E-02
GO:0046939	biological process	nucleotide phosphorylation	4	5.16E-02
GO:0016226	biological process	iron-sulfur cluster assembly	3	5.72E-02
GO:0022607	biological process	cellular component assembly	5	5.16E-02
GO:0031163	biological process	metallo-sulfur cluster assembly	3	5.72E-02
GO:0044265	biological process	cellular macromolecule catabolic process	4	8.41E-02
GO:0005198	molecular function	structural molecule activity	20	2.08E-11
GO:0006457	biological process	protein folding	6	1.29E-03
GO:0003755	molecular function	peptidyl-prolyl cis-trans isomerase activity	3	5.72E-02
GO:0016859	molecular function	cis-trans isomerase activity	3	5.72E-02
GO:0003985	molecular function	acetyl-CoA C-acetyltransferase activity	4	2.67E-02
GO:0003988	molecular function	acetyl-CoA C-acyltransferase activity	4	5.16E-02
GO:0016453	molecular function	C-acetyltransferase activity	4	2.67E-02
GO:0006566	biological process	threonine metabolic process	3	3.43E-02
GO:0009088	biological process	threonine biosynthetic process	3	3.43E-02
GO:0019843	molecular function	rRNA binding	14	3.44E-08
GO:0005840	cellular component	ribosome	19	2.10E-10
GO:0003735	molecular function	structural constituent of ribosome	19	1.10E-10
GO:0015934	cellular component	large ribosomal subunit	4	5.96E-03
GO:0044391	cellular component	ribosomal subunit	11	2.85E-11
GO:0044422	cellular component	organelle part	11	1.67E-09
GO:0044446	cellular component	intracellular organelle part	11	7.12E-10
GO:0015935	cellular component	small ribosomal subunit	7	2.36E-08

### B.3. High confidence protein identification



**Figure B. 1 Heatmaps representing high stringency NP-MS data**

The heatmaps shows the clustering of the biological replicate experiments. The red colouring is indicative of a higher Lfq intensity whereas green represents a lower protein abundance whilst grey indicates the absence of the protein from that immunoprecipitated sample. The heatmap shows the clustering of the anti-RNA polymerase immunoprecipitation on the right and control immunoprecipitations (clustering left and center) shows very little non-specific binding, as is indicated by the grey colouring.



**Table B. 4 High stringency identified proteins in *Mycobacterium smegmatis***

UniProt Annotation	MSMEG Annotation	Rv Homologue	Gene Name	Protein Name	Gene Ontology
<b>High Confidence</b>					
A0QNE0	MSMEG_0005	Rv0005	<i>gyrB</i>	DNA	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; magnesium ion binding
P48354	MSMEG_0006	Rv0006	<i>gyrA</i>	DNA gyrase subunit A	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; response to antibiotic
A0QNF5	MSMEG_0023	Rv0008c	<i>cwsA</i>	Cell wall synthesis protein CwsA	cell cycle; cell division; cell wall biogenesis; integral component of plasma membrane; protein localization; regulation of cell shape
A0QNI9	MSMEG_0059	Rv3868	<i>eccA1</i>	ATPase, AAA family protein	ATP binding
A0QNJ7	MSMEG_0067			Uncharacterized protein	
A0QNN0	MSMEG_0102			Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QNZ3	MSMEG_0216			3-hydroxyacyl-CoA dehydrogenase	oxidoreductase activity
A0QP06	MSMEG_0229	Rv0189c	<i>ilvD</i>	Dihydroxy-acid dehydratase	4 iron, 4 sulfur cluster binding; dihydroxy-acid dehydratase activity; isoleucine biosynthetic process; metal ion binding; valine biosynthetic process
A0QP11	MSMEG_0234	Rv0198c		Metallopeptidase	metalloendopeptidase activity
A0QP15	MSMEG_0238			CoA binding domain protein	cofactor binding
A0QP20	MSMEG_0243			Uncharacterized protein	
A0QP32	MSMEG_0255	Rv0211	<i>pckG</i>	Phosphoenolpyruvate carboxykinase	cytoplasm; gluconeogenesis; GTP binding; manganese ion binding; phosphoenolpyruvate carboxykinase (GTP) activity
A0QPE7	MSMEG_0372	Rv0242c		3-oxoacyl-acyl-carrier protein reductase	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity
A0QPG3	MSMEG_0388		<i>tylF</i>	Macrocin-O-methyltransferase	macrocin O-methyltransferase activity
A0QPH5	MSMEG_0400			Peptide synthetase	catalytic activity
Q3L891	MSMEG_0402			Linear gramicidin synthetase subunit D	isomerase activity; oxidoreductase activity
Q3L887	MSMEG_0406	Rv0244c	<i>fadE5</i>	Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; butyryl-CoA dehydrogenase activity
Q3L885	MSMEG_0408			Type I modular polyketide synthase	erythronolide synthase activity; oxidoreductase activity; phosphopantetheine binding; zinc ion binding

A0QPJ0	MSMEG_0415	Rv0245		Flavin reductase-like, FMN-binding protein	FMN binding; riboflavin reductase (NADPH) activity
A0QPN1	MSMEG_0456			DNA topoisomerase	ATP binding; chromosome; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change
A0QPV4	MSMEG_0531			putative Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QQF9	MSMEG_0741			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0QQH2	MSMEG_0754	Rv0360c		Uncharacterized protein	
A0QQP1	MSMEG_0825	Rv0422c	<i>thiD</i>	Phosphomethylpyrimidine kinase	ATP binding; phosphomethylpyrimidine kinase activity; thiamine biosynthetic process
A0QQP2	MSMEG_0826	Rv0423c	<i>thiC</i>	Thiamine biosynthesis protein ThiC	4 iron, 4 sulfur cluster binding; lyase activity; thiamine biosynthetic process; thiamine diphosphate biosynthetic process; zinc ion binding
A0QQS8	MSMEG_0863	Rv0439c		Short chain dehydrogenase	oxidoreductase activity
A0QQW8	MSMEG_0903	Rv0462	<i>lpdA</i>	Dihydrolipoyl dehydrogenase	cell redox homeostasis; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
Q3I5Q7	MSMEG_0919	Rv0475		Uncharacterized protein	cell adhesion; cell surface; heparin binding; pathogenesis
A0QR13	MSMEG_0948			Uncharacterized protein	
A0QR19	MSMEG_0954	Rv0511	<i>hemD</i>	Uroporphyrinogen-III synthase	methyltransferase activity; tetrapyrrole biosynthetic process; uroporphyrinogen-III synthase activity
P0CH00	MSMEG_2299	Rv3051c	<i>nrpE2</i>	Ribonucleoside-diphosphate reductase subunit alpha 2	ATP binding; DNA replication; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
P0CH37	MSMEG_2317	Rv3045	<i>adhC2</i>	NADP-dependent alcohol dehydrogenase C 2	alcohol dehydrogenase (NADP+) activity; zinc ion binding
A0QRA8	MSMEG_2329	Rv3038c		Methyltransferase type 11	methyltransferase activity
A0QRB8	MSMEG_1060			Uncharacterized protein	
A0QRM0	MSMEG_1165	Rv0566c		Uncharacterized protein	
A0QRV8	MSMEG_1255			DNA helicase	ATP binding; ATP-dependent DNA helicase activity; DNA binding
A0QS36	MSMEG_1334	Rv0634c		Metallo-beta-lactamase family protein	hydroxyacylglutathione hydrolase activity
A0QS42	MSMEG_1342	Rv0637		Uncharacterized protein	
A0QS49	MSMEG_1350		<i>cmaA1</i>	Cyclopropane-fatty-acyl-phospholipid synthase 1	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
A0QSB1	MSMEG_1416	Rv0688		FAD-dependent pyridine nucleotide-disulfide oxidoreductase	cell redox homeostasis; ferredoxin-NAD+ reductase activity; flavin adenine dinucleotide binding
A0QSL0	MSMEG_1516			Thioredoxin reductase	oxidoreductase activity; phosphorelay signal transduction system
A0QSN7	MSMEG_1543			Eptc-inducible aldehyde dehydrogenase	aldehyde dehydrogenase (NAD) activity

A0QSU3	MSMEG_1602	Rv3411c	<i>guaB</i>	Inosine-5'-monophosphate dehydrogenase	adenyl nucleotide binding; GMP biosynthetic process; IMP dehydrogenase activity; metal ion binding
A0QSV0	MSMEG_1610	Rv3396c	<i>guaA</i>	GMP synthetase	ATP binding; glutamine metabolic process; GMP biosynthetic process; GMP synthase (glutamine-hydrolyzing) activity; pyrophosphatase activity
A0QSX4	MSMEG_1635	Rv3368c		Nitroreductase family protein	oxidoreductase activity
A0QT08	MSMEG_1670	Rv3318	<i>sdhA</i>	Succinate dehydrogenase flavoprotein subunit	electron transport chain; flavin adenine dinucleotide binding; succinate dehydrogenase activity; tricarboxylic acid cycle
A0QT17	MSMEG_1679		<i>amiB1</i>	Amidohydrolase AmiB1	hydrolase activity
A0QT18	MSMEG_1680			Uncharacterized protein	
A0QTE7	MSMEG_1813	Rv3280	<i>accD5</i>	Propionyl-CoA carboxylase beta chain 5 AccD5	propionyl-CoA carboxylase activity
A0QTF4	MSMEG_1821	Rv3274c	<i>fadE25</i>	Putative acyl-CoA dehydrogenase fadE25	acyl-CoA dehydrogenase activity; butyryl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QTK2	MSMEG_1874	Rv3246c	<i>mtrA</i>	DNA-binding response regulator MtrA	cytoplasm; DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
P71533	MSMEG_1881	Rv3240c	<i>secA1</i>	Protein translocase subunit SecA 1	ATP binding; cytoplasm; intracellular protein transmembrane transport; plasma membrane; protein import; protein targeting
A0QTP2	MSMEG_1914	Rv3223c	<i>sigH</i>	ECF RNA polymerase sigma factor SigH	DNA binding; DNA-templated transcription, initiation; sequence-specific DNA binding transcription factor activity; sigma factor activity
A0QTQ8	MSMEG_1930	Rv3211		DEAD/DEAH box helicase	ATP binding; ATP-dependent helicase activity; nucleic acid binding
A0QTR5	MSMEG_1937	Rv3206c	<i>MoeB</i>	Molybdopterin biosynthesis protein MoeB	catalytic activity
A0QU53	MSMEG_2080	Rv3140	<i>fadE23</i>	Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity
A0QU61	MSMEG_2089	Rv3102c	<i>ftsE</i>	Cell division ATP-binding protein FtsE	ATPase activity; ATP binding; cell division
A0QUE0	MSMEG_2174			DNA helicase	ATP binding; ATP-dependent DNA helicase activity; DNA binding
A0QR89	MSMEG_2308 MSMEG_1028			Geranylgeranyl reductase	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor
A0QUX6	MSMEG_2372	Rv3003c	<i>ilvB</i>	Acetolactate synthase	acetolactate synthase activity; flavin adenine dinucleotide binding; isoleucine biosynthetic process; magnesium ion binding; thiamine pyrophosphate binding; valine biosynthetic process
A0QUY6	MSMEG_2382	Rv2993c		5-carboxymethyl-2-hydroxyimuconate delta-isomerase	isomerase activity
A0QUY9	MSMEG_2387	Rv2988c	<i>leuC</i>	3-isopropylmalate dehydratase large subunit	3-isopropylmalate dehydratase activity; 4 iron, 4 sulfur cluster binding; leucine biosynthetic process; metal ion binding
A0QUZ0	MSMEG_2388	Rv2987c	<i>leuD</i>	3-isopropylmalate dehydratase small subunit	3-isopropylmalate dehydratase activity; 3-isopropylmalate dehydratase complex; leucine biosynthetic process

A0QV32	MSMEG_2430	Rv2916c	<i>ffh</i>	Signal recognition particle protein	7S RNA binding; GTPase activity; GTP binding; signal recognition particle; SRP-dependent cotranslational protein targeting to membrane
A0QV38	MSMEG_2436	Rv2908c		Uncharacterized protein	RNA binding
A0QVM4	MSMEG_2625	Rv2841c	<i>nusA</i>	Transcription termination/antitermination protein NusA	regulation of DNA-templated transcription, termination; RNA binding; sequence-specific DNA binding transcription factor activity
A0QVU2	MSMEG_2695	Rv2744c		35 kDa protein	
Q59560	MSMEG_2723	Rv2737c	<i>recA</i>	Protein RecA	ATP binding; cytoplasm; damaged DNA binding; DNA-dependent ATPase activity; DNA recombination; DNA repair; single-stranded DNA binding; SOS response
A0QVZ5	MSMEG_2752	Rv2710	<i>mysB</i>	RNA polymerase sigma factor	DNA binding; DNA-templated transcription, initiation; sequence-specific DNA binding transcription factor activity; sigma factor activity
A0QW02	MSMEG_2758	Rv2703	<i>rpoD</i>	RNA polymerase sigma factor SigA	cytoplasm; DNA binding; sequence-specific DNA binding transcription factor activity; sigma factor activity; transcription initiation from bacterial-type RNA polymerase promoter
A0QW25	MSMEG_2782	Rv2676c		Uncharacterized protein	
A0QW71	MSMEG_2839			Transcriptional accessory protein	DNA binding; DNA repair; hydrolase activity, acting on ester bonds; RNA binding
A0QWG8	MSMEG_2937	Rv2606c	<i>pdxS</i>	Pyridoxal 5'-phosphate synthase subunit	lyase activity; pyridoxal phosphate biosynthetic process
A0QWQ4	MSMEG_3025	Rv2555c	<i>alaS</i>	Alanine-tRNA ligase	alanine-tRNA ligase activity; alanyl-tRNA aminoacylation; ATP binding; cytoplasm; tRNA binding; zinc ion binding
A0QWR4	MSMEG_3035	Rv2534c	<i>efp</i>	Elongation factor P	cytoplasm; peptide biosynthetic process; translation elongation factor activity
A0QWR5	MSMEG_3036	Rv2533c	<i>nusB</i>	N utilization substance protein B homolog	DNA-templated transcription, termination; regulation of transcription, DNA-templated; RNA binding
A0QWS4	MSMEG_3046	Rv1383	<i>carA</i>	Carbamoyl-phosphate synthase small chain	'de novo' UMP biosynthetic process; arginine biosynthetic process; ATP binding; carbamoyl phosphate biosynthetic process; carbamoyl-phosphate synthase (glutamine-hydrolyzing) activity; glutamine catabolic process
A0QWT1	MSMEG_3053	Rv1390	<i>rpoZ</i>	DNA-directed RNA polymerase subunit omega	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QWT7	MSMEG_3059	Rv1400c		Esterase	hydrolase activity
A0QWV0	MSMEG_3072	Rv1415	<i>ribAB</i>	Riboflavin biosynthesis protein RibBA	3,4-dihydroxy-2-butanone-4-phosphate synthase activity; GTP binding; GTP cyclohydrolase II activity; magnesium ion binding; manganese ion binding; riboflavin biosynthetic process; zinc ion binding
A0QWV9	MSMEG_3081	Rv1423	<i>whiA</i>	Putative sporulation transcription regulator WhiA	DNA binding; regulation of sporulation; regulation of transcription, DNA-templated; transcription, DNA-templated

A0QWW4	MSMEG_3086	Rv1438	<i>tpiA</i>	Triosephosphate isomerase	cytoplasm; gluconeogenesis; glycolytic process; pentose-phosphate shunt; triose-phosphate isomerase activity
A0QWZ9	MSMEG_3122	Rv1461	<i>sufB</i>	FeS assembly protein SufB	iron-sulfur cluster assembly
A0QX00	MSMEG_3123	Rv1462	<i>sufD</i>	FeS assembly protein SufD	iron-sulfur cluster assembly
A0QX01	MSMEG_3124	Rv1463	<i>sufC</i>	FeS assembly ATPase SufC	ATPase activity; ATP binding; transport
P71534	MSMEG_3150	Rv1483	<i>fabG</i>	3-oxoacyl-[acyl-carrier-protein] reductase FabG	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity; fatty acid elongation; NADP binding
A0QX55	MSMEG_3178	Rv1547	<i>dnaE1</i>	DNA-directed DNA polymerase	3'-5' exonuclease activity; DNA binding; DNA-directed DNA polymerase activity; DNA replication
A0QX95	MSMEG_3219	Rv1611	<i>trpC</i>	Indole-3-glycerol phosphate synthase	indole-3-glycerol-phosphate synthase activity; tryptophan biosynthetic process
A0QYB0	MSMEG_3595			Uncharacterized protein	
A0QYD6	MSMEG_3621	Rv1854c	<i>ndh</i>	Membrane NADH dehydrogenase NdhA	flavin adenine dinucleotide binding; NADH dehydrogenase activity
A0QYE7	MSMEG_3632	Rv1844c	<i>gnd</i>	6-phosphogluconate dehydrogenase, decarboxylating	NADP binding; pentose-phosphate shunt; phosphogluconate dehydrogenase (decarboxylating) activity
A0QYF7	MSMEG_3642	Rv1832	<i>gcvP</i>	Glycine dehydrogenase	glycine decarboxylation via glycine cleavage system; glycine dehydrogenase (decarboxylating) activity; pyridoxal phosphate binding
A0QYQ7	MSMEG_3746	Rv1699	<i>pyrG</i>	CTP synthase	'de novo' CTP biosynthetic process; ATP binding; CTP synthase activity; glutamine metabolic process
A0QYR3	MSMEG_3754	Rv1691	<i>#REF!</i>	TPR-repeat-containing protein	
A0QYS6	MSMEG_3770	Rv1658	<i>argG</i>	Argininosuccinate synthase	arginine biosynthetic process; argininosuccinate synthase activity; ATP binding; cytoplasm
A0QYU8	MSMEG_3793	Rv1641	<i>infC</i>	Translation initiation factor IF-3	cytoplasm; translation initiation factor activity
A0QYW4	MSMEG_3808	Rv1638	<i>uvrA</i>	Excinuclease ABC subunit A	ATPase activity; ATP binding; cytoplasm; DNA binding; excinuclease ABC activity; excinuclease repair complex; nucleotide-excision repair; SOS response; zinc ion binding
A0QYZ2	MSMEG_3839	Rv1629	<i>polA</i>	DNA polymerase I	3'-5' exonuclease activity; DNA binding; DNA-dependent DNA replication; DNA-directed DNA polymerase activity
A0QZ11	MSMEG_3858	Rv2050	<i>rbpA</i>	RNA polymerase-binding protein RbpA	bacterial-type RNA polymerase core enzyme binding; positive regulation of transcription, DNA-templated; response to antibiotic; transcription, DNA-templated
A0QZ49	MSMEG_3897	Rv2112c	<i>dop</i>	Pup deamidase/depupylase	ATP binding; hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides; metal ion binding; modification-dependent protein catabolic process; peptidase activity; proteasomal protein catabolic process; protein pupylation

A0QZ54	MSMEG_3902	Rv2115c	<i>mpa</i>	Proteasome-associated ATPase	ATPase activity; ATP binding; modification-dependent protein catabolic process; proteasomal protein catabolic process; proteasome-activating nucleotidase complex; protein unfolding
A0QZX6	MSMEG_4185	Rv2124c	<i>meth</i>	Methionine synthase	cobalamin binding; intracellular; methionine synthase activity; pteridine-containing compound metabolic process; S-adenosylmethionine-homocysteine S-methyltransferase activity; zinc ion binding
A0QZY0	MSMEG_4189	Rv2130c	<i>mshC</i>	Mycothioli ligase	ATP binding; cysteine-glucosaminylinositol ligase activity; mycothiol biosynthetic process; zinc ion binding
A0R006	MSMEG_4217	Rv2145c	<i>wag31</i>	Cell wall synthesis protein Wag31	cell cycle; cell division; cytoplasm; regulation of cell shape
A0R012	MSMEG_4222	Rv2150c	<i>ftsZ</i>	Cell division protein FtsZ	barrier septum assembly; cell division site; cytoplasm; FtsZ-dependent cytokinesis; GTPase activity; GTP binding; protein complex; protein polymerization
A0R029	MSMEG_4240	Rv2173		Polyprenyl synthetase	isoprenoid biosynthetic process; transferase activity
A0R033	MSMEG_4244	Rv2178c	<i>aroG</i>	Phospho-2-dehydro-3-deoxyheptonate aldolase	3-deoxy-7-phosphoheptulonate synthase activity; aromatic amino acid family biosynthetic process; chorismate biosynthetic process; manganese ion binding; plasma membrane
A0R061	MSMEG_4272	Rv2204c		Iron-sulfur cluster assembly accessory protein	iron-sulfur cluster assembly; iron-sulfur cluster binding; structural molecule activity
A0R069	MSMEG_4281	Rv2213	<i>pepA</i>	Probable cytosol aminopeptidase	aminopeptidase activity; cytoplasm; manganese ion binding; metalloexopeptidase activity
A0R072	MSMEG_4283	Rv2215	<i>sucB</i>	2-oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase	dihydrolipoyllysine-residue succinyltransferase activity
A0R090	MSMEG_4301			Acyl-CoA synthase	catalytic activity
A0R095	MSMEG_4306			Uncharacterized protein	
A0R0A1	MSMEG_4313			Glyoxalase/bleomycin resistance protein/dioxygenase	dioxygenase activity
A0R0B0	MSMEG_4323	Rv2241	<i>aceE</i>	Pyruvate dehydrogenase E1 component	glycolytic process; pyruvate dehydrogenase (acetyl-transferring) activity
A0R0B6	MSMEG_4329	Rv2247	<i>accD6</i>	Propionyl-CoA carboxylase beta chain	propionyl-CoA carboxylase activity
A0R0C7	MSMEG_4340	Rv2259	<i>adhE2</i>	NAD/mycothiol-dependent formaldehyde dehydrogenase	oxidoreductase activity; zinc ion binding
A0R0R9	MSMEG_4485	Rv2357c	<i>glyS</i>	Glycine-tRNA ligase	ATP binding; cytoplasm; glycine-tRNA ligase activity; glycyl-tRNA aminoacylation
A0R0T8	MSMEG_4504	Rv2373c	<i>dnaJ</i>	Chaperone protein DnaJ	ATP binding; cytoplasm; DNA replication; protein folding; response to heat; zinc ion binding
A0R0W1	MSMEG_4527	Rv2391	<i>sirA</i>	Ferredoxin sulfite reductase	4 iron, 4 sulfur cluster binding; heme binding; metal ion binding; sulfite reductase (ferredoxin) activity

