Deciphering the Impact of *rpoB* Mutations on the Gene Expression Profile of *Mycobacterium tuberculosis*

---

by

Juanelle du Plessis

Thesis presented in fulfillment of the requirements for the degree of
Master of Medical Science (Molecular Biology) in the
Faculty of Medicine and Health Sciences at
Stellenbosch University

Supervisor: Prof Thomas Calldo Victor

Co-supervisor: Prof Robin Mark Warren

March 2014

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.

ABSTRACT

Mycobacterium tuberculosis is the etiological agent for tuberculosis, an infectious disease which is one of the leading causes of morbidity and mortality in developing countries. The emergence of drug resistant tuberculosis has negatively impacted the efficacy of current treatment regimens and threatens to undermine tuberculosis control programs worldwide. Rifampicin forms the backbone of the World Health Organization's recommended treatment regimen for the treatment of drug susceptible tuberculosis. Resistance to rifampicin is caused by mutations in the 81 bp core region of the *rpoB* gene which encodes the β subunit of RNA polymerase. Numerous studies have shown that mutations at codons 531 and 526 are the most frequent in clinical isolates, yet little is known concerning the mechanistic effect of these mutations on the fidelity of RNA polymerase. In the present study, we aimed to determine the influence of *rpoB* mutations on the gene expression profile of *M. tuberculosis* cultured *in vitro*. To accomplish this, rifampicin resistant clinical isolates and spontaneous mutants (selected *in vitro* from H37Rv and a drug-sensitive clinical strain) harbouring *rpoB* H526Y and S531L mutations were subjected to whole genome sequencing and genome-wide transcriptional profiling.

When comparing the transcription profile of H37Rv to the *in vitro rpoB* mutants, a large proportion of the differentially expressed genes were found to encode for proteins involved in intermediary metabolism and respiration; and cell wall and cell processes. The majority of these differentially expressed genes were downregulated. Prominent differential expression in the same functional categories was also evident when comparing the clinical isolates with these mutations; however, a greater number of genes were differentially expressed in this case. Furthermore, expression of genes that are part of the WhiB7 regulon were found to be upregulated in the *rpoB*526 mutants, and downregulated in the *rpoB*531 mutants. These

findings indicate that both the position of the *rpoB* mutation, as well as the genetic background of the strain, play an important role in the gene expression profile of *rpoB* mutants. Surprisingly, transcriptional profiling of cultures that were exposed to the critical concentration of rifampicin for 24 hours did not exhibit significant differential gene expression. Whole genome sequencing, followed by bioinformatic analysis, revealed that the *in vitro* mutants harbour synonymous and non-synonymous single nucleotide polymorphisms in addition to the respective *rpoB* mutations. This suggests that the mycobacterial genome is constantly evolving, challenging previous assumptions of relatively static mycobacterial genomes.

The findings from this research have provided novel insight into understanding the influence of resistance-conferring mutations on the biology of *M. tuberculosis* and have shown that further studies are urgently needed to better understand the complex physiology of this pathogen. This knowledge will be critical for the success of future drug development endeavours.

OPSOMMING

Mycobacterium tuberculosis is die etiologiese agent vir tuberkulose, een van die grootste oorsake van morbiditeit en sterftes in ontwikkelende lande. Die verskyning van middelweerstandige tuberkulose het 'n negatiewe impak op die effektiwiteit van die huidige behandeling van tuberkulose en dreig om tuberkulose beheerprogramme wêreldwyd te ondermyn. Weerstandigheid teen rifampisien, een van die eerste-lyn anti-tuberkulose middels, word veroorsaak deur mutasies in die 81 bp kerngedeelte van die *rpoB* geen. Die mees algemene mutasies in kliniese isolate word gevind in kodons 531 en 526; alhoewel daar min inligting beskikbaar is oor die effek van hierdie mutasies op die funksie van RNS polimerase. Die doel van hierdie studie was om die effek van verskillende *rpoB* mutasies op die geen-uitdrukkingsprofiel van *M. tuberculosis* te bepaal. Rifampisien-weerstandige kliniese isolate en *in vitro* mutante (geselekteer vanaf H37Rv en 'n rifampisien-sensitiewe kliniese isolaat) met *rpoB* H526Y en S531L mutasies was vir hierdie doel geselekteer en gebruik om die heel genoom volgorde te bepaal en geenuitdrukking te kwantifiseer.

Vergelyking van die H37Rv transkripsieprofiel met *in vitro* geselekteerde *rpoB* mutante het getoon dat 'n groot aantal gene wat differensieel uitgedruk is, vir proteïene kodeer wat betrokke is in selwand-prosesse en intermediêre metabolisme en respirasie. Die meerderheid van hierdie differensieel uitgedrukte gene was afgereguleer. ’n Soortgelyke verskynsel is ook waargeneem in kliniese isolate, met die verskil dat 'n groter aantal gene in hierdie geval differensieel-uitgedruk was. Gene wat deel vorm van die WhiB7 regulon is opgereguleer in die *rpoB*526 mutante, terwyl dit afgereguleer was in die *rpoB*531 mutante. Hierdie resultate is ’n baie sterk aanduiding dat beide die posisie van die *rpoB* mutasie, asook die genetiese agtergrond van die organisme, ’n belangrike rol speel in die uitdrukking van gene in rifampisin weerstandige *M. tuberculosis*. Kulture wat aan die kritiese konsentrasie van

rifampisien vir 24 uur blootgestel is, het teenverwagting geen differensiële geen-uitdrukking getoon nie. Verder het die heel genoom volgorde bepaling van die *in vitro* mutante getoon dat sinonieme en nie-sinonieme enkel-nukleotied polimorfismes teenwoordig is in die onderskeie *rpoB* mutante. Dit dui daarop dat die mikobakteriële genoom voortdurend verander, moontlik nie so stadig as wat voorheen vermoed is nie.

Die bevindinge van hierdie navorsing bied nuwe insigte om die invloed van mutasies wat middel-weerstandigheid veroorsaak op die biologie van *M. tuberculosis* te verstaan. Dit het ook getoon dat verdere studies dringend nodig is om die komplekse fisiologie van hierdie patogeen te verstaan. Hierdie kennis sal van kardinale belang wees vir die sukses van toekomstige ondernemings om nuwe anti-tuberkulose middels te ontwikkel.

AKNOWLEDGEMENTS

First of all, I would like to extend a special word of thanks and gratitude to my supervisors; Prof Tommie Victor and Prof Rob Warren – your guidance, advice and intellectual stimulation have been invaluable over the past two years. Dr Lynthia Paul, Dr Monique Williams and Prof Sam Sampson – I appreciate your insightful comments, suggestions and questions. Margaretha, Lizma and Melanie – thank you for all your help and advice with lab work and data analysis. Also, a special thanks to everyone in Prof David Sherman's lab at Seattle Biomedical Research Institute, especially Dr Tige Rustad and Dr Kyle Minch for their help during my time in Seattle.

I am grateful for the financial support provided by the Harry Crossley Foundation; the Second Stella and Paul Lowenstein Charitable and Educational Trust over the past two years, as well as the Southern African Global Infectious Diseases training program for funding my research visit to Seattle BioMed. The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

To Annika and all my friends and colleagues in the department – thanks for all the laughs and fun times; and to my best friend, Rachel, thanks for always listening to what I have to say, even when you have no idea what I'm talking about. I would like to thank my wonderful parents, Jean and Estelle, and my sisters, Christelle and Rochelle, who have encouraged me throughout my postgraduate studies. With each passing year I realize more and more how fortunate I am to have such a supportive family.

Above all, I owe everything to my Heavenly Father. I feel blessed to be able to do something that I enjoy and find comfort in knowing that He watches over me.

The fear of the LORD is the beginning of wisdom,
and the knowledge of the holy is understanding.
Proverbs 9:10

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

°C	:	degree Celsius
α	:	alpha
β	:	beta
µg	:	microgram
µL	:	microliter
µm	:	micrometre
σ	:	sigma
ω	:	omega
A	:	alanine
Å	:	angstrom
ADC	:	albumin dextrose catalase
ADP	:	adenosine diphosphate
BAM	:	Binary Alignment/Map
bp	:	base pair
BSC	:	biosafety cabinet
BSL-3	:	Biosafety Level 3
cDNA	:	complementary DNA
Cy	:	cyanine
D	:	aspartic acid
DBC	:	DNA-binding channel
ddH_2O	:	double distilled water
DEPC	:	diethylpyrocarbonate

DNA	:	deoxyribonucleic acid
dNTP	:	deoxyribonucleotide triphosphate
E	:	glutamic acid
ECF	:	extracytoplasmic function
EDTA	:	ethylenediaminetetraacetic acid
Escherichia coli	:	*E. coli*
F	:	phenylalanine
g	:	gram
GATK	:	Genome Analysis Toolkit
gDNA	:	genomic DNA
GDP	:	guanosine diphosphate
GTP	:	guanosine triphosphate
H	:	histidine
in/dels	:	insertions or deletions
IS	:	insertion sequence
kDa	:	kilo-Dalton
L	:	leucine
M. tuberculosis	:	*Mycobacterium tuberculosis*
MDR	:	multidrug-resistant
MIC	:	minimum inhibitory concentration
M	:	molar
mg	:	milligram
mL	:	millilitre

mM	:	millimolar
mRNA	:	messenger RNA
ng	:	nanogram
nM	:	nanomolar
OADC	:	oleic albumin dextrose catalase
OD	:	optical density
OFL	:	ofloxacin
P	:	proline
PCR	:	polymerase chain reaction
PE	:	proline-glutamate
phage	:	bacteriophage
PPE	:	proline-proline-glutamate
ppGpp	:	guanosine tetraphosphate
pppGpp	:	guanosine pentaphosphate
R	:	arginine
RFLP	:	restriction fragment length polymorphism
RIF	:	rifampicin
RMA	:	robust multichip average
RPc	:	closed promoter complex
RPo	:	open promoter complex
RNA	:	ribonucleic acid
RNAp	:	RNA polymerase
RRDR	:	Rifampicin Resistance Determining Region

rRNA	:	ribosomal RNA
S	:	serine
SAM	:	Sequence Alignment/Map
SIFT	:	Sorting Intolerant From Tolerant
sRNA	:	small RNA
SNP	:	single nucleotide polymorphism
TB	:	tuberculosis
TBDB	:	TB Database
TBE	:	Tris/Borate/EDTA
TE	:	Tris/EDTA
Tm	:	melting temperature
tRNA	:	transfer RNA
V	:	volt
VCF	:	variant caller format
WT	:	wild-type
XDR	:	extensively drug-resistant
T	:	threonine
Y	:	tyrosine

Chapter 1

General Introduction

1.1 Background

Mycobacterium tuberculosis, the aetiological agent of tuberculosis (TB), was first described by Robert Koch in 1882 (1). Over the past 130 years, this pathogenic bacterial species has become one of the largest global health problems and continues to be one of the leading causes of morbidity and mortality in low income countries. TB was the cause of approximately 1.4 million deaths in 2011 and the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) TB poses a major threat to TB control programs worldwide (2). This is especially true for countries such as South Africa where more than 60% of the total TB budget is used to control drug resistant TB (3). With a population of over 50 million, the prevalence of TB in South Africa was estimated to be more than 760/100,000 in 2011 (2). According to data from the National Institute for Communicable Diseases, the frequency of MDR-TB among new cases was 7.8% in the same year (4). MDR-TB is defined as *M. tuberculosis* which has gained resistance to rifampicin (RIF) and isoniazid, two of the most powerful drugs used to treat TB. Treatment of MDR-TB requires the use of highly toxic drugs and has poor outcomes, with less than 50% of patients being treated successfully (2).

Drug resistance in *M. tuberculosis* primarily arises as a result of chromosomal mutations in genes which encode drug targets or activators. Resistance to RIF is associated with mutations in the RIF resistance-determining region (RRDR) of the *rpoB* gene which encodes for the β-subunit of RNA polymerase (RNAp), the primary enzyme responsible for transcription in prokaryotes. Numerous amino acid changes in the RRDR have been found to confer RIF-resistance, however S450L and H445Y mutations are the most frequent in clinical isolates, often referred to as S531L and H526Y mutations, respectively, as these are the corresponding codon positions in the gene sequence of *E. coli*. (5).

Recent work from our lab has demonstrated that closely related drug resistant *M. tuberculosis* isolates with a mutation in the *rpoB* gene displays varying levels of RIF resistance. It was found that this resistance could largely be reversed with the addition of efflux pump inhibitors, suggesting that the activation of efflux pumps is involved (6). A second observation from the same study showed that RIF was able to induce tolerance to ofloxacin (OFL) in RIF mono-resistant strains, which is one of the most important drugs used to treat MDR-TB. Strains with the *rpoB*526 mutation pre-exposed to RIF showed a limited increase in the minimum inhibitory concentration (MIC) of OFL after 7 days. Surprisingly, strains with the *rpoB*531 mutation showed induction of OFL tolerance after only 24 hours. These *in vitro* experiments suggest that certain *rpoB* gene mutations condition the bacterium to becoming more tolerant to other anti-TB drugs when exposed to RIF, which may lead to an increased risk of transmission of these strains. This could explain why clinical strains with the *rpoB*531 mutation are more prevalent than those with the *rpoB*526 mutation, however little is known concerning the effect of these mutations on the enzymatic function of RNA polymerase.

A study by Bisson *et al*. demonstrated an upregulation of the phthiocerol dimycocerosate biosynthetic pathway upon investigating the proteome and metabolome of RIF-resistant *M. tuberculosis* compared to wild-type (WT) isolates (7). This finding suggests that a single mutation in the *rpoB* gene can have a considerable impact on gene expression in *M. tuberculosis*. In this study, we aim to describe the transcription profile of *M. tuberculosis* with two of the most commonly-found *rpoB* mutations, S531L and H526Y, as well as the influence of these mutations on gene expression during exposure to RIF. Additionally, the genomes of the *in vitro* mutants used in this study will be sequenced to determine whether short insertions/deletions (in/dels) or additional single nucleotide polymorphisms (SNPs) were

acquired in conjunction with the respective *rpoB* mutations. This will allow us to confirm whether altered gene expression can be attributed to the *rpoB* mutations alone, or whether other genetic variations may be influencing the transcriptomic data we aim to generate in this study.

1.2 Hypothesis

Mutations in the β-subunit of RNA polymerase lead to diverse changes in the gene expression profile of *Mycobacterium tuberculosis.*

1.3 Aims

1. Investigate the effect of *rpoB*526 and *rpoB*531 mutations on gene expression in clinical isolates and *in vitro* selected mutants of *M. tuberculosis.*
2. Determine the transcriptomic response of RIF mono-resistant *in vitro* selected mutants and clinical isolates of *M. tuberculosis* after 24 hours of RIF exposure.
3. Determine whether the *in vitro* selected mutants used in this study harbour additional SNPs or short in/dels compared to the respective progenitors.

Chapter 2

Literature Review

The β and β' Subunits of Mycobacterial RNA Polymerase: Regulatory Hotspot and Drug Target

2.1 Abstract

This review aims to discuss the bacterial and non-bacterial factors that interact with the β and β' subunits of RNA polymerase at the level of transcription initiation with the view to collate the current knowledge regarding the regulation and inhibition of RNA polymerase in *Mycobacterium tuberculosis*. These interactions include sigma factors, small RNAs and guanosine tetra- and pentaphosphates, which have all been shown to play a role in regulating the pathway from DNA to messenger RNA. Furthermore, numerous antimicrobial agents and bacteriophage-encoded transcriptional regulators can bind to RNA polymerase in the same region in order to inhibit transcription in prokaryotes. Under certain conditions, *M. tuberculosis* is able to circumvent these inhibitory mechanisms by gaining mutations that prevent binding to RNA polymerase. It is evident that certain resistance-conferring mutations in *M. tuberculosis* aid bacterial survival in the face of antibiotic pressure, however little is known concerning their mechanistic effect on RNA polymerase or their influence on other RNA polymerase-ligand interactions. Review of the literature indicated that certain aspects of transcriptional regulation in *M. tuberculosis* is considerably understudied and that relying on studies done in model organisms such as *Escherichia coli* may not be the ideal way to gain a better understanding of mechanisms of regulation in pathogenic mycobacteria. Additional studies are urgently needed to fill such knowledge gaps and provide novel insights into the link between transcriptional regulation and mycobacterial physiology.

2.2 Introduction

As defined in the central dogma of molecular biology, transcription and translation are the fundamental processes involved in the expression of genetic information. DNA-dependant RNA polymerase (RNAp) is the primary enzyme responsible for transcription in prokaryotes. This large (~400 kDa), multi-subunit enzyme facilitates transcription of genes by binding to double-stranded DNA in order to synthesize RNA. Subsequently, messenger RNA (mRNA) is translated into proteins, completing the transformation of genetic information from gene to functional gene product. As an exception to this rule, ribozymes and several other RNA species, such as transfer RNA (tRNA) and ribosomal RNA (rRNA), are functional RNAs that are not translated into proteins.

To aid bacterial survival, the types and quantities of proteins present within the cell are altered to allow the bacterium to adapt to its environment. Transcriptional regulation is the primary process through which gene expression is controlled and consists of numerous networks which employ different regulatory mechanisms. Primary examples of these mechanisms include classical regulators of transcription such as repressors and activators, which bind to promoter regions of DNA in order to regulate the transcription of operons. In most cases, repressors prevent binding of RNAp to promoter regions; whereas activators enhance transcription by facilitating the link between the α subunits of RNAp and the promoter region. Additionally, enhancers are short DNA sequences which bind to DNA without interacting with RNAp and are able to activate genes from a distance. This review will discuss transcriptional regulators, antibacterial compounds and bacteriophage-encoded proteins that interact with the β and β' subunits of RNAp in *M. tuberculosis* in order to regulate or inhibit enzymatic activity at the level of transcription initiation (Figure 2.1).