A0R152	MSMEG_4626	Rv2444c	<i>rne</i>	Ribonuclease E	cytoplasm; metal ion binding; mRNA processing; ribonuclease activity; RNA binding; rRNA processing
A0R171	MSMEG_4646	Rv2455c		Pyruvate synthase	2-oxoglutarate synthase activity
A0R196	MSMEG_4671	Rv2457c	<i>clpX</i>	ATP-dependent Clp protease ATP-binding subunit ClpX	ATP binding; protein folding; zinc ion binding
A0R197	MSMEG_4672	Rv2460c	<i>clpP</i>	Endopeptidase Clp	cytoplasm; serine-type endopeptidase activity
A0R199	MSMEG_4674	Rv2462c	<i>tig</i>	Trigger factor	cell cycle; cell division; cytoplasm; peptidyl-prolyl cis-trans isomerase activity; protein folding; protein transport
A0R1C3	MSMEG_4700	Rv2477c		ATPase component of ABC transporter	ATPase activity; ATP binding
A0R1H3	MSMEG_4753	Rv2521		Antioxidant, AhpC/TSA family protein	peroxidase activity; peroxiredoxin activity
A0R1Y7	MSMEG_4920	Rv1323	<i>fadA4</i>	Probable acetyl-CoA acetyltransferase	acetyl-CoA C-acetyltransferase activity
A0R218	MSMEG_4954	Rv1297	<i>rho</i>	Transcription termination factor Rho	ATP binding; DNA-templated transcription, termination; helicase activity; regulation of transcription, DNA-templated; RNA binding; RNA-dependent ATPase activity
A0R221	MSMEG_4957	Rv1294		Homoserine dehydrogenase	amino acid binding; homoserine dehydrogenase activity; isoleucine biosynthetic process; methionine biosynthetic process; NADP binding; threonine biosynthetic process
A0R242	MSMEG_4979	Rv1285	<i>cysD</i>	Sulfate adenylyltransferase subunit 2	ATP binding; hydrogen sulfide biosynthetic process; sulfate adenylyltransferase (ATP) activity; sulfate assimilation; sulfate reduction
A0R248	MSMEG_4985	Rv1284		Carbonic anhydrase	carbonate dehydratase activity; zinc ion binding
A0R2A4	MSMEG_5042	Rv1253	<i>deaD</i>	ATP-dependent RNA helicase DeaD	ATP binding; ATP-dependent RNA helicase activity; cellular response to cold; cytoplasm; ribosomal large subunit assembly; RNA binding; RNA catabolic process
A0R2D0	MSMEG_5068	Rv1229c	<i>mrp</i>	ATP-binding Mrp protein	ATP binding
A0R2J0	MSMEG_5132	Rv1165	<i>typA</i>	GTP-binding translation elongation factor TypA	GTPase activity; GTP binding; translation elongation factor activity
A0R2Q5	MSMEG_5197			Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0R2T0	MSMEG_5222	Rv1112	<i>ychF</i>	Ribosome-binding ATPase YchF	ATPase activity; ATP binding; GTP binding; ribosomal large subunit binding; ribosome binding
A0R2V6	MSMEG_5248	Rv1094	<i>desA2</i>	Acyl-acyl-carrier protein desaturase DesA2	acyl-[acyl-carrier-protein] desaturase activity; fatty acid metabolic process
A0R2V7	MSMEG_5249	Rv1093	<i>glyA</i>	Serine hydroxymethyltransferase	cytoplasm; glycine biosynthetic process from serine; glycine hydroxymethyltransferase activity; pyridoxal phosphate binding; tetrahydrofolate interconversion
A0R3B5	MSMEG_5412			Uncharacterized protein	extracellular space

A0R3C5	MSMEG_5423	Rv1020	<i>mfd</i>	Transcription-repair-coupling factor	ATP binding; cytoplasm; damaged DNA binding; helicase activity; regulation of transcription, DNA-templated; transcription-coupled nucleotide-excision repair, DNA damage recognition
A0R3D6	MSMEG_5435	Rv1013		Putative ligase	ligase activity
A0R3D9	MSMEG_5438	Rv1010	<i>ksgA</i>	Ribosomal RNA small subunit methyltransferase A	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase activity; cytoplasm; RNA binding; rRNA (adenine-N6,N6-)-dimethyltransferase activity
A0R3L1	MSMEG_5512	Rv0958		Magnesium chelatase	ATP binding; regulation of transcription, DNA-templated
A0R3L4	MSMEG_5515	Rv0957	<i>purH</i>	Bifunctional purine biosynthesis protein PurH	'de novo' IMP biosynthetic process; IMP cyclohydrolase activity; phosphoribosylaminoimidazolecarboxamide formyltransferase activity
A0R3Y5	MSMEG_5639	Rv0905	<i>echA6</i>	Enoyl-CoA hydratase	enoyl-CoA hydratase activity; isomerase activity
A0R409	MSMEG_5664			Peptidyl-prolyl cis-trans isomerase	peptidyl-prolyl cis-trans isomerase activity; protein folding
A0R425	MSMEG_5680	Rv0887c		Glyoxalase family protein	dioxygenase activity
A0R441	MSMEG_5696	Rv0871	<i>cspB</i>	'Cold-shock' DNA-binding domain protein	cytoplasm; DNA binding; regulation of transcription, DNA-templated
A0R451	MSMEG_5706	Rv0861c		DNA or RNA helicase of superfamily protein II	ATP binding; ATP-dependent DNA helicase activity; DNA binding; nucleotide-excision repair
A0R461	MSMEG_5715			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0R465	MSMEG_5720	Rv0860	<i>fadB</i>	Enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase	3-hydroxyacyl-CoA dehydrogenase activity; coenzyme binding; fatty acid metabolic process
A0R466	MSMEG_5721	Rv0859		Acetyl-CoA acetyltransferase	acetyl-CoA C-acyltransferase activity
A0R4B3	MSMEG_5773	Rv0824c	<i>des</i>	Acyl-acyl-carrier protein desaturase DesA1	acyl-[acyl-carrier-protein] desaturase activity; fatty acid metabolic process
A0R4G4	MSMEG_5824	Rv0803	<i>purL</i>	Phosphoribosylformylglycinamide synthase subunit PurL	'de novo' IMP biosynthetic process; ATP binding; cytoplasm; magnesium ion binding; phosphoribosylformylglycinamide synthase activity
A0R4H6	MSMEG_5837			Glutathione peroxidase	glutathione peroxidase activity; response to oxidative stress
A0R4L1	MSMEG_5872	Rv0757	<i>phoP</i>	DNA-binding response regulator PhoP	DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R4S6	MSMEG_5937	Rv3534c	<i>bphI-2</i>	4-hydroxy-2-oxovalerate aldolase 2	4-hydroxy-2-oxovalerate aldolase activity; aromatic compound catabolic process; manganese ion binding
A0R4S7	MSMEG_5939	Rv3535c	<i>mhpF</i>	Acetaldehyde dehydrogenase 2	acetaldehyde dehydrogenase (acetylating) activity; aromatic compound catabolic process; NAD binding
A0R4T0	MSMEG_5943	Rv3538		MaoC-like dehydratase	3alpha,7alpha,12alpha-trihydroxy-5beta-cholest-24-enoyl-CoA hydratase activity; 3-hydroxyacyl-CoA dehydrogenase activity
A0R561	MSMEG_6077	Rv3583c	<i>carD</i>	RNA polymerase-binding transcription factor CarD	

A0R566	MSMEG_6082	Rv3588c	<i>cynT</i>	Carbonic anhydrase	carbonate dehydratase activity; carbon utilization; zinc ion binding
A0R577	MSMEG_6094	Rv3598c	<i>lysS</i>	Lysine-tRNA ligase	ATP binding; cytoplasm; lysine-tRNA ligase activity; lysyl-tRNA aminoacylation; magnesium ion binding; nucleic acid binding
A0R5D9	MSMEG_6157	Rv3646c	<i>topA</i>	DNA topoisomerase 1	DNA binding; DNA topoisomerase type I activity; DNA topological change; magnesium ion binding
A0R5M2	MSMEG_6241			ATPase associated with various cellular activities, AAA-5	ATPase activity; ATP binding
A0R5N8	MSMEG_6257	Rv3709c	<i>ask</i>	Aspartokinase	amino acid binding; aspartate kinase activity; lysine biosynthetic process via diaminopimelate; threonine biosynthetic process
A0R616	MSMEG_6391	Rv3799c	<i>accD4</i>	Propionyl-CoA carboxylase beta chain 4 accD4	propionyl-CoA carboxylase activity
A0R617	MSMEG_6392	Rv3800c		Polyketide synthase	biosynthetic process; hydrolase activity, acting on ester bonds; phosphopantetheine binding; transferase activity
A0R628	MSMEG_6403	Rv3808c	<i>glfT2</i>	Galactofuranosyl transferase GIfT2	capsule polysaccharide biosynthetic process; cell wall macromolecule biosynthetic process; cell wall organization; membrane; metal ion binding; transferase activity; transferase activity, transferring glycosyl groups
A0R635	MSMEG_6410	Rv3818		Putative Rieske 2Fe-2S iron-sulfur protein	2 iron, 2 sulfur cluster binding; metal ion binding; oxidoreductase activity
A0R656	MSMEG_6431	Rv3849	<i>espR</i>	Uncharacterized protein	sequence-specific DNA binding
A0R683	MSMEG_6458	Rv3858c	<i>gltD</i>	Glutamate synthase, small subunit	flavin adenine dinucleotide binding; glutamate biosynthetic process; iron-sulfur cluster binding; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor
A0R684	MSMEG_6459	Rv3859c		Ferredoxin-dependent glutamate synthase 1	glutamate biosynthetic process; glutamate synthase (ferredoxin) activity
A0R696	MSMEG_6471			Glycine/D-amino acid oxidase	oxidoreductase activity
A0R6D6	MSMEG_6511			Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0R6D7	MSMEG_6512			Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0R6Q9	MSMEG_6638	Rv1133c	<i>metE</i>	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity; methionine biosynthetic process; zinc ion binding
A0R7J3	MSMEG_6938	Rv3917c	<i>parB</i>	Putative chromosome-partitioning protein parB	DNA binding
A0R7J6	MSMEG_6941	Rv3920c		Single-stranded nucleic acid binding R3H	nucleic acid binding
<b>Low Confidence</b>					
A0QNF6	MSMEG_0024	Rv0009	<i>ppiA</i>	Peptidyl-prolyl cis-trans isomerase	peptidyl-prolyl cis-trans isomerase activity; protein folding
A0QQC8	MSMEG_0709	Rv0350	<i>dnaK</i>	Chaperone protein DnaK	ATP binding; protein folding

A0QQX6	MSMEG_0911	Rv0467	<i>aceA</i>	Isocitrate lyase	carboxylic acid metabolic process; isocitrate lyase activity
A0QQX8	MSMEG_0913	Rv0469	<i>umaA</i>	Methoxy mycolic acid synthase 1	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
A0QS44	MSMEG_1345	Rv0639	<i>nusG</i>	Transcription termination/antitermination protein NusG	DNA-templated transcription, elongation; DNA-templated transcription, termination; regulation of DNA-templated transcription, elongation; transcription antitermination
A0QS46	MSMEG_1347	Rv0641	<i>rplA</i>	50S ribosomal protein L1	large ribosomal subunit; regulation of translation; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QS63	MSMEG_1365	Rv0652	<i>rplL</i>	50S ribosomal protein L7/L12	ribosome; structural constituent of ribosome; translation
P60281	MSMEG_1367	Rv0667	<i>rpoB</i>	DNA-directed RNA polymerase subunit beta	DNA binding; DNA-directed RNA polymerase activity; response to antibiotic; ribonucleoside binding; transcription, DNA-templated
A0QS66	MSMEG_1368	Rv0668	<i>rpoC</i>	DNA-directed RNA polymerase subunit beta'	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QS97	MSMEG_1399	Rv0683	<i>rpsG</i>	30S ribosomal protein S7	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation; tRNA binding
A0QS98	MSMEG_1401	Rv0685	<i>tuf</i>	Elongation factor Tu	cytoplasm; GTPase activity; GTP binding; translation elongation factor activity
A0QSD0	MSMEG_1435	Rv0700	<i>rpsJ</i>	30S ribosomal protein S10	ribosome; structural constituent of ribosome; translation; tRNA binding
A0QSD1	MSMEG_1436	Rv0701	<i>rplC</i>	50S ribosomal protein L3	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSD4	MSMEG_1439	Rv0704	<i>rplB</i>	50S ribosomal protein L2	large ribosomal subunit; rRNA binding; structural constituent of ribosome; transferase activity; translation
A0QSD5	MSMEG_1440	Rv0705	<i>rpsS</i>	30S ribosomal protein S19	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSD6	MSMEG_1441	Rv0706	<i>rplV</i>	50S ribosomal protein L22	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSD7	MSMEG_1442	Rv0707	<i>rpsC</i>	30S ribosomal protein S3	mRNA binding; rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSF9	MSMEG_1465	Rv0714	<i>rplN</i>	50S ribosomal protein L14	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSG1	MSMEG_1467	Rv0716	<i>rplE</i>	50S ribosomal protein L5	ribosome; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QSG6	MSMEG_1472	Rv0721	<i>rpsE</i>	30S ribosomal protein S5	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSG8	MSMEG_1474	Rv0723	<i>rplO</i>	50S ribosomal protein L15	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSH8	MSMEG_1484	Rv0733	<i>adk</i>	Adenylate monophosphate kinase	adenylate kinase activity; AMP salvage; ATP binding; cytoplasm
A0QSL3	MSMEG_1519	Rv3462c	<i>infA</i>	Translation initiation factor IF-1	cytoplasm; translation initiation factor activity

A0QSL6	MSMEG_1522	Rv3459c	<i>rpsK</i>	30S ribosomal protein S11	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSL7	MSMEG_1523	Rv3458c	<i>rpsD</i>	30S ribosomal protein S4	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation
A0QSL8	MSMEG_1524	Rv3457c	<i>rpoA</i>	DNA-directed RNA polymerase subunit alpha	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QSL9	MSMEG_1525	Rv3456c	<i>rplQ</i>	50S ribosomal protein L17	ribosome; structural constituent of ribosome; translation
A0QSS4	MSMEG_1583	Rv3417c	<i>groL2</i>	60 kDa chaperonin 2	ATP binding; cytoplasm; protein refolding
A0QSZ1	MSMEG_1652	Rv3340	<i>metC</i>	O-acetylhomoserine sulfhydrylase	O-acetylhomoserine aminocarboxypropyltransferase activity; pyridoxal phosphate binding
A0QSZ3	MSMEG_1654	Rv0066c	<i>icd2</i>	Isocitrate dehydrogenase, NADP-dependent	isocitrate dehydrogenase (NADP+) activity; metal ion binding; tricarboxylic acid cycle
A0QTE1	MSMEG_1807	Rv3285	<i>accA3</i>	Acetyl-/propionyl-coenzyme A carboxylase alpha chain	ATP binding; biotin carboxylase activity; metal ion binding
A0QU54	MSMEG_2081	Rv3139		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QUV6	MSMEG_2351	Rv3029c	<i>etfB</i>	Electron transfer flavoprotein beta subunit	electron carrier activity
A0QUX8	MSMEG_2374	Rv3001c	<i>ilvC</i>	Ketol-acid reductoisomerase	coenzyme binding; isoleucine biosynthetic process; ketol-acid reductoisomerase activity; valine biosynthetic process
A0QUY2	MSMEG_2378	Rv2996c	<i>serA</i>	D-3-phosphoglycerate dehydrogenase	amino acid binding; L-serine biosynthetic process; NAD binding; phosphoglycerate dehydrogenase activity
Q9ZHC5	MSMEG_2389	Rv2986c	<i>hup</i>	DNA-binding protein HU homolog (Hlp)	cell wall; chromosome condensation; DNA binding; extracellular region
A0QVB8	MSMEG_2519	Rv2890c	<i>rpsB</i>	30S ribosomal protein S2	small ribosomal subunit; structural constituent of ribosome; translation
A0QVQ5	MSMEG_2656	Rv2783c	<i>pnp</i>	Polyribonucleotide nucleotidyltransferase	3'-5'-exoribonuclease activity; cytoplasm; magnesium ion binding; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity; RNA binding; RNA processing
A0QWS8	MSMEG_3050	Rv1388	<i>mihF</i>	Integration host factor	nucleic acid binding
A0QWY0	MSMEG_3103	Rv1449c	<i>tkt</i>	Transketolase	metal ion binding; transketolase activity
A0QXA3	MSMEG_3227	Rv1617	<i>pyk</i>	Pyruvate kinase	glycolytic process; magnesium ion binding; potassium ion binding; pyruvate kinase activity
A0QYW6	MSMEG_3811	Rv1636		Universal stress protein family protein, putative	cytoplasm; response to stress
A0QYY6	MSMEG_3833	Rv1630	<i>rpsA</i>	30S ribosomal protein S1	nucleotidyltransferase activity; ribosome; RNA binding; structural constituent of ribosome; translation
A0R0B4	MSMEG_4327	Rv2245	<i>kasA</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 1	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R0B5	MSMEG_4328	Rv2246	<i>kasB</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 2	3-oxoacyl-[acyl-carrier-protein] synthase activity

A0R1H7	MSMEG_4757	Rv2524c		Fatty acid synthase	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process; fatty acid synthase complex; nucleotide binding
A0R1V9	MSMEG_4891	Rv2428		Alkyl hydroperoxide reductase subunit C	peroxidase activity; peroxiredoxin activity
A0R2B1	MSMEG_5049	Rv1248c	<i>kgd</i>	Multifunctional 2-oxoglutarate metabolism enzyme	2-hydroxy-3-oxoadipate synthase activity; 2-oxoglutarate decarboxylase activity; dihydrolipoylysine-residue succinyltransferase activity; metal ion binding; oxoglutarate dehydrogenase (succinyl-transferring) activity; thiamine pyrophosphate binding; tricarboxylic acid cycle
A0R2U8	MSMEG_5240	Rv1098c	<i>fumC</i>	Fumarate hydratase class II	fumarate hydratase activity; fumarate metabolic process; tricarboxylic acid cycle; tricarboxylic acid cycle enzyme complex
A0R2X1	MSMEG_5263	Rv1080c	<i>greA</i>	Transcription elongation factor GreA	DNA binding; regulation of DNA-templated transcription, elongation; transcription, DNA-templated
A0R2Y1	MSMEG_5273	Rv1074c	<i>fadA3</i>	Acetyl-CoA acetyltransferase	transferase activity, transferring acyl groups other than amino-acyl groups
A0R3B8	MSMEG_5415	Rv1023	<i>eno</i>	Enolase	cell surface; extracellular region; glycolytic process; magnesium ion binding; phosphopyruvate hydratase activity; phosphopyruvate hydratase complex
A0R417	MSMEG_5672	Rv0896	<i>gltA</i>	Citrate synthase	cellular carbohydrate metabolic process; citrate (Si)-synthase activity; cytoplasm; tricarboxylic acid cycle
A0R574	MSMEG_6091	Rv3596c	<i>clpC1</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein metabolic process
A0R5E1	MSMEG_6159	Rv3648c	<i>cspA</i>	Probable cold shock protein A	cytoplasm; DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R5H1	MSMEG_6189	Rv3676	<i>crp</i>	Crp/Fnr familytranscriptional regulator	DNA binding; intracellular; sequence-specific DNA binding transcription factor activity; transcription, DNA-templated
A0R5M3	MSMEG_6242			Alcohol dehydrogenase, iron-containing	1,3-propanediol dehydrogenase activity; alcohol dehydrogenase (NAD) activity; metal ion binding
A0R618	MSMEG_6393	Rv3801c	<i>fadD32</i>	Acyl-CoA synthase	ligase activity
A0R729	MSMEG_6759			Glycerol kinase	ATP binding; glycerol-3-phosphate metabolic process; glycerol catabolic process; glycerol kinase activity
A0R730	MSMEG_6761			Glycerol-3-phosphate dehydrogenase	glycerol-3-phosphate dehydrogenase complex; glycerol-3-phosphate metabolic process; sn-glycerol-3-phosphate:ubiquinone-8 oxidoreductase activity
Q9AFI5	MSMEG_6896	Rv0054	<i>ssb</i>	Single-stranded DNA-binding protein	DNA replication; single-stranded DNA binding
A0R7F9	MSMEG_6897	Rv0053	<i>rpsF</i>	30S ribosomal protein S6	ribosome; rRNA binding; structural constituent of ribosome; translation
A0R7G6	MSMEG_6904	Rv0046c	<i>ino1</i>	Inositol-3-phosphate synthase	inositol-3-phosphate synthase activity; inositol biosynthetic process; phospholipid biosynthetic process

**Table B. 5 High stringency protein identification gene ontology enrichment data**

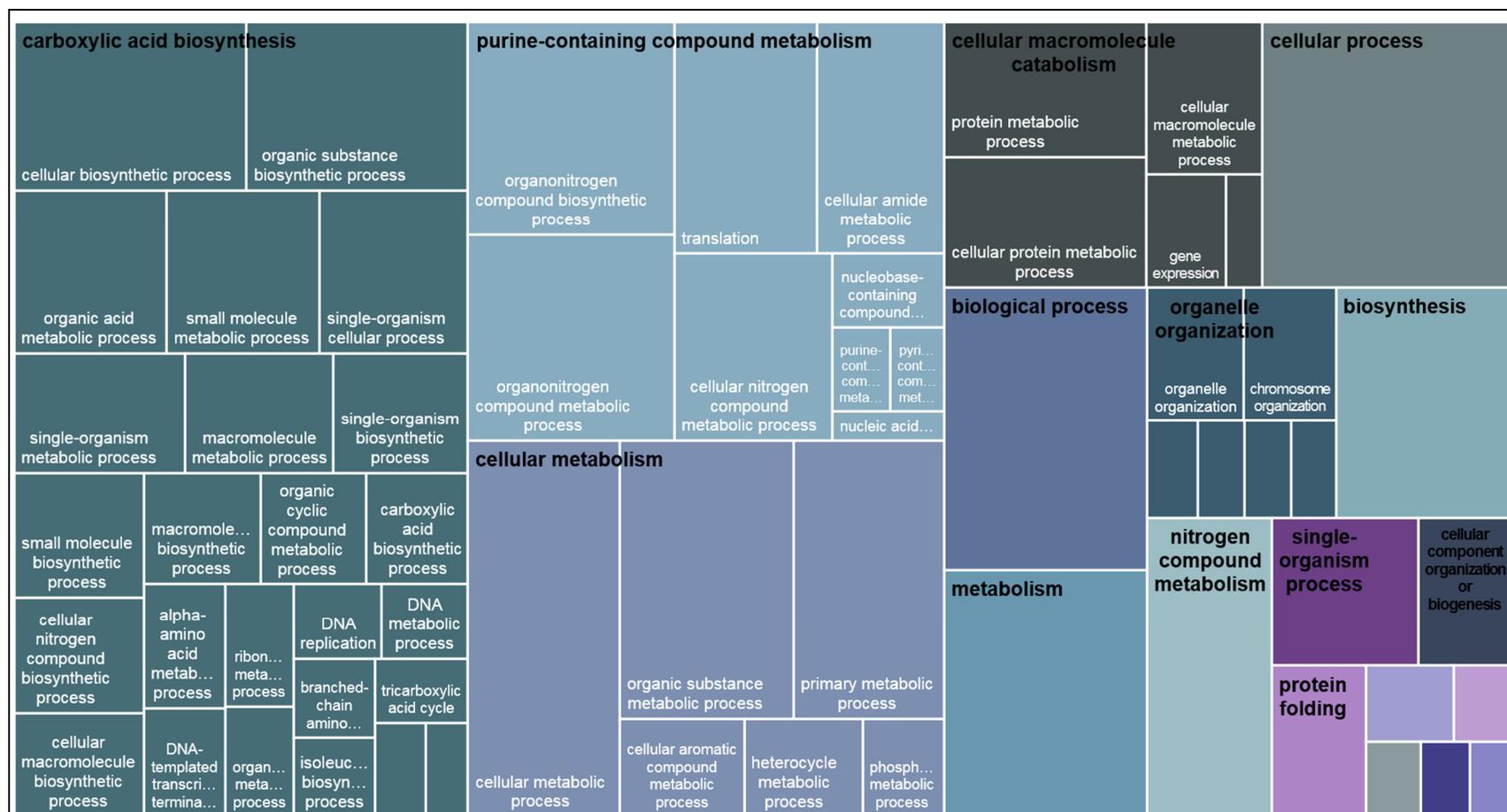
GOID	Ontology	Term	Number of identifications	p-value
GO:0000166	molecular function	nucleotide binding	69	2.38E-13
GO:0001882	molecular function	nucleoside binding	50	6.16E-12
GO:0001883	molecular function	purine nucleoside binding	49	1.97E-11
GO:0003674	molecular function	molecular function	220	6.75E-22
GO:0005488	molecular function	binding	149	6.83E-21
GO:0005524	molecular function	ATP binding	43	4.45E-09
GO:0017076	molecular function	purine nucleotide binding	49	2.07E-11
GO:0030554	molecular function	adenyl nucleotide binding	43	4.68E-09
GO:0032549	molecular function	ribonucleoside binding	50	6.16E-12
GO:0032550	molecular function	purine ribonucleoside binding	49	1.97E-11
GO:0032553	molecular function	ribonucleotide binding	50	3.50E-10
GO:0032555	molecular function	purine ribonucleotide binding	49	1.97E-11
GO:0032559	molecular function	adenyl ribonucleotide binding	43	4.45E-09
GO:0035639	molecular function	purine ribonucleoside triphosphate binding	49	1.97E-11
GO:0036094	molecular function	small molecule binding	74	1.14E-14
GO:0097159	molecular function	organic cyclic compound binding	121	3.90E-18
GO:0097367	molecular function	carbohydrate derivative binding	52	3.33E-11
GO:1901265	molecular function	nucleoside phosphate binding	69	2.38E-13
GO:1901363	molecular function	heterocyclic compound binding	121	3.90E-18
GO:0003676	molecular function	nucleic acid binding	66	1.11E-08
GO:0003824	molecular function	catalytic activity	166	4.63E-09
GO:0003916	molecular function	DNA topoisomerase activity	4	3.54E-03
GO:0003918	molecular function	DNA topoisomerase type II (ATP-hydrolyzing) activity	3	3.81E-02
GO:0008094	molecular function	DNA-dependent ATPase activity	7	3.24E-04
GO:0016462	molecular function	pyrophosphatase activity	23	2.27E-03
GO:0016817	molecular function	hydrolase activity, acting on acid anhydrides	23	2.66E-03
GO:0016818	molecular function	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	23	2.39E-03
GO:0017111	molecular function	nucleoside-triphosphatase activity	22	2.91E-03
GO:0061505	molecular function	DNA topoisomerase II activity	3	3.81E-02
GO:0006139	biological process	nucleobase-containing compound metabolic process	50	2.77E-03
GO:0006259	biological process	DNA metabolic process	14	9.31E-03
GO:0006265	biological process	DNA topological change	4	3.54E-03
GO:0006725	biological process	cellular aromatic compound metabolic process	59	1.55E-04
GO:0006807	biological process	nitrogen compound metabolic process	92	2.17E-12
GO:0006996	biological process	organelle organization	7	9.70E-05
GO:0008150	biological process	biological process	145	1.28E-18