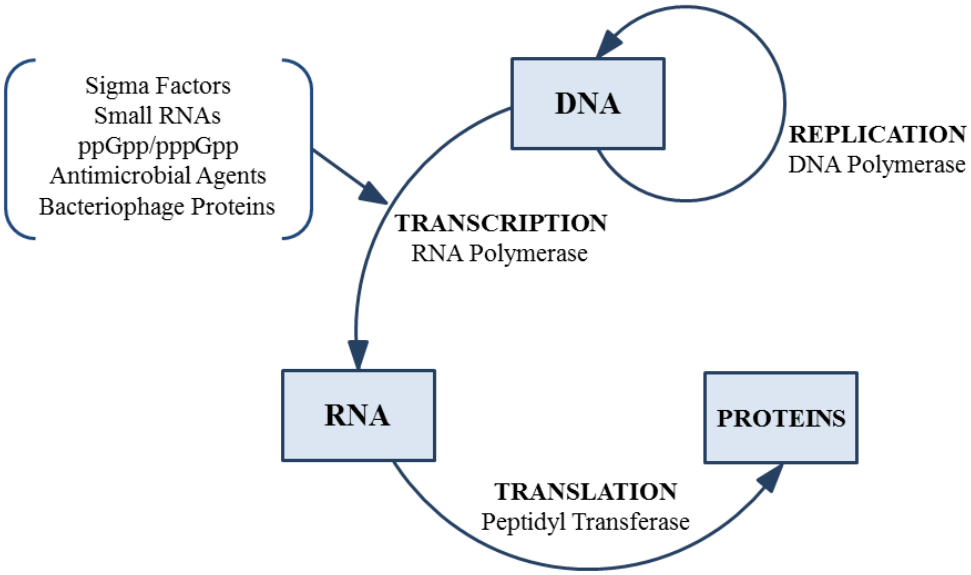

Figure 2.1 An illustration of the central dogma of molecular biology, the primary enzymes involved in each phase and the bacterial and non-bacterial factors discussed in this literature review that interact with the β and β' subunits of RNAp.

2.3 The Structure and Function of RNA Polymerase

RNAp consists of 5 subunits ($\alpha_2\beta\beta'\omega$) and one of 13 sigma (σ) factors ($\sigma^{A\text{-}M}$) in *M. tuberculosis* (Figure 2.2) (8). Two identical alpha (α) chains are transcribed from *rpoA* and consist of an N-terminal and C-terminal domain, connected by a flexible linker (9, 10). The C-terminal domain interacts with transcription factors and stabilizes the RNAp-promoter complex by binding to regions upstream of certain promoters called UP-elements. The beta chains, β and β', are the largest subunits of the enzyme and are encoded by *rpoB* and *rpoC*, respectively. Together, these subunits are said to resemble crab pincers as they appear to clamp the DNA strand during the process of transcription. The β' and β subunits form the DNA-binding channel (DBC), containing the active-centre cleft for ribonucleotide polymerization (11). Important characteristics of the β' subunit include the zipper, which separates the RNA-DNA hybrid, and the rudder, which divides the DBC and the smaller

secondary channel where ribonucleotides enter and join the growing strand of RNA. The dimerization of the omega (ω) subunit, encoded by *rpoZ*, is the first step in the assembly of the subunits which make up the core enzyme, thereafter binding of the dissociable σ factor allows the enzyme to recognize and bind to specific promoter regions (12).

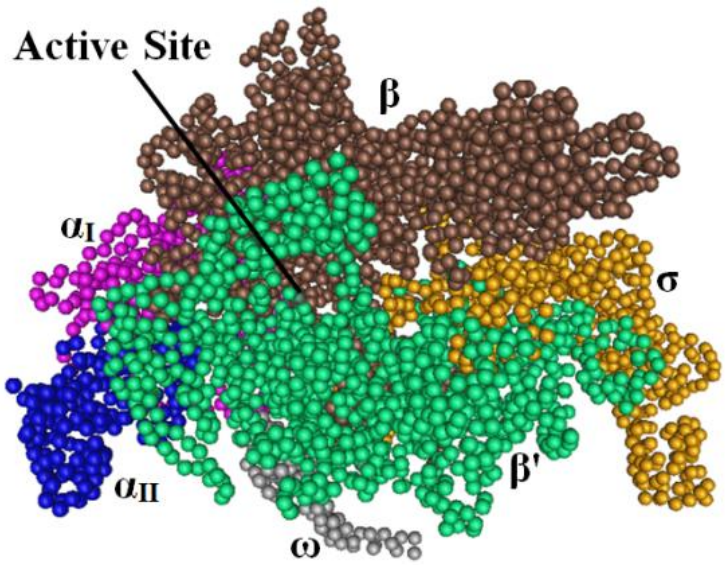

Figure 2.2 The crystal structure of *E. coli* σ^{70} RNAp illustrating the six subunits ($\alpha_2\beta\beta'\omega\sigma$) that form the holoenzyme and the active site located between the β and β' subunits (13). Illustration generated using Cn3D version 4.3 (14).

The transcription cycle consists of three main steps: initiation, elongation and termination. During initiation, a single σ factor binds to the RNAp core to form a holoenzyme, which aids promoter recognition by decreasing the affinity of RNAp for non-specific DNA. The unbound RNAp holoenzyme, referred to as the closed promoter complex (RPc), binds to regions -10 and -35 of the promoter region and undergoes numerous conformational changes to form the transcriptionally active open promoter complex (RPo). Subsequently, promoter DNA interacts with the DBC of RNAp, enabling the DNA strand to curve at the -10 region (15). Once

loading of the DNA strand has taken place, double-stranded DNA at the transcription start site is melted, leading to the formation of a transcription bubble and the initiation of ribonucleotide polymerization (Figure 2.3). The σ factor dissociates from the complex once a 9-10 nucleotide strand of RNA has been synthesized, thereafter the core enzyme continues with elongation of the RNA transcript.

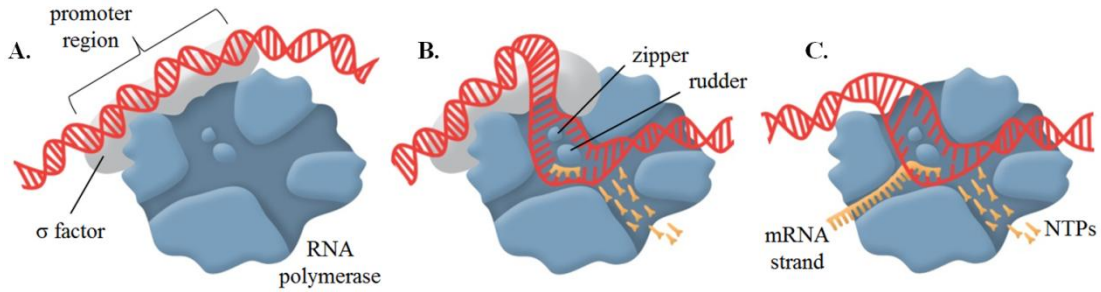

Figure 2.3 An illustration of transcription initiation in prokaryotes: A) The σ factor recognizes and binds to regions -10 and -35 of the promoter region; B) DNA is loaded into the DBC by curving at the -10 region; C) The σ factor dissociates once a 9-10 nucleotide strand has been synthesized and elongation of the mRNA continues.

2.4 Bacterial Transcriptional Regulators

2.4.1 Sigma Factors

Different σ factors bind to RNAp and alter the affinity of the enzyme for specific promoter regions (16). Adaptation of the bacterium to external stress or internal stimuli is made possible through binding of specific σ factors, which lead to diverse transcriptional responses. Two main families of σ factors have been described in prokaryotes, grouped according to their molecular weights: RpoN-like σ^{54} and RpoS-like σ^{70}, which are 54 kDa and 70 kDa, respectively. All 13 σ factors in *M. tuberculosis* are part of the σ^{70} family which consists of

four conserved regions. Domains 2.3-2.4, 3.0 and 4.2 have been found to bind to promoter regions -10, extended -10 and -35, respectively (Figure 2.4).

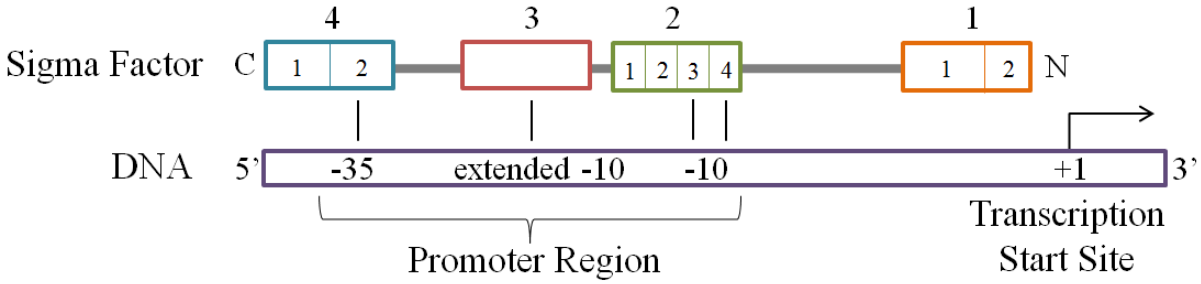

Figure 2.4 An illustration of the DNA promoter region and domains 1-4 found within σ^{70}.

The primary σ factor in *M. tuberculosis* is σ^{A}, which is responsible for transcription of housekeeping genes that are involved in essential cellular pathways. When the bacterium undergoes stress, the expression of alternative σ factors is induced, which allows them to compete with σ^{A} for binding to RNAp, thereby directing transcription from different promoter regions (Table 2.1). Ten of the σ factors in *M. tuberculosis* (all except σ^{A}, σ^{B} and σ^{F}) belong to the extracytoplasmic function (ECF) subfamily of σ factors which are characterized based on their sequence homology. Members of this family have been found to regulate proteins with extracytoplasmic functions or proteins involved in adaptation to extracytoplasmic stresses such as oxidative stress, heat shock or surface stress (17). Additional mechanisms of regulation include anti-σ factors, which bind to σ factors to prevent them from binding to the RNAp core, and anti-anti-σ factors are able to inhibit anti-σ factors to counteract this mechanism.

Table 2.1 Sigma factors found in *M. tuberculosis* with known external stress and internal stimuli that induce their expression.

Sigma Factor	Factors which induce expression	References
σ^A	Human macrophage infection	(18)
σ^B	Heat-shock, surface stress, low aeration, oxidative stress, stationary phase, nutrient starvation	(19–21)
σ^C	Early phase of oxygen depleted dormancy	(22)
σ^D	Nutrient starvation	(21)
σ^E	Heat-shock, surface stress, human macrophage infection, nutrient starvation	(18, 19, 21, 23)
σ^F	Antibiotic stress, nutrient starvation, oxidative stress, cold-shock, anaerobic metabolism, stationary phase	(21, 24)
σ^G	Human macrophage infection	(18)
σ^H	Heat-shock, human macrophage infection, oxidative stress	(19, 23, 25)
σ^I	Mild cold-shock, late stationary phase	(19, 26)
σ^J	Late stationary phase, human macrophage infection, oxidative stress	(26–28)
σ^K, σ^L, σ^M	Unknown	-

2.4.2 Guanosine Nucleotides

Guanosine tetra- and pentaphosphates (ppGpp/pppGpp), collectively referred to as guanosine nucleotides in this review, are synthesized by the phosphorylation of GDP and GTP, respectively. In *E. coli* and other Gram-negative prokaryotes, the enzymes RelA and SpoT modulate levels of (p)ppGpp in the cell, however only a single bifunctional enzyme (Rel_{Mtb}) exists in *M. tuberculosis*, which is responsible for both the synthesis and hydrolysis of (p)ppGpp (29, 30). This dual-function enzyme comprises of 738 amino acid residues and two catalytic domains, with amino acid fragments 1-394 and 1-181 shown to be responsible for (p)ppGpp synthetase and hydrolase activity, respectively (31).

Guanosine nucleotides are transcriptional regulators which form part of the stringent response in most prokaryotes and accumulate when amino acid availability is low (32, 33). Stringently controlled genes are highly-expressed during optimal growth conditions; however their transcription is discontinued in nutrient-limited environments in order to conserve cellular resources. When the stringent response is triggered by the presence of uncharged tRNA, RelA activity is induced, leading to an increase in the level of (p)ppGpp. Guanosine nucleotides, together with a protein called DksA, bind directly to RNAp and regulate transcription initiation in *E. coli* (34). Transcription from promoters which control expression of genes encoding rRNA, ribosomal proteins and regulators of protein translation are downregulated due to the limited supply of amino acids (35). Conversely, transcription from genes that encode alternative σ factors and amino acid biosynthesis are induced (36). It has also been found that guanosine nucleotides are able to regulate RNAp activity indirectly in *E. coli* by σ factor competition (37). This mechanism prevents RNAp from binding of the σ factor which is responsible for expression of housekeeping genes. As a result, alternative σ factors can bind to RNAp and drive transcription of regulons which are essential for survival. In both cases, SpoT hydrolase activity allows for decreased levels of guanosine nucleotides once ideal environmental conditions return, which restores transcription of genes that are needed for bacterial replication (Figure 2.5). A study by Dahl and colleagues demonstrated that p/ppGpp regulates the expression of 159 genes in *M. tuberculosis* during starvation (38). A number of these genes were found to be linked to persistence and the repression of metabolism which allow for an increased chance of survival in adverse conditions.

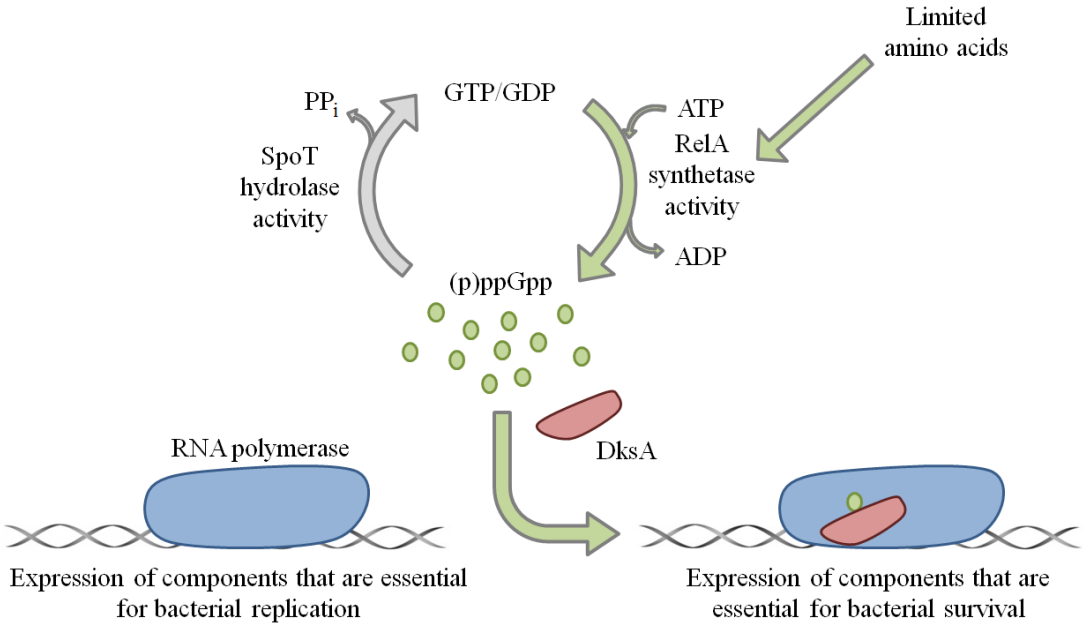

Figure 2.5 The role of (p)ppGpp in the stringent response in *E. coli*. The alarmone (p)ppGpp, synthesized in abundance as a response to limited amino acids, binds directly to bacterial RNAp together with DksA in order to regulate transcription. Once suitable environmental conditions return, SpoT hydrolase activity degrades guanosine nucleotides, allowing expression of genes that are essential for replication.

Guanosine nucleotides are able to regulate transcription by binding to RNAp at a site near the N-terminus of the β' subunit and the C-terminus of the β subunit (39). It has been found that RIF-resistant *E. coli* mutants, harbouring mutations at codons 331 and 332 in *rpoB*, exhibit a 20-fold increase in intracellular sensitivity to guanosine nucleotides (40, 41). However, the levels of ppGpp were approximately 10 times lower in these *rpoB* mutants compared to the WT strains, indicating an effect of *rpoB* mutations on ppGpp metabolism. It would be noteworthy to assess whether the same phenomenon is seen in strains of *M. tuberculosis* harbouring *rpoB* mutations. A homolog for DksA has not been identified in *M. tuberculosis*, however a transcriptional regulator named CarD has been found to share many similarities (42). CarD is induced by starvation as well as DNA damage in mycobacteria and is able to complement a Δ*dksA* strain of *E. coli* (43). However CarD is thought to function in a different

mechanism to DksA, as it binds to a different site on RNAp (N terminus of the β subunit). Furthermore, expression of CarD is vital for mycobacterial viability under all conditions, whereas DksA is not essential when *E. coli* is growing in a nutrient-rich environment.

The stringent response is one of the factors which contribute towards long term survival of *M. tuberculosis* (38). For this reason, our knowledge of guanosine nucleotides and transcriptional regulators which induce the stringent response play an important role in understanding persistent TB infections. Numerous studies have contributed to our knowledge of the global role of this alarmone in *E. coli*, however many aspects of (p)ppGpp synthesis and function in *M. tuberculosis* remain to be investigated.

2.4.3 Small RNA

It has long been known that small RNA (sRNA) plays a major role in posttranscriptional modulation, mainly through binding to complementary segments of target mRNA, thereby changing the stability of the mRNA and altering translation. However, it has recently been discovered that a number of sRNAs are linked to transcriptional regulation in prokaryotes. 6S sRNA in *E. coli* has been found to repress transcription from σ^{70}–dependent promoters by binding to RNAp during stationary phase (44, 45). For this reason, 6S RNA is thought to play a role in the transcriptional shift from exponential to stationary phase in prokaryotes. Furthermore, 6S RNA has been shown to act as a transcription template when binding to RNAp to initiate *de novo* synthesis of RNA molecules (46). However, it is unknown whether these transcripts are involved in transcriptional regulation or 6S RNA function. Results from a study by Arnvig *et al* suggests that MTS2823 in *M. tuberculosis* is functionally homologous to 6S RNA as it is able to modulate the downregulation of genes that are involved in

exponential growth (47). The mechanism through which this occurs remains unclear, however characterisation of MTS2823 may reveal a link between the modulation of transcription by sRNA and the downregulation of genes, which is characteristic of persistent *M. tuberculosis* infections (22, 48).

In a study which aimed to identify proteins that bind to sRNA in *E. coli*, it was revealed that a number of different sRNAs bind to the β subunit of RNAp, cleaving and extending the 3' end of the sRNA strands (49). The authors suggested that sRNA binds to the active cleft of RNAp, which is able to execute RNA-dependent polymerase activity. As the majority of essential cellular processes and machinery remain conserved among prokaryotes, one could propose that sRNA plays a similarly important role in mycobacteria, however further studies are needed to investigate this.

sRNA can also regulate transcription indirectly, by increasing translation of σ factors which bind to RNAp, bringing about a transcriptional response. In *E. coli*, the *rpoS* gene encodes for σ^S, which binds to RNAp and drives transcription of certain genes in response to external stress stimuli or entry into stationary phase (50, 51). During exponential growth, the 5' end of *rpoS* mRNA forms a hairpin-loop which obstructs the Shine Delgarno ribosome-biding site, thereby preventing initiation of *rpoS* translation (52). As an increased need for σ^S arises, DsrA sRNA, with the help of the Hfq chaperone, binds to a region of the hairpin-loop, thereby exposing the Shine Delgarno ribosome-binding site and enabling translation of *rpoS* mRNA (Figure 2.6) (53). In the absence of DsrA, RprA sRNA has been shown to play a similar role in the positive regulation of σ^S translation, suggesting that multiple sRNAs can interact with the same mRNA (54).

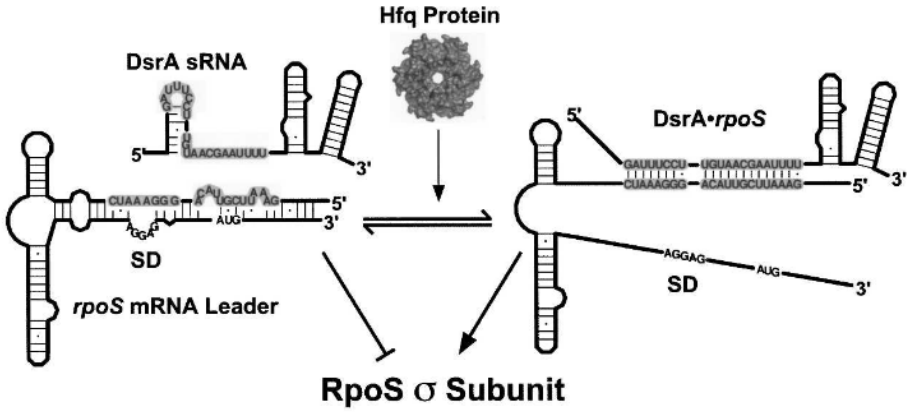

Figure 2.6 Regulation of *rpoS* translation by DsrA sRNA and Hfq in *E. coli*. SD: Shine Delgarno ribosome-binding site. Illustration adapted from Soper and Woodson (55).

An Hfq homolog has not been identified in *M. tuberculosis*; however it may be possible that sRNA regulates the expression of σ factors in a comparable way. Furthermore, DsrA has been shown to play a direct role in transcription, by overcoming H-NS silencing of the gene which encodes RcsA, a positive regulator of capsular polysaccharide synthesis in *E. coli* (56). This indicates that a single sRNA can influence the regulation of both transcription and translation in diverse ways. As the majority of these regulatory pathways have only been investigated in *E. coli* and other prokaryotes, additional studies in *M. tuberculosis* are needed to assess the influence of sRNA on transcriptional regulation in this species. Recently, a number of sRNAs have been identified and characterised in *M. tuberculosis*, however further investigation will be necessary to determine the functional role of these non-coding RNAs in mycobacterial gene regulation (57, 58).

2.5 Non-Bacterial Transcriptional Regulators

Bacteriophages (phages) are viruses which infect and kill bacterial cells. Once inside the bacterial host, transcription of early viral genes takes place by means of the host RNAp, the

primary enzyme responsible for gene transcription in prokaryotes. Thereafter, host transcription is inhibited and viral transcription of middle and late viral genes continue via phage-encoded RNAp. One of the ways in which phages inhibit host transcription is though phage-encoded proteins which bind to bacterial RNAp. A well-studied example of this interaction is that of T7 phage protein, Gp2, which effectively binds to and inhibits RNAp activity in *E. coli* (59–62). The binding of Gp2 to the β' subunit of RNAp, coded by the *rpoC* gene, has been found to be σ factor dependent, indicating that Gp2 may also play a role in the regulation of bacterial transcription (63).

Numerous studies have been done to investigate the mechanism of action of this small (~7 kDa), yet potent transcriptional inhibitor. Residues R56 and R58, located on the surface of Gp2, have been found to be important in binding of Gp2 to the β' subunit of RNAp (64). A multiple sequence alignment of 33 phage proteins that are known homologs of Gp2 revealed a high level of conservation at amino acids R56 and R58, indicating the importance of these residues in protein function (65). It was also found that substitutions that cause a charge reversal at these positions lead to decreased activity *in vitro* and *in vivo*, which suggests that an electrostatic component may be involved. Once Gp2 is bound to RNAp, a strip of 7 negatively charged amino acid side chains (E24, E34, D37, E38, E41, E44, E53,) face the downstream DBC (Figure 2.7). This is thought to aid inhibition of RNAp by either repelling negatively charged DNA or preventing conformational changes needed for stable RPo formation (64).

An alternative method used by phages to gain control of the transcription machinery within the host is through covalent modification such as ADP-ribosylation. An example of this mechanism is the interaction between AsiA, a 10 kDa T4 phage-encoded protein, and RNAp

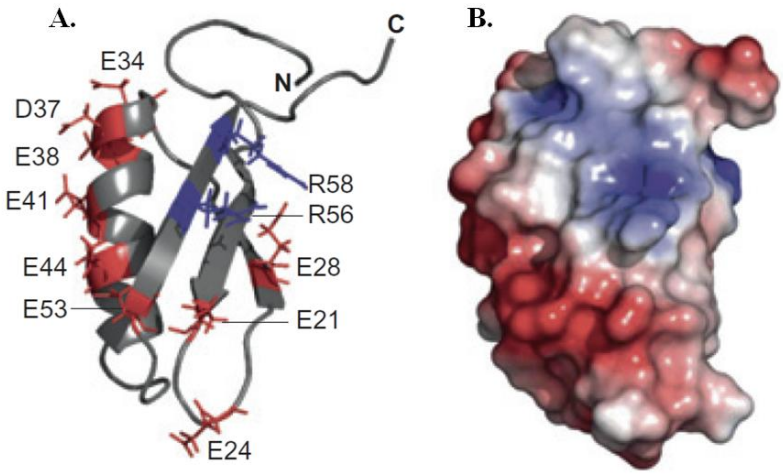

Figure 2.7 A. Ribbon structure of Gp2, illustrating negatively-charged amino acid side chains in red and positively-charged amino acid side chains in blue. B. Illustration of the basic electrostatic surface distribution on the molecular surface of Gp2. Adapted from Cámara *et al* (64).

in *E. coli* (66, 67). Firstly, the early viral genes *mod* and *alt* encode for ADP-ribosylases which modify α subunits of bacterial RNAp. Thereafter, AsiA binds to the modified enzyme to induce transcription of middle viral genes (68). Alternate σ factor gp55 is one of the middle genes that are transcribed which, in turn, induces transcription of late viral genes (69). At this stage, AsiA acts as an anti-σ factor by binding to σ^{70} and allowing phage-encoded gp55 to bind instead (70). This process of transcription from early to late viral genes is dependent on AsiA, revealing its role as a temporal regulator of transcription in *E. coli*.

Phage-mediated RNAp inhibition and regulation has not been studied in *M. tuberculosis* therefore it is unknown whether similar mechanisms exist for phages that are specific for mycobacteria, termed mycobacteriophages. However, it is likely that studies similar to the ones described here could provide novel insights into RNAp inhibition as a means of TB treatment.

2.6 Antimicrobial Agents

A number of compounds are known to inhibit RNAp in bacteria; however rifampicin (RIF) was the first drug to be used for the treatment of *M. tuberculosis* infections. RIF is a broad-spectrum antibiotic which inhibits mycobacterial transcription by binding to the β-subunit of RNAp (71). RIF binds at a position at least 12 Å away from the active site of RNAp and prevents extension of RNA transcripts longer than 2-3 nucleotides (72). RIF resistant isolates demonstrate reduced binding of RIF to RNAp due to mutations in the *rpoB* gene (71) (Table 2.2). Clusters of these mutations fall within the 81 bp sequence called the "Rifampicin Resistance Determining Region" (RRDR) (5). In a systematic review by Ramaswamy and Musser in 1998, it was found that approximately 95% of RIF-resistant isolates have point mutations or short insertions/deletions in this region of *rpoB* (5).

Table 2.2 High confidence *rpoB* mutations linked to RIF-resistance in *M. tuberculosis*. Adapted from the TB Drug Resistance Mutation Database (TBDReaMDB) (73).

Polymorphism	Amino Acid Change	Primary Reference
CTG/CCG	L511P	Kapur et al, 1994 (74)
CAA/CTA	Q513L	Kapur et al, 1994 (74)
GAC/GTC	D516V	Kapur et al, 1994 (74)
GAC/TAC	D516Y	Kapur et al, 1994 (74)
TCG/TTG	S522L	Bodmer et al, 1995 (75)
CAC/AAC	H526N	Ramaswamy et al, 2004 (76)
CAC/GAC	H526D	Kapur et al, 1994 (74)
CAC/TAC	H526Y	Kapur et al, 1994 (74)
CAC/CGC	H526R	Kim et al, 1997 (77)
CAC/CTC	H526L	Kapur et al, 1994 (74)
TCG/TTG	S531L	Donnabella et al, 1994 (78)
TCG/TGG	S531W	Bodmer et al, 1995 (75)
CTG/CCG	L533P	Moghazeh et al, 1996 (79)

A study which investigated the mechanistic effect of different *rpoB* mutations, based on structural models of the interaction between RIF and RNAp, revealed that certain *rpoB* mutations not only prevent binding of RIF, but also reduce the affinity of RNAp for RIF molecules (80). In addition to hindering the intermolecular hydrogen bonds between RIF and its binding site, S531L and H526D mutations lead to the formation of a structural obstruction and a charge repulsion, respectively (Figure 2.8). As a result, it appears that these particular mutations aid the bacterium in a multi-pronged mechanism which may account for the high level of RIF-resistance which has previously been associated with S531L mutations (81).

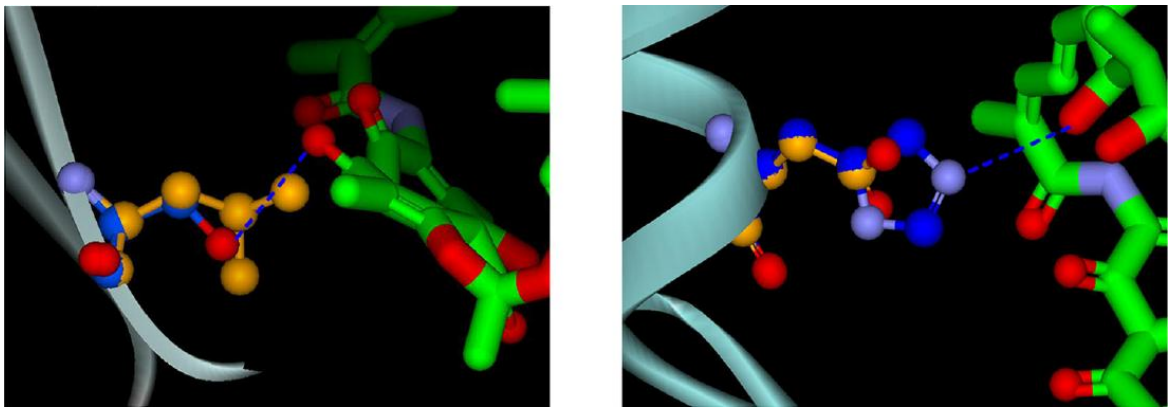

Figure 2.8 Structural models illustrating the interaction between a RIF molecule (depicted in green and red) and RNAp with S531L and H526D mutations, respectively. Hydrogen bonds between WT complexes are indicated by blue dotted lines. Illustration adapted from Pang *et al* (80).

In some cases, mutations which confer drug resistance can be harmful to the bacterium and may cause a fitness cost given that they affect important cellular functions and can cause a metabolic burden. In numerous studies, S531L or H526Y were found to be the most frequent RIF-resistance conferring mutations among RIF resistant clinical isolates and *in vitro* mutants, suggesting that these mutations have the smallest impact on mycobacterial fitness (82, 83). Gagneux and colleagues investigated the fitness cost inferred by different *rpoB* mutations in

both RIF-resistant clinical isolates of *M. tuberculosis* and laboratory-generated *rpoB* mutants (84). Fitness cost was calculated based on the mutant's ability to compete for limited resources when in the same environment as RIF-sensitive strains. No fitness cost was detected in the majority of the clinical isolates harbouring a S531L mutation, which may be why this particular *rpoB* mutation is more prevalent in clinical isolates. However, a fitness cost was observed in all *in vitro* generated mutants with an *rpoB* mutation, albeit very slight, as in the case of the S531L mutant. This indicates that there are mechanisms through which *M. tuberculosis* is able to ameliorate a fitness cost when in the host, possibly due to additional compensatory mutations which aid bacterial survival *in vivo*.

Gene mutations leading to a deleterious effect, as is the case with many drug resistance-conferring mutations, can have an array of consequences, such as reduced growth and virulence which would ultimately lead to decreased survival. The fate of a deleterious mutation includes reversion to the WT form by means of the bacterial DNA repair mechanism; or its continuance within the population by gaining compensatory mutations which reduce the deleterious effect without causing a loss of the drug resistant phenotype, referred to as compensatory evolution. Various studies have described compensatory mutations in RIF resistant strains of *M. tuberculosis*, one of which suggested that the acquisition of specific mutations in *rpoA* and *rpoC* can result in the emergence of high fitness multidrug-resistant (MDR) strains (Figure 2.9) (85, 86). Although these mutations were shown to occur at high frequencies in clinical isolates, the authors of this study specify that it is unknown how these mutations influence the success of the bacterium. More recently, De Vos and colleagues investigated the epidemiological relevance of *rpoC* mutations in clinical isolates of *M. tuberculosis* and revealed that 23.5% of the RIF-resistant strains included in the study harboured a non-synonymous polymorphism in *rpoC* and that these putative

compensatory mutations were significantly associated with the *rpoB* S531L mutation (87). Various models were proposed to explain the relationship between this particular *rpoB* mutation and *rpoC* mutations; however additional studies are needed to elucidate the underlying mechanisms that are involved.

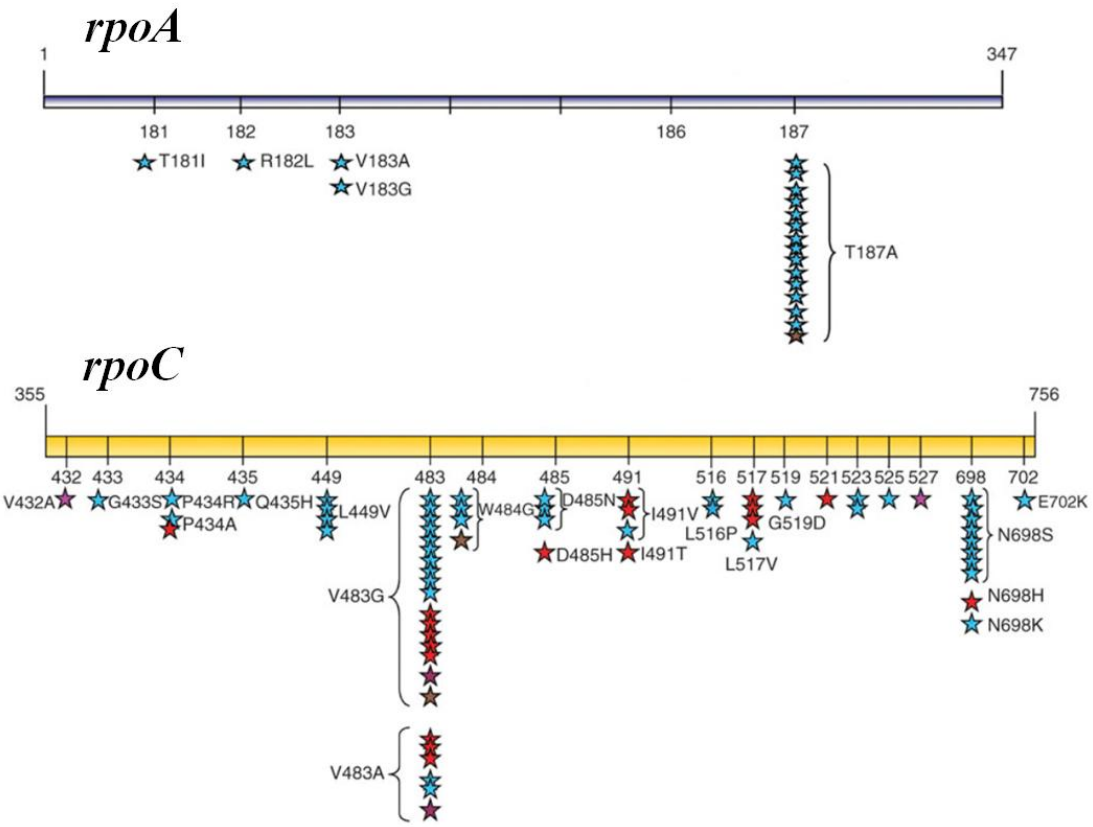

Figure 2.9 Mutations identified in *rpoA* and *rpoC* in a number of MDR strains of *M. tuberculosis*. Colours represent different lineages, indicating that the same mutations are present across different strain backgrounds. Adapted from Comas *et al* (85).

A number of other natural antibiotics that target RNAp have been discovered, one of which, called lipiarmycin, has recently been approved for the treatment of *Clostridium difficile* infections (88). Lipiarmycin is a macrocyclic antibiotic which has been shown to have activity against RIF-resistant *M. tuberculosis* (89, 90). The compound targets region 3.2 of σ^{70} as well

as a region called the switch-2 element on the β′ subunit of RNAp, thereby inhibiting formation of the RPo complex (91). Myxopyronin, corallopyronin and ripostatin are other drugs which bind to the switch-2 element of RNAp, indicating that this may be a promising drug target for treating *M. tuberculosis* infections (Figure 2.10) (92).

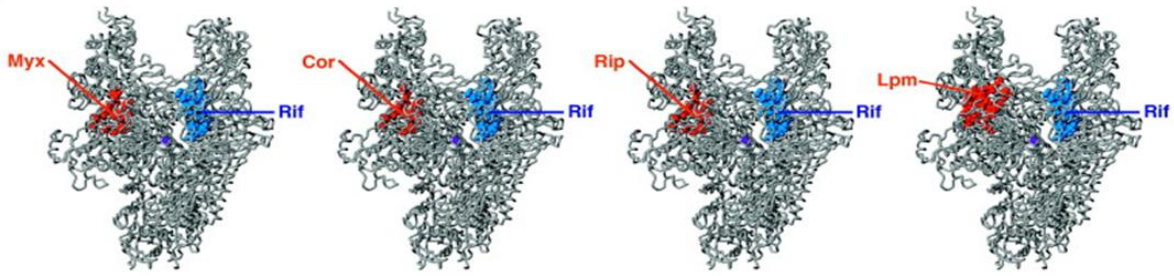

Figure 2.10 Target sites of myxopyronin (Myx), corallopyronin (Cor), ripostatin (Rip) and lipiarmycin (Lpm) in the switch region of the β′ subunit of RNAp, relative to the region where rifamycin (Rif) resistance-conferring substitutions occur on the β subunit. Adapted from Srivastava *et al* (92).