GO:0008152	biological process	metabolic process	122	2.11E-16
GO:0009987	biological process	cellular process	133	8.61E-22
GO:0016043	biological process	cellular component organization	13	3.40E-05
GO:0034641	biological process	cellular nitrogen compound metabolic process	76	5.73E-10
GO:0043170	biological process	macromolecule metabolic process	61	2.84E-06
GO:0044237	biological process	cellular metabolic process	117	1.45E-18
GO:0044238	biological process	primary metabolic process	101	7.10E-14
GO:0044260	biological process	cellular macromolecule metabolic process	60	2.77E-06
GO:0046483	biological process	heterocycle metabolic process	58	2.38E-04
GO:0051276	biological process	chromosome organization	6	1.44E-04
GO:0071103	biological process	DNA conformation change	5	1.77E-03
GO:0071704	biological process	organic substance metabolic process	115	7.39E-16
GO:0071840	biological process	cellular component organization or biogenesis	15	3.01E-05
GO:0090304	biological process	nucleic acid metabolic process	38	8.94E-02
GO:1901360	biological process	organic cyclic compound metabolic process	59	2.38E-04
GO:0006260	biological process	DNA replication	7	8.47E-03
GO:0009058	biological process	biosynthetic process	90	1.15E-13
GO:0009059	biological process	macromolecule biosynthetic process	48	9.21E-05
GO:0034645	biological process	cellular macromolecule biosynthetic process	48	6.18E-05
GO:0044249	biological process	cellular biosynthetic process	85	6.32E-13
GO:1901576	biological process	organic substance biosynthetic process	85	2.20E-12
GO:0005575	cellular component	cellular component	75	5.52E-05
GO:0005622	cellular component	intracellular	67	2.17E-20
GO:0005623	cellular component	cell	73	8.74E-13
GO:0005694	cellular component	chromosome	3	5.78E-02
GO:0043226	cellular component	organelle	22	1.17E-12
GO:0043228	cellular component	non-membrane-bounded organelle	22	5.05E-13
GO:0043229	cellular component	intracellular organelle	22	8.74E-13
GO:0043232	cellular component	intracellular non-membrane-bounded organelle	22	5.05E-13
GO:0044424	cellular component	intracellular part	63	1.38E-22
GO:0044464	cellular component	cell part	71	1.96E-12
GO:0005737	cellular component	cytoplasm	61	1.38E-22
GO:0043167	molecular function	ion binding	45	1.51E-03
GO:0043169	molecular function	cation binding	39	7.00E-03
GO:0046872	molecular function	metal ion binding	39	5.69E-03
GO:0050896	biological process	response to stimulus	15	7.42E-02
GO:0044699	biological process	single-organism process	70	1.70E-07
GO:0044763	biological process	single-organism cellular process	66	2.63E-08
GO:0044085	biological process	cellular component biogenesis	8	4.19E-02
GO:0008104	biological process	protein localization	4	8.28E-02

GO:0003995	molecular function	acyl-CoA dehydrogenase activity	9	7.57E-02
GO:0050660	molecular function	flavin adenine dinucleotide binding	13	7.81E-02
GO:0050662	molecular function	coenzyme binding	19	5.76E-02
GO:0006082	biological process	organic acid metabolic process	37	1.58E-08
GO:0006549	biological process	isoleucine metabolic process	4	9.87E-03
GO:0009081	biological process	branched-chain amino acid metabolic process	4	9.87E-03
GO:0009082	biological process	branched-chain amino acid biosynthetic process	4	9.87E-03
GO:0009097	biological process	isoleucine biosynthetic process	4	9.87E-03
GO:0016053	biological process	organic acid biosynthetic process	18	4.44E-04
GO:0044281	biological process	small molecule metabolic process	49	1.58E-08
GO:0044283	biological process	small molecule biosynthetic process	25	8.97E-06
GO:0044710	biological process	single-organism metabolic process	55	3.80E-07
GO:0044711	biological process	single-organism biosynthetic process	36	8.97E-06
GO:1901564	biological process	organonitrogen compound metabolic process	57	5.25E-14
GO:1901566	biological process	organonitrogen compound biosynthetic process	50	1.82E-14
GO:1901605	biological process	alpha-amino acid metabolic process	16	6.69E-04
GO:1901607	biological process	alpha-amino acid biosynthetic process	13	4.72E-04
GO:0006520	biological process	cellular amino acid metabolic process	18	7.58E-04
GO:0006573	biological process	valine metabolic process	3	5.78E-02
GO:0008652	biological process	cellular amino acid biosynthetic process	12	3.05E-03
GO:0009099	biological process	valine biosynthetic process	3	3.81E-02
GO:0019752	biological process	carboxylic acid metabolic process	32	1.83E-07
GO:0043436	biological process	oxoacid metabolic process	32	4.49E-07
GO:0046394	biological process	carboxylic acid biosynthetic process	17	3.04E-04
GO:0005525	molecular function	GTP binding	7	7.17E-03
GO:0019001	molecular function	guanyl nucleotide binding	7	7.17E-03
GO:0032561	molecular function	guanyl ribonucleotide binding	7	7.17E-03
GO:0030145	molecular function	manganese ion binding	5	5.83E-03
GO:0008270	molecular function	zinc ion binding	14	1.12E-02
GO:0006790	biological process	sulfur compound metabolic process	7	3.24E-02
GO:0044271	biological process	cellular nitrogen compound biosynthetic process	54	1.91E-05
GO:0006793	biological process	phosphorus metabolic process	18	3.22E-03
GO:0006796	biological process	phosphate-containing compound metabolic process	17	7.17E-03
GO:0019637	biological process	organophosphate metabolic process	17	4.51E-03
GO:0051186	biological process	cofactor metabolic process	12	5.22E-02
GO:0004386	molecular function	helicase activity	7	7.17E-03
GO:0006163	biological process	purine nucleotide metabolic process	9	1.66E-02
GO:0006753	biological process	nucleoside phosphate metabolic process	12	2.61E-02
GO:0009116	biological process	nucleoside metabolic process	9	2.38E-02
GO:0009117	biological process	nucleotide metabolic process	12	2.42E-02

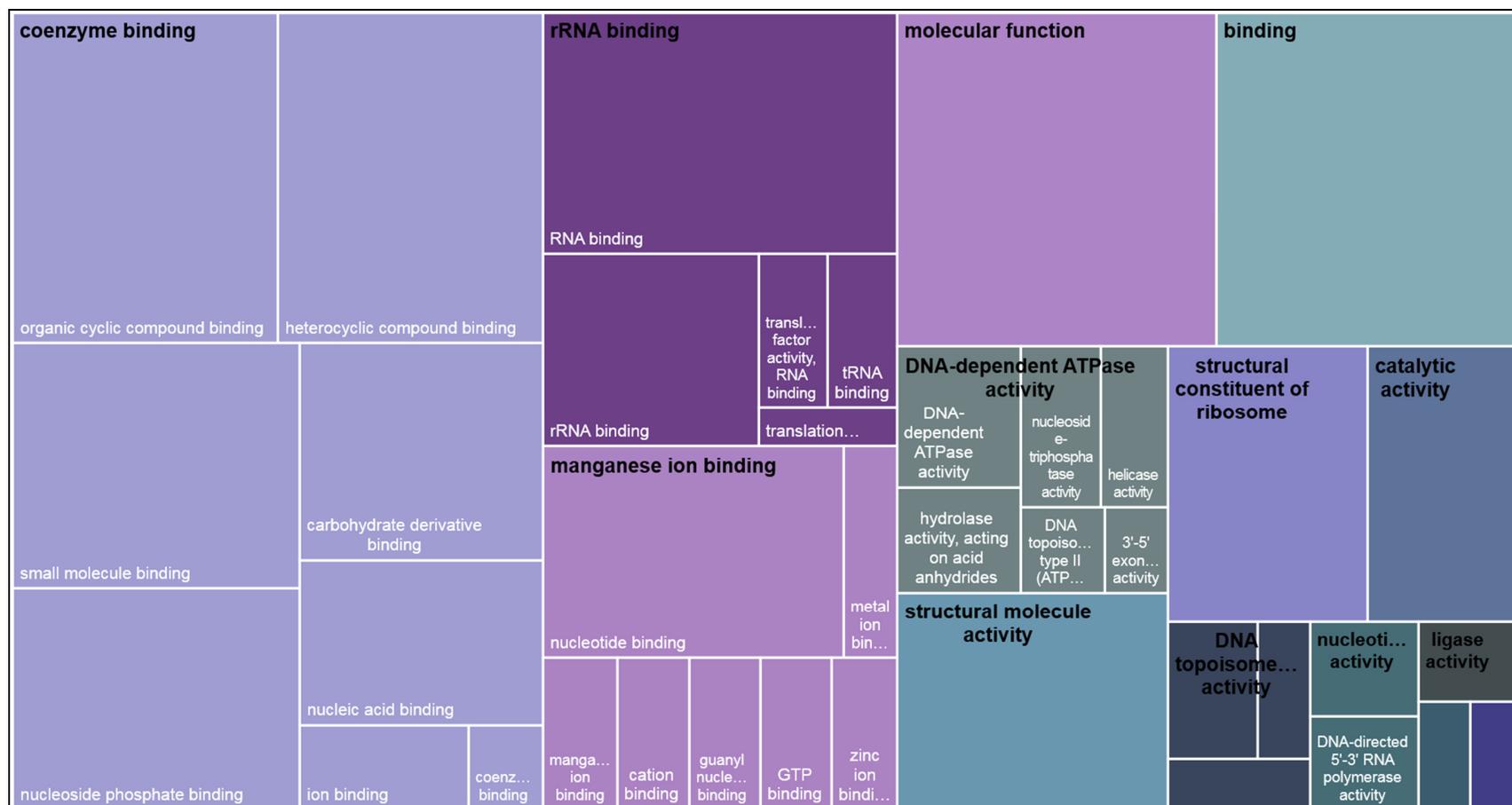
GO:0009119	biological process	ribonucleoside metabolic process	9	1.00E-02
GO:0009123	biological process	nucleoside monophosphate metabolic process	10	4.44E-03
GO:0009126	biological process	purine nucleoside monophosphate metabolic process	9	2.70E-03
GO:0009150	biological process	purine ribonucleotide metabolic process	9	9.31E-03
GO:0009156	biological process	ribonucleoside monophosphate biosynthetic process	6	9.50E-02
GO:0009161	biological process	ribonucleoside monophosphate metabolic process	10	2.66E-03
GO:0009167	biological process	purine ribonucleoside monophosphate metabolic process	9	2.70E-03
GO:0009259	biological process	ribonucleotide metabolic process	11	2.39E-03
GO:0009260	biological process	ribonucleotide biosynthetic process	7	4.58E-02
GO:0019693	biological process	ribose phosphate metabolic process	12	1.79E-03
GO:0042278	biological process	purine nucleoside metabolic process	7	4.58E-02
GO:0046128	biological process	purine ribonucleoside metabolic process	7	4.58E-02
GO:0046390	biological process	ribose phosphate biosynthetic process	7	5.21E-02
GO:0055086	biological process	nucleobase-containing small molecule metabolic process	12	5.73E-02
GO:0072521	biological process	purine-containing compound metabolic process	9	3.12E-02
GO:1901135	biological process	carbohydrate derivative metabolic process	15	3.18E-02
GO:1901657	biological process	glycosyl compound metabolic process	10	9.74E-03
GO:1901659	biological process	glycosyl compound biosynthetic process	6	8.50E-02
GO:0016874	molecular function	ligase activity	16	2.38E-02
GO:0006091	biological process	generation of precursor metabolites and energy	9	8.26E-03
GO:0006099	biological process	tricarboxylic acid cycle	5	1.37E-02
GO:0009060	biological process	aerobic respiration	5	2.99E-02
GO:0004658	molecular function	propionyl-CoA carboxylase activity	3	8.54E-02
GO:0010467	biological process	gene expression	42	1.42E-03
GO:0005829	cellular component	cytosol	3	5.78E-02
GO:0032991	cellular component	macromolecular complex	28	2.19E-09
GO:0044444	cellular component	cytoplasmic part	25	7.34E-15
GO:0044445	cellular component	cytosolic part	3	3.81E-02
GO:0003723	molecular function	RNA binding	31	9.57E-18
GO:0030529	cellular component	intracellular ribonucleoprotein complex	20	2.84E-12
GO:1990904	cellular component	ribonucleoprotein complex	20	2.84E-12
GO:0006353	biological process	DNA-templated transcription, termination	4	1.79E-03
GO:0072524	biological process	pyridine-containing compound metabolic process	6	3.88E-02
GO:0006412	biological process	translation	22	4.76E-11
GO:0006518	biological process	peptide metabolic process	23	4.66E-11
GO:0019538	biological process	protein metabolic process	25	3.08E-09
GO:0043043	biological process	peptide biosynthetic process	23	1.69E-11
GO:0043603	biological process	cellular amide metabolic process	23	5.73E-10

GO:0043604	biological process	amide biosynthetic process	23	2.36E-11
GO:0044267	biological process	cellular protein metabolic process	24	5.77E-09
GO:0000049	molecular function	tRNA binding	5	7.94E-03
GO:0003746	molecular function	translation elongation factor activity	3	8.54E-02
GO:0008135	molecular function	translation factor activity, RNA binding	5	7.94E-03
GO:0003899	molecular function	DNA-directed RNA polymerase activity	4	9.87E-03
GO:0016779	molecular function	nucleotidyltransferase activity	9	9.31E-03
GO:0034062	molecular function	RNA polymerase activity	4	9.87E-03
GO:0006096	biological process	glycolytic process	4	3.12E-02
GO:0006165	biological process	nucleoside diphosphate phosphorylation	4	3.12E-02
GO:0006757	biological process	ATP generation from ADP	4	3.12E-02
GO:0009132	biological process	nucleoside diphosphate metabolic process	4	6.56E-02
GO:0009135	biological process	purine nucleoside diphosphate metabolic process	4	3.12E-02
GO:0009179	biological process	purine ribonucleoside diphosphate metabolic process	4	3.12E-02
GO:0009185	biological process	ribonucleoside diphosphate metabolic process	4	3.12E-02
GO:0044724	biological process	single-organism carbohydrate catabolic process	5	8.14E-02
GO:0046031	biological process	ADP metabolic process	4	3.12E-02
GO:0046939	biological process	nucleotide phosphorylation	4	3.12E-02
GO:0016226	biological process	iron-sulfur cluster assembly	3	3.81E-02
GO:0022607	biological process	cellular component assembly	5	2.99E-02
GO:0031163	biological process	metallo-sulfur cluster assembly	3	3.81E-02
GO:0008408	molecular function	3'-5' exonuclease activity	3	8.54E-02
GO:0044265	biological process	cellular macromolecule catabolic process	4	5.23E-02
GO:0005198	molecular function	structural molecule activity	20	2.17E-12
GO:0006457	biological process	protein folding	7	3.59E-05
GO:0003755	molecular function	peptidyl-prolyl cis-trans isomerase activity	3	3.81E-02
GO:0016859	molecular function	cis-trans isomerase activity	3	3.81E-02
GO:0005840	cellular component	ribosome	19	2.07E-11
GO:0015934	cellular component	large ribosomal subunit	5	1.05E-04
GO:0044391	cellular component	ribosomal subunit	11	7.95E-12
GO:0044422	cellular component	organelle part	11	4.31E-10
GO:0044446	cellular component	intracellular organelle part	11	1.80E-10
GO:0019843	molecular function	rRNA binding	14	5.92E-09
GO:0003735	molecular function	structural constituent of ribosome	19	1.22E-11
GO:0015935	cellular component	small ribosomal subunit	6	1.53E-06



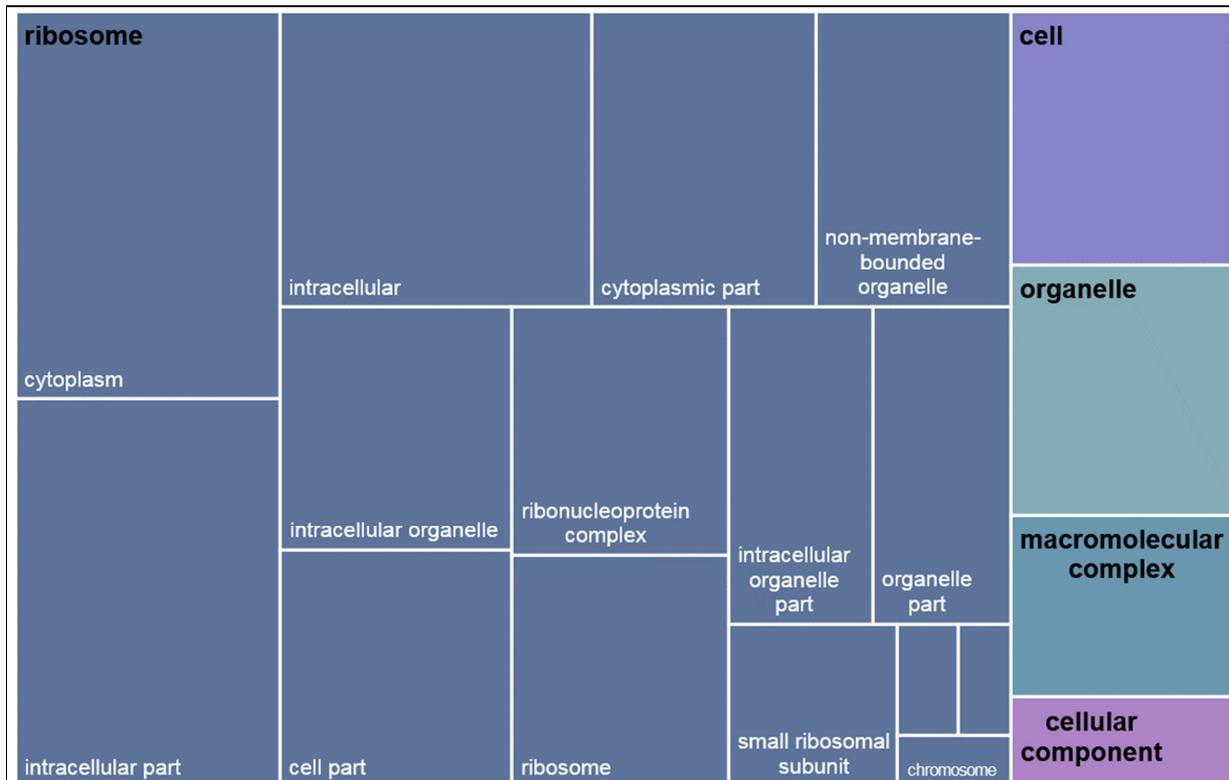
**Figure B. 3 Visualization of enriched biological function GO identities**

The hierarchical map of biological processes shows parent GO terms in black and child GO terms in white. The treemap shows carboxylic acid biosynthesis to be the most abundant parent GO identity which umbrellas the child GO identities DNA metabolic process, DNA replication and DNA templated transcription. Other enriched GO identities include cellular metabolism, purine-containing compound metabolism, organelle organization.



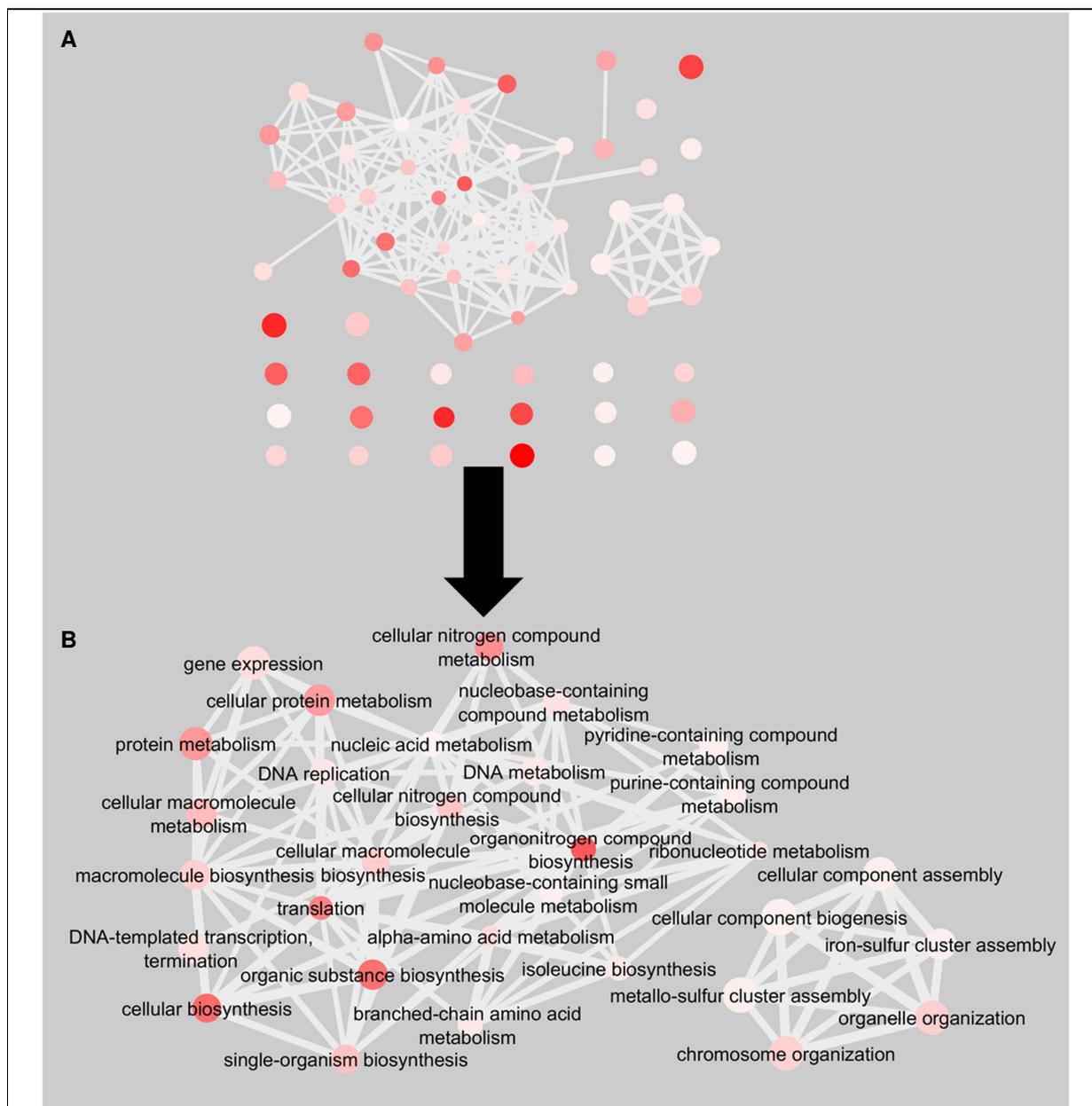
**Figure B. 4 Hierarchical clustering of enriched molecular function GO identities**

The treemap displays enriched parent GO identities in black and child GO terms in white. Hierarchical mapping shows that coenzyme binding was the most abundant parent GO term within the functional category molecular function and umbrellas GO identities nucleic acid binding, nucleoside phosphate binding. Other enriched GO terms included rRNA binding, DNA -dependent ATPase activity and DNA topoisomerase activity.