2.7 Knowledge Gaps

2.7.1 Transcriptional Regulation

Conditions such as cellular stress or a change in growth phase drives prokaryotes to rapidly modulate gene expression in order to survive. Numerous transcriptional regulators interact with RNAp in order to regulate gene expression as it is the primary enzyme responsible for transcription in prokaryotes. This review aimed to highlight a number of transcriptional regulators that function at the level of transcription initiation by interacting with the β and β' subunits of RNAp. However, gene regulation is far more complex than the aspects discussed here. A better understanding of regulatory proteins that function at all levels of transcription is necessary to fully comprehend RNAp regulation in bacteria. Unfortunately many of these mechanisms remain unclear or unknown in *M. tuberculosis*, and prove to be a major obstacle

in our understanding of host-pathogen interactions. Based on the aspects described in this review, it is evident that well-described mechanisms of transcriptional regulation in *E. coli* appear to differ from corresponding mechanisms in *M. tuberculosis*. This is evident when we compare the factors involved in transcriptional regulation mediated by σ factors, the stringent response and sRNA (Table 2.3).

Table 2.3 A comparison of the factors that influence transcriptional regulation in the model organism, *E. coli*, and pathogenic *M. tuberculosis*.

Factors influencing transcriptional regulation	***E. coli***	***M. tuberculosis***
σ factor families	σ^{54} and σ^{70}	σ^{70}
Number of σ factors	7	13
Number of anti-σ factors	3	5
Number of anti-anti-σ factors	0	2
p/ppGpp synthesis and hydrolysis	RelA and SpoT enzymes	Bifunctional Rel_{Mtb} enzyme
Transcription factor involved in stringent response	DksA binds to secondary channel of RNAp; not essential when *E. coli* is growing in nutrient-rich environments	CarD binds to N-terminus of β subunit; vital for mycobacterial viability under all conditions
Effect of *rpoB* mutations on the stringent response	Mutations at codons 331 and 332 lead to 20-fold increase in intracellular sensitivity to p/ppGpp	unknown
sRNA that modulates RNAp	6S represses transcription from σ^{70}–dependent promoters	MTS2823 facilitates downregulation of genes involved in exponential growth
sRNA chaperone	Hfq	unknown
Regulation of σ factors by sRNA	DsrA sRNA binds to hairpin-loop of *rpoS* mRNA	unknown
Bacteriophage-encoded inhibitors of RNAp	Gp2 well-characterised	unknown

Gaps in our understanding of *M. tuberculosis* physiology is highlighted in the table as it demonstrates a number of features which have not been studied or where functional homologs have not been identified in *M. tuberculosis*. Although *E. coli* has greatly improved our understanding of bacterial physiology, additional studies in *M. tuberculosis* are urgently needed to gain a better understanding of pathogenic bacterial characteristics such as drug tolerance, virulence, persistence and hypermutability. Considering the importance of (p)ppGpp and sRNA in transcriptional regulation, it is surprising how little is known about the role they play in mycobacteria. Further research could reveal the function of these poorly-understood regulators and the effect they have on RNAp and transcriptional regulation in *M. tuberculosis*.

Additionally, it is evident that the emergence of RIF-resistant TB has negatively affected TB control programs and the lack of a large repertoire of novel anti-TB drugs exacerbates this problem. A new approach to finding a sustainable solution for treating TB infections could emerge from investigating the mechanisms employed by phages to inhibit bacterial RNAp. The identification and characterization of inhibitory proteins in mycobacteriophages similar to the studies done in *E. coli* would be particularly useful in providing novel insight into RNAp as a drug target. Identifying compounds which mimic these inhibitory phage proteins would be beneficial for drug discovery and development in the global battle against drug-resistant TB. From past experience we know that if new RNAp-targeting drugs should be introduced into the current TB treatment regimen, it would be worthwhile to determine the effects of known resistance-conferring mutations prior to treating patients.

2.7.2 Understanding Mechanisms of Drug Resistance

It is uncertain how mutations that confer RIF resistance affect RNAp function and whether they cause structural changes that can lead to a change in the interactions between RNAp and transcriptional regulators. It is possible that these mutations affect protein function by changing the affinity for certain transcriptional regulators, just as certain *rpoB* mutations lead to decreased affinity of RIF for RNAp, however additional studies are needed to investigate this. A better understanding of the physiological changes that take place as a result of resistance-conferring mutations is necessary to determine their influence on protein function. Additionally, it is central to our understanding to determine how compensatory mutations in RNAp are able to ameliorate the fitness cost inferred by RIF resistance-conferring mutations.

Chapter 3

Materials and Methods

3.1 Mycobacterial Strain Selection and Cultivation

Clinical isolates were selected from a reference database which is maintained at Stellenbosch University and linked to a sample bank of well-characterized drug resistant *M. tuberculosis* isolates. The selected isolates included two RIF mono-resistant clinical isolates with different *rpoB* mutations, H526Y (n=1; C526) and S531L (n=1; C531), as well as a drug sensitive clinical isolate with no known resistance-conferring mutations, M0 (Table 3.1). All three isolates were characterized by IS*6110*-based restriction fragment length polymorphism (RFLP) and were shown to belong to closely related Beijing clusters 11427, 209 and 208, respectively.

Both the reference strain, *M. tuberculosis* H37Rv, and the aforementioned drug-sensitive clinical isolate were used to generate *in vitro* mutants by spontaneous mutation using the Luria–Delbrück fluctuation assay [previously done by M. de Vos (PhD Thesis, 2013) and D. Willemse (MSc Thesis, 2013)]. Targeted sequencing was used to confirm the position of *rpoB* mutations in approximately 150 spontaneous mutants; thereafter stocks of the mutants were stored at -80ºC. For the purpose of this study, two RIF mono-resistant progeny from each progenitor strain were selected; one with a H526Y mutation and another with a S531L mutation. The *in vitro* mutants (n=4) as well as the drug-sensitive progenitors (n=2) were included in this study with the aim of investigating the effect of different *rpoB* mutations on the transcriptome of *M. tuberculosis.*

Table 3.1 Clinical isolates, WT progenitors and *in vitro* mutants selected for transcriptomic analysis.

Sample	Genotype	Sample Type	*rpoB* Mutation
C526[*]	Beijing cluster 11427	clinical isolate	H526Y
C531[*]	Beijing cluster 209	clinical isolate	S531L
M526[Δ]	Beijing cluster 208	*in vitro* mutant	H526Y
M531[Δ]	Beijing cluster 208	*in vitro* mutant	S531L
M0[*]	Beijing cluster 208	clinical isolate (progenitor)	WT
H526[‡]	H37Rv	*in vitro* mutant	H526Y
H531[‡]	H37Rv	*in vitro* mutant	S531L
H37Rv	H37Rv	reference strain (progenitor)	WT

Legend to Table 3.1:

***** Samples selected from the clinical isolate sample bank: C526 (Isolate ID R1415); C531 (Isolate ID R721); M0 (Isolate ID K636)

[Δ] Samples selected from the K636 *in vitro* mutant depository: M526 (Sample ID N25S4); M531 (Sample ID N36S1)

[‡] Samples selected from the H37Rv *in vitro* mutant depository: H526 (Sample ID 74); H531 (Sample ID 75)

Clinical isolates were cultured on Löwenstein-Jensen slants in a Biosafety Level 3 (BSL-3) laboratory for 4-6 weeks (Appendix B); thereafter a scrape of the culture was inoculated into 5 mL of Middlebrook 7H9 broth (Becton, Dickinson and Company, MD, USA) enriched with albumin dextrose catalase (ADC), glycerol and Tween 80 (Merck, Darmstadt, Germany). Similarly, *in vitro* selected mutant colonies were picked from a culture inoculated onto Middlebrook 7H10 agar (Becton, Dickinson and Company, MD, USA) containing 2 µg/mL RIF and supplemented with oleic albumin dextrose catalase (OADC). The liquid cultures were incubated at 37°C for 2 weeks, thereafter 10 mL of ADC-enriched Middlebrook 7H9 broth was inoculated with 1 mL of culture and incubated at 37ºC until mid-log phase (OD_{600} = 0.7 - 0.8). Cultures were inspected for contamination by inoculating blood agar plates and the Ziehl-Neelsen differential staining method. *M. tuberculosis* does not grow on blood agar

therefore if no contamination is present, no growth should be observed on the plates after incubation at 37ºC for two days. The Ziehl-Neelsen staining method is used to detect acid-fast bacilli, such as *M. tuberculosis*, by light microscopy. The primary stain, fuchsin, binds to mycolic acid within the mycobacterial cell wall. Subsequent decolourization with acid does not remove the primary stain from mycobacteria and counterstaining with methylene blue allows for the identification of possible contaminants. Stocks of these cultures were stored at -80°C in 1 mL aliquots.

ADC-enriched Middlebrook 7H9 broth (50 mL) was inoculated in duplicate with 1 mL culture and the growth was monitored by measuring the optical density of the cultures over a period of 21 days. Optical density was measured using the Novaspec II spectrophotometer (Pharmacia Biotech, England, UK) with 1 mL cuvettes. This was done in order to generate growth curves for each of the strains to determine the number of days needed for cultures to reach mid-log phase (OD_{600} = 0.7 - 0.8). Once this was known, identical cultures were prepared to perform RNA isolation.

3.2 RNA Isolation and Microarray Analysis

3.2.1 Rifampicin Exposure Conditions

At mid-log phase, cultures for all of the samples were incubated for an additional 24 hours in the absence of RIF, whereas additional cultures for M526, M531, C526 and C531 were exposed to 2 µg/mL of RIF (Sigma-Aldrich, St Louis, Germany) for 24 hours. The abovementioned cultures were set up on different days as biological triplicates in order to increase statistical power in subsequent data analysis steps.

RNA Isolation

RNA isolation was done using the FastRNA® Blue kit (MP Biomedicals LLC, CA, USA) according to the protocol provided by the manufacturer (Appendix C). In summary, 10 mL of each culture was centrifuged at 1,500 x g for 10 minutes at 4°C and the pellet resuspended in 1 mL of RNA*pro* solution provided in the kit. Suspensions were transferred to a 2 mL screw-cap tube containing Lysing Matrix B and processed in a reciprocal shaker (Hybaid, UK) for 4 consecutive runs of 25 seconds at a setting of 6.5 W with intermittent cooling on ice for 1 minute. Tubes were decontaminated and removed from the BSL-3 laboratory and centrifuged at 12,000 x g for 20 minutes at 4°C. The supernatant was retained and transferred to 300 µL of chloroform (Sigma-Aldrich, St Louis, Germany), followed by a 5 minute incubation step at room temperature. Thereafter, samples were centrifuged at 12,000 x g for 10 minutes at 4°C and the upper phase was transferred to a tube containing 500 µL of cooled absolute ethanol (Sigma-Aldrich, St Louis, Germany). Samples were mixed by inversion and incubated at -20°C for at least one hour followed by centrifugation at 12,000 x g for 30 minutes at 4°C. Once the supernatant was aspirated, 500 µL of cooled 75% ethanol (Sigma-Aldrich, St Louis, Germany) was used to wash the pellet which was air dried for 5 min and re-suspended in 50 µL of diethylpyrocarbonate (DEPC) treated water (Ambion, Applied Biosystems, CA, USA). Samples were incubated for 5 minutes at room temperature to ensure efficient resuspension. Thereafter, aliquots were prepared and either stored at -80°C or treated with DNase as described below.

3.2.3 DNase Treatment and Phenol:Chloroform Purification

To remove residual genomic DNA (gDNA), samples were treated with RNase-free DNase in the following reaction: 10 µL of RNA, 10 µL of DEPC-treated H_2O, 4 µL of DNase Buffer

and 4 µL of RQ1 DNase (Promega, WI, USA). The reaction mixture was incubated at 37°C for 30 minutes, thereafter 172 µL of DEPC-treated water and 200 µL of cooled phenol:chlorophorm (4:1, v/v) (Sigma-Aldrich, St Louis, Germany) were added. After mixing by vortex, the samples were placed on ice for 10 minutes and thereafter centrifuged for 10 minutes at room temperature. The upper phase was retained and added to a mixture of 0.1 volumes RNase-free sodium acetate (pH 5.2) (Sigma-Aldrich, St Louis, Germany) and 2.5 volumes cooled absolute ethanol and incubated overnight at 4°C. Samples were centrifuged at 12,000 x g for 30 minutes at 4°C and the resulting pellet was washed with 500 µL of cooled 75% ethanol. Thereafter, the ethanol was decanted and the pellet was air-dried for 5 minutes before resuspending in 15 µL of DEPC-treated water. Aliquots of 5 µL were prepared for the purpose of quality assessment and the remaining RNA was stored at -80°C.

3.2.4 RNA Quality Assessment

RNA quality was determining by the A_{260}/A_{230} and A_{260}/A_{280} ratios, using the NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, MA, USA). The Experion RNA StdSens analysis kit and automated electrophoresis station (Bio-Rad Laboratories Inc., CA, USA) were used according to the manufacturer's instructions to quantify samples and evaluate RNA integrity based on the clarity of the 16s and 23s rRNA bands.

Samples were tested for gDNA contamination by polymerase chain reaction (PCR) amplification of the *katG* gene using a 0.05 mM concentration of the published primer set RTB59/RTB38 (Table 3.2) in the following reaction: 10x Buffer, 2 mM $MgCl_2$, 0.4 mM dNTPs, HotStarTaq DNA polymerase (Qiagen, Hilden, Germany) and RNA as template (93). *M. tuberculosis* H37Rv DNA was used as the template for positive controls and water was

added to negative controls. PCR amplification was performed under the following conditions: 95°C for 15 min; 35 cycles of 94°C for 1 min, 62°C for 1 min and 72°C for 1 min; followed by 72°C for 10 min. The PCR products were separated by electrophoresis using a 1.5% agarose (Sigma-Aldrich, St Louis, Germany) gel run in 1x TBE buffer (pH 8.3) at 150V for 1 hour and visualized with a Kodak Digital Science Electrophoresis Documentation and Analysis System (Vilber Lourmat, France).

Table 3.2 The primer set used for amplification of genomic DNA to determine the presence of DNA contamination in isolated RNA samples.

Name	Sequence	Product Size (bp)	Annealing Tm (°C)
RTB59	5' TGGCCGCGGCGGTCGACATT 3'	419	62
RTB38	5' GGTCAGTGGCCAGCATCGTC 3'		

3.2.5 Aminoallyl-Tagged cDNA Synthesis and Probe Preparation

Aminoallyl-tagged complementary DNA (cDNA) synthesis and probe preparation was done according to the protocol available from The Institute for Genomic Research (94). A volume of 2 µL random hexamer primers (3 µg/µL) was added to 1 µg of total RNA to give a final volume of 18.5 µL. After mixing by vortex, samples were incubated for 10 minutes at 70°C and placed on dry ice for 1 minute. Thereafter, each sample was added to the following reaction mix: 6 µL of 5x first strand buffer, 3 µL of 0.1 M DTT, 1.2 µL of 50 x aminoallyl-dNTP mix, 2 µL of SuperScript II reverse transcriptase (Invitrogen, CA, USA); and incubated at 42°C overnight. A volume of 10 µL of 1 M sodium hydroxide and 10 µL of 0.5 M EDTA was added to the mix and incubated at 65°C for 15 minutes to hydrolyse the RNA. The reaction was neutralized by adding 25 µL of Tris-HCl (pH 7.4) and samples underwent two purification steps to remove unincorporated nucleotides and unreacted Cy dye, respectively.

This was done using a QIAquick PCR purification kit (Qiagen, Hilden, Germany) according to the instructions provided by the manufacturer with certain modifications. The buffers supplied in the QIAquick kit contain free amines which compete with the Cy dye esters in the coupling reaction, therefore phosphate wash buffer (5 mM KPO_4, 80% EtOH) and phosphate elution buffer (4 mM KPO_4) were used in the first purification step.

Purification step 1 followed by Cy Dye coupling:

A volume of 300 µL phosphate buffer (pH 8.5-8.7) was added to the cDNA samples before centrifuging at 18,000 x g for 1 minute in a column. Samples were washed twice with 750 µL of phosphate wash buffer, followed by a final centrifugation step to remove excess buffer. Columns were transferred to a collection tube and incubated for 1 minute at room temperature with 50 µL of phosphate elution buffer. Thereafter, samples were centrifuged and placed in a centrifugal evaporator (SpeedVac, Thermo Fisher Scientific, MA, USA) for 1 hour. Subsequently, samples were resuspended in 4.5 µL of 0.1 M sodium carbonate buffer and coupled to Cy dyes by adding 4.5 µL of N-hydroxysuccinimide ester Cy3 or Cy5 (Amersham, NJ, USA) and incubating in the dark for 1 hour at room temperature.

Purification step 2:

A volume of 35 µL 100 mM sodium acetate (pH 5.2) and 500 µL of Buffer PB were added to the labelled cDNA and transferred to a column to be centrifuged at 18,000 x g for 1 minute. Samples were washed twice with 750 µL of Buffer PE, followed by a final centrifugation step to dry the columns before transferring to the collection tubes. Probes were eluted by adding 25 µL of Buffer EB and incubating for 1 minute at room temperature before centrifuging. This step was repeated before placing samples in a centrifugal evaporator for approximately 1 hour or until completely dry.

3.2.6 Microarray Hybridization

The following steps were done using the NimbleGen Hybridization kit (Roche Applied Science, Germany) according to the manufacturer's instructions. Each set of probes were resuspended in the following reaction mix: 4.5 µL of 2x Hybridization Buffer, 1.8 µL of Hybridization Component A, 0.2 µL of Alignment Oligo and 2 µL of water. After mixing by vortex, samples were placed in a thermocycler and incubated at 95°C for 5 minutes, followed by a second incubation step at 42°C for at least 5 minutes.

A 12x135K NimbleGen microarray slide (Roche Applied Science, Germany) was assembled using the Precision Mixer Alignment Tool and placed on a 42°C heating block for 5 minutes. Each slide comprises of 12 identical arrays of 135,000 60-mer oligonucleotide probes covering approximately 4000 open reading frames in the *M. tuberculosis* H37Rv genome. A volume of 6 µL of each sample was loaded into the chambers and the mixer-slide assembly was placed in a NimbleGen hybridization system at 42°C for 16 to 20 hours.

After incubating in the hybridization system, the mixer-slide assembly was placed in the Mixer Disassembly Tool in 270 mL of Wash I which was preheated to 42°C. While submerged, the mixer was removed from the slide and washed with Wash I, II and III prepared according to the manufacturer's instructions. The slide was dried with a slide spinner (Sigma-Aldrich, St Louis, Germany) and scanned at 5 µm resolution using a GenePix 400B scanner (Molecular Devices, CA, USA) with GenePix software version 6.0.

3.2.7 Transcriptomic Analysis

NimbleScan software (Roche Applied Science, Germany) was used to burst each scanned image into 24 individual arrays and calculate spot intensities. Data was imported for robust multichip average (RMA) normalization and statistical analysis using ArrayStar Version 4.1.0 (DNASTAR Inc., WI, USA). Following Benjamini Hochberg multiple testing correction, a Student's t-test p-value of <0.05 was used to determine statistical significance and a mean $\log_2$ fold change of >1.0 across 3 replicates was used as a cut-off for differential expression. A log-transformed fold change of 1.0 corresponds to a non-transformed fold change of 2.0, which is the commonly used threshold for differential gene expression. Furthermore, a linear correlation test was done to determine R^2 values among sample replicates.

3.2.7.1 The Effect of Different *rpoB* Mutations on Global Gene Expression

Firstly, *in vitro* selected mutants and progenitors were compared to investigate the effect of a particular *rpoB* mutation on the expression profile of *M. tuberculosis*. In order to do this, the following expression ratios were calculated:

- H37Rv compared to the *in vitro* selected mutants generated from H37Rv (H526/H37Rv; H531/H37Rv)
- M0 compared to the *in vitro* selected mutants generated from M0 (M526/M0; M531/M0)

3.2.7.2 The Effect of Genetic Background on Gene Expression in *rpoB* Mutants

Furthermore, the following ratios were calculated in order to determine whether the same transcriptomic trend can be seen in the expression profiles of *rpoB* mutants with different genetic backgrounds:

- M526 compared to M531 (*in vitro* mutants generated from M0)
- H526 compared to H531 (*in vitro* mutants generated from H37Rv)
- C526 compared to C531 (RIF-resistant clinical isolates)

3.2.7.3 Transcriptomic Response to Rifampicin Exposure

Expression profiles for all RIF exposed and control samples were compared to determine the effect of exposing RIF mono-resistant *M. tuberculosis* to 2 µg/mL of RIF for 24 hours. The ratios calculated for this comparison were as follows:

- *in vitro* mutants exposed to RIF compared to the unexposed controls (M526 exposed/M526 unexposed; M531 exposed/M531 unexposed)
- clinical isolates exposed to RIF compared to the unexposed controls (C526 exposed/C526 unexposed; C531 exposed /C531 unexposed)

3.3 DNA Isolation and Whole Genome Sequencing

3.3.1 DNA Isolation

For the purpose of whole genome sequencing, DNA was extracted from samples M0, M526, M531, H37Rv, H526 and H531 as previously described (95). In summary, 100 µL of each isolate stock culture was inoculated onto two petri dishes (140 mm diameter) of OADC-

enriched 7H10 Middlebrook agar and incubated for two weeks at 37°C. The petri dishes were placed in a pre-warmed fan oven for 1 hour at 80°C in order to heat-inactivate the bacteria. Thereafter, colonies were scraped from the surface of the media and transferred to a 50 mL tube containing 6 mL of extraction buffer (25 mM EDTA, 50 mM Tris-HCl pH 7.4, 5% sodium glutamate) and ± 20 glass beads with a diameter of 5 mm. The cellular suspension was mixed by vortex to disrupt bacterial colonies whereafter 50 mg of lysozyme (Roche Applied Science, Germany) and 25 µg of RNase A (Roche Applied Science, Germany) were added and incubated for 2 hours at 37°C. Subsequently, 650 µL of 10x Proteinase K buffer (100 nM Tris-HCl (pH 7.8), 50 mM EDTA, 5% sodium dodecyl sulphate) and 300 µL of Proteinase K (10 mg/mL) were added and incubated for 16 hours at 45°C. An equal volume of phenol/chloroform/isoamyl alcohol (24:23:1) was added and the suspension incubated at room temperature with mixing every 30 minutes over a period of 2 hours. After centrifuging for 20 minutes at 1800 x g, the aqueous phase was aspirated and added to 5 mL of chloroform/isoamyl alcohol (24:1). The suspension was centrifuged again for 20 minutes at 1800 x g and the top phase transferred to 600 µL of 3 M sodium acetate (pH 5.2) and 7 mL of ice-cold isopropanol. Precipitated DNA was immediately collected on a glass rod and incubated for 10 minutes in 1 mL of 70% ethanol, thereafter redissolved in 300 µL of TE buffer (1mM EDTA, 10 mM Tris-HCl pH 8.0) and stored at -20°C. DNA quality and quantity were determined spectrophotometrically using the NanoDrop-1000.

3.3.2 Whole Genome Sequencing

Whole genome sequencing of the drug-sensitive clinical isolate (M0), reference strain (H37Rv) and *in vitro* selected mutants (H531, H526, M526) was done at the Central Analytical Facility at Stellenbosch University using the Ion Torrent PGM Sequencer (Life

Technologies, CT, USA). The computational analysis for *in vitro* selected mutant M531 differed slightly from the pipeline described below (Figure 3.1) since it was sequenced using version 1.5 of the Illumina sequencing platform (Illumina, CA, USA) [M. de Vos (PhD thesis, 2013)].

3.3.3 Quality Assessment

FASTQ files for each of the genomes were generated and FastQC software (Babraham Institute, UK) was used to assess the quality of the raw sequence reads. FastQC software evaluates various aspects of the raw data and produces results in HTML format. A number of parameters are considered during quality control analysis such as sequence quality, sequence content and the level of sequence duplication. FastQC generates a detailed report which provides a warning if quality scores are below that of the pre-set parameters. Furthermore, basic statistics such as read length, total number of reads processed and percentage GC content are also included in the report.

3.3.4 Computational Analysis

Based on the abovementioned quality assessment, the raw sequence reads were trimmed by 27 bp using PRINSEQ-lite (http://prinseq.sourceforge.net). This was done to remove poor quality sequences at the end of longer reads which may negatively influence sequence alignment and downstream analysis. The reference genome, *M. tuberculosis* H37Rv (obtained from TB Database, TBDB; http://genome.tbdb.org) was indexed using Tmap. This mapping software, specifically designed for Ion Torrent data, was also used to align the sequence reads to the reference genome using the "mapall" function. The resulting file is in Sequence

Alignment/Map (SAM) format, which was converted to Binary Alignment/Map (BAM) format using SAMtools software (http://samtools.sourceforge.net). Likewise, SAMtools was used to sort and index the BAM files as well as assess the percentage mapped reads using the "flagstat" function. Genome Analysis Toolkit (GATK) was used for calling in/dels within the genome and two variant callers, SAMtools and GATK, were used to identify SNPs (highlighted in green in Figure 3.1). The resulting files were provided in variant caller format (VCF). The pipelines for the SNP callers are identical with the exception of a local realignment step when using GATK (highlighted in red in Figure 3.1). This step was necessary to correct for misalignments across deletions, an error which may occur when the alignment is performed.

The SNPs identified by both the callers were extracted from the respective VCF files using in-house python scripts (M. Daya, Department of Biomedical Sciences, Stellenbosch University). This was done in order to minimize the number of false positives identified by the individual SNP callers. As only GATK was used for calling in/dels, it was not necessary to overlap the variants in this way. In-house Perl scripts (M. de Vos, Department of Biomedical Sciences, Stellenbosch University) were used to annotate the identified SNPs and in/dels and those with a quality value of more than 1000 were investigated using GenomeView software (www.genomeview.org).

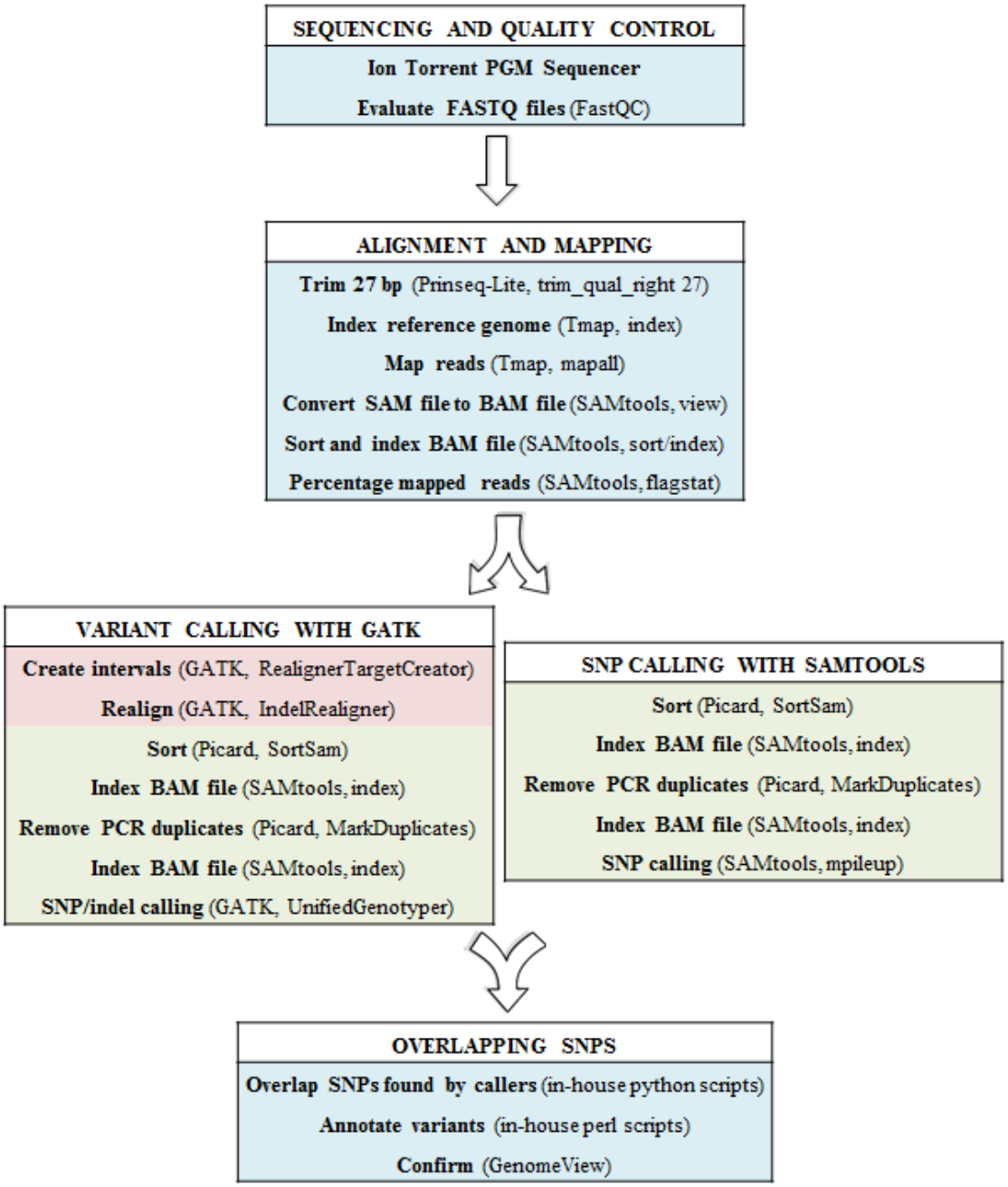

Figure 3.1 Summary of the computational analysis pipeline used to identify variants in whole genome sequence data (software, function).

3.3.5 Validation of Variants

To confirm the presence of high-confidence SNPs, PCR primers were designed to amplify the genomic regions containing the respective SNPs (Table 3.3). PCR amplification was done as described in section 3.2.4 and the resulting amplicons were purified and sequenced using the

Sanger sequencing method on the ABI PRISM DNA Sequencer, Model 377 at the Central Analytical Facility, Stellenbosch University. DNA sequences were aligned to the reference genome of *M. tuberculosis* H37Rv using BioEdit version 7.0.5.3 (available from www.mbio.ncsu.edu/bioedit/bioedit.html). SNPs that were identified by whole genome sequencing and validated by Sanger sequencing were considered to be present in the genome of the respective *M. tuberculosis* isolates.

Table 3.3 Primers used to validate high confidence SNPs identified through whole genome sequencing of *in vitro* selected mutants (H526; H531; M526; M531).

Gene		Primer Sequence	Product Size (bp)	T_m (ºC)
TB16.3	Forward:	5' GATCCAGGCGAGGTGATG 3'	544	60.16
	Reverse:	5' CGGTGGATACCACCAACACT 3'		60.59
lipM	Forward:	5' GAGCGAATGGTGGTCAAACA 3'	372	62.05
	Reverse:	5' CAACCCTAGCGACCTCAACC 3'		61.96
trmB	Forward:	5' GTAGGGCTACGTCCCGACAC 3'	431	61.86
	Reverse:	5' AAGACTCGAACACCGCACAA 3'		61.83

3.3.6 Bioinformatic Analysis of Non-Synonymous SNPs

To determine whether validated non-synonymous SNPs influence protein structure, models of the respective proteins were generated using SWISS-MODEL and visualized using the PyMOL Molecular Graphics System, Version 1.1 Schrödinger, LLC (96). A QMEAN Z-score was generated by the SWISS-MODEL workspace and was used to estimate the absolute quality of the protein models. This score takes 4 statistical potential scores into account (torsion potential, solvation potential, all-atom interaction potential and C-β interaction potential) and ranges from 0 to 1; with a higher score indicating a more reliable model (97). Amino acid sequences were also submitted to SIFT (Sorting Intolerant From Tolerant;

http://sift.bii.a-star.edu.sg) to predict whether the respective polymorphisms influence protein function (98). SIFT generates a score based on sequence homology and amino acid properties which represents the normalized probability that a particular amino acid change is tolerated, with a score of <0.05 indicating deleterious polymorphisms.

Chapter 4

Results

4.1 Growth Curves

Growth curves were constructed for each of the isolates to determine the number of days required for the cultures to reach mid-log phase (Figure 4.1). Cultures were incubated for approximately 5 to 13 days before reaching an OD_{600} of 0.7 - 0.8 (highlighted in blue shading on graph). It was found that WT H37Rv demonstrated a faster growth rate as compared to the *in vitro mutants* H531 and H526. A similar trend was seen for the clinical progenitor isolate, M0, as compared to the *in vitro* mutant M531. Clinical isolates, C531 and C526, showed a slower growth rate, taking between 9 and 14 days to reach mid-log phase.

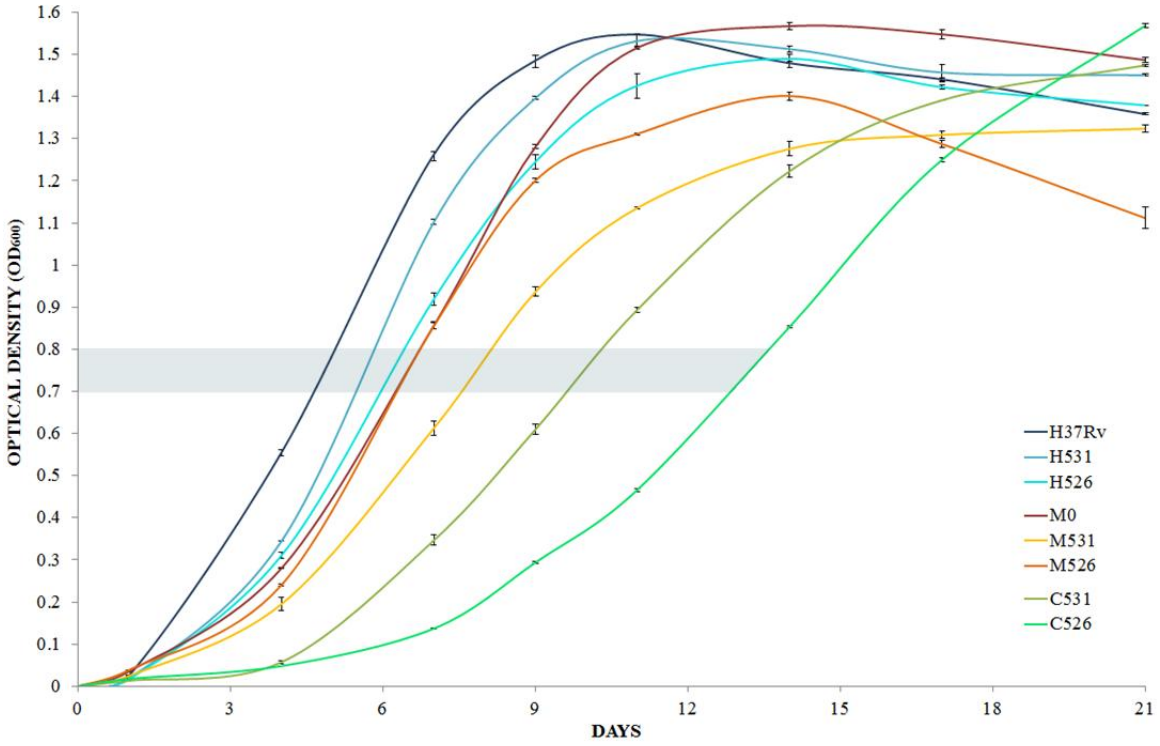

Figure 4.1 Growth curves for progenitors (H37Rv; M0); *in vitro* selected mutants (H526; H531; M526; M531) and clinical isolates (C526; C531). Error bars represent standard deviation between duplicates.

4.2 RNA Isolation and Microarray Analysis

4.2.1 RNA Quality Assessment

Figure 4.2a shows the integrity of the isolated RNA after fractionation using the Experion automated electrophoresis system. The RNA was assessed for gDNA contamination by PCR amplification of the *katG* gene. No PCR products were present after fractionation by agarose gel electrophoresis, confirming the absence of DNA contamination in the isolated RNA (Figure 4.2b). The concentrations for the purified RNA samples (n=36) ranged from 95.68 - 880.87 ng/µL (Appendix D).

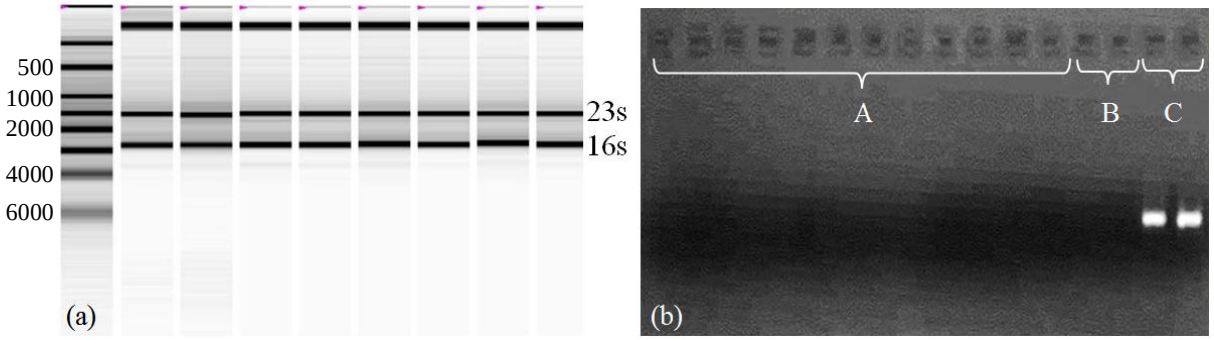

Figure 4.2 (a) Fractionation of purified RNA to determine its integrity according to the clarity of the 16s and 23s rRNA bands. (b) Electrophoretic fractionation of the *katG* PCR amplification product. A: RNA samples; B: negative controls C: positive controls.