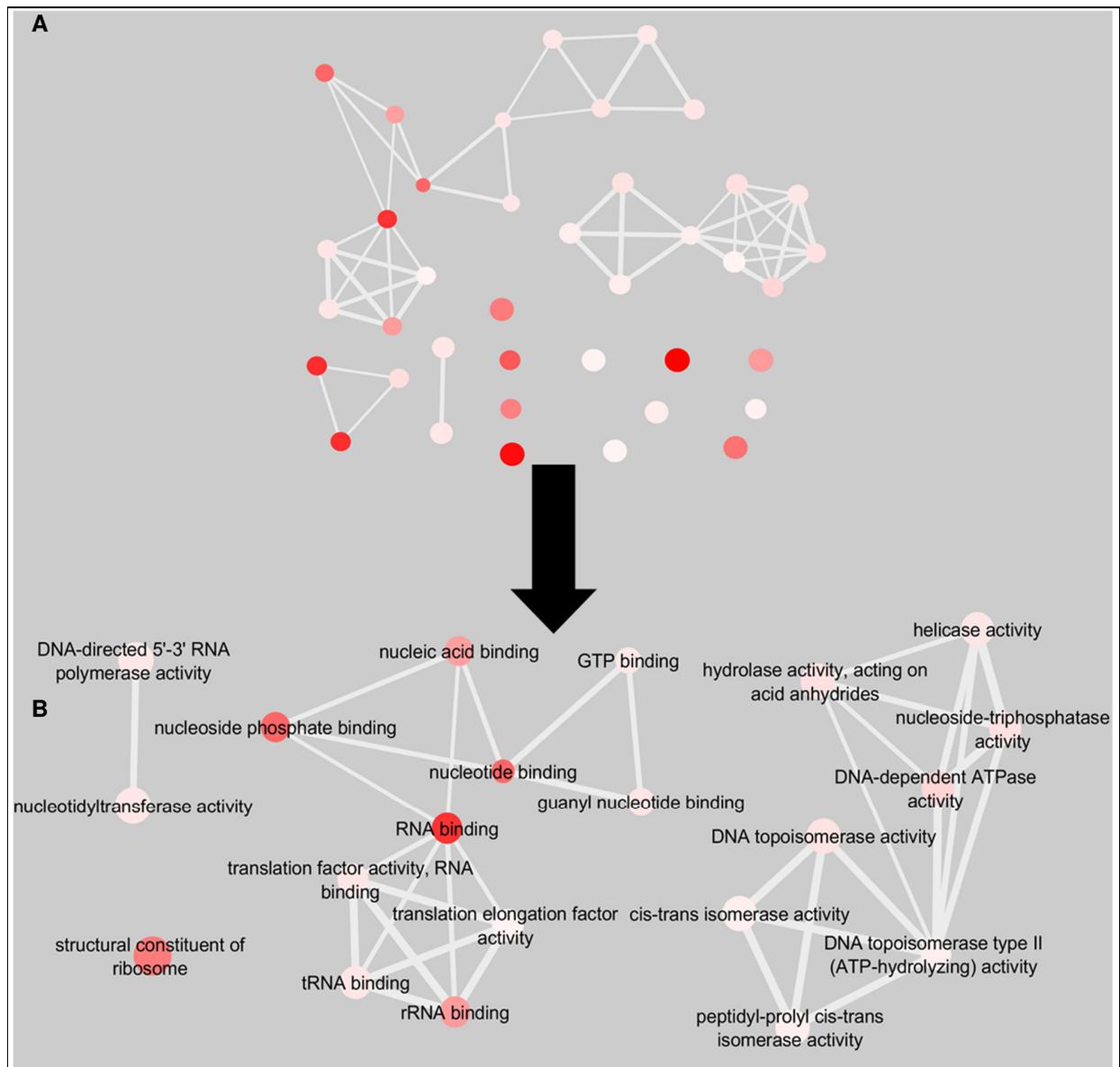
**Figure B. 5 Treemap representation of enriched cellular component GO identities**

Hierarchical mapping of enriched GO identities categorised as cellular components shows parent GO terms in black and child GO terms in white. The treemap shows the enrichment of GO terms associated with the ribosome which umbrellas ribonucleoprotein complex and chromosome.



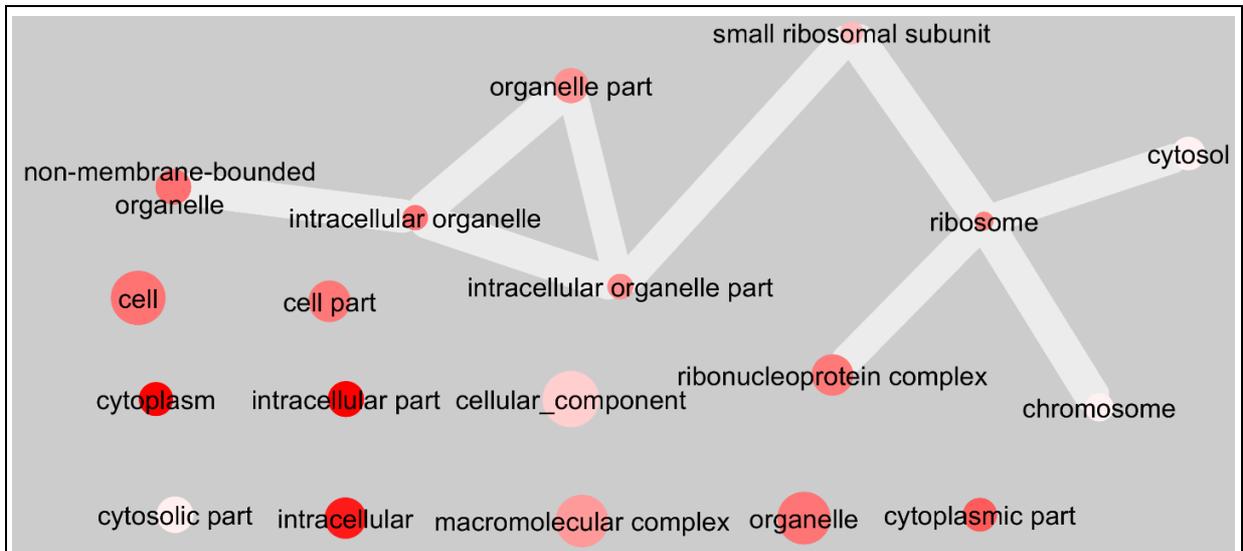
**Figure B. 6 Network representation of biological processes GO identities**

The figure shows the GO identities associated with biological processes as an interacting network. Each GO identity is represented as a node and interactions between nodes are represented by edges (lines). The colour intensity of each node is indicative of the  $-\log_{10} p$ -value, the darker the more significant the enrichment. All GO terms associated with biological process are depicted in A and B shows GO IDs associated with DNA, chromosome, transcription, gene, RNA, ribosome, translation, nucleotide and their first connecting neighbours.



**Figure B. 7 Network representation of Molecular function GO identities**

GO identities associated with molecular function are depicted in A as nodes with interactions being represented as connecting lines. The colour intensity of each node is indicative of the  $-\log_{10} p$ -value, the darker the more significant the enrichment. GO identities associated with DNA, chromosome, transcription, gene, RNA, ribosome, translation, nucleotide and their first connecting neighbours are shown in B.



**Figure B. 8 Network representation of cellular component GO identities**

GO identities associated with cellular components are depicted in A as nodes with interactions being represented as connecting lines. The colour intensity of each node is indicative of the  $-\log_{10} p$ -value, the darker the more significant the enrichment. GO terms associated with DNA, chromosome, transcription, gene, RNA, ribosome, translation, nucleotide and their first connecting neighbours are shown in B.



**B.4. High Stringency vs. Lower Stringency Protein Identification****Table B. 6 Proteins unique to high stringency protein identification**

UniProt Annotation	MSMEG Annotation	Gene Name	Protein Name	Gene Ontology
A0R7F9	MSMEG_6897	<i>rpsF</i>	30S ribosomal protein S6	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSH8	MSMEG_1484	<i>adk</i>	Adenylate monophosphate kinase	adenylate kinase activity; AMP salvage; ATP binding; cytoplasm
A0QSS4	MSMEG_1583	<i>groL2</i>	GroEL protein 2	ATP binding; cytoplasm; protein refolding
A0R2U8	MSMEG_5240	<i>fumC</i>	Fumarate hydratase class II	fumarate hydratase activity; fumarate metabolic process; tricarboxylic acid cycle; tricarboxylic acid cycle enzyme complex
A0QSF9	MSMEG_1465	<i>rplN</i>	50S ribosomal protein L14	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation

**Table B. 7 Proteins unique to lower stringency identification**

UniProt Annotation	MSMEG Annotation	Gene Name	Protein Name	Gene Ontology
A0QWQ0	MSMEG_3020		Uncharacterized protein	
A0R239	MSMEG_4976		Isochorismatase hydrolase	hydrolase activity
A0QPN2	MSMEG_0457		DNA gyrase subunit B-like protein	ATP binding; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change
A0QSZ6	MSMEG_1657	<i>trpS</i>	Tryptophan-tRNA ligase	ATP binding; cytoplasm; tryptophan-tRNA ligase activity; tryptophanyl-tRNA aminoacylation
A0QYD4	MSMEG_3619		Short chain dehydrogenase	oxidoreductase activity
A0QNM0	MSMEG_0092		Probable transcriptional regulatory protein	DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R2X8	MSMEG_5270	<i>cbs</i>	Cystathionine beta-synthase	adenyl nucleotide binding; cystathionine beta-synthase activity; cysteine biosynthetic process via cystathionine; cytoplasm
A0R7J0	MSMEG_6934	<i>trx</i>	Thioredoxin	cell redox homeostasis; glycerol ether metabolic process; protein disulfide oxidoreductase activity
A0QQB0	MSMEG_0690		Iron-sulfur cluster-binding protein	iron-sulfur cluster binding
A0QS41	MSMEG_1341		MaoC family protein	
A0QPI4	MSMEG_0409		Condensation domain protein	
A0R0Q9	MSMEG_4474		Acyl-CoA oxidase	acyl-CoA dehydrogenase activity; acyl-CoA oxidase activity; fatty acid beta-oxidation; flavin adenine dinucleotide binding; peroxisome
A0R7J4	MSMEG_6939		Chromosome partitioning protein ParA	

A0QYG0	MSMEG_3645		Uncharacterized protein	nuclease activity
A0QS81	MSMEG_1383	<i>nfo</i>	Endonuclease IV	deoxyribonuclease IV (phage-T4-induced) activity; DNA binding; DNA repair; zinc ion binding
A0R2Y5	MSMEG_5277	<i>echA8</i>	Enoyl-CoA hydratase/isomerase	enoyl-CoA hydratase activity; isomerase activity
A0QYB1	MSMEG_3596		AAA ATPase	
A0R2Q7	MSMEG_5199		Acetyl-CoA acetyltransferase	acetyl-CoA C-acyltransferase activity
A0QU00	MSMEG_2026		Short chain dehydrogenase	oxidoreductase activity
A0QRU1	MSMEG_1238		Type III restriction enzyme	ATP binding; DNA binding; Type I site-specific deoxyribonuclease activity
A0QS96	MSMEG_1398	<i>rpsL</i>	30S ribosomal protein S12	rRNA binding; small ribosomal subunit; structural constituent of ribosome; translation; tRNA binding
A0R4J1	MSMEG_5852	<i>purD</i>	Phosphoribosylamine-glycine ligase	de novo' IMP biosynthetic process; ATP binding; magnesium ion binding; manganese ion binding; phosphoribosylamine-glycine ligase activity; purine nucleobase biosynthetic process
A0QWY3	MSMEG_3106		Quinone oxidoreductase	NADPH:quinone reductase activity; zinc ion binding
A0QT19	MSMEG_1681		Endoribonuclease L-PSP superfamily protein	
A0QVZ3	MSMEG_2750	<i>ideR</i>	Iron-dependent repressor IdeR	DNA binding; sequence-specific DNA binding transcription factor activity; transition metal ion binding
A0QQS3	MSMEG_0859		Transcriptional regulator, TetR family	DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R7K1	MSMEG_6947	<i>dnaA</i>	Chromosomal replication initiator protein DnaA	ATP binding; cytoplasm; DNA replication initiation; DNA replication origin binding; regulation of DNA replication
A0R1G3	MSMEG_4742		Clavaldehyde dehydrogenase	oxidoreductase activity
A0QQP0	MSMEG_0824		Uncharacterized protein	
A0R2W9	MSMEG_5261	<i>mca</i>	Mycothiol S-conjugate amidase	hydrolase activity; mycothiol-dependent detoxification; mycothiol metabolic process; zinc ion binding
A0R7F7	MSMEG_6895	<i>rpsR2</i>	30S ribosomal protein S18 2	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QNY0	MSMEG_0203		IS1096, tnpR protein	
A0QX20	MSMEG_3143	<i>acnA</i>	Aconitate hydratase A (RNA-binding protein)	aconitate hydratase activity; iron-sulfur cluster binding; metal ion binding; tricarboxylic acid cycle
A0R3M3	MSMEG_5524	<i>sucD</i>	Succinyl-CoA ligase [ADP-forming] subunit alpha	ATP binding; ATP citrate synthase activity; cofactor binding; succinate-CoA ligase (ADP-forming) activity
A0QUX1	MSMEG_2367	<i>gatB</i>	amidotransferase subunit B	ATP binding; carbon-nitrogen ligase activity, with glutamine as amido-N-donor; translation
A0R5N7	MSMEG_6256	<i>asd</i>	Aspartate-semialdehyde dehydrogenase	'de novo' L-methionine biosynthetic process; aspartate-semialdehyde dehydrogenase activity; cytoplasm; diaminopimelate biosynthetic process; isoleucine biosynthetic process; lysine biosynthetic process via diaminopimelate; N-acetyl-gamma-glutamyl-phosphate reductase activity; NAD binding; NADP binding; threonine biosynthetic process

A0QP16	MSMEG_0239		O-acetylhomoserine/O-acetylserine sulfhydrylase	O-acetylhomoserine aminocarboxypropyltransferase activity; pyridoxal phosphate binding
A0QX15	MSMEG_3138	<i>trx</i>	<i>Mycobacterium smegmatis</i> str. MC2 155, complete genome (Thioredoxin) (Thioredoxin TrxB1)	cell redox homeostasis; glycerol ether metabolic process; protein disulfide oxidoreductase activity
A0QVE0	MSMEG_2541	<i>frr</i>	Ribosome-recycling factor	cytoplasm; translational termination

## Addendum C: Nucleoproteins of the exponential and stationary growth phase of *M. smegmatis*

### C.1. Exponential Phase

Table C. 1 Exponential Phase Protein Identifications

UniProt Annotation	MSMEG Annotation	RV homologue	Gene Name	Protein Names	Gene Ontology
<b>High Confidence Proteins</b>					
A0QNE0	MSMEG_0005		<i>gyrB</i>	DNA gyrase subunit B	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; magnesium ion binding
P48354	MSMEG_0006	Rv0006	<i>gyrA</i>	DNA gyrase subunit A	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; response to antibiotic
A0QNJ7	MSMEG_0067	Rv3876		Uncharacterized protein	
A0QNN0	MSMEG_0102	Rv0154c		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QP06	MSMEG_0229		<i>ilvD</i>	Dihydroxy-acid dehydratase	4 iron, 4 sulfur cluster binding; dihydroxy-acid dehydratase activity; isoleucine biosynthetic process; metal ion binding; valine biosynthetic process
A0QP32	MSMEG_0255	Rv0211	<i>pckG</i>	Phosphoenolpyruvate carboxykinase	cytoplasm; gluconeogenesis; GTP binding; manganese ion binding; phosphoenolpyruvate carboxykinase (GTP) activity
A0QPE7	MSMEG_0372	Rv0242c	<i>fabG4</i>	3-oxoacyl-acyl-carrier protein reductase FabG4	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity
A0QPH5	MSMEG_0400			Peptide synthetase	catalytic activity
Q3L891	MSMEG_0402		<i>mps2</i>	Non-ribosomal peptide synthetase	isomerase activity; oxidoreductase activity
Q3L885	MSMEG_0408		<i>pks</i>	Type I polyketide synthase	erythronolide synthase activity; oxidoreductase activity; phosphopantetheine binding; zinc ion binding
A0QPJ0	MSMEG_0415	Rv0245		NADH-fmn oxidoreductase	FMN binding; riboflavin reductase (NADPH) activity
A0QPN1	MSMEG_0456			DNA topoisomerase	ATP binding; chromosome; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change
A0QQC8	MSMEG_0709	Rv0350	<i>dnaK</i>	Heat shock protein 70	ATP binding; protein folding
A0QQF9	MSMEG_0741			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen

A0QQH2	MSMEG_0754	Rv0360c		Uncharacterized protein	
A0QQU2	MSMEG_0876			Short-chain dehydrogenase/reductase SDR	oxidoreductase activity
A0QQW8	MSMEG_0903	Rv0462	<i>lpdA</i>	Dihydrolipoyl dehydrogenase	cell redox homeostasis; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
A0QQX6	MSMEG_0911	Rv0467	<i>aceA</i>	Isocitrate lyase	carboxylic acid metabolic process; isocitrate lyase activity
A0QQX8	MSMEG_0913	Rv0469	<i>umaA</i>	Methoxy mycolic acid synthase 1	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
Q3I5Q7	MSMEG_0919	Rv0475		Uncharacterized protein	cell adhesion; cell surface; heparin binding; pathogenesis
A0QR19	MSMEG_0954	Rv0511	<i>hemD</i>	Uroporphyrinogen-III synthase	methyltransferase activity; tetrapyrrole biosynthetic process; uroporphyrinogen-III synthase activity
A0QR94	MSMEG_1033 MSMEG_2313		<i>nrdF2</i>	Ribonucleoside-diphosphate reductase subunit beta	deoxyribonucleoside diphosphate metabolic process; deoxyribonucleotide biosynthetic process; DNA replication; metal ion binding; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor; ribonucleoside-diphosphate reductase complex
P0CH37	MSMEG_1037 MSMEG_2317		<i>adhC2</i>	NADP-dependent alcohol dehydrogenase C 2	alcohol dehydrogenase (NADP+) activity; zinc ion binding
A0QRA8	MSMEG_1049 MSMEG_2329			Methyltransferase type 11	methyltransferase activity
A0QS44	MSMEG_1345	Rv0639	<i>nusG</i>	Transcription termination/antitermination protein	DNA-templated transcription, elongation; DNA-templated transcription, termination; regulation of DNA-templated transcription, elongation; transcription antitermination
A0QS45	MSMEG_1346	Rv0640	<i>rplK</i>	50S ribosomal protein L11	large ribosomal subunit rRNA binding; ribosome; structural constituent of ribosome; translation
A0QS49	MSMEG_1350		<i>omaA1</i>	Mycolic acid synthase UmaA	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
A0QSL0	MSMEG_1516			Thioredoxin reductase	oxidoreductase activity; phosphorelay signal transduction system
A0QSL3	MSMEG_1519	Rv3462c	<i>infA</i>	Translation initiation factor IF-1	cytoplasm; translation initiation factor activity
A0QSL8	MSMEG_1524	Rv3457c	<i>rpoA</i>	DNA-directed RNA polymerase subunit alpha	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QSU3	MSMEG_1602	Rv3411c	<i>guaB</i>	IMP dehydrogenase	adenyl nucleotide binding; GMP biosynthetic process; IMP dehydrogenase activity; metal ion binding
A0QSV0	MSMEG_1610	Rv3396c	<i>guaA</i>	Glutamine amidotransferase	ATP binding; glutamine metabolic process; GMP biosynthetic process; GMP synthase (glutamine-hydrolyzing) activity; pyrophosphatase activity
A0QSZ3	MSMEG_1654	Rv0066c	<i>icd2</i>	Isocitrate dehydrogenase (NADP)	isocitrate dehydrogenase (NADP+) activity; metal ion binding; tricarboxylic acid cycle
A0QT08	MSMEG_1670	Rv3318	<i>sdhA</i>	Succinate dehydrogenase flavoprotein subunit	electron transport chain; flavin adenine dinucleotide binding; succinate dehydrogenase activity; tricarboxylic acid cycle
A0QT17	MSMEG_1679		<i>amiB1</i>	Amidohydrolase AmiB1	hydrolase activity
A0QT18	MSMEG_1680			Uncharacterized protein	

A0QT19	MSMEG_1681			Endoribonuclease L-PSP superfamily protein	
A0QTE7	MSMEG_1813	Rv3280	<i>accD5</i>	Propionyl-CoA carboxylase beta chain 5	propionyl-CoA carboxylase activity
A0QTK2	MSMEG_1874	Rv3246c	<i>mtrA</i>	DNA-binding response regulator PhoP	cytoplasm; DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
P71533	MSMEG_1881	Rv3240c	<i>secA1</i>	Protein translocase subunit	ATP binding; cytoplasm; intracellular protein transmembrane transport; plasma membrane; protein import; protein targeting
A0QTP2	MSMEG_1914	Rv3223c	<i>sigH</i>	RNA polymerase sigma-H factor	DNA binding; DNA-templated transcription, initiation; sequence-specific DNA binding transcription factor activity; sigma factor activity
A0QTR5	MSMEG_1937	Rv3116 Rv3206c	<i>MoeB</i>	Molybdopterin biosynthesis protein	catalytic activity
A0QU52	MSMEG_2079			Alcohol dehydrogenase	oxidoreductase activity; zinc ion binding
A0QU53	MSMEG_2080	Rv3140	<i>fadE23</i>	Acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity
A0QU54	MSMEG_2081	Rv3139		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QUE0	MSMEG_2174			DNA helicase	ATP binding; ATP-dependent DNA helicase activity; DNA binding
P0CH00	MSMEG_2299 MSMEG_1019		<i>nrpE2</i>	Ribonucleoside-diphosphate reductase subunit alpha 2	ATP binding; DNA replication; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
A0QR89	MSMEG_2308 MSMEG_1028			Geranylgeranyl reductase	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor
A0QR91	MSMEG_2310 MSMEG_1030			Cyclohexanone monooxygenase	cyclohexanone monooxygenase activity
A0QUX1	MSMEG_2367	Rv3009c	<i>gatB</i>	Asp/Glu-ADT subunit B	ATP binding; carbon-nitrogen ligase activity, with glutamine as amido-N-donor; translation
A0QUX6	MSMEG_2372	Rv3003c	<i>ilvB</i>	Acetolactate synthase	acetolactate synthase activity; flavin adenine dinucleotide binding; isoleucine biosynthetic process; magnesium ion binding; thiamine pyrophosphate binding; valine biosynthetic process
A0QUY2	MSMEG_2378		<i>serA</i>	D-3-phosphoglycerate dehydrogenase	amino acid binding; L-serine biosynthetic process; NAD binding; phosphoglycerate dehydrogenase activity
A0QUZ0	MSMEG_2388	Rv2987c	<i>leuD</i>	Isopropylmalate isomerase	3-isopropylmalate dehydratase activity; 3-isopropylmalate dehydratase complex; leucine biosynthetic process
A0QV32	MSMEG_2430	Rv2916c	<i>ffh</i>	Signal recognition particle protein	7S RNA binding; GTPase activity; GTP binding; signal recognition particle; SRP-dependent cotranslational protein targeting to membrane
A0QVH9	MSMEG_2580	Rv2868c	<i>ispG</i>	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity; 4 iron, 4 sulfur cluster binding; iron ion binding; isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway; terpenoid biosynthetic process