4.2.2 The Effect of *rpoB* Mutations on Global Gene Expression

Expression ratios were calculated in order to evaluate the effect of different *rpoB* mutations on the gene expression profiles of the respective *M. tuberculosis* strains. Two sets of WT progenitors and *in vitro* selected spontaneous mutants were available for this comparison: H37Rv/H526/H531 and M0/M526/M531, respectively. A cross-correlation revealed the similarity between the expression profiles, with R^2 values closer to 1 indicating greater similarity between samples. Table 4.1 illustrates the R^2 values for these samples, ranging from

0.86 - 0.96. Clinical isolates C526 and C531 were not included in this analysis as WT progenitor strains were not available for these isolates.

Table 4.1 Cross-correlation R^2 values for each of the WT progenitors and *in vitro* selected mutants once biological replicate datasets were grouped. Red shading indicates R^2 values closer to 1.00.

	H37Rv	**H526**	**H531**	**M0**	**M526**	**M531**
H37Rv	-----	0.95	0.96	0.84	0.73	0.71
H526	0.95	-----	0.93	0.84	0.75	0.70
H531	0.96	0.93	-----	0.82	0.74	0.73
M0	0.84	0.84	0.82	-----	0.90	0.86
M526	0.73	0.75	0.74	0.90	-----	0.90
M531	0.71	0.70	0.73	0.86	0.90	-----

4.2.2.1 Drug-Sensitive Clinical Isolate and *In Vitro* Selected Mutants

Data in table 4.1 showed that progenitor M0 varied considerably from *in vitro* mutants M526 and M531 (R^2 values 0.86-0.90), unlike the H37Rv progenitor and its corresponding mutant clones, which were very similar (R^2 values 0.93-0.96). Due to this discrepancy, dataset M0 was excluded from subsequent analyses.

4.2.2.2 H37Rv and *In Vitro* Selected Mutants

Comparing WT H37Rv to each of the *rpoB* mutants revealed a total of 65 and 100 genes which were differentially expressed with a p-value of >0.05 and a $\log_2$-transformed fold change of >1.0 (equivalent to a non-transformed fold change of >2.0) in H526 and H531,

respectively. The majority of these differentially expressed genes were found to be downregulated in both *rpoB* mutants (Figure 4.3)

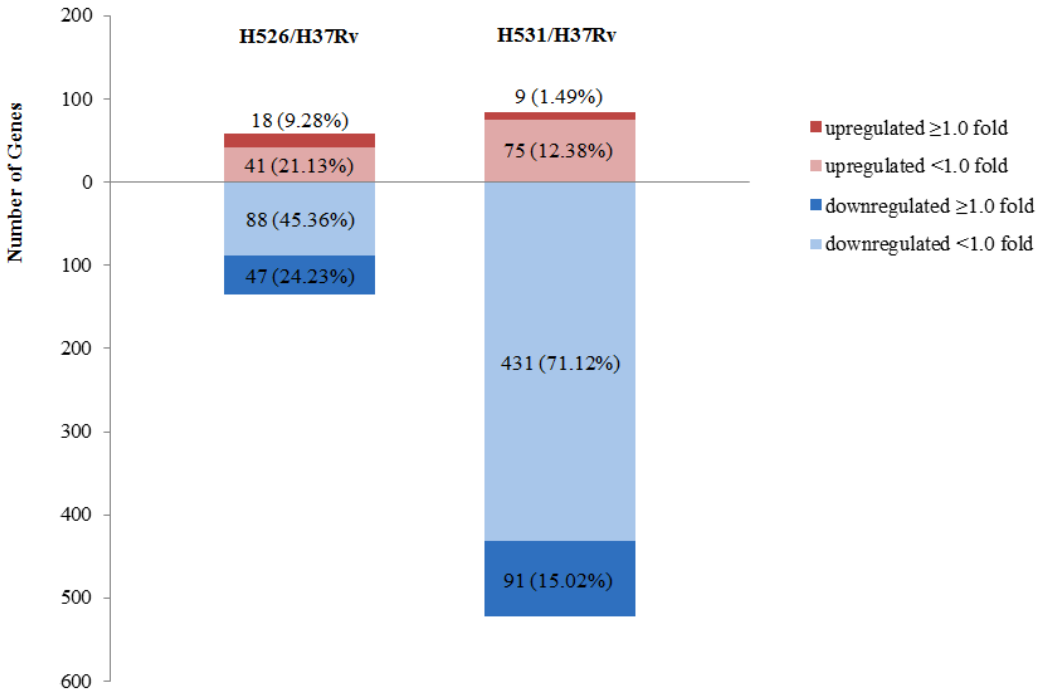

Figure 4.3 Distribution of significantly differentially expressed genes in *rpoB* mutants (H526, H531) with reference to WT progenitor H37Rv, grouped according to fold change.

Subsequently, differentially expressed genes with a $\log_2$ fold change of >1.0 were grouped into functional categories according to TubercuList (http://tuberculist.epfl.ch/) to determine the groups of genes whose transcription were most affected by the *rpoB* mutations (Figure 4.4). The majority of the differentially expressed genes were found to encode for hypothetical proteins, followed by proteins involved in intermediary metabolism and respiration; and cell wall and cell processes. In 7 of the 10 functional categories, a greater number of genes were downregulated in H531 compared to H526, which correlates with the representation of downregulated genes in Figure 4.3.

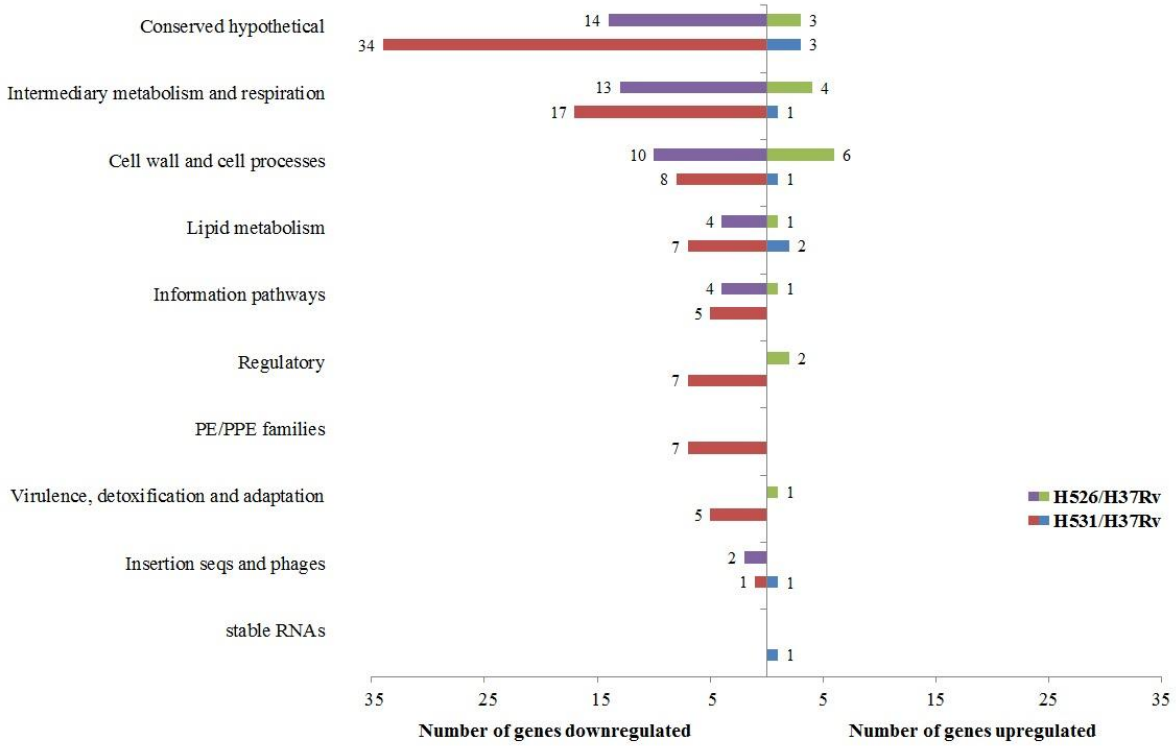

Figure 4.4 Significantly up- and downregulated genes (log_2 fold change >1.0) in *rpoB* mutants (H526, H531) relative to expression in the WT progenitor (H37Rv) grouped according to functional categories.

It was noted that a number of transcriptional regulators were differentially expressed in each of the *rpoB* mutants. Table 4.2 provides an overview of 2 regulators which were upregulated in H526 and 7 which were downregulated in H531, as depicted in the "Regulatory" category in Figure 4.4. Interestingly, WhiB7 was the only transcriptional regulator that was differentially expressed in both mutants, albeit in opposite directions (Table 4.2).

Table 4.2 Differentially expressed transcriptional regulators identified when comparing expression profiles of *rpoB*526 and *rpoB*531 mutants with that of WT H37Rv.

	LOCUS	FOLD CHANGE (log_2)	FUNCTION
H526	**Rv1816**	+1.24	-
	Rv3197A (*whiB7*)	+1.34	Plays a role in redox homeostasis and intrinsic drug resistance (99)
H531	**Rv0792c**	-1.26	-
	Rv1167c	-1.09	In an operon with *lipX* (PE11)
	Rv1395	-1.64	Involved in cytochrome P450 regulation (100)
	Rv2912c	-1.74	In an operon with Rv2913c which encodes D-amino acid hydrolase
	Rv3082c (*virS*)	-1.16	Substrate of PknK; regulates the *mymA* operon (101, 102)
	Rv3197A (*whiB7*)	-2.03	Plays a role in redox homeostasis and intrinsic drug resistance (99)
	Rv3833	-1.07	-

When comparing the expression profiles of the two *rpoB* mutants, it was found that only 8 genes were differentially expressed in both (Figure 4.5). Rv0789c-Rv0791c (hypothetical proteins), Rv1131 (probable methylcitrate synthase) and Rv3453 (conserved transmembrane protein) were downregulated in both mutants, whereas Rv3197A (*whiB7*), Rv1258c (*tap*) and Rv2416c (*eis*) were upregulated in H526 and downregulated in H531. This was a striking observation as WhiB7 regulates both *tap* and *eis* in *M. tuberculosis*. Other genes which have been found to play a role in the WhiB7 regulon include Rv2725c (*hflx*), Rv1988 (*erm*) and Rv2301 (*cut2*) (103). Upon closer inspection it was discovered that these genes followed the same trend as *whiB7* in the two *rpoB* mutants (Table 4.3)

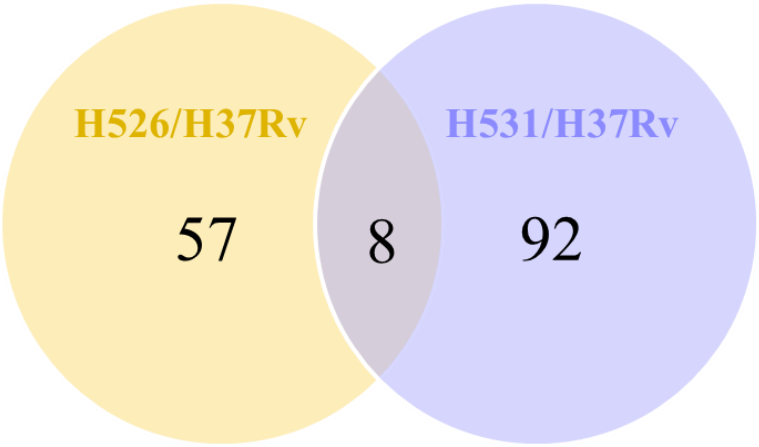

Figure 4.5 Venn diagram illustrating the distribution of significantly differentially expressed genes in different *rpoB* mutants (H526, H531) compared to WT progenitor H37Rv.

Table 4.3 Differential expression of genes that are part of the WhiB7 regulon.

			FOLD CHANGE (log_2)	
LOCUS	**GENE**	**GENE PRODUCT**	**H526/H37Rv**	**H531/H37Rv**
Rv1258c	*tap*	membrane transport protein	1.15	-2.05
Rv1988	*erm*	methyltransferase	(0.99)	-0.90
Rv2301	*cut2*	cutinase	(0.38)	-0.86
Rv2416c	*eis*	enhanced intracellular survival protein	1.44	-1.79
Rv2725c	*hflx*	GTP-binding protein	0.98	-0.40
Rv3197A	*whib7*	transcriptional regulator	1.34	-2.03

Legend to Table 4.3: Values in parentheses indicate an expression ratio with a p-value of >0.05.

In order to investigate other cellular pathways which may be influenced by the two different *rpoB* mutations, the lists of differentially expressed genes were submitted to KEGG Mapper (http://www.genome.jp/kegg/tool/map_pathway1.html). To gain a better perspective of global gene expression in the mutants, a log_2 fold change threshold of >0.5 was used in order to include a greater number of genes in this analysis. The KEGG PATHWAY database includes 7 global metabolic maps for *M. tuberculosis*. Table 4.4 gives an overview of the number of

differentially expressed genes involved in each of these major pathways. A closer look at the sub-categories of the global pathways (listed in Appendix E) revealed a vast number of pathways that were downregulated in H531. Pathways where at least 10 genes were involved include: limonene and pinene degradation (n=15); propanoate metabolism (n=13); butanoate metabolism (n=13); geraniol degradation (n=11); valine, leucine and isoleucine degradation (n=11); fatty acid metabolism (n=10) and tryptophan metabolism (n=10).

Table 4.4 The number of differentially expressed genes involved in each of the global metabolic pathways according to the KEGG PATHWAY database for *M. tuberculosis.*

Global Pathway	**Number of Differentially Expressed Genes** (log_2 fold change > 0.50; p-value < 0.05)			
	↑H526	**↓H526**	**↑H531**	**↓H531**
Metabolic pathways	6	29	13	65
Biosynthesis of secondary metabolites	3	21	9	42
Microbial metabolism in diverse environments	2	10	4	37
Carbon metabolism	2	7	2	11
Biosynthesis of amino acids	-	6	5	6
2-Oxocarboxylis acid metabolism	-	-	-	4
Degradation of aromatic compounds	-	-	-	-

4.2.3 The Effect of Genetic Background on Gene Expression in *rpoB* Mutants

To determine whether the respective *rpoB* mutations caused similar transcriptomic changes in both *in vitro* selected mutants and closely related clinical isolates, the expression profiles for all the *rpoB* mutant sets (H526/H531; M526/M531 and C526/C531) were compared and analysed (Figure 4.6). As progenitors were not included in this analysis, gene expression in the *rpoB*526 mutants is relative to the gene expression in the *rpoB*531 mutant, unless stated otherwise.

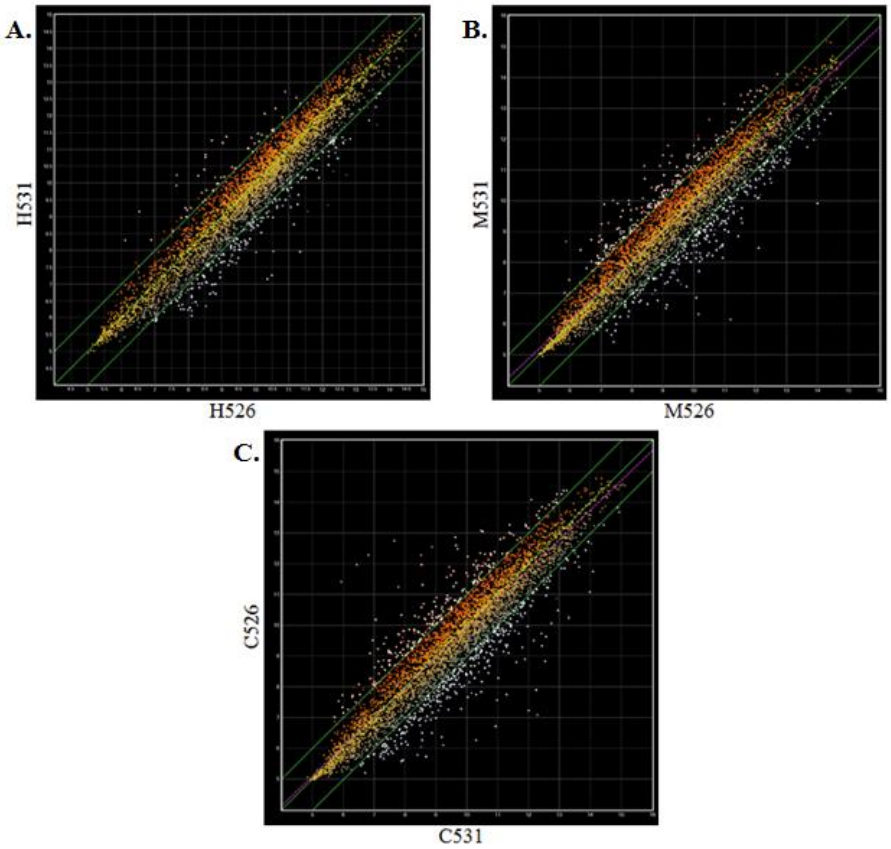

Figure 4.6 Scatter plots illustrating global gene expression when comparing data sets A: H526/H531; B: M526/M531 and C: C526/C531. White dots bordering parallel green lines represent significantly differentially expressed genes (p-value <0.05; $\log_2$ fold change >1.0).

Figure 4.7 shows 218 (H526/H531); 477 (M526/M531); and 552 (C526/C531) genes were differentially expressed in each of the comparisons, respectively. This indicates that a greater number of genes were differentially expressed in the clinical isolates (C526; C531) and *in vitro* mutants (M526; M531) generated from a drug-sensitive clinical isolate compared to the *in vitro* mutants generated from H37Rv.