A0QVM4	MSMEG_2625	Rv2841c	<i>nusA</i>	Transcription termination/antitermination protein	regulation of DNA-templated transcription, termination; RNA binding; sequence-specific DNA binding transcription factor activity
A0QVQ5	MSMEG_2656	Rv2783c	<i>pnp</i>	Polyribonucleotide nucleotidyltransferase	3'-5'-exoribonuclease activity; cytoplasm; magnesium ion binding; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity; RNA binding; RNA processing
A0QVU2	MSMEG_2695	Rv2744c		35 kDa protein	
Q59560	MSMEG_2723	Rv2737c	<i>recA</i>	Recombinase A	ATP binding; cytoplasm; damaged DNA binding; DNA-dependent ATPase activity; DNA recombination; DNA repair; single-stranded DNA binding; SOS response
A0QW02	MSMEG_2758	Rv2703	<i>rpoD</i>	RNA polymerase sigma factor SigA	cytoplasm; DNA binding; sequence-specific DNA binding transcription factor activity; sigma factor activity; transcription initiation from bacterial-type RNA polymerase promoter
A0QW71	MSMEG_2839			Transcriptional accessory protein	DNA binding; DNA repair; hydrolase activity, acting on ester bonds; RNA binding
A0QWT1	MSMEG_3053	Rv1390	<i>rpoZ</i>	DNA-directed RNA polymerase subunit omega	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QWT3	MSMEG_3055	Rv1392	<i>metK</i>	Methionine adenosyltransferase	ATP binding; cytoplasm; magnesium ion binding; methionine adenosyltransferase activity; one-carbon metabolic process; S-adenosylmethionine biosynthetic process
A0QWV0	MSMEG_3072	Rv1415	<i>ribAB</i>	Riboflavin biosynthesis protein	3,4-dihydroxy-2-butanone-4-phosphate synthase activity; GTP binding; GTP cyclohydrolase II activity; magnesium ion binding; manganese ion binding; riboflavin biosynthetic process; zinc ion binding
A0QWV9	MSMEG_3081	Rv1423	<i>whiA</i>	Putative sporulation transcription regulator	DNA binding; regulation of sporulation; regulation of transcription, DNA-templated; transcription, DNA-templated metal ion binding; transketolase activity
A0QWY0	MSMEG_3103	Rv1449c	<i>tkt</i>	Transketolase	
A0QWY3	MSMEG_3106	Rv1454c		Quinone oxidoreductase	NADPH:quinone reductase activity; zinc ion binding
A0QX20	MSMEG_3143	Rv1475c	<i>acnA</i>	Aconitate hydratase A	aconitate hydratase activity; iron-sulfur cluster binding; metal ion binding; tricarboxylic acid cycle
A0QX24	MSMEG_3147		<i>moxR</i>	ATPase, MoxR family protein	ATPase activity; ATP binding
P71534	MSMEG_3150	Rv1483	<i>fabG</i>	Beta-ketoacyl-ACP reductase	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity; fatty acid elongation; NADP binding
P42829	MSMEG_3151	Rv1484	<i>inhA</i>	NADH-dependent enoyl-ACP reductase	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process; response to antibiotic
A0QYD6	MSMEG_3621	Rv1854c	<i>ndh</i>	Membrane NADH dehydrogenase	flavin adenine dinucleotide binding; NADH dehydrogenase activity
A0QYE7	MSMEG_3632	Rv1844c	<i>gnd</i>	6-phosphogluconate dehydrogenase, decarboxylating	NADP binding; pentose-phosphate shunt; phosphogluconate dehydrogenase (decarboxylating) activity
A0QYF7	MSMEG_3642	Rv1832	<i>gcvP</i>	Glycine dehydrogenase	glycine decarboxylation via glycine cleavage system; glycine dehydrogenase (decarboxylating) activity; pyridoxal phosphate binding
A0QYS6	MSMEG_3770	Rv1658	<i>argG</i>	Argininosuccinate synthase	arginine biosynthetic process; argininosuccinate synthase activity; ATP binding; cytoplasm

A0QYZ2	MSMEG_3839	Rv1629	<i>polA</i>	DNA polymerase I	3'-5' exonuclease activity; DNA binding; DNA-dependent DNA replication; DNA-directed DNA polymerase activity
A0QZ11	MSMEG_3858	Rv2050	<i>rbpA</i>	RNA polymerase-binding protein	bacterial-type RNA polymerase core enzyme binding; positive regulation of transcription, DNA-templated; response to antibiotic; transcription, DNA-templated
A0QZ54	MSMEG_3902	Rv2115c	<i>mpa</i>	Proteasome-associated ATPase	ATPase activity; ATP binding; modification-dependent protein catabolic process; proteasomal protein catabolic process; proteasome-activating nucleotidase complex; protein unfolding
A0QZX6	MSMEG_4185	Rv2124c	<i>metH</i>	Methionine synthase	cobalamin binding; intracellular; methionine synthase activity; pteridine-containing compound metabolic process; S-adenosylmethionine-homocysteine S-methyltransferase activity; zinc ion binding
A0R029	MSMEG_4240	Rv2173		Polyprenyl synthetase	isoprenoid biosynthetic process; transferase activity
A0R033	MSMEG_4244	Rv2178c	<i>aroG</i>	Phospho-2-keto-3-deoxyheptonate aldolase	3-deoxy-7-phosphoheptulonate synthase activity; aromatic amino acid family biosynthetic process; chorismate biosynthetic process; manganese ion binding; plasma membrane
A0R061	MSMEG_4272	Rv2204c		Iron-sulfur cluster assembly accessory protein	iron-sulfur cluster assembly; iron-sulfur cluster binding; structural molecule activity
A0R069	MSMEG_4281	Rv2213	<i>pepA</i>	Probable cytosol aminopeptidase	aminopeptidase activity; cytoplasm; manganese ion binding; metalloexopeptidase activity
A0R095	MSMEG_4306	Rv2229c		Uncharacterized protein	
A0R0A1	MSMEG_4313			Glyoxalase/bleomycin resistance protein/dioxygenase	dioxygenase activity
A0R0B0	MSMEG_4323	Rv2241	<i>aceE</i>	Pyruvate dehydrogenase E1 component	glycolytic process; pyruvate dehydrogenase (acetyl-transferring) activity
A0R0B2	MSMEG_4325	Rv2243	<i>fabD</i>	Malonyl CoA-acyl carrier protein transacylase	[acyl-carrier-protein] S-malonyltransferase activity; fatty acid biosynthetic process
A0R0B5	MSMEG_4328	Rv2246	<i>kasB</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 1 KasA	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R0B6	MSMEG_4329	Rv2247	<i>accD6</i>	Acetyl/propionyl-CoA carboxylase	propionyl-CoA carboxylase activity
A0R0R9	MSMEG_4485	Rv2357c	<i>glyS</i>	Glycine-tRNA ligase	ATP binding; cytoplasm; glycine-tRNA ligase activity; glycyl-tRNA aminoacylation
A0R0T8	MSMEG_4504	Rv2373c	<i>dnaJ</i>	Chaperone protein	ATP binding; cytoplasm; DNA replication; protein folding; response to heat; zinc ion binding
A0R0W1	MSMEG_4527	Rv2391		Ferredoxin sulfite reductase	4 iron, 4 sulfur cluster binding; heme binding; metal ion binding; sulfite reductase (ferredoxin) activity
A0R196	MSMEG_4671	Rv2457c	<i>clpX</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein folding; zinc ion binding
A0R197	MSMEG_4672	Rv2460c	<i>clpP</i>	Endopeptidase Clp	cytoplasm; serine-type endopeptidase activity
A0R199	MSMEG_4674	Rv2462c	<i>tig</i>	Trigger factor	cell cycle; cell division; cytoplasm; peptidyl-prolyl cis-trans isomerase activity; protein folding; protein transport
A0R1C3	MSMEG_4700	Rv2477c		ATPase component of ABC transporter	ATPase activity; ATP binding

A0R221	MSMEG_4957	Rv1294		Homoserine dehydrogenase	amino acid binding; homoserine dehydrogenase activity; isoleucine biosynthetic process; methionine biosynthetic process; NADP binding; threonine biosynthetic process
A0R239	MSMEG_4976			Isochorismatase hydrolase	hydrolase activity
A0R2A4	MSMEG_5042	Rv1253	<i>deaD</i>	ATP-dependent RNA helicase DeaD	ATP binding; ATP-dependent RNA helicase activity; cellular response to cold; cytoplasm; ribosomal large subunit assembly; RNA binding; RNA catabolic process
A0R2B1	MSMEG_5049	Rv1248c	<i>kgd</i>	Alpha-ketoglutarate decarboxylase	2-hydroxy-3-oxoadipate synthase activity; 2-oxoglutarate decarboxylase activity; dihydrolipoyllysine-residue succinyltransferase activity; metal ion binding; oxoglutarate dehydrogenase (succinyl-transferring) activity; thiamine pyrophosphate binding; tricarboxylic acid cycle
A0R2U8	MSMEG_5240	Rv1098c	<i>fumC</i>	Fumarate hydratase class II	fumarate hydratase activity; fumarate metabolic process; tricarboxylic acid cycle; tricarboxylic acid cycle enzyme complex
A0R2X1	MSMEG_5263	Rv1080c	<i>greA</i>	Transcription elongation factor	DNA binding; regulation of DNA-templated transcription, elongation; transcription, DNA-templated
A0R2Y1	MSMEG_5273	Rv1074c	<i>fadA3</i>	Acetyl-CoA acetyltransferase	transferase activity, transferring acyl groups other than amino-acyl groups
A0R3C5	MSMEG_5423	Rv1020	<i>mfd</i>	Transcription-repair-coupling factor	ATP binding; cytoplasm; damaged DNA binding; helicase activity; regulation of transcription, DNA-templated; transcription-coupled nucleotide-excision repair, DNA damage recognition
A0R3D6	MSMEG_5435	Rv1013		Putative ligase	ligase activity
A0R3D9	MSMEG_5438	Rv1010	<i>ksgA</i>	Ribosomal RNA small subunit methyltransferase A	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase activity; cytoplasm; RNA binding; rRNA (adenine-N6,N6-)-dimethyltransferase activity
A0R3H1	MSMEG_5471	Rv0993	<i>galU</i>	UTP-glucose-1-phosphate uridylyltransferase	biosynthetic process; UTP:glucose-1-phosphate uridylyltransferase activity
A0R3L1	MSMEG_5512	Rv0958		Magnesium chelatase	ATP binding; regulation of transcription, DNA-templated
A0R3L4	MSMEG_5515	Rv0957	<i>purH</i>	Bifunctional purine biosynthesis protein	'de novo' IMP biosynthetic process; IMP cyclohydrolase activity; phosphoribosylaminoimidazolecarboxamide formyltransferase activity
A0R3M4	MSMEG_5525	Rv0951	<i>sucC</i>	Succinyl-CoA ligase	ATP binding; magnesium ion binding; manganese ion binding; succinate-CoA ligase (ADP-forming) activity; tricarboxylic acid cycle
A0R3Y5	MSMEG_5639	Rv0905	<i>echA6</i>	Enoyl-CoA hydratase	enoyl-CoA hydratase activity; isomerase activity
A0R417	MSMEG_5672	Rv0896	<i>gltA</i>	Citrate synthase	cellular carbohydrate metabolic process; citrate (Si)-synthase activity; cytoplasm; tricarboxylic acid cycle
A0R425	MSMEG_5680	Rv0887c		Glyoxalase family protein	dioxygenase activity
A0R441	MSMEG_5696	Rv0871		'Cold-shock' DNA-binding domain protein	cytoplasm; DNA binding; regulation of transcription, DNA-templated
A0R451	MSMEG_5706	Rv0861c		DNA or RNA helicase of superfamily protein II	ATP binding; ATP-dependent DNA helicase activity; DNA binding; nucleotide-excision repair

A0R461	MSMEG_5715			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0R4D0	MSMEG_5790	Rv3118 Rv0814c		Uncharacterized protein	
A0R4H6	MSMEG_5837			Glutathione peroxidase	glutathione peroxidase activity; response to oxidative stress
A0R4L1	MSMEG_5872	Rv0757	<i>phoP</i>	DNA-binding response regulator PhoP	DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R4S6	MSMEG_5937	Rv3534c	<i>bphI-2</i>	4-hydroxy-2-oxopentanoate aldolase 2	4-hydroxy-2-oxovalerate aldolase activity; aromatic compound catabolic process; manganese ion binding
A0R561	MSMEG_6077	Rv3583c	<i>carD</i>	RNA polymerase-binding transcription factor CarD	
A0R574	MSMEG_6091		<i>clpC1</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein metabolic process
A0R5D9	MSMEG_6157	Rv3646c	<i>topA</i>	DNA topoisomerase 1	DNA binding; DNA topoisomerase type I activity; DNA topological change; magnesium ion binding
A0R5E1	MSMEG_6159		<i>cspA</i>	Probable cold shock protein A	cytoplasm; DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R5M3	MSMEG_6242			Iron-containing alcohol dehydrogenase	1,3-propanediol dehydrogenase activity; alcohol dehydrogenase (NAD) activity; metal ion binding
A0R5N7	MSMEG_6256	Rv3708c	<i>asd</i>	Aspartate-semialdehyde dehydrogenase	'de novo' L-methionine biosynthetic process; aspartate-semialdehyde dehydrogenase activity; cytoplasm; diaminopimelate biosynthetic process; isoleucine biosynthetic process; lysine biosynthetic process via diaminopimelate; N-acetyl-gamma-glutamyl-phosphate reductase activity; NAD binding; NADP binding; threonine biosynthetic process
A0R5N8	MSMEG_6257	Rv3709c	<i>ask</i>	Aspartokinase	amino acid binding; aspartate kinase activity; lysine biosynthetic process via diaminopimelate; threonine biosynthetic process
A0R5Q2	MSMEG_6271	Rv3710	<i>leuA</i>	Alpha-isopropylmalate synthase	2-isopropylmalate synthase activity; leucine biosynthetic process
A0R616	MSMEG_6391	Rv3799c	<i>accD4</i>	Propionyl-CoA carboxylase beta chain 4	propionyl-CoA carboxylase activity
A0R617	MSMEG_6392	Rv3800c		Polyketide synthase	biosynthetic process; hydrolase activity, acting on ester bonds; phosphopantetheine binding; transferase activity
A0R618	MSMEG_6393		<i>fadD32</i>	Acyl-CoA synthase	ligase activity
A0R635	MSMEG_6410	Rv3818		Putative Rieske 2Fe-2S iron-sulfur protein	2 iron, 2 sulfur cluster binding; metal ion binding; oxidoreductase activity
A0R656	MSMEG_6431	Rv3849		Uncharacterized protein	sequence-specific DNA binding
A0R683	MSMEG_6458	Rv3858c	<i>gltD</i>	Glutamate synthase, small subunit	flavin adenine dinucleotide binding; glutamate biosynthetic process; iron-sulfur cluster binding; oxidoreductase activity, acting on the CH-NH2 group of donors, NAD or NADP as acceptor
A0R684	MSMEG_6459			Ferredoxin-dependent glutamate synthase 1	glutamate biosynthetic process; glutamate synthase (ferredoxin) activity

A0R6Q9	MSMEG_6638	Rv1133c	<i>metE</i>	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity; methionine biosynthetic process; zinc ion binding
Q9AFI5	MSMEG_6896	Rv0054	<i>ssb</i>	Single-stranded DNA-binding protein	DNA replication; single-stranded DNA binding
A0R7G6	MSMEG_6904	Rv0046c	<i>ino1</i>	Inositol-3-phosphate synthase	inositol-3-phosphate synthase activity; inositol biosynthetic process; phospholipid biosynthetic process

**Table C. 2 Exponential phase protein identification gene ontology enrichment data**

<b>GOID</b>	<b>Ontology</b>	<b>Term</b>	<b>Number of identifications</b>	<b>p-value</b>
GO:0000166	molecular function	nucleotide binding	41	5.92E-08
GO:0001882	molecular function	nucleoside binding	27	1.54E-05
GO:0001883	molecular function	purine nucleoside binding	27	1.54E-05
GO:0003674	molecular function	molecular function	129	8.88E-14
GO:0005488	molecular function	binding	86	1.40E-11
GO:0005524	molecular function	ATP binding	24	1.63E-04
GO:0017076	molecular function	purine nucleotide binding	27	1.54E-05
GO:0030554	molecular function	adenyl nucleotide binding	24	1.67E-04
GO:0032549	molecular function	ribonucleoside binding	27	1.54E-05
GO:0032550	molecular function	purine ribonucleoside binding	27	1.54E-05
GO:0032553	molecular function	ribonucleotide binding	28	4.03E-05
GO:0032555	molecular function	purine ribonucleotide binding	27	1.54E-05
GO:0032559	molecular function	adenyl ribonucleotide binding	24	1.63E-04
GO:0035639	molecular function	purine ribonucleoside triphosphate binding	27	1.54E-05
GO:0036094	molecular function	small molecule binding	45	2.98E-09
GO:0097159	molecular function	organic cyclic compound binding	65	8.01E-08
GO:0097367	molecular function	carbohydrate derivative binding	29	1.54E-05
GO:1901265	molecular function	nucleoside phosphate binding	41	5.92E-08
GO:1901363	molecular function	heterocyclic compound binding	65	8.01E-08
GO:0003676	molecular function	nucleic acid binding	30	6.06E-02
GO:0003824	molecular function	catalytic activity	106	2.50E-09
GO:0003916	molecular function	DNA topoisomerase activity	4	5.41E-04
GO:0003918	molecular function	DNA topoisomerase type II (ATP-hydrolyzing) activity	3	1.05E-02
GO:0008094	molecular function	DNA-dependent ATPase activity	6	2.20E-04
GO:0016462	molecular function	pyrophosphatase activity	13	8.93E-02
GO:0016817	molecular function	hydrolase activity, acting on acid anhydrides	13	9.64E-02
GO:0016818	molecular function	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	13	9.18E-02
GO:0061505	molecular function	DNA topoisomerase II activity	3	1.05E-02
GO:0006139	biological process	nucleobase-containing compound metabolic process	33	3.48E-03
GO:0006259	biological process	DNA metabolic process	13	1.78E-04
GO:0006265	biological process	DNA topological change	4	5.41E-04
GO:0006725	biological process	cellular aromatic compound metabolic process	37	1.50E-03
GO:0006807	biological process	nitrogen compound metabolic process	52	2.24E-06
GO:0006996	biological process	organelle organization	6	9.01E-05
GO:0008150	biological process	biological process	87	7.24E-12
GO:0008152	biological process	metabolic process	76	9.21E-12
GO:0009987	biological process	cellular process	77	7.24E-12

GO:0016043	biological process	cellular component organization	8	3.48E-03
GO:0034641	biological process	cellular nitrogen compound metabolic process	37	5.87E-03
GO:0043170	biological process	macromolecule metabolic process	32	2.07E-02
GO:0044237	biological process	cellular metabolic process	70	1.47E-11
GO:0044238	biological process	primary metabolic process	64	2.00E-10
GO:0044260	biological process	cellular macromolecule metabolic process	31	2.73E-02
GO:0046483	biological process	heterocycle metabolic process	36	2.72E-03
GO:0051276	biological process	chromosome organization	5	2.48E-04
GO:0071103	biological process	DNA conformation change	4	3.73E-03
GO:0071704	biological process	organic substance metabolic process	71	4.47E-11
GO:0071840	biological process	cellular component organization or biogenesis	8	2.17E-02
GO:0090304	biological process	nucleic acid metabolic process	28	1.04E-02
GO:1901360	biological process	organic cyclic compound metabolic process	37	1.90E-03
GO:0006260	biological process	DNA replication	7	3.79E-04
GO:0006261	biological process	DNA-dependent DNA replication	3	4.96E-02
GO:0009058	biological process	biosynthetic process	53	4.93E-08
GO:0044249	biological process	cellular biosynthetic process	48	1.16E-06
GO:1901576	biological process	organic substance biosynthetic process	49	9.23E-07
GO:0005622	cellular component	intracellular	34	3.64E-08
GO:0005623	cellular component	cell	37	9.47E-05
GO:0005694	cellular component	chromosome	3	1.77E-02
GO:0044424	cellular component	intracellular part	31	9.52E-09
GO:0044464	cellular component	cell part	36	1.29E-04
GO:0005737	cellular component	cytoplasm	29	3.70E-08
GO:0000287	molecular function	magnesium ion binding	7	9.64E-02
GO:0043167	molecular function	ion binding	34	4.63E-05
GO:0043169	molecular function	cation binding	30	2.14E-04
GO:0046872	molecular function	metal ion binding	30	1.78E-04
GO:0050896	biological process	response to stimulus	13	5.18E-03
GO:0016491	molecular function	oxidoreductase activity	41	8.86E-03
GO:0050662	molecular function	coenzyme binding	13	8.08E-02
GO:0006082	biological process	organic acid metabolic process	29	2.98E-09
GO:0006549	biological process	isoleucine metabolic process	4	1.57E-03
GO:0009081	biological process	branched-chain amino acid metabolic process	4	1.57E-03
GO:0009082	biological process	branched-chain amino acid biosynthetic process	4	1.57E-03
GO:0009097	biological process	isoleucine biosynthetic process	4	1.57E-03
GO:0016053	biological process	organic acid biosynthetic process	16	8.65E-06
GO:0044281	biological process	small molecule metabolic process	37	2.98E-09
GO:0044283	biological process	small molecule biosynthetic process	19	7.03E-06
GO:0044699	biological process	single-organism process	48	2.45E-07
GO:0044710	biological process	single-organism metabolic process	40	9.05E-08

GO:0044711	biological process	single-organism biosynthetic process	27	2.52E-06
GO:0044763	biological process	single-organism cellular process	45	1.18E-07
GO:1901564	biological process	organonitrogen compound metabolic process	25	1.01E-03
GO:1901566	biological process	organonitrogen compound biosynthetic process	21	1.30E-03
GO:1901605	biological process	alpha-amino acid metabolic process	12	7.34E-04
GO:1901607	biological process	alpha-amino acid biosynthetic process	10	5.41E-04
GO:0006520	biological process	cellular amino acid metabolic process	13	1.31E-03
GO:0008652	biological process	cellular amino acid biosynthetic process	10	7.34E-04
GO:0019752	biological process	carboxylic acid metabolic process	23	1.08E-06
GO:0043436	biological process	oxoacid metabolic process	23	1.96E-06
GO:0046394	biological process	carboxylic acid biosynthetic process	16	1.59E-06
GO:0030145	molecular function	manganese ion binding	6	2.46E-05
GO:0046914	molecular function	transition metal ion binding	15	3.17E-02
GO:0016740	molecular function	transferase activity	30	8.08E-02
GO:0019842	molecular function	vitamin binding	5	4.45E-02
GO:0008270	molecular function	zinc ion binding	9	6.53E-02
GO:0006457	biological process	protein folding	4	1.27E-02
GO:0008610	biological process	lipid biosynthetic process	8	9.94E-03
GO:0003723	molecular function	RNA binding	8	7.16E-02
GO:0051716	biological process	cellular response to stimulus	8	9.64E-02
GO:0016874	molecular function	ligase activity	11	4.52E-02
GO:0006091	biological process	generation of precursor metabolites and energy	8	1.30E-03
GO:0006099	biological process	tricarboxylic acid cycle	7	4.27E-06
GO:0009060	biological process	aerobic respiration	7	1.47E-05
GO:0015980	biological process	energy derivation by oxidation of organic compounds	7	1.48E-03
GO:0045333	biological process	cellular respiration	7	2.04E-04
GO:0055114	biological process	oxidation-reduction process	7	1.04E-02
GO:0004658	molecular function	propionyl-CoA carboxylase activity	3	2.61E-02
GO:0016421	molecular function	CoA carboxylase activity	3	4.96E-02
GO:0016885	molecular function	ligase activity, forming carbon-carbon bonds	3	8.05E-02
GO:0016597	molecular function	amino acid binding	3	9.64E-02
GO:0050661	molecular function	NADP binding	4	5.09E-02
GO:0043648	biological process	dicarboxylic acid metabolic process	6	2.31E-02
GO:0043650	biological process	dicarboxylic acid biosynthetic process	4	3.06E-02
GO:0006555	biological process	methionine metabolic process	3	4.96E-02
GO:0009086	biological process	methionine biosynthetic process	3	3.70E-02
GO:0006566	biological process	threonine metabolic process	3	5.79E-03
GO:0009088	biological process	threonine biosynthetic process	3	5.79E-03

## C.2. Stationary Phase

Table C. 3 Stationary Phase Protein Identification

UniProt Annotation	MSMEG Annotation	RV homologue	Gene Name	Protein Names	Gene Ontology
<b>High Confidence Proteins</b>					
A0QNE0	MSMEG_0005		<i>gyrB</i>	DNA gyrase subunit B	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; magnesium ion binding
P48354	MSMEG_0006	Rv0006	<i>gyrA</i>	DNA gyrase subunit A	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; response to antibiotic
A0QNF6	MSMEG_0024	Rv0009	<i>ppiA</i>	Peptidyl-prolyl cis-trans isomerase	peptidyl-prolyl cis-trans isomerase activity; protein folding
A0QNN0	MSMEG_0102	Rv0154c		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QNZ3	MSMEG_0216			3-hydroxyacyl-CoA dehydrogenase	oxidoreductase activity
A0QP06	MSMEG_0229		<i>ilvD</i>	Dihydroxy-acid dehydratase	4 iron, 4 sulfur cluster binding; dihydroxy-acid dehydratase activity; isoleucine biosynthetic process; metal ion binding; valine biosynthetic process
A0QPE7	MSMEG_0372	Rv0242c	<i>fabG4</i>	3-oxoacyl-acyl-carrier protein reductase	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity
A0QPH5	MSMEG_0400			Peptide synthetase	catalytic activity
Q3L891	MSMEG_0402		<i>mps2</i>	Linear gramicidin synthetase subunit D	isomerase activity; oxidoreductase activity
A0QPJ0	MSMEG_0415	Rv0245		NADH-fmn oxidoreductase	FMN binding; riboflavin reductase (NADPH) activity
A0QPN1	MSMEG_0456			DNA topoisomerase	ATP binding; chromosome; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change
A0QQB0	MSMEG_0690	Rv0338c		Iron-sulfur cluster-binding protein	iron-sulfur cluster binding
A0QQW8	MSMEG_0903	Rv0462	<i>lpdA</i>	Dihydrolipoyl dehydrogenase	cell redox homeostasis; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
A0QQX6	MSMEG_0911	Rv0467	<i>aceA</i>	Isocitrate lyase	carboxylic acid metabolic process; isocitrate lyase activity
A0QQX8	MSMEG_0913	Rv0469	<i>umaA</i>	Methoxy mycolic acid synthase 1	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
Q3I5Q7	MSMEG_0919	Rv0475		Uncharacterized protein	cell adhesion; cell surface; heparin binding; pathogenesis
A0QR19	MSMEG_0954	Rv0511	<i>hemD</i>	Uroporphyrinogen-III synthase	methyltransferase activity; tetrapyrrole biosynthetic process; uroporphyrinogen-III synthase activity

P0CH37	MSMEG_1037 MSMEG_2317		<i>adhC2</i>	NADP-dependent alcohol dehydrogenase C 2	alcohol dehydrogenase (NADP+) activity; zinc ion binding
A0QRA8	MSMEG_1049 MSMEG_2329			Methyltransferase type 11	methyltransferase activity
A0QRB8	MSMEG_1060			Uncharacterized protein	
A0QS44	MSMEG_1345	Rv0639	<i>nusG</i>	Transcription termination/antitermination protein	DNA-templated transcription, elongation; DNA-templated transcription, termination; regulation of DNA-templated transcription, elongation; transcription antitermination
A0QS46	MSMEG_1347	Rv0641	<i>rplA</i>	50S ribosomal protein L1	large ribosomal subunit; regulation of translation; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QS62	MSMEG_1364	Rv0651	<i>rplJ</i>	50S ribosomal protein L10	large ribosomal subunit rRNA binding; ribosome; ribosome biogenesis; structural constituent of ribosome; translation
A0QSB1	MSMEG_1416	Rv0688		FAD-dependent pyridine nucleotide-disulfide oxidoreductase	cell redox homeostasis; ferredoxin-NAD+ reductase activity; flavin adenine dinucleotide binding
A0QSD3	MSMEG_1438	Rv0703	<i>rplW</i>	50S ribosomal protein L23	nucleotide binding; ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSD8	MSMEG_1443	Rv0708	<i>rplP</i>	50S ribosomal protein L16	ribosome; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QSE0	MSMEG_1445	Rv0710	<i>rpsQ</i>	30S ribosomal protein S17	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSF9	MSMEG_1465	Rv0714	<i>rplN</i>	50S ribosomal protein L14	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSG1	MSMEG_1467	Rv0716	<i>rplE</i>	50S ribosomal protein L5	ribosome; rRNA binding; structural constituent of ribosome; translation; tRNA binding
A0QSG3	MSMEG_1469	Rv0718	<i>rpsH</i>	30S ribosomal protein S8	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSG8	MSMEG_1474	Rv0723	<i>rplO</i>	50S ribosomal protein L15	large ribosomal subunit; rRNA binding; structural constituent of ribosome; translation
A0QSL0	MSMEG_1516			Thioredoxin reductase	oxidoreductase activity; phosphorelay signal transduction system
A0QSL3	MSMEG_1519	Rv3462c	<i>infA</i>	Translation initiation factor IF-1	cytoplasm; translation initiation factor activity
A0QSL6	MSMEG_1522	Rv3459c	<i>rpsK</i>	30S ribosomal protein S11	ribosome; rRNA binding; structural constituent of ribosome; translation
A0QSP8	MSMEG_1556	Rv3443c	<i>rplM</i>	50S ribosomal protein L13	ribosome; structural constituent of ribosome; translation
A0QSS3	MSMEG_1582	Rv3418c	<i>groS</i>	GroES protein	ATP binding; cytoplasm; protein folding
A0QSS4	MSMEG_1583	Rv3417c	<i>groL2</i>	GroEL protein 2	ATP binding; cytoplasm; protein refolding
A0QSU3	MSMEG_1602	Rv3411c	<i>guaB</i>	IMP dehydrogenase	adenyl nucleotide binding; GMP biosynthetic process; IMP dehydrogenase activity; metal ion binding
A0QSU4	MSMEG_1603	Rv3410c		Uncharacterized oxidoreductase	oxidoreductase activity
A0QSV0	MSMEG_1610	Rv3396c	<i>guaA</i>	Glutamine amidotransferase	ATP binding; glutamine metabolic process; GMP biosynthetic process; GMP synthase (glutamine-hydrolyzing) activity; pyrophosphatase activity

A0QSZ3	MSMEG_1654	Rv0066c	<i>icd2</i>	Isocitrate dehydrogenase	isocitrate dehydrogenase (NADP+) activity; metal ion binding; tricarboxylic acid cycle
A0QT08	MSMEG_1670	Rv3318	<i>sdhA</i>	Succinate dehydrogenase flavoprotein subunit	electron transport chain; flavin adenine dinucleotide binding; succinate dehydrogenase activity; tricarboxylic acid cycle
A0QT17	MSMEG_1679		<i>amiB1</i>	Amidohydrolase AmiB1	hydrolase activity
A0QT18	MSMEG_1680			Uncharacterized protein	
A0QT19	MSMEG_1681			Endoribonuclease L-PSP superfamily protein	
A0QT20	MSMEG_1682			Flavin-containing monooxygenase FMO	monooxygenase activity
A0QTE7	MSMEG_1813	Rv3280	<i>accD5</i>	Propionyl-CoA carboxylase beta chain 5	propionyl-CoA carboxylase activity
A0QTK2	MSMEG_1874	Rv3246c	<i>mtrA</i>	DNA-binding response regulator	cytoplasm; DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
A0QTK6	MSMEG_1878	Rv3241c		Sigma 54 modulation protein/ribosomal protein S30EA	primary metabolic process; ribosome
P71533	MSMEG_1881	Rv3240c	<i>secA1</i>	Protein translocase subunit	ATP binding; cytoplasm; intracellular protein transmembrane transport; plasma membrane; protein import; protein targeting
A0QTP2	MSMEG_1914	Rv3223c	<i>sigH</i>	RNA polymerase sigma factor SigH	DNA binding; DNA-templated transcription, initiation; sequence-specific DNA binding transcription factor activity; sigma factor activity
A0QTR5	MSMEG_1937	Rv3116 Rv3206c		Molybdopterin biosynthesis protein MoeB	catalytic activity
A0QU00	MSMEG_2026			Short-chain dehydrogenase/reductase SDR	oxidoreductase activity
A0QU52	MSMEG_2079			Alcohol dehydrogenase	oxidoreductase activity; zinc ion binding
A0QU53	MSMEG_2080	Rv3140	<i>fadE23</i>	Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity
A0QU54	MSMEG_2081	Rv3139		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QUE0	MSMEG_2174			DNA helicase	ATP binding; ATP-dependent DNA helicase activity; DNA binding
P0CH00	MSMEG_2299 MSMEG_1019		<i>nrdE2</i>	Ribonucleoside-diphosphate reductase subunit alpha 2	ATP binding; DNA replication; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
A0QR91	MSMEG_2310 MSMEG_1030			Cyclohexanone monooxygenase	cyclohexanone monooxygenase activity
A0QRA5	MSMEG_2326 MSMEG_1046			ABC transporter ATP-binding protein	ATPase activity; ATP binding
A0QUV6	MSMEG_2351	Rv3029c	<i>etfB</i>	Electron transfer flavoprotein beta subunit	electron carrier activity
A0QUV7	MSMEG_2352	Rv3028c	<i>etfA</i>	Electron transfer flavoprotein, alpha subunit	electron carrier activity; flavin adenine dinucleotide binding
A0QUX6	MSMEG_2372	Rv3003c	<i>ilvB</i>	Acetolactate synthase	acetolactate synthase activity; flavin adenine dinucleotide binding; isoleucine biosynthetic process; magnesium ion binding; thiamine pyrophosphate binding; valine biosynthetic process

A0QUY2	MSMEG_2378		<i>serA</i>	D-3-phosphoglycerate dehydrogenase	amino acid binding; L-serine biosynthetic process; NAD binding; phosphoglycerate dehydrogenase activity
A0QUY9	MSMEG_2387	Rv2988c	<i>leuC</i>	3-isopropylmalate dehydratase large subunit	3-isopropylmalate dehydratase activity; 4 iron, 4 sulfur cluster binding; leucine biosynthetic process; metal ion binding
A0QV14	MSMEG_2412		<i>pyc</i>	Pyruvate carboxylase	ATP binding; biotin carboxylase activity; gluconeogenesis; metal ion binding; pyruvate carboxylase activity
A0QV32	MSMEG_2430	Rv2916c	<i>ffh</i>	Signal recognition particle protein	7S RNA binding; GTPase activity; GTP binding; signal recognition particle; SRP-dependent cotranslational protein targeting to membrane
A0QV37	MSMEG_2435	Rv2909c	<i>rpsP</i>	30S ribosomal protein S16	ribosome; structural constituent of ribosome; translation
A0QV42	MSMEG_2440	Rv2904c	<i>rplS</i>	50S ribosomal protein L19	ribosome; structural constituent of ribosome; translation
A0QVB9	MSMEG_2520	Rv2889c	<i>tsf</i>	Elongation factor Ts (EF-Ts)	cytoplasm; translation elongation factor activity
A0QVH9	MSMEG_2580	Rv2868c	<i>ispG</i>	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (flavodoxin)	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity; 4 iron, 4 sulfur cluster binding; iron ion binding; isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway; terpenoid biosynthetic process
A0QVM4	MSMEG_2625	Rv2841c	<i>nusA</i>	Transcription termination/antitermination protein	regulation of DNA-templated transcription, termination; RNA binding; sequence-specific DNA binding transcription factor activity
A0QVM7	MSMEG_2628	Rv2839c	<i>infB</i>	Translation initiation factor IF-2	cytoplasm; GTPase activity; GTP binding; translation initiation factor activity
A0QVQ5	MSMEG_2656	Rv2783c	<i>pnp</i>	Polyribonucleotide nucleotidyltransferase	3'-5'-exoribonuclease activity; cytoplasm; magnesium ion binding; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity; RNA binding; RNA processing
A0QVU2	MSMEG_2695	Rv2744c		35 kDa protein	
Q59560	MSMEG_2723	Rv2737c	<i>recA</i>	Recombinase A	ATP binding; cytoplasm; damaged DNA binding; DNA-dependent ATPase activity; DNA recombination; DNA repair; single-stranded DNA binding; SOS response
A0QW02	MSMEG_2758	Rv2703	<i>rpoD</i>	RNA polymerase sigma factor SigA	cytoplasm; DNA binding; sequence-specific DNA binding transcription factor activity; sigma factor activity; transcription initiation from bacterial-type RNA polymerase promoter
A0QW25	MSMEG_2782	Rv2676c		Uncharacterized protein	
A0QW71	MSMEG_2839			Transcriptional accessory protein	DNA binding; DNA repair; hydrolase activity, acting on ester bonds; RNA binding
A0QWS8	MSMEG_3050	Rv1388	<i>mihF</i>	Integration host factor	nucleic acid binding
A0QWT1	MSMEG_3053	Rv1390	<i>rpoZ</i>	DNA-directed RNA polymerase subunit omega	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QWT3	MSMEG_3055	Rv1392	<i>metK</i>	Methionine adenosyltransferase	ATP binding; cytoplasm; magnesium ion binding; methionine adenosyltransferase activity; one-carbon metabolic process; S-adenosylmethionine biosynthetic process
A0QWV0	MSMEG_3072	Rv1415	<i>ribAB</i>	Riboflavin biosynthesis protein	3,4-dihydroxy-2-butanone-4-phosphate synthase activity; GTP binding; GTP cyclohydrolase II activity; magnesium ion binding;

					manganese ion binding; riboflavin biosynthetic process; zinc ion binding
A0QWV9	MSMEG_3081	Rv1423	<i>whiA</i>	Putative sporulation transcription regulator	DNA binding; regulation of sporulation; regulation of transcription, DNA-templated; transcription, DNA-templated
A0QX20	MSMEG_3143	Rv1475c	<i>acnA</i>	Aconitate hydratase A	aconitate hydratase activity; iron-sulfur cluster binding; metal ion binding; tricarboxylic acid cycle
A0QX24	MSMEG_3147		<i>moxR</i>	ATPase, MoxR family protein	ATPase activity; ATP binding
P42829	MSMEG_3151	Rv1484	<i>inhA</i>	NADH-dependent enoyl-ACP reductase	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process; response to antibiotic
A0QXA3	MSMEG_3227	Rv1617	<i>pyk</i>	Pyruvate kinase	glycolytic process; magnesium ion binding; potassium ion binding; pyruvate kinase activity
A0QYB0	MSMEG_3595			Uncharacterized protein	
A0QYB1	MSMEG_3596			AAA ATPase	
A0QYD6	MSMEG_3621	Rv1854c	<i>ndh</i>	Membrane NADH dehydrogenase	flavin adenine dinucleotide binding; NADH dehydrogenase activity
A0QYE7	MSMEG_3632	Rv1844c	<i>gnd</i>	6-phosphogluconate dehydrogenase, decarboxylating	NADP binding; pentose-phosphate shunt; phosphogluconate dehydrogenase (decarboxylating) activity
A0QYF7	MSMEG_3642	Rv1832	<i>gcvP</i>	Glycine dehydrogenase	glycine decarboxylation via glycine cleavage system; glycine dehydrogenase (decarboxylating) activity; pyridoxal phosphate binding
A0QYG9	MSMEG_3654	Rv1821	<i>secA2</i>	Protein translocase subunit	ATP binding; cytoplasm; intracellular protein transmembrane transport; plasma membrane; protein import; protein targeting
A0QYP9	MSMEG_3738	Rv1713	<i>engA</i>	GTP-binding protein EngA	GTP binding; ribosome biogenesis; transferase activity
A0QYS6	MSMEG_3770	Rv1658	<i>argG</i>	Argininosuccinate synthase	arginine biosynthetic process; argininosuccinate synthase activity; ATP binding; cytoplasm
A0QYW6	MSMEG_3811	Rv1636		Universal stress protein family protein, putative	cytoplasm; response to stress
A0QYZ2	MSMEG_3839	Rv1629	<i>polA</i>	DNA polymerase I	3'-5' exonuclease activity; DNA binding; DNA-dependent DNA replication; DNA-directed DNA polymerase activity
A0QZ11	MSMEG_3858	Rv2050	<i>rbpA</i>	RNA polymerase-binding protein	bacterial-type RNA polymerase core enzyme binding; positive regulation of transcription, DNA-templated; response to antibiotic; transcription, DNA-templated
A0QZ54	MSMEG_3902	Rv2115c	<i>mpa</i>	Proteasome-associated ATPase	ATPase activity; ATP binding; modification-dependent protein catabolic process; proteasomal protein catabolic process; proteasome-activating nucleotidase complex; protein unfolding
A0QZ83	MSMEG_3932	Rv2031c	<i>hspX</i>	Heat shock protein hspX	
A0QZ96	MSMEG_3945			Universal stress protein family protein	response to stress
A0QZA1	MSMEG_3950			Universal stress protein	response to stress
A0QZX6	MSMEG_4185	Rv2124c	<i>methH</i>	Methionine synthase	cobalamin binding; intracellular; methionine synthase activity; pteridine-containing compound metabolic process; S-adenosylmethionine-homocysteine S-methyltransferase activity; zinc ion binding

A0R012	MSMEG_4222	Rv2150c	<i>ftsZ</i>	Cell division protein FtsZ	barrier septum assembly; cell division site; cytoplasm; FtsZ-dependent cytokinesis; GTPase activity; GTP binding; protein complex; protein polymerization
A0R029	MSMEG_4240	Rv2173		Polyprenyl synthetase	isoprenoid biosynthetic process; transferase activity
A0R0A1	MSMEG_4313			Glyoxalase/bleomycin resistance protein/dioxygenase	dioxygenase activity
A0R0B0	MSMEG_4323	Rv2241	<i>aceE</i>	Pyruvate dehydrogenase E1 component	glycolytic process; pyruvate dehydrogenase (acetyl-transferring) activity
A0R0B2	MSMEG_4325	Rv2243	<i>fabD</i>	Malonyl CoA-acyl carrier protein transacylase	[acyl-carrier-protein] S-malonyltransferase activity; fatty acid biosynthetic process
A0R0B3	MSMEG_4326	Rv2244	<i>acpM</i>	Meromycolate extension acyl carrier protein	ACP phosphopantetheine attachment site binding involved in fatty acid biosynthetic process; cytoplasm
A0R0B4	MSMEG_4327	Rv2245	<i>kasA</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 1	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R0B5	MSMEG_4328	Rv2246	<i>kasB</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 1	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R0B6	MSMEG_4329	Rv2247	<i>accD6</i>	Acetyl/propionyl-CoA carboxylase (Beta subunit)	propionyl-CoA carboxylase activity
A0R0C7	MSMEG_4340	Rv2259	<i>adhE2</i>	Zinc-dependent alcohol dehydrogenase	oxidoreductase activity; zinc ion binding
A0R0I3	MSMEG_4396			Isochorismatase hydrolase	hydrolase activity
A0R0I8	MSMEG_4401			Phosphonoacetaldehyde hydrolase	hydrolase activity
A0R0Q9	MSMEG_4474			Acyl-CoA oxidase	acyl-CoA dehydrogenase activity; acyl-CoA oxidase activity; fatty acid beta-oxidation; flavin adenine dinucleotide binding; peroxisome
A0R0R9	MSMEG_4485	Rv2357c	<i>glyS</i>	Glycine-tRNA ligase	ATP binding; cytoplasm; glycine-tRNA ligase activity; glycyl-tRNA aminoacylation
A0R0T8	MSMEG_4504	Rv2373c	<i>dnaJ</i>	Chaperone protein DnaJ	ATP binding; cytoplasm; DNA replication; protein folding; response to heat; zinc ion binding
A0R102	MSMEG_4571	Rv2412	<i>rpsT</i>	30S ribosomal protein S20	ribosome; rRNA binding; structural constituent of ribosome; translation
A0R171	MSMEG_4646	Rv2455c		Pyruvate synthase	2-oxoglutarate synthase activity
A0R196	MSMEG_4671	Rv2457c	<i>clpX</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein folding; zinc ion binding
A0R197	MSMEG_4672	Rv2460c	<i>clpP</i>	Endopeptidase Clp	cytoplasm; serine-type endopeptidase activity
A0R199	MSMEG_4674	Rv2462c	<i>tig</i>	Trigger factor	cell cycle; cell division; cytoplasm; peptidyl-prolyl cis-trans isomerase activity; protein folding; protein transport
A0R1C3	MSMEG_4700	Rv2477c		ABC-transporter protein, ATP binding component	ATPase activity; ATP binding
A0R1V9	MSMEG_4891	Rv2428	<i>ahpC</i>	Alkyl hydroperoxide reductase subunit C	peroxidase activity; peroxiredoxin activity
A0R218	MSMEG_4954	Rv1297	<i>rho</i>	Transcription termination factor	ATP binding; DNA-templated transcription, termination; helicase activity; regulation of transcription, DNA-templated; RNA binding; RNA-dependent ATPase activity