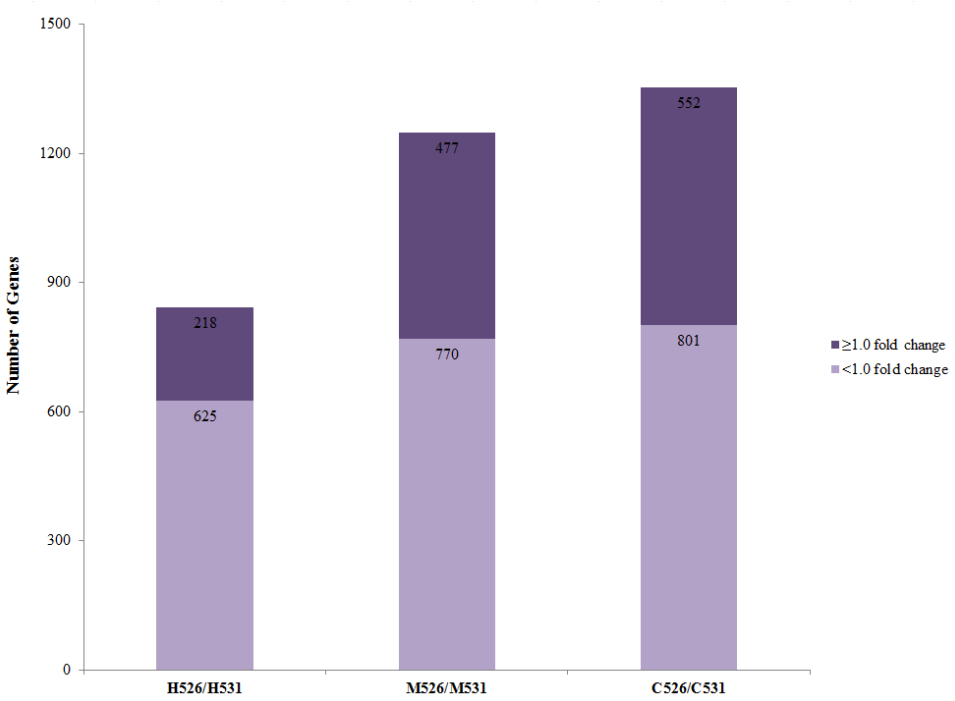

Figure 4.7 Distribution of significantly differentially expressed genes in *rpoB*526 mutants relative to expression in *rpoB*531 mutants across different genetic backgrounds.

Figure 4.8 gives an illustration of these genes when grouped according to functional categories. Similar to the previous progenitor-mutant analysis, the majority of differentially expressed genes were in the conserved hypothetical, intermediary metabolism and respiration, and cell wall and cell processes categories across all three comparisons. In 9 of the 10 functional categories, a larger number of genes were differentially expressed in mutant sets C526/C531 and M526/M531 compared to mutant set H526/H531.

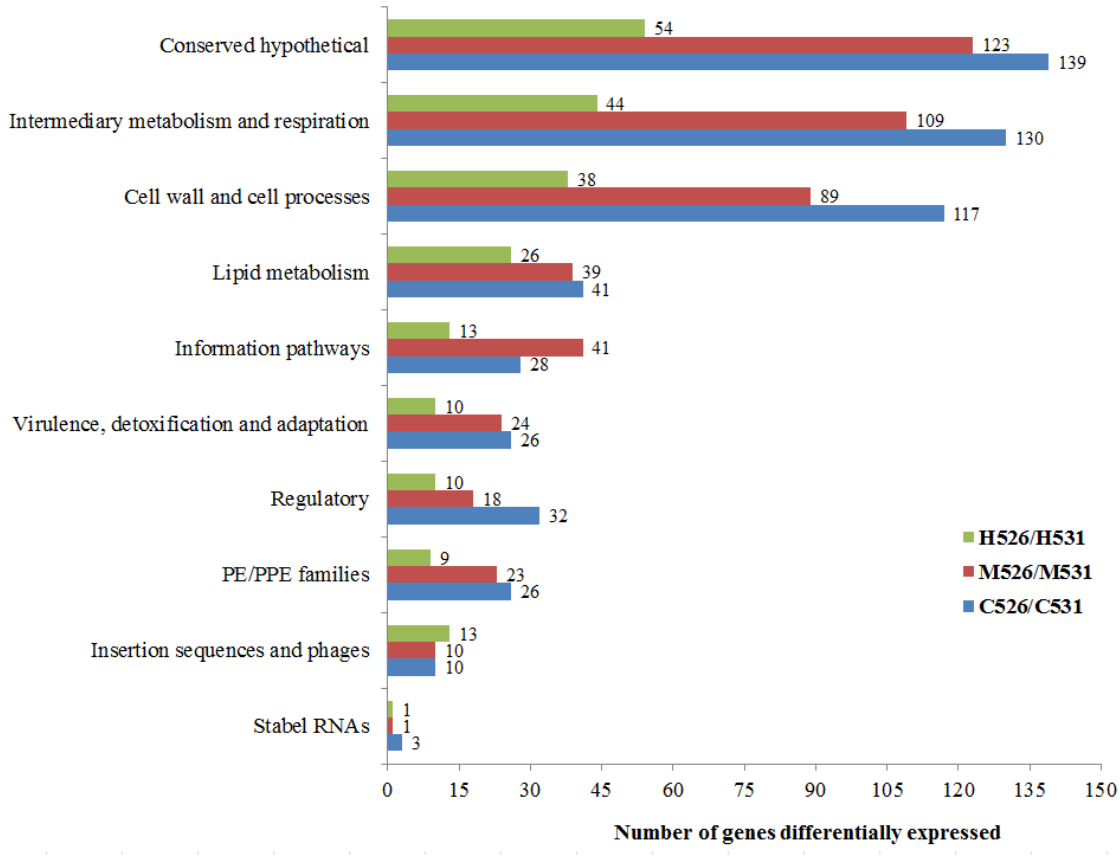

Figure 4.8 Distribution of differentially expressed genes when comparing *rpoB*526 and *rpoB*531 mutants across different genetic backgrounds, grouped according to functional categories.

A three-way comparison of the datasets showed that 28 genes were commonly differentially expressed (up- or down-regulated in both sets) in all *rpoB*526 mutants relative to the *rpoB*531 mutants (Figure 4.9). Only genes which were consistently upregulated or consistently downregulated were included in overlapping subsets to demonstrate a common trend among samples with the same *rpoB* mutation. Differentially expressed genes in the H526/H531 comparison had 62 and 58 genes in common with mutant sets M526/M531 and C526/C531, respectively. However, 149 genes were commonly differentially expressed in M526/M531 and C526/C531 gene sets, which is an indication of the similarity between samples with a clinical genetic background.

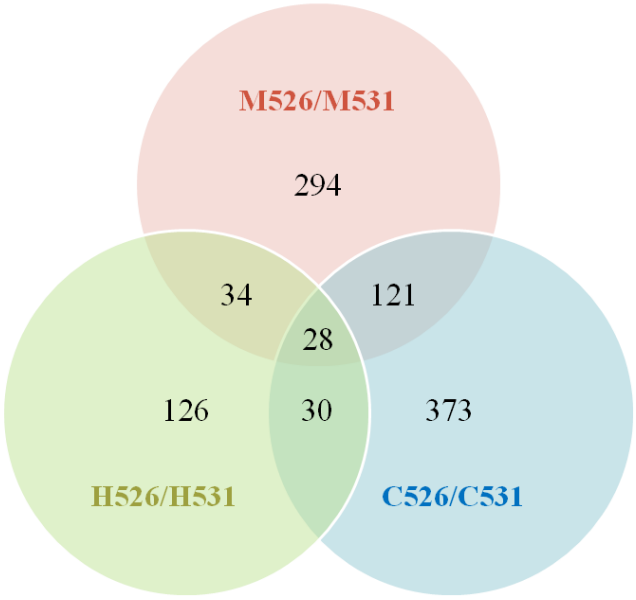

Figure 4.9 Distribution of significantly differentially expressed genes across three sets of *rpoB* mutants (H526/H531; M526/M531 and C526/C531). Only genes which were up/downregulated in all compared datasets were included in overlapping subsets.

Two operons were found in the group of 28 genes which were differentially expressed in all mutant sets, both playing a role in intermediary metabolism and respiration (Table 4.5). The operon comprising of Rv1121 (*zwf1*) and Rv1122 (*gnd2*) was upregulated in the *rpoB*531 mutants, whereas Rv3001c (*ilvC*), Rv3002c (*ilvN*) and Rv3003c (*ilvB1*) were downregulated.

Table 4.5 Commonly differentially expressed genes when comparing all three sets of *rpoB* mutants (H526/H531; M526/M531; C526/C531). Operons referred to in text are highlighted in red.

FC*	LOCUS	PRODUCT	EXPRESSION RATIO (log_2)		
			H526 / H531	**M526 / M531**	**C526 / C531**
1	Rv0950c	conserved hypothetical protein	-1.01	-2.02	-1.99
	Rv1535	hypothetical protein	+1.52	+1.24	+1.39
	Rv1883c	conserved hypothetical protein	+1.01	+1.59	+1.35
	Rv1907c	hypothetical protein	+1.33	+1.62	+2.20
	Rv1947	hypothetical protein	-1.33	-1.70	-2.53
	Rv2271	conserved hypothetical protein	-1.12	-1.21	-2.67
	Rv2960c	hypothetical protein	-1.11	-1.12	-1.21
	Rv3074	conserved hypothetical protein	+1.42	+1.03	+1.72
	Rv3288c	hypothetical protein (UsfY)	+1.25	+1.17	+1.25
	Rv3647c	conserved hypothetical protein	+1.74	+1.71	+2.45
2	Rv1121	glucose-6-phosphate 1-dehydrogenase (Zwf1)	-1.27	-1.80	-1.73
	Rv1122	6-phosphogluconate dehydrogenase (Gnd2)	-1.18	-1.97	-1.96
	Rv1412	riboflavin synthase alpha chain (RibC)	+1.30	+1.02	+1.15
	Rv2918c	protein-pii uridylyltransferase (GlnD)	-1.17	-1.31	-1.25
	Rv3001c	ketol-acid reductoisomerase (IlvC)	+1.51	+1.68	+1.65
	Rv3002c	acetolactate synthase small subunit (IlvN)	+1.49	+1.74	+1.58
	Rv3003c	acetolactate synthase large subunit (IlvB1)	+1.65	+1.97	+1.92
	Rv3247c	thymidylate kinase (Tmk)	+1.24	+2.78	+1.56
3	Rv0686	membrane protein	+1.72	+1.90	+1.73
	Rv1184c	hypothetical exported protein	-1.04	-2.29	-3.41
	Rv1686c	conserved membrane protein	+1.34	+1.01	+1.65
	Rv2253	secreted protein	+1.61	+3.83	+1.42
	Rv3289c	transmembrane protein	+1.34	+2.37	+2.74
4	Rv1080c	transcription elongation factor (GreA)	+1.33	+1.90	+1.38
	Rv3202c	ATP-dependent DNA helicase	+1.48	+1.23	+1.10
5	Rv0891c	transcriptional regulator	-1.11	-1.64	-2.98
	Rv3197A	transcriptional regulator whib-like (WhiB7)	+3.38	+5.02	+2.56
6	Rv3385c	possible antitoxin (VapB46)	+1.15	+1.29	+1.70

Legend to Table 4.5: *Functional Categories: 1 - Conserved hypothetical; 2 - Intermediary metabolism and respiration; 3 - Cell wall and cell processes; 4 - Information pathways; 5 – Regulatory; 6 - Virulence, detoxification and adaptation.

As highlighted in Table 4.5, *whiB7* is downregulated by a $\log_2$ fold change of 2.57-5.02 in all three *rpoB531* mutants. Parallel to the findings in the progenitor-mutant analysis, genes which have been shown to form part of the WhiB7 regulon display a similar trend in the *rpoB* mutants across different genetic backgrounds (Table 4.6).

Table 4.6 Expression ratios of genes that are part of the WhiB7 regulon when comparing samples with different *rpoB* mutations.

			EXPRESSION RATIO ($\log_2$)		
LOCUS	**GENE**	**GENE PRODUCT**	**H526 / H531**	**M526 / M531**	**C526 / C531**
Rv1258c	*tap*	membrane transport protein	3.21	2.52	(0.61)
Rv1988	*erm*	methyltransferase	1.89	1.40	0.87
Rv2301	*cut2*	cutinase	1.23	(0.53)	0.69
Rv2416c	*eis*	enhanced intracellular survival protein	3.23	3.24	(0.55)
Rv2725c	*hflx*	GTP-binding protein	1.38	1.39	0.86

Legend to Table 4.6: Values in parentheses indicate an expression ratio with a p-value of >0.05.

4.2.4 Transcriptomic Response to Rifampicin Exposure

Expression profiles of all RIF exposed (M526; M531; C526; C531) and control samples were compared to determine the effect of exposing RIF-resistant clinical isolates and *in vitro* selected mutants to the critical concentration (2µg/mL) of RIF. No significant differential expression was observed in any of the comparisons as demonstrated in the scatterplots in Figure 4.10.

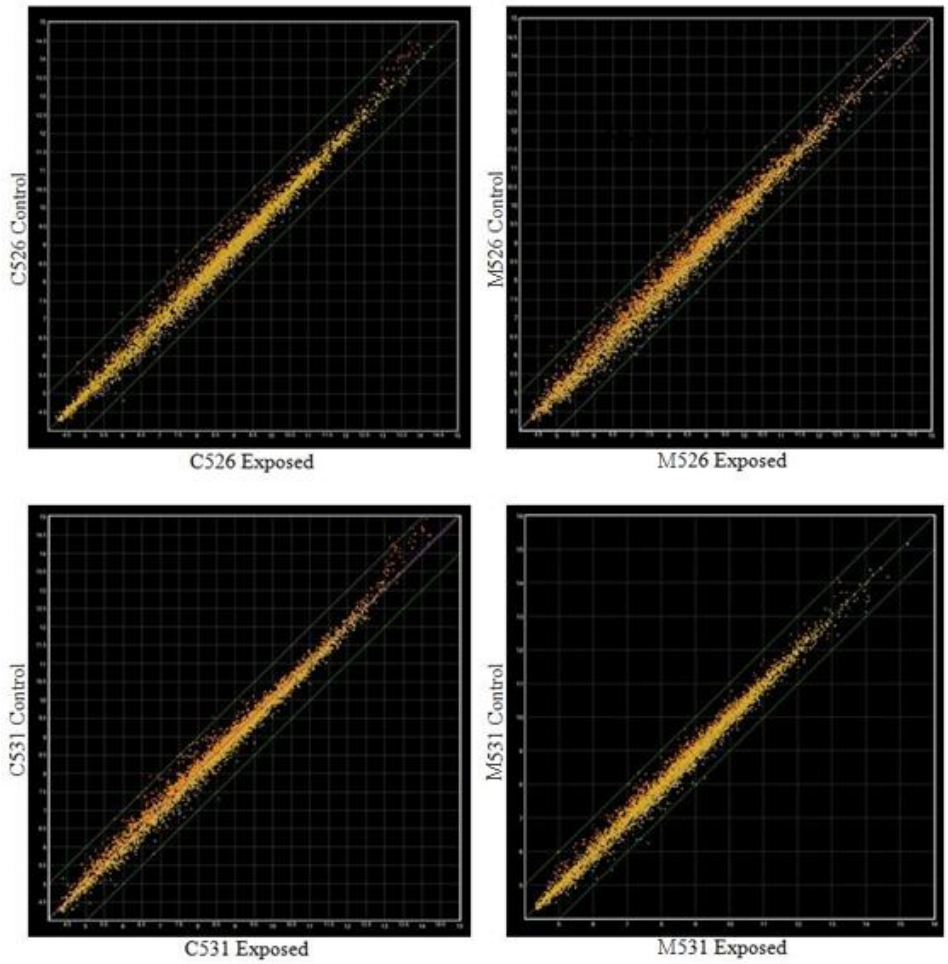

Figure 4.10 Scatterplots illustrating no significantly differential gene expression in cultures exposed to 2 µg/mL RIF for 24 hours compared to unexposed cultures.

4.3 Whole Genome Sequencing

4.3.1 Quality Control

Whole genome sequencing of the drug-sensitive progenitor strains (M0; H37Rv) and *in vitro* selected mutants (H531; H526; M526) was done to determine whether the *in vitro* mutants harbour additional SNPs (besides known *rpoB* mutations) or short in/dels which may influence gene expression. FastQC reports were generated for each the sequenced genomes in order to assess the quality of the raw sequence reads (Appendix F). Once quality control

parameters and basic statistics were examined and found to be satisfactory, data was used for subsequent computational analysis (Table 4.7).

Table 4.7 Basic statistics for whole genome sequences provided by SAMTools and FastQC.

	M0	**M526**	**M531$^{\Delta}$**	**H37Rv**	**H526**	**H531**
GC Content (%)	62	62	63	62	63	63
Mapping Coverage (%)	96.71	95.85	70.98	98.65	98.27	98.73
Number of Reads Mapped	4582019	4098689	18541342	4368072	5276570	5199461
Mean Read Length (bp)	202	177	100	197	203	210
Coverage Depth* (fold)	209.81	164.45	420.29	195.06	242.81	247.51

**Legend to Table 4.8: ***Coverage depth = (number of reads mapped x mean read length)/4411532

$^{\Delta}$Sequenced on the Illumina sequencing platform (all other samples sequenced using the Ion Torrent PGM Sequencer)

4.3.2 Computational Analysis

The genomes were aligned to the H37Rv reference genome prior to comparing mutant genomes to the respective progenitors (Table 4.8). SNP analysis confirmed that the *rpoB* mutation was present in M526 with no other SNPs or in/dels being found in this sample when compared to M0. For M531, an additional synonymous SNP was observed in *TB16.3* (Rv2185c). Furthermore, a non-synonymous SNP was found in each of the *in vitro* mutants generated from H37Rv: a P379L polymorphism in Rv2284 (esterase) was identified in H526 and a Y109D polymorphism in Rv0208c (probable methyltransferase) was seen in H531. Once confirmed in GenomeView, these high-confidence SNPs were validated by PCR amplification and Sanger sequencing of the target regions, as described in section 3.3.5 (Table 4.9).

Table 4.8 Selection criteria used for the identification of SNPs and in/dels in *in vitro* selected mutants (M526; M531; H526; H531) compared to the respective progenitors (M0; H37Rv).

	M526	**M531**	**H526**	**H531**
Sequencing Platform	Ion Torrent	Illumina	Ion Torrent	Ion Torrent
SNPs called by GATK and SAMtools (Compared to H37Rv)	1383	1531	98	93
Unique to mutant (SNPs not found in progenitor)	83	38	11	9
Quality value ≥1000	30	9	3	2
Confirmed in GenomeView	1	2	2	2
Validated by Sanger sequencing	1	2	2	2
In/dels called by GATK (Compared to H37Rv)	2579	139	54352	53358
Unique to mutant (In/dels not found in progenitor)	982	5	26821	26343
Quality value ≥1000	2	5	1926	2368
Confirmed in GenomeView	0	0	0	0

Table 4.9 Validated SNPs identified through bioinformatic analysis of *in vitro* selected mutants (H526; H531; M526; M531).