A0R221	MSMEG_4957	Rv1294		Homoserine dehydrogenase	amino acid binding; homoserine dehydrogenase activity; isoleucine biosynthetic process; methionine biosynthetic process; NADP binding; threonine biosynthetic process
A0R239	MSMEG_4976			Isochorismatase hydrolase	hydrolase activity
A0R248	MSMEG_4985	Rv1284		Carbonic anhydrase	carbonate dehydratase activity; zinc ion binding
A0R2A4	MSMEG_5042	Rv1253	<i>deaD</i>	ATP-dependent RNA helicase DeaD	ATP binding; ATP-dependent RNA helicase activity; cellular response to cold; cytoplasm; ribosomal large subunit assembly; RNA binding; RNA catabolic process
A0R2T0	MSMEG_5222	Rv1112	<i>ychF</i>	Ribosome-binding ATPase	ATPase activity; ATP binding; GTP binding; ribosomal large subunit binding; ribosome binding
A0R2V2	MSMEG_5244		<i>devR</i>	Two component transcriptional regulatory protein devr	DNA binding; phosphorelay signal transduction system; sequence-specific DNA binding transcription factor activity; transcription, DNA-templated
A0R2V4	MSMEG_5246			Uncharacterized protein	oxidoreductase activity
A0R2V6	MSMEG_5248	Rv1094	<i>desA2</i>	Acyl-acyl-carrier protein desaturase DesA2	acyl-[acyl-carrier-protein] desaturase activity; fatty acid metabolic process
A0R2X1	MSMEG_5263	Rv1080c	<i>greA</i>	Transcription elongation factor GreA	DNA binding; regulation of DNA-templated transcription, elongation; transcription, DNA-templated
A0R2Y1	MSMEG_5273	Rv1074c	<i>fadA3</i>	Acetyl-CoA acetyltransferase	transferase activity, transferring acyl groups other than amino-acyl groups
A0R3B8	MSMEG_5415	Rv1023	<i>eno</i>	Enolase	cell surface; extracellular region; glycolytic process; magnesium ion binding; phosphopyruvate hydratase activity; phosphopyruvate hydratase complex
A0R3C5	MSMEG_5423	Rv1020	<i>mfd</i>	Transcription-repair-coupling factor	ATP binding; cytoplasm; damaged DNA binding; helicase activity; regulation of transcription, DNA-templated; transcription-coupled nucleotide-excision repair, DNA damage recognition
A0R3D2	MSMEG_5431	Rv1015c	<i>rplY</i>	50S ribosomal protein L25	5S rRNA binding; ribosome; structural constituent of ribosome; translation
A0R3D6	MSMEG_5435	Rv1013		Putative ligase	ligase activity
A0R3D9	MSMEG_5438	Rv1010	<i>ksgA</i>	Ribosomal RNA small subunit methyltransferase A	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase activity; cytoplasm; RNA binding; rRNA (adenine-N6,N6-)-dimethyltransferase activity
A0R3H1	MSMEG_5471	Rv0993	<i>galU</i>	UTP-glucose-1-phosphate uridylyltransferase	biosynthetic process; UTP:glucose-1-phosphate uridylyltransferase activity
A0R3L1	MSMEG_5512	Rv0958		Magnesium chelatase	ATP binding; regulation of transcription, DNA-templated
A0R3L4	MSMEG_5515	Rv0957	<i>purH</i>	Bifunctional purine biosynthesis protein	'de novo' IMP biosynthetic process; IMP cyclohydrolase activity; phosphoribosylaminoimidazolecarboxamide formyltransferase activity
A0R3M3	MSMEG_5524	Rv0952	<i>sucD</i>	Succinyl-CoA ligase [ADP-forming] subunit alpha	ATP binding; ATP citrate synthase activity; cofactor binding; succinate-CoA ligase (ADP-forming) activity
A0R3M4	MSMEG_5525	Rv0951	<i>sucC</i>	Succinyl-CoA ligase [ADP-forming] subunit beta	ATP binding; magnesium ion binding; manganese ion binding; succinate-CoA ligase (ADP-forming) activity; tricarboxylic acid cycle

A0R3Y5	MSMEG_5639	Rv0905	<i>echA6</i>	Enoyl-CoA hydratase	enoyl-CoA hydratase activity; isomerase activity
A0R417	MSMEG_5672	Rv0896	<i>gltA</i>	Citrate synthase	cellular carbohydrate metabolic process; citrate (Si)-synthase activity; cytoplasm; tricarboxylic acid cycle
A0R441	MSMEG_5696	Rv0871		'Cold-shock' DNA-binding domain protein	cytoplasm; DNA binding; regulation of transcription, DNA-templated
A0R461	MSMEG_5715			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0R4C9	MSMEG_5789	Rv3117 Rv0815c		Putative thiosulfate sulfurtransferase	thiosulfate sulfurtransferase activity
A0R4D0	MSMEG_5790	Rv3118 Rv0814c		Uncharacterized protein	
A0R4H6	MSMEG_5837			Glutathione peroxidase	glutathione peroxidase activity; response to oxidative stress
A0R4J1	MSMEG_5852	Rv0772	<i>purD</i>	Phosphoribosylamine-glycine ligase	de novo' IMP biosynthetic process; ATP binding; magnesium ion binding; manganese ion binding; phosphoribosylamine-glycine ligase activity; purine nucleobase biosynthetic process
A0R4L1	MSMEG_5872	Rv0757	<i>phoP</i>	DNA-binding response regulator PhoP	DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R574	MSMEG_6091		<i>clpC1</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein metabolic process
A0R577	MSMEG_6094	Rv3598c	<i>lysS</i>	Lysine-tRNA ligase	ATP binding; cytoplasm; lysine-tRNA ligase activity; lysyl-tRNA aminoacylation; magnesium ion binding; nucleic acid binding
A0R5D9	MSMEG_6157	Rv3646c	<i>topA</i>	DNA topoisomerase 1	DNA binding; DNA topoisomerase type I activity; DNA topological change; magnesium ion binding
A0R5E1	MSMEG_6159		<i>cspA</i>	Probable cold shock protein A	cytoplasm; DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R5H1	MSMEG_6189	Rv3676		Crp/Fnr family transcriptional regulator	DNA binding; intracellular; sequence-specific DNA binding transcription factor activity; transcription, DNA-templated
A0R5H5	MSMEG_6193	Rv3679		Anion-transporting ATPase	arsenite-transmembrane transporting ATPase activity; ATP binding
A0R5M3	MSMEG_6242			Iron-containing alcohol dehydrogenase	1,3-propanediol dehydrogenase activity; alcohol dehydrogenase (NAD) activity; metal ion binding
A0R5N7	MSMEG_6256	Rv3708c	<i>asd</i>	Aspartate-semialdehyde dehydrogenase	'de novo' L-methionine biosynthetic process; aspartate-semialdehyde dehydrogenase activity; cytoplasm; diaminopimelate biosynthetic process; isoleucine biosynthetic process; lysine biosynthetic process via diaminopimelate; N-acetyl-gamma-glutamyl-phosphate reductase activity; NAD binding; NADP binding; threonine biosynthetic process
A0R5N8	MSMEG_6257	Rv3709c	<i>ask</i>	Aspartokinase	amino acid binding; aspartate kinase activity; lysine biosynthetic process via diaminopimelate; threonine biosynthetic process
A0R616	MSMEG_6391	Rv3799c	<i>accD4</i>	Propionyl-CoA carboxylase beta chain	propionyl-CoA carboxylase activity
A0R617	MSMEG_6392	Rv3800c		Polyketide synthase	biosynthetic process; hydrolase activity, acting on ester bonds; phosphopantetheine binding; transferase activity
A0R618	MSMEG_6393		<i>fadD32</i>	Acyl-CoA synthase	ligase activity

A0R628	MSMEG_6403	Rv3808c	<i>glfT2</i>	Galactofuranosyl transferase GlfT2	capsule polysaccharide biosynthetic process; cell wall macromolecule biosynthetic process; cell wall organization; membrane; metal ion binding; transferase activity; transferase activity, transferring glycosyl groups
A0R635	MSMEG_6410	Rv3818		Putative Rieske 2Fe-2S iron-sulfur protein	2 iron, 2 sulfur cluster binding; metal ion binding; oxidoreductase activity
A0R684	MSMEG_6459			Ferredoxin-dependent glutamate synthase 1	glutamate biosynthetic process; glutamate synthase (ferredoxin) activity
Q9AFI5	MSMEG_6896	Rv0054	<i>ssb</i>	Single-stranded DNA-binding protein	DNA replication; single-stranded DNA binding
A0R7F9	MSMEG_6897	Rv0053	<i>rpsF</i>	30S ribosomal protein S6	ribosome; rRNA binding; structural constituent of ribosome; translation
A0R7G6	MSMEG_6904	Rv0046c	<i>ino1</i>	Inositol-3-phosphate synthase	inositol-3-phosphate synthase activity; inositol biosynthetic process; phospholipid biosynthetic process
A0R7G8	MSMEG_6907	Rv0044c		Mmcl protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen

**Table C. 4 Stationary phase protein identification gene ontology enrichment data**

<b>GOID</b>	<b>Ontology</b>	<b>Term</b>	<b>Number of identifications</b>	<b>p-value</b>
GO:0000166	molecular function	nucleotide binding	53	3.69E-11
GO:0001882	molecular function	nucleoside binding	37	6.88E-09
GO:0001883	molecular function	purine nucleoside binding	37	6.88E-09
GO:0003674	molecular function	molecular function	161	1.89E-16
GO:0005488	molecular function	binding	108	1.49E-14
GO:0005524	molecular function	ATP binding	32	7.82E-07
GO:0017076	molecular function	purine nucleotide binding	37	6.88E-09
GO:0030554	molecular function	adenyl nucleotide binding	32	8.18E-07
GO:0032549	molecular function	ribonucleoside binding	37	6.88E-09
GO:0032550	molecular function	purine ribonucleoside binding	37	6.88E-09
GO:0032553	molecular function	ribonucleotide binding	38	3.83E-08
GO:0032555	molecular function	purine ribonucleotide binding	37	6.88E-09
GO:0032559	molecular function	adenyl ribonucleotide binding	32	7.82E-07
GO:0035639	molecular function	purine ribonucleoside triphosphate binding	37	6.88E-09
GO:0036094	molecular function	small molecule binding	55	3.17E-11
GO:0097159	molecular function	organic cyclic compound binding	88	6.04E-13
GO:0097367	molecular function	carbohydrate derivative binding	39	1.30E-08
GO:1901265	molecular function	nucleoside phosphate binding	53	3.69E-11
GO:1901363	molecular function	heterocyclic compound binding	88	6.04E-13
GO:0003676	molecular function	nucleic acid binding	46	2.67E-05
GO:0003824	molecular function	catalytic activity	115	1.62E-04
GO:0003916	molecular function	DNA topoisomerase activity	4	1.24E-03
GO:0003918	molecular function	DNA topoisomerase type II (ATP-hydrolyzing) activity	3	1.89E-02
GO:0008094	molecular function	DNA-dependent ATPase activity	5	8.97E-03
GO:0016462	molecular function	pyrophosphatase activity	18	6.20E-03
GO:0016817	molecular function	hydrolase activity, acting on acid anhydrides	18	7.07E-03
GO:0016818	molecular function	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	18	6.45E-03
GO:0017111	molecular function	nucleoside-triphosphatase activity	17	9.99E-03
GO:0061505	molecular function	DNA topoisomerase II activity	3	1.89E-02
GO:0006139	biological process	nucleobase-containing compound metabolic process	37	1.58E-02
GO:0006259	biological process	DNA metabolic process	11	2.36E-02
GO:0006265	biological process	DNA topological change	4	1.24E-03
GO:0006725	biological process	cellular aromatic compound metabolic process	39	3.71E-02
GO:0006807	biological process	nitrogen compound metabolic process	66	1.58E-08
GO:0006996	biological process	organelle organization	6	2.61E-04
GO:0008150	biological process	biological process	108	1.49E-14
GO:0008152	biological process	metabolic process	88	1.76E-11

GO:000987	biological process	cellular process	95	1.49E-14
GO:0016043	biological process	cellular component organization	9	3.06E-03
GO:0034641	biological process	cellular nitrogen compound metabolic process	55	5.65E-07
GO:0043170	biological process	macromolecule metabolic process	48	6.63E-06
GO:0044237	biological process	cellular metabolic process	83	2.51E-12
GO:0044238	biological process	primary metabolic process	80	4.01E-13
GO:0044260	biological process	cellular macromolecule metabolic process	47	8.16E-06
GO:0046483	biological process	heterocycle metabolic process	40	1.86E-02
GO:0051276	biological process	chromosome organization	5	6.92E-04
GO:0071103	biological process	DNA conformation change	4	8.97E-03
GO:0071704	biological process	organic substance metabolic process	84	1.32E-11
GO:0071840	biological process	cellular component organization or biogenesis	11	1.05E-03
GO:1901360	biological process	organic cyclic compound metabolic process	40	2.75E-02
GO:0006260	biological process	DNA replication	6	1.16E-02
GO:0006261	biological process	DNA-dependent DNA replication	3	7.32E-02
GO:0009058	biological process	biosynthetic process	64	3.93E-09
GO:0009059	biological process	macromolecule biosynthetic process	39	5.44E-05
GO:0034645	biological process	cellular macromolecule biosynthetic process	39	3.90E-05
GO:0044249	biological process	cellular biosynthetic process	60	1.67E-08
GO:1901576	biological process	organic substance biosynthetic process	61	1.58E-08
GO:0005575	cellular component	cellular component	61	8.77E-06
GO:0005622	cellular component	intracellular	57	1.65E-20
GO:0005623	cellular component	cell	60	7.65E-13
GO:0005694	cellular component	chromosome	3	2.95E-02
GO:0043226	cellular component	organelle	21	5.20E-14
GO:0043228	cellular component	non-membrane-bounded organelle	20	2.81E-13
GO:0043229	cellular component	intracellular organelle	21	4.26E-14
GO:0043232	cellular component	intracellular non-membrane-bounded organelle	20	2.81E-13
GO:0044424	cellular component	intracellular part	52	1.11E-20
GO:0044464	cellular component	cell part	58	2.51E-12
GO:0005737	cellular component	cytoplasm	50	1.21E-20
GO:0000287	molecular function	magnesium ion binding	11	1.24E-03
GO:0043167	molecular function	ion binding	32	2.46E-02
GO:0043169	molecular function	cation binding	28	5.15E-02
GO:0046872	molecular function	metal ion binding	28	4.56E-02
GO:0050896	biological process	response to stimulus	16	1.24E-03
GO:0006457	biological process	protein folding	6	1.14E-04
GO:0006082	biological process	organic acid metabolic process	28	1.19E-06
GO:0006549	biological process	isoleucine metabolic process	4	3.83E-03
GO:0009081	biological process	branched-chain amino acid metabolic process	4	3.83E-03

GO:0009082	biological process	branched-chain amino acid biosynthetic process	4	3.83E-03
GO:0009097	biological process	isoleucine biosynthetic process	4	3.83E-03
GO:0016053	biological process	organic acid biosynthetic process	11	6.72E-02
GO:0044281	biological process	small molecule metabolic process	35	1.03E-05
GO:0044283	biological process	small molecule biosynthetic process	14	4.57E-02
GO:0044699	biological process	single-organism process	50	5.71E-05
GO:0044710	biological process	single-organism metabolic process	38	3.12E-04
GO:0044711	biological process	single-organism biosynthetic process	22	2.27E-02
GO:0044763	biological process	single-organism cellular process	49	2.58E-06
GO:1901564	biological process	organonitrogen compound metabolic process	39	1.10E-08
GO:1901566	biological process	organonitrogen compound biosynthetic process	33	2.49E-08
GO:1901605	biological process	alpha-amino acid metabolic process	10	5.95E-02
GO:1901607	biological process	alpha-amino acid biosynthetic process	8	5.28E-02
GO:0006520	biological process	cellular amino acid metabolic process	12	3.28E-02
GO:0008652	biological process	cellular amino acid biosynthetic process	8	6.43E-02
GO:0019752	biological process	carboxylic acid metabolic process	23	4.31E-05
GO:0043436	biological process	oxoacid metabolic process	23	7.91E-05
GO:0046394	biological process	carboxylic acid biosynthetic process	11	2.95E-02
GO:0044271	biological process	cellular nitrogen compound biosynthetic process	38	2.17E-03
GO:0010467	biological process	gene expression	33	2.22E-03
GO:0006353	biological process	DNA-templated transcription, termination	3	1.89E-02
GO:0005840	cellular component	ribosome	17	2.64E-11
GO:0015934	cellular component	large ribosomal subunit	3	2.95E-02
GO:0030529	cellular component	intracellular ribonucleoprotein complex	18	2.51E-12
GO:0032991	cellular component	macromolecular complex	21	4.59E-07
GO:0044444	cellular component	cytoplasmic part	20	2.43E-12
GO:1990904	cellular component	ribonucleoprotein complex	18	2.51E-12
GO:0003723	molecular function	RNA binding	23	7.08E-13
GO:0019843	molecular function	rRNA binding	13	2.59E-09
GO:0003735	molecular function	structural constituent of ribosome	16	1.97E-10
GO:0005198	molecular function	structural molecule activity	16	3.40E-10
GO:0006412	biological process	translation	18	1.71E-09
GO:0006518	biological process	peptide metabolic process	18	6.88E-09
GO:0019538	biological process	protein metabolic process	20	6.25E-08
GO:0043043	biological process	peptide biosynthetic process	18	3.54E-09
GO:0043603	biological process	cellular amide metabolic process	18	4.36E-08
GO:0043604	biological process	amide biosynthetic process	18	4.89E-09
GO:0044267	biological process	cellular protein metabolic process	19	1.71E-07
GO:0006163	biological process	purine nucleotide metabolic process	7	4.56E-02
GO:0009123	biological process	nucleoside monophosphate metabolic process	7	4.89E-02

GO:0009126	biological process	purine nucleoside monophosphate metabolic process	7	1.21E-02
GO:0009150	biological process	purine ribonucleotide metabolic process	7	3.12E-02
GO:0009161	biological process	ribonucleoside monophosphate metabolic process	7	3.61E-02
GO:0009167	biological process	purine ribonucleoside monophosphate metabolic process	7	1.21E-02
GO:0009259	biological process	ribonucleotide metabolic process	7	8.02E-02
GO:0019693	biological process	ribose phosphate metabolic process	8	4.57E-02
GO:0072521	biological process	purine-containing compound metabolic process	7	7.19E-02
GO:0016874	molecular function	ligase activity	14	9.99E-03
GO:0006091	biological process	generation of precursor metabolites and energy	8	6.03E-03
GO:0006099	biological process	tricarboxylic acid cycle	5	4.13E-03
GO:0009060	biological process	aerobic respiration	5	8.97E-03
GO:0045333	biological process	cellular respiration	5	4.47E-02
GO:0004658	molecular function	propionyl-CoA carboxylase activity	3	4.06E-02
GO:0016421	molecular function	CoA carboxylase activity	3	7.32E-02
GO:0016885	molecular function	ligase activity, forming carbon-carbon bonds	4	1.18E-02
GO:0006886	biological process	intracellular protein transport	3	4.06E-02
GO:0008104	biological process	protein localization	4	3.27E-02
GO:0015031	biological process	protein transport	4	2.09E-02
GO:0033036	biological process	macromolecule localization	4	4.57E-02
GO:0034613	biological process	cellular protein localization	3	5.47E-02
GO:0045184	biological process	establishment of protein localization	4	2.67E-02
GO:0046907	biological process	intracellular transport	3	4.06E-02
GO:0051641	biological process	cellular localization	3	5.47E-02
GO:0051649	biological process	establishment of localization in cell	3	4.06E-02
GO:0070727	biological process	cellular macromolecule localization	3	5.47E-02
GO:1902580	biological process	single-organism cellular localization	3	5.47E-02
GO:1902582	biological process	single-organism intracellular transport	3	4.06E-02
GO:0006605	biological process	protein targeting	3	4.06E-02
GO:0031406	molecular function	carboxylic acid binding	4	2.67E-02
GO:0006090	biological process	pyruvate metabolic process	5	1.29E-02
GO:0032787	biological process	monocarboxylic acid metabolic process	9	1.90E-02
GO:0005525	molecular function	GTP binding	6	9.99E-03
GO:0019001	molecular function	guanyl nucleotide binding	6	9.99E-03
GO:0032561	molecular function	guanyl ribonucleotide binding	6	9.99E-03
GO:0006950	biological process	response to stress	9	6.52E-02
GO:0006566	biological process	threonine metabolic process	3	1.06E-02
GO:0009088	biological process	threonine biosynthetic process	3	1.06E-02

## C.3. Exponential and Stationary Phase Proteins

Table C. 5 Proteins identified in both the exponential and stationary growth phase using NP-MS

UniProt	MSMEG Annotation	RV homologue	Gene Name	Protein Names	Gene Ontology
A0QNE0	MSMEG_0005	Rv0005	<i>gyrB</i>	DNA gyrase subunit B	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; magnesium ion binding
P48354	MSMEG_0006	Rv0006	<i>gyrA</i>	DNA gyrase subunit A	ATP binding; chromosome; cytoplasm; DNA binding; DNA-dependent DNA replication; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change; response to antibiotic
A0QNN0	MSMEG_0102	Rv0154c		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QP06	MSMEG_0229	Rv0189c	<i>ilvD</i>	Dihydroxy-acid dehydratase	4 iron, 4 sulfur cluster binding; dihydroxy-acid dehydratase activity; isoleucine biosynthetic process; metal ion binding; valine biosynthetic process
A0QPE7	MSMEG_0372	Rv0242c	<i>fabG4</i>	3-oxoacyl-acyl-carrier protein reductase	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity
A0QPH5	MSMEG_0400			Peptide synthetase	catalytic activity
Q3L891	MSMEG_0402		<i>mps2</i>	Linear gramicidin synthetase subunit D	isomerase activity; oxidoreductase activity
A0QPJ0	MSMEG_0415	Rv0245		Flavin reductase-like, FMN-binding protein	FMN binding; riboflavin reductase (NADPH) activity
A0QPN1	MSMEG_0456			DNA topoisomerase	ATP binding; chromosome; DNA binding; DNA topoisomerase type II (ATP-hydrolyzing) activity; DNA topological change
A0QQW8	MSMEG_0903	Rv0462	<i>lpdA</i>	Dihydrolipoyl dehydrogenase	cell redox homeostasis; dihydrolipoyl dehydrogenase activity; flavin adenine dinucleotide binding
A0QQX6	MSMEG_0911	Rv0467	<i>aceA</i>	Isocitrate lyase	carboxylic acid metabolic process; isocitrate lyase activity
A0QQX8	MSMEG_0913	Rv0469	<i>umaA</i>	Mycolic acid synthase	cyclopropane-fatty-acyl-phospholipid synthase activity; lipid biosynthetic process
Q3I5Q7	MSMEG_0919	Rv0475		Uncharacterized protein	cell adhesion; cell surface; heparin binding; pathogenesis
A0QR19	MSMEG_0954	Rv0511	<i>hemD</i>	Uroporphyrinogen-III synthase	methyltransferase activity; tetrapyrrole biosynthetic process; uroporphyrinogen-III synthase activity
P0CH00	MSMEG_1019 MSMEG_2299	Rv3051c	<i>nrdE2</i>	Ribonucleoside-diphosphate reductase subunit alpha 2	ATP binding; DNA replication; ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor
A0QR91	MSMEG_1030 MSMEG_2310	Rv3049c		Cyclohexanone monooxygenase	cyclohexanone monooxygenase activity
P0CH37	MSMEG_1037 MSMEG_2317	Rv3045	<i>adhC2</i>	NADP-dependent alcohol dehydrogenase C 2	alcohol dehydrogenase (NADP+) activity; zinc ion binding

A0QRA8	MSMEG_1049 MSMEG_2329	Rv3038c		Methyltransferase type 11	methyltransferase activity
A0QS44	MSMEG_1345	Rv0639	<i>nusG</i>	Transcription termination/antitermination protein	DNA-templated transcription, elongation; DNA-templated transcription, termination; regulation of DNA-templated transcription, elongation; transcription antitermination
A0QSL0	MSMEG_1516			Thioredoxin reductase	oxidoreductase activity; phosphorelay signal transduction system
A0QSL3	MSMEG_1519	Rv3462c	<i>infA</i>	Translation initiation factor IF-1	cytoplasm; translation initiation factor activity
A0QSU3	MSMEG_1602	Rv3411c	<i>guaB</i>	Inosine-5'-monophosphate dehydrogenase	adenyl nucleotide binding; GMP biosynthetic process; IMP dehydrogenase activity; metal ion binding
A0QSV0	MSMEG_1610	Rv3396c	<i>guaA</i>	Glutamine amidotransferase	ATP binding; glutamine metabolic process; GMP biosynthetic process; GMP synthase (glutamine-hydrolyzing) activity; pyrophosphatase activity
A0QSZ3	MSMEG_1654	Rv0066c	<i>icd2</i>	Isocitrate dehydrogenase	isocitrate dehydrogenase (NADP+) activity; metal ion binding; tricarboxylic acid cycle
A0QT08	MSMEG_1670	Rv3318	<i>sdhA</i>	Succinate dehydrogenase flavoprotein subunit	electron transport chain; flavin adenine dinucleotide binding; succinate dehydrogenase activity; tricarboxylic acid cycle
A0QT17	MSMEG_1679		<i>amiB1</i>	Amidohydrolase AmiB1	hydrolase activity
A0QT18	MSMEG_1680			Uncharacterized protein	
A0QT19	MSMEG_1681			Endoribonuclease L-PSP superfamily protein	
A0QTE7	MSMEG_1813	Rv3280	<i>accD5</i>	Propionyl-CoA carboxylase beta chain	propionyl-CoA carboxylase activity
A0QTK2	MSMEG_1874	Rv3246c	<i>mtrA</i>	DNA-binding response regulator	cytoplasm; DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
P71533	MSMEG_1881	Rv3240c	<i>secA1</i>	Protein translocase subunit SecA 1	ATP binding; cytoplasm; intracellular protein transmembrane transport; plasma membrane; protein import; protein targeting
A0QTP2	MSMEG_1914	Rv3223c	<i>sigH</i>	RNA polymerase sigma-H factor	DNA binding; DNA-templated transcription, initiation; sequence-specific DNA binding transcription factor activity; sigma factor activity
A0QTR5	MSMEG_1937	Rv3116 Rv3206c		Molybdopterin biosynthesis protein MoeB	catalytic activity
A0QU52	MSMEG_2079			Alcohol dehydrogenase	oxidoreductase activity; zinc ion binding
A0QU53	MSMEG_2080	Rv3140	<i>fadE23</i>	Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity
A0QU54	MSMEG_2081	Rv3139		Putative acyl-CoA dehydrogenase	acyl-CoA dehydrogenase activity; flavin adenine dinucleotide binding
A0QUE0	MSMEG_2174			DNA helicase	ATP binding; ATP-dependent DNA helicase activity; DNA binding
A0QUX6	MSMEG_2372	Rv3003c	<i>ilvB</i>	Acetolactate synthase	acetolactate synthase activity; flavin adenine dinucleotide binding; isoleucine biosynthetic process; magnesium ion binding; thiamine pyrophosphate binding; valine biosynthetic process

A0QUY2	MSMEG_2378	Rv2996c	<i>serA</i>	D-3-phosphoglycerate dehydrogenase	amino acid binding; L-serine biosynthetic process; NAD binding; phosphoglycerate dehydrogenase activity
A0QV32	MSMEG_2430	Rv2916c	<i>ffh</i>	Signal recognition particle protein	7S RNA binding; GTPase activity; GTP binding; signal recognition particle; SRP-dependent cotranslational protein targeting to membrane
A0QVH9	MSMEG_2580	Rv2868c	<i>ispG</i>	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase activity; 4 iron, 4 sulfur cluster binding; iron ion binding; isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway; terpenoid biosynthetic process
A0QVM4	MSMEG_2625	Rv2841c	<i>nusA</i>	Transcription termination/antitermination protein	regulation of DNA-templated transcription, termination; RNA binding; sequence-specific DNA binding transcription factor activity
A0QVQ5	MSMEG_2656	Rv2783c	<i>pnp</i>	Polyribonucleotide nucleotidyltransferase	3'-5'-exoribonuclease activity; cytoplasm; magnesium ion binding; mRNA catabolic process; polyribonucleotide nucleotidyltransferase activity; RNA binding; RNA processing
A0QVU2	MSMEG_2695	Rv2744c		35 kDa protein	
Q59560	MSMEG_2723	Rv2737c	<i>recA</i>	Recombinase A	ATP binding; cytoplasm; damaged DNA binding; DNA-dependent ATPase activity; DNA recombination; DNA repair; single-stranded DNA binding; SOS response
A0QW02	MSMEG_2758	Rv2703	<i>rpoD</i>	RNA polymerase sigma factor SigA	cytoplasm; DNA binding; sequence-specific DNA binding transcription factor activity; sigma factor activity; transcription initiation from bacterial-type RNA polymerase promoter
A0QW71	MSMEG_2839			Transcriptional accessory protein	DNA binding; DNA repair; hydrolase activity, acting on ester bonds; RNA binding
A0QWT1	MSMEG_3053	Rv1390	<i>rpoZ</i>	DNA-directed RNA polymerase subunit omega	DNA binding; DNA-directed RNA polymerase activity; transcription, DNA-templated
A0QWT3	MSMEG_3055	Rv1392	<i>metK</i>	Methionine adenosyltransferase	ATP binding; cytoplasm; magnesium ion binding; methionine adenosyltransferase activity; one-carbon metabolic process; S-adenosylmethionine biosynthetic process
A0QWV0	MSMEG_3072	Rv1415	<i>ribAB</i>	Riboflavin biosynthesis protein	3,4-dihydroxy-2-butanone-4-phosphate synthase activity; GTP binding; GTP cyclohydrolase II activity; magnesium ion binding; manganese ion binding; riboflavin biosynthetic process; zinc ion binding
A0QWV9	MSMEG_3081	Rv1423	<i>whiA</i>	Putative sporulation transcription regulator	DNA binding; regulation of sporulation; regulation of transcription, DNA-templated; transcription, DNA-templated
A0QX20	MSMEG_3143	Rv1475c	<i>acnA</i>	Aconitate hydratase A (RNA-binding protein)	aconitate hydratase activity; iron-sulfur cluster binding; metal ion binding; tricarboxylic acid cycle
A0QX24	MSMEG_3147	Rv1479	<i>moxR</i>	ATPase, MoxR family protein	ATPase activity; ATP binding
P42829	MSMEG_3151	Rv1484	<i>inhA</i>	NADH-dependent enoyl-ACP reductase	enoyl-[acyl-carrier-protein] reductase (NADH) activity; fatty acid biosynthetic process; response to antibiotic
A0QYD6	MSMEG_3621	Rv1854c	<i>ndh</i>	NADH dehydrogenase	flavin adenine dinucleotide binding; NADH dehydrogenase activity
A0QYE7	MSMEG_3632	Rv1844c	<i>gnd</i>	6-phosphogluconate dehydrogenase, decarboxylating	NADP binding; pentose-phosphate shunt; phosphogluconate dehydrogenase (decarboxylating) activity

A0QYF7	MSMEG_3642	Rv1832	<i>gcvP</i>	Glycine dehydrogenase	glycine decarboxylation via glycine cleavage system; glycine dehydrogenase (decarboxylating) activity; pyridoxal phosphate binding
A0QYS6	MSMEG_3770	Rv1658	<i>argG</i>	Argininosuccinate synthase	arginine biosynthetic process; argininosuccinate synthase activity; ATP binding; cytoplasm
A0QYZ2	MSMEG_3839	Rv1629	<i>polA</i>	DNA polymerase I	3'-5' exonuclease activity; DNA binding; DNA-dependent DNA replication; DNA-directed DNA polymerase activity
A0QZ11	MSMEG_3858	Rv2050	<i>rbpA</i>	RNA polymerase-binding protein	bacterial-type RNA polymerase core enzyme binding; positive regulation of transcription, DNA-templated; response to antibiotic; transcription, DNA-templated
A0QZ54	MSMEG_3902	Rv2115c	<i>mpa</i>	Proteasome-associated ATPase	ATPase activity; ATP binding; modification-dependent protein catabolic process; proteasomal protein catabolic process; proteasome-activating nucleotidase complex; protein unfolding
A0QZX6	MSMEG_4185	Rv2124c	<i>methH</i>	Methionine synthase	cobalamin binding; intracellular; methionine synthase activity; pteridine-containing compound metabolic process; S-adenosylmethionine-homocysteine S-methyltransferase activity; zinc ion binding
A0R029	MSMEG_4240	Rv2173		Polyprenyl synthetase	isoprenoid biosynthetic process; transferase activity
A0R0A1	MSMEG_4313			Glyoxalase/bleomycin resistance protein/dioxygenase	dioxygenase activity
A0R0B0	MSMEG_4323	Rv2241	<i>aceE</i>	Pyruvate dehydrogenase E1 component	glycolytic process; pyruvate dehydrogenase (acetyl-transferring) activity
A0R0B2	MSMEG_4325	Rv2243	<i>fabD</i>	Malonyl CoA-acyl carrier protein transacylase	[acyl-carrier-protein] S-malonyltransferase activity; fatty acid biosynthetic process
A0R0B5	MSMEG_4328	Rv2246	<i>kasB</i>	3-oxoacyl-(Acyl-carrier-protein) synthase 1	3-oxoacyl-[acyl-carrier-protein] synthase activity
A0R0B6	MSMEG_4329	Rv2247	<i>accD6</i>	Propionyl-CoA carboxylase beta chain	propionyl-CoA carboxylase activity
A0R0R9	MSMEG_4485	Rv2357c	<i>glyS</i>	Glycine-tRNA ligase	ATP binding; cytoplasm; glycine-tRNA ligase activity; glycyl-tRNA aminoacylation
A0R0T8	MSMEG_4504	Rv2373c	<i>dnaJ</i>	Chaperone protein DnaJ	ATP binding; cytoplasm; DNA replication; protein folding; response to heat; zinc ion binding
A0R196	MSMEG_4671	Rv2457c	<i>clpX</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein folding; zinc ion binding
A0R197	MSMEG_4672	Rv2460c	<i>clpP</i>	ATP-dependent Clp protease proteolytic subunit	cytoplasm; serine-type endopeptidase activity
A0R199	MSMEG_4674	Rv2462c	<i>tig</i>	Trigger factor (TF)	cell cycle; cell division; cytoplasm; peptidyl-prolyl cis-trans isomerase activity; protein folding; protein transport
A0R1C3	MSMEG_4700	Rv2477c		ABC-transporter protein, ATP binding component	ATPase activity; ATP binding
A0R221	MSMEG_4957	Rv1294		Homoserine dehydrogenase	amino acid binding; homoserine dehydrogenase activity; isoleucine biosynthetic process; methionine biosynthetic process; NADP binding; threonine biosynthetic process
A0R239	MSMEG_4976			Isochorismatase hydrolase	hydrolase activity
A0R2A4	MSMEG_5042	Rv1253	<i>deaD</i>	ATP-dependent RNA helicase	ATP binding; ATP-dependent RNA helicase activity; cellular response to cold; cytoplasm; ribosomal large subunit assembly; RNA binding; RNA catabolic process

A0R2X1	MSMEG_5263	Rv1080c	<i>greA</i>	Transcription elongation factor GreA	DNA binding; regulation of DNA-templated transcription, elongation; transcription, DNA-templated
A0R2Y1	MSMEG_5273	Rv1074c	<i>fadA3</i>	Acetyl-CoA acetyltransferase	transferase activity, transferring acyl groups other than amino-acyl groups
A0R3C5	MSMEG_5423	Rv1020	<i>mfd</i>	Transcription-repair-coupling factor	ATP binding; cytoplasm; damaged DNA binding; helicase activity; regulation of transcription, DNA-templated; transcription-coupled nucleotide-excision repair, DNA damage recognition
A0R3D6	MSMEG_5435	Rv1013		Putative ligase	ligase activity
A0R3D9	MSMEG_5438	Rv1010	<i>ksgA</i>	Ribosomal RNA small subunit methyltransferase A	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase activity; cytoplasm; RNA binding; rRNA (adenine-N6,N6-)-dimethyltransferase activity
A0R3H1	MSMEG_5471	Rv0993	<i>galU</i>	UTP-glucose-1-phosphate uridylyltransferase	biosynthetic process; UTP:glucose-1-phosphate uridylyltransferase activity
A0R3L1	MSMEG_5512	Rv0958		Magnesium chelatase	ATP binding; regulation of transcription, DNA-templated
A0R3L4	MSMEG_5515	Rv0957	<i>purH</i>	Bifunctional purine biosynthesis protein PurH	'de novo' IMP biosynthetic process; IMP cyclohydrolase activity; phosphoribosylaminoimidazolecarboxamide formyltransferase activity
A0R3M4	MSMEG_5525	Rv0951	<i>sucC</i>	Succinyl-CoA ligase	ATP binding; magnesium ion binding; manganese ion binding; succinate-CoA ligase (ADP-forming) activity; tricarboxylic acid cycle
A0R3Y5	MSMEG_5639	Rv0905	<i>echA6</i>	Enoyl-CoA hydratase	enoyl-CoA hydratase activity; isomerase activity
A0R417	MSMEG_5672	Rv0896	<i>gltA</i>	Citrate synthase	cellular carbohydrate metabolic process; citrate (Si)-synthase activity; cytoplasm; tricarboxylic acid cycle
A0R441	MSMEG_5696	Rv0871		'Cold-shock' DNA-binding domain protein	cytoplasm; DNA binding; regulation of transcription, DNA-templated
A0R461	MSMEG_5715			Uncharacterized protein	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
A0R4D0	MSMEG_5790	Rv3118 Rv0814c		Uncharacterized protein	
A0R4H6	MSMEG_5837	#N/A		Glutathione peroxidase	glutathione peroxidase activity; response to oxidative stress
A0R4L1	MSMEG_5872	Rv0757	<i>phoP</i>	DNA-binding response regulator PhoP	DNA binding; phosphorelay signal transduction system; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R574	MSMEG_6091	Rv3596c	<i>clpC1</i>	ATP-dependent Clp protease ATP-binding subunit	ATP binding; protein metabolic process
A0R5D9	MSMEG_6157	Rv3646c	<i>topA</i>	DNA topoisomerase 1	DNA binding; DNA topoisomerase type I activity; DNA topological change; magnesium ion binding
A0R5E1	MSMEG_6159	Rv3648c	<i>cspA</i>	Probable cold shock protein A	cytoplasm; DNA binding; regulation of transcription, DNA-templated; transcription, DNA-templated
A0R5M3	MSMEG_6242			Alcohol dehydrogenase, iron-containing	1,3-propanediol dehydrogenase activity; alcohol dehydrogenase (NAD) activity; metal ion binding
A0R5N7	MSMEG_6256	Rv3708c	<i>asd</i>	Aspartate-semialdehyde dehydrogenase	'de novo' L-methionine biosynthetic process; aspartate-semialdehyde dehydrogenase activity; cytoplasm;

A0R5N8	MSMEG_6257	Rv3709c	<i>ask</i>	Aspartokinase	diaminopimelate biosynthetic process; isoleucine biosynthetic process; lysine biosynthetic process via diaminopimelate; N-acetyl-gamma-glutamyl-phosphate reductase activity; NAD binding; NADP binding; threonine biosynthetic process
A0R616	MSMEG_6391	Rv3799c	<i>accD4</i>	Propionyl-CoA carboxylase beta chain	amino acid binding; aspartate kinase activity; lysine biosynthetic process via diaminopimelate; threonine biosynthetic process propionyl-CoA carboxylase activity
A0R617	MSMEG_6392	Rv3800c		Polyketide synthase	biosynthetic process; hydrolase activity, acting on ester bonds; phosphopantetheine binding; transferase activity
A0R618	MSMEG_6393	Rv3801c	<i>fadD32</i>	Acyl-CoA synthase	ligase activity
A0R635	MSMEG_6410	Rv3818		Putative Rieske 2Fe-2S iron-sulfur protein	2 iron, 2 sulfur cluster binding; metal ion binding; oxidoreductase activity
A0R684	MSMEG_6459	Rv3859c		Ferredoxin-dependent glutamate synthase 1	glutamate biosynthetic process; glutamate synthase (ferredoxin) activity
Q9AFI5	MSMEG_6896	Rv0054	<i>ssb</i>	Single-stranded DNA-binding protein	DNA replication; single-stranded DNA binding
A0R7G6	MSMEG_6904	Rv0046c	<i>ino1</i>	Inositol-3-phosphate synthase	inositol-3-phosphate synthase activity; inositol biosynthetic process; phospholipid biosynthetic process

**Table C. 6 Gene ontology enrichment of uniquely identified exponential growth phase proteins**

<b>GOID</b>	<b>Ontology</b>	<b>Term</b>	<b>Number of identifications</b>	<b>p-value</b>
GO:0005488	molecular function	binding	20	3.22E-02
GO:0008152	biological process	metabolic process	17	4.80E-02
GO:0009058	biological process	biosynthetic process	13	6.60E-02
GO:0044281	biological process	small molecule metabolic process	10	1.95E-02
GO:0044283	biological process	small molecule biosynthetic process	7	1.64E-02
GO:0044710	biological process	single-organism metabolic process	11	3.20E-02
GO:0044711	biological process	single-organism biosynthetic process	9	1.64E-02
GO:0030145	molecular function	manganese ion binding	4	8.17E-04
GO:0043167	molecular function	ion binding	11	3.00E-02
GO:0043169	molecular function	cation binding	11	1.64E-02
GO:0046872	molecular function	metal ion binding	11	1.64E-02
GO:0046914	molecular function	transition metal ion binding	7	3.58E-02
GO:0044237	biological process	cellular metabolic process	15	9.36E-02
GO:0005829	cellular component	cytosol	2	4.86E-02
GO:0044444	cellular component	cytoplasmic part	4	5.17E-02
GO:0044445	cellular component	cytosolic part	2	3.91E-02
GO:0006082	biological process	organic acid metabolic process	8	1.95E-02
GO:0006551	biological process	leucine metabolic process	2	5.53E-02
GO:0009098	biological process	leucine biosynthetic process	2	5.53E-02
GO:0016053	biological process	organic acid biosynthetic process	6	1.64E-02
GO:0019752	biological process	carboxylic acid metabolic process	7	3.30E-02
GO:0043436	biological process	oxoacid metabolic process	7	3.58E-02
GO:0046394	biological process	carboxylic acid biosynthetic process	6	1.64E-02

**Table C. 7 Gene ontology enrichment of uniquely identified stationary growth phase proteins**

<b>GOID</b>	<b>Ontology</b>	<b>Term</b>	<b>Number of identifications</b>	<b>p-value</b>
GO:0003674	molecular function	molecular function	60	4.93E-04
GO:0006457	biological process	protein folding	3	2.07E-02
GO:0008150	biological process	biological process	39	6.93E-04
GO:0009987	biological process	cellular process	34	6.59E-04
GO:0005488	molecular function	binding	42	2.36E-05
GO:0000166	molecular function	nucleotide binding	18	9.30E-03
GO:0001882	molecular function	nucleoside binding	14	5.53E-03
GO:0001883	molecular function	purine nucleoside binding	14	5.53E-03
GO:0005524	molecular function	ATP binding	11	9.22E-02
GO:0017076	molecular function	purine nucleotide binding	14	5.53E-03
GO:0030554	molecular function	adenyl nucleotide binding	11	9.28E-02
GO:0032549	molecular function	ribonucleoside binding	14	5.53E-03
GO:0032550	molecular function	purine ribonucleoside binding	14	5.53E-03
GO:0032553	molecular function	ribonucleotide binding	14	1.58E-02
GO:0032555	molecular function	purine ribonucleotide binding	14	5.53E-03
GO:0032559	molecular function	adenyl ribonucleotide binding	11	9.22E-02
GO:0035639	molecular function	purine ribonucleoside triphosphate binding	14	5.53E-03
GO:0036094	molecular function	small molecule binding	18	1.75E-02
GO:0097159	molecular function	organic cyclic compound binding	34	1.37E-04
GO:0097367	molecular function	carbohydrate derivative binding	14	1.58E-02
GO:1901265	molecular function	nucleoside phosphate binding	18	9.30E-03
GO:1901363	molecular function	heterocyclic compound binding	34	1.37E-04
GO:0005575	cellular component	cellular component	31	2.91E-05
GO:0005622	cellular component	intracellular	29	8.49E-14
GO:0005623	cellular component	cell	30	3.56E-09
GO:0005737	cellular component	cytoplasm	27	5.41E-15
GO:0005840	cellular component	ribosome	17	4.96E-18
GO:0015934	cellular component	large ribosomal subunit	3	2.84E-03
GO:0030529	cellular component	intracellular ribonucleoprotein complex	17	4.96E-18
GO:0032991	cellular component	macromolecular complex	19	7.24E-13
GO:0043226	cellular component	organelle	18	4.96E-18
GO:0043228	cellular component	non-membrane-bounded organelle	17	5.10E-17
GO:0043229	cellular component	intracellular organelle	18	4.96E-18
GO:0043232	cellular component	intracellular non-membrane-bounded organelle	17	5.10E-17
GO:0044391	cellular component	ribosomal subunit	3	1.75E-02
GO:0044422	cellular component	organelle part	3	3.48E-02
GO:0044424	cellular component	intracellular part	27	2.57E-14
GO:0044444	cellular component	cytoplasmic part	19	2.70E-18
GO:0044446	cellular component	intracellular organelle part	3	2.96E-02

GO:0044464	cellular component	cell part	29	9.61E-09
GO:1990904	cellular component	ribonucleoprotein complex	17	4.96E-18
GO:0003676	molecular function	nucleic acid binding	20	9.19E-03
GO:0003723	molecular function	RNA binding	16	7.24E-13
GO:0019843	molecular function	rRNA binding	13	1.00E-14
GO:0003735	molecular function	structural constituent of ribosome	16	6.18E-17
GO:0005198	molecular function	structural molecule activity	16	1.03E-16
GO:0000049	molecular function	tRNA binding	3	2.07E-02
GO:0006412	biological process	translation	17	2.52E-15
GO:0006518	biological process	peptide metabolic process	17	1.00E-14
GO:0006807	biological process	nitrogen compound metabolic process	24	2.04E-02
GO:0008152	biological process	metabolic process	29	4.04E-02
GO:0009058	biological process	biosynthetic process	24	6.37E-03
GO:0009059	biological process	macromolecule biosynthetic process	21	1.37E-04
GO:0010467	biological process	gene expression	20	1.39E-04
GO:0019538	biological process	protein metabolic process	17	4.13E-12
GO:0034641	biological process	cellular nitrogen compound metabolic process	23	4.05E-03
GO:0034645	biological process	cellular macromolecule biosynthetic process	21	1.14E-04
GO:0043043	biological process	peptide biosynthetic process	17	5.16E-15
GO:0043170	biological process	macromolecule metabolic process	21	4.64E-03
GO:0043603	biological process	cellular amide metabolic process	17	7.00E-14
GO:0043604	biological process	amide biosynthetic process	17	6.50E-15
GO:0044237	biological process	cellular metabolic process	28	1.24E-02
GO:0044238	biological process	primary metabolic process	29	8.91E-04
GO:0044249	biological process	cellular biosynthetic process	23	6.67E-03
GO:0044260	biological process	cellular macromolecule metabolic process	21	3.40E-03
GO:0044267	biological process	cellular protein metabolic process	17	1.80E-12
GO:0044271	biological process	cellular nitrogen compound biosynthetic process	21	7.88E-04
GO:0071704	biological process	organic substance metabolic process	28	2.45E-02
GO:1901564	biological process	organonitrogen compound metabolic process	21	6.82E-07
GO:1901566	biological process	organonitrogen compound biosynthetic process	19	2.74E-07
GO:1901576	biological process	organic substance biosynthetic process	24	4.05E-03
GO:0006090	biological process	pyruvate metabolic process	3	6.20E-02
GO:0032787	biological process	monocarboxylic acid metabolic process	5	7.36E-02
GO:0005525	molecular function	GTP binding	4	1.58E-02
GO:0019001	molecular function	guanyl nucleotide binding	4	1.58E-02
GO:0032561	molecular function	guanyl ribonucleotide binding	4	1.58E-02