Sample	**Locus**	**Gene**	**Gene Product**	**SNP**	**Amino Acid Change**
M526	Rv0667	*rpoB*	β-subunit RNA polymerase	C → T	H445Y*
M531	Rv0667	*rpoB*	β-subunit RNA polymerase	C → T	S450L*
	Rv2185c	*TB16.3*	hypothetical protein	A → C	G101G
H526	Rv0667	*rpoB*	β-subunit RNA polymerase	C → T	H445Y*
	Rv2284	*lipM*	esterase	C → T	P379L
H531	Rv0667	*rpoB*	β-subunit RNA polymerase	C → T	S450L*
	Rv0208c	*trmB*	putative methyltransferase	T → G	Y109D

Legend to Table 4.10: *Codon positions 445 and 450 in *M. tuberculosis* correspond to codon position 526 and 531 in the gene sequence of *E. coli*.

A protein model was generated to determine whether the non-synonymous SNP in *trmB* (Rv0208c) could influence the structure of the gene product, a putative methyltransferase (Figure 4.11). The sequence identity for this model was 34.74%, with a QMEAN4 Z-score of 0.66 which indicates that this is a medium-accuracy model. The sequence identity and QMEAN4 Z-score for the LipM model was much lower (26.39% and 0.33, respectively) therefore a reliable 3D protein model could not be generated to visualize the position of the polymorphism in *lipM*. Lastly, amino acid sequences were submitted to SIFT which predicted that the polymorphisms in both *trmB* and *lipM* would affect protein function with scores of 0.02 and 0.00, respectively.

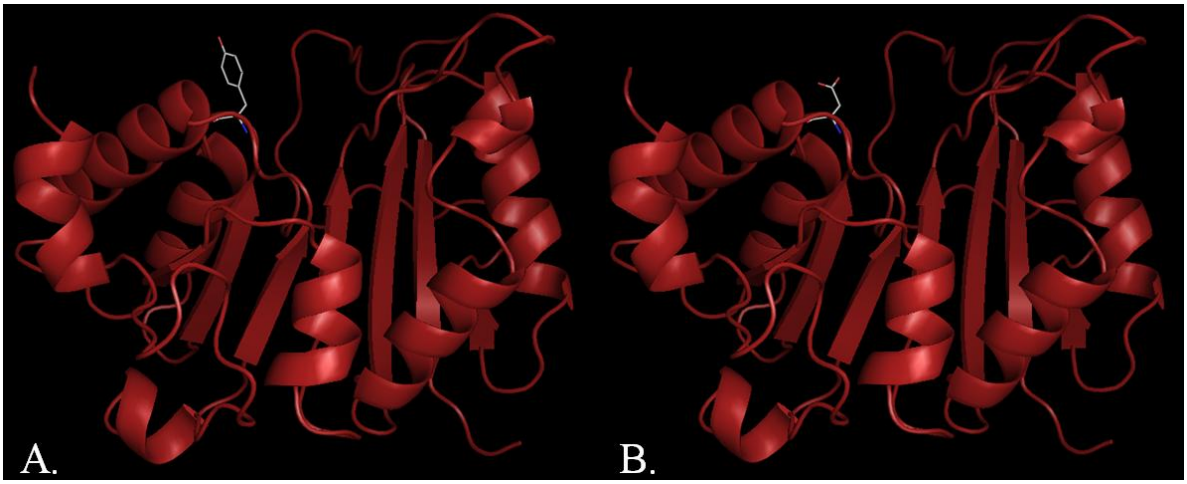

Figure 4.11 Protein model illustrating the amino acid change, tyrosine (A) to aspartic acid (B), caused by the polymorphism in *trmB*. Amino acid side chains are illustrated in white.

Chapter 5

Discussion

The present study is the first of its kind to investigate the effect of *rpoB* mutations on global gene expression in both clinical isolates and *in vitro* mutants of RIF-resistant *M. tuberculosis*. This research encompassed the following: i) transcriptional profiling of *M. tuberculosis* with different *rpoB* mutations to determine the effect of these mutations on gene expression; ii) exposing RIF mono-resistant isolates to 2 µg/mL of RIF to investigate the effect of this drug on gene expression and iii) whole genome sequencing of all *in vitro* generated mutants used in this study to assess whether additional SNPs or in/dels may have influenced gene expression. The findings from this study have provided novel insight into understanding the influence of different *rpoB* gene mutations on the transcriptome of *M. tuberculosis*.

5.1 Deciphering the Impact of *rpoB* Mutations on Gene Expression

The growth curves constructed for each of the isolates revealed that WT H37Rv displayed a faster growth rate compared to both the H526Y and S531L *in vitro* mutants generated from this strain. Likewise, a similar trend was seen for the drug-sensitive clinical isolate and *in vitro* selected S531L mutant. This finding demonstrates the fitness cost inferred by *rpoB* mutations when the bacterium is not cultured in the presence of RIF (84). When comparing the RIF mono-resistant clinical isolates to the *in vitro* generated mutants, it was apparent that the clinical strains grow at a much slower rate. This may be because the laboratory strain of *M. tuberculosis* has adapted to *in vitro* growth conditions. Alternatively, this may be due to the existence of additional mutations in the clinical strains, acquired while in the human host. These mutations help the bacterium to adapt and survive in harsh environments; however they can lead to an increased fitness cost when grown in a system without cellular stress such as the host immune response and antibiotic pressure.

Comparison of H37Rv gene expression to each of the *rpoB* mutants revealed that the majority of differentially expressed genes, excluding hypothetical proteins, belonged to the intermediary metabolism/respiration, and cell wall/cell processes functional categories. In both cases, as well as all other functional categories, the majority of the genes involved were downregulated in the mutant strains. This trend was amplified in H531, possibly due to the transcriptional regulators which were downregulated in the *rpoB*531 mutant. Of the 7 differentially expressed regulators, WhiB7 is the best described in literature and has been linked to both redox homeostasis and intrinsic drug resistance (99). WhiB7 expression is induced upon exposure to fatty acids and antibiotics, such as erythromycin and tetracycline, and plays an important role in regulating gene expression by binding to region 4 of the housekeeping σ factor, σ^A (103, 104). However, it is uncertain whether the global trend of downregulation seen in the *rpoB*531 mutant can be accounted for by the downregulation of a single transcriptional regulator, considering that *whiB7* was upregulated in the *rpoB*526 mutant, which also displayed a vast number of downregulated genes. It is possible that the effect of WhiB7 is amplified due to other regulators which were downregulated in the *rpoB*531 mutant, and not in the *rpoB*526 mutant, however these transcriptional regulators have not been studied in detail.

Differential expression of 57 and 92 genes were unique to the H526Y and S531L mutants of *M. tuberculosis* H37Rv, respectively. Based on these findings, we suggest that the position of the *rpoB* mutation has a substantial effect on gene expression. Surprisingly, very few genes were differentially expressed in both the *rpoB* mutants, suggesting that each mutation has a distinct effect on the transcriptional properties of RNAp. Five genes were downregulated in both mutants: three hypothetical proteins (Rv0789c-Rv0791c), citrate synthase (Rv1131), a conserved membrane protein (Rv3453); and three genes were upregulated in the *rpoB*526

mutant and downregulated in the *rpoB*531 mutant: *whiB7*, *tap* and *eis*. Furthermore, *hflx*, *erm* and *cut2*, which also play a role in the WhiB7 regulon, followed the same trend in both *in vitro* generated mutants, albeit at a lower level of expression. These same genes were downregulated in the *rpoB*531 mutants with a clinical genetic background, which lends support to the notion that WhiB7 plays a major role in the downregulation of metabolism and cellular processes we observe in the *rpoB*531 mutants.

It is also possible that other factors, not accounted for when evaluating gene expression of transcriptional regulators, could be playing a role. If mutated RNAp influences RNAp-ligand or RNAp-DNA interactions, this may be the cause of the differences we observe when comparing the transcriptomes of different *rpoB* mutants. Based on factors which interact with the β and β' subunits of RNAp, we aimed to systematically evaluate the origin of the downregulation of metabolic pathways observed in the *rpoB* mutants, especially those with a S531L mutation. Firstly, as p/ppGpp and CarD directly bind to RNAp in order to mediate the stringent response in *M. tuberculosis*, we evaluated whether an *rpoB* mutation may influence this interaction (Appendix G). Secondly, it is possible that *rpoB* mutations influence the affinity of RNAp for alternate σ factors, which may drive a particular transcriptional response (Appendix H). However, after closer investigation, we can conclude that neither of these scenarios appeared to fully explain the expression profiles of the *rpoB* mutants in this study. At this stage, an appealing speculation is one which includes two scenarios: *rpoB*531 mutations lead to the induction of a tapered stringent response as well as differential expression of genes which are part of the Whib7 regulon, both of which lead to the global change in gene expression we observe in the *rpoB* mutants.

It is also possible that factors beyond our current understanding can account for the link between *rpoB* mutations and mycobacterial gene expression. A primary example includes the interaction between RNAp and sRNA. Unfortunately, our knowledge of sRNA in mycobacteria is limited, which prevents us from fully understanding this aspect of transcriptional regulation. However, a recent study revealed that MTS2823 sRNA plays an important role in mediating the repression of genes involved in exponential growth in *M. tuberculosis* (47). MTS2823 was overexpressed to investigate its role in stationary phase, which lead to widespread downregulation of a vast number of genes, primarily those involved in energy metabolism and macromolecular synthesis. Although the regulatory mechanism of MTS2823 sRNA remains unclear, it is thought that it is comparable to 6S RNA in *E. coli*, which binds directly to RNAp (44, 45). It is possible that a sRNA-RNAp interaction such as this is influenced by the presence of an *rpoB* mutation, which may cause differential expression of genes under the control of regulatory sRNA. To further support this theory, a study by Windbichler *et al* observed a number of sRNAs that bind to the β subunit of RNAp in *E. coli* (49).

Alternatively, RNAp processes such as transcription elongation and termination could also be affected by the presence of an *rpoB* mutation. A study in *E. coli* has revealed that a H526Y mutation in *rpoB* decreases transcription termination at the tryptophan operon attenuator (105). As the termination efficiency was not reduced at operons which do not form transcript termination structures, the authors suggested that particular *rpoB* mutations lead to reduced efficiency of termination only when mutated RNAp encounters these structures. If the β subunit does indeed play an active role in the recognition of termination signals, this could indirectly affect the level of gene expression of certain operons when particular *rpoB*

mutations are present. Further studies are needed to investigate these findings in *M. tuberculosis*.

As mutations at codon 531 and 526 of *rpoB* lead to altered gene expression in *M. tuberculosis* H37Rv, we aimed to confirm whether the same trend is observed when comparing strains with different genetic backgrounds. In contrast to the H37Rv mutant set, a greater number of genes were differentially expressed when comparing the *rpoB* mutants with a clinical genetic background. Similar to our findings from evaluating the growth curves in this study, this is thought to be due to the genetic variations in clinical strains of *M. tuberculosis* which have had the opportunity to adapt in the human host.

When grouping differentially expressed genes according to functional categories, the majority were found to be in the intermediary metabolism/respiration, and cell wall/cell processes categories, comparable to our findings in the progenitor-mutant analysis. In 9 of the 10 functional categories, a larger number of genes were differentially expressed in mutant sets with a clinical genetic background compared to the H37Rv mutant set, which emphasizes the difference between strains with a clinical background and laboratory strains such as H37Rv. This is a reminder that research based on the use of reference strains may not be a true representation of *M. tuberculosis* isolated from the host, despite the array of advantages that working with such strains have to offer.

Two operons were identified in the 28 genes which were commonly differentially expressed when comparing all three mutant sets. The operon comprising of *ilvC*, *ilvN* and *ilvB1* displayed downregulation in the S531L mutants, while *zwf1* and *gnd2* exhibited upregulation.

The *ilvC* gene encodes for ketol-acid reductoisomerase, while *ilvN* and *ilvB1*encode for the small and large subunits of acetolactate synthase, respectively. Acetolactate synthase catalyses the first step of the branched-chain amino acid biosynthesis pathway. A Δ*ilvB1* deletion mutant of *M. tuberculosis* exhibits auxotrophy of all three branched-chain amino acids, namely valine, leucine and isoleucine (106). Furthermore, the mutant strain displays diminished growth *in vitro* and attenuated virulence *in vivo*. Downregulation of these genes in the *rpoB*531 mutants could be an indication of decreased branched-chain amino acid biosynthesis. This supports our hypothesis that the full stringent response is not triggered in *rpoB*531 mutants, as increased amino acid biosynthesis is expected once the stringent response is initiated.

Zwf1 and Gnd2 are enzymes which play a role in the pentose phosphate pathway and are responsible for the production of NADPH. Deletion of the gene which encodes Gnd1 has been shown to lead to increased sensitivity to oxidants in *Saccharomyces cerevisiae*, therefore it is considered that a sufficient production of NADPH is necessary to protect cells from oxidative stress (107, 108). Furthermore, mutations in *zwf1* and *gnd2* are associated with decreased tolerance of *S. cerevisiae* to furfural, whereas overexpression of *zwf1* allows the microorganism to grow in toxic concentrations of this inhibitor, emphasizing the vital role of NADPH in stress tolerance (109). We hypothesize that differential expression of these genes in the *rpoB* mutants may influence the redox balance within the cell. Furthermore, a reductive shift such as this could contribute to the change in *whiB7* expression which was observed, as Whib7 has been shown to play a role in redox homeostasis (99).

5.2 Transcriptomic Response of *rpoB* Mutants to Rifampicin Exposure

A recent study by Louw *et al* indicated that exposing RIF mono-resistant strains to the critical concentration of RIF was able to induce tolerance to ofloxacin (OFL) (6). Strains with the *rpoB*526 mutation pre-exposed to RIF showed an increase in the MIC of OFL after 7 days and strains with the *rpoB*531 mutation showed induction of OFL tolerance after only 24 hours. This was thought to be linked to the activation of efflux pumps, as the addition of efflux pump inhibitors restored OFL susceptibility in MDR strains of *M. tuberculosis*. If *rpoB*531 mutations are able to condition the bacterium to become tolerant to OFL, this could explain why these mutations are more prevalent in RIF resistant clinical strains. To investigate this hypothesis, we exposed RIF mono-resistant clinical isolates and *in vitro* generated mutants with different *rpoB* mutations to the critical concentration of RIF for 24 hours. Surprisingly, no significant differential expression was observed when comparing the gene expression profiles of RIF exposed and unexposed samples. This remained true for strains with the *rpoB*531 and *rpoB*526 mutations, indicating that the presence or position of *rpoB* mutations does not influence the transcriptional response of *M. tuberculosis* to the critical concentration of RIF used in this experiment.

Other studies have demonstrated the effect of RIF on gene expression in *M. tuberculosis*, however much higher concentrations of the drug were used in all of these instances, usually correlating to the MICs for the individual strains. As previous studies have shown that blood serum concentrations of RIF can be as high as 8 - 24 µg/mL, we suggest that higher concentrations of RIF be tested to examine the influence of a clinically relevant concentration of RIF on gene expression in *M. tuberculosis* (110, 111).

5.3 Whole Genome Sequencing of *In Vitro* Generated Mutants

Bioinformatic analysis of whole genome sequence data for the *rpoB*526 mutant generated from H37Rv revealed a non-synonymous SNP in a gene which encodes an esterase, *lipM*. LipM is one of 21 annotated proteins in the Lip family of enzymes in *M. tuberculosis* which are thought to play a role in the hydrolysis of host lipids to store trialglycerol as an energy source for the cell (8, 112). Transcriptomic data revealed that the SNP did not influence expression of this gene, however the SIFT program predicted that the SNP would have an effect on protein function. Unfortunately, an accurate structural model could not be generated for this protein, therefore further investigation of the polymorphism proved difficult.

In silico modelling of *trmB* to visualize the SNP identified in the *rpoB*531 mutant generated from H37Rv, revealed that the tyrosine to aspartic acid amino acid change at codon 109 is on the surface of the protein. Although the polymorphism does not appear to alter the structure of the protein, SIFT predicted that it would have an impact on protein function. According to information on TBDB, 4 different polymorphisms have been described in this gene; however this particular SNP has not been reported. The *trmB* gene codes for a tRNA-methyltransferase which is essential for mycobacterial growth (113). The function of this enzyme is to catalyse the N^7-methylguanine modification at position 46 in the variable loop of tRNA during tRNA maturation. Expression of this gene was slightly lower (0.79 fold) in the *rpoB*531 mutant of H37Rv compared to the *rpoB*526 mutant and no differential expression was seen when comparing the *in vitro* mutant set without this SNP. This suggests that the polymorphism may modify gene expression or enzyme activity. A study by Fu and colleagues demonstrated induction of *trmB* after exposure to 5µg/mL isoniazid and 10µg/mL ethionamide (114). Although the authors did not speculate about the reason for this, further investigation may reveal a link between *trmB* and drug resistance in *M. tuberculosis*.

At this stage, we are not able to establish the significance of the synonymous polymorphism identified in Rv2185c (TB16.3) in the *rpoB*531 mutant generated from the drug-sensitive clinical isolate. Previously, the identification of synonymous SNPs was disregarded as it was assumed that their effect on protein function was inconsequential. Over the past decade, however, research based on the Human Genome Project has revealed that these polymorphisms occur at a higher frequency than previously thought and that they can, in fact, have an impact on the structure and function of proteins (115). Synonymous SNPs, also referred to as silent polymorphisms, have since been linked to alterations in mRNA structure and stability, as well as protein translation and folding. In a study which investigated ethambutol resistance in *M. tuberculosis*, a synonymous SNP in *aftA* was found to increase the expression of a downstream gene, *embC,* thereby increasing the level of ethambutol resistance in laboratory strains (116). This is an indication that silent polymorphisms are also able to influence gene expression of surrounding genes. Unfortunately, Rv2185c does not appear to be a part of an operon and gene expression of this gene did not differ between the mutant and progenitor strains. Rv2185c encodes for a hypothetical protein to which a function or cellular pathway has not been assigned; therefore further investigation of this SNP will first require elucidation of the gene product as well as its role in mycobacterial physiology.

Chapter 6

Limitations and Future Studies

Interestingly, the growth curves for the *in vitro* mutants generated from H37Rv do not appear to correlate to the degree of metabolic downregulation in these strains. The *rpoB*526 mutant displayed the slowest growth rate when compared to H37Rv, whereas the *rpoB*531 mutant exhibited the highest number of downregulated pathways. This is an indication that investigating only the transcriptome in these mutants is insufficient when considering the impact of *rpoB* mutations on the physiology of *M. tuberculosis*. Comprehensive studies, which consider the entire process from gene expression to functional gene product, are necessary to gain a better perspective of the effects of mutations on the physiology of *M. tuberculosis*. Furthermore, *M. tuberculosis* which has acquired *rpoB* mutations *in vitro* may not be the ideal model for strains which have acquired RIF-resistance while in the host. Ideally, isolates used in future studies would include a progenitor and *rpoB* mutant from the same patient, obtained before and after acquiring RIF-resistance *in vivo*.

A study by Gagneux *et al*. revealed that the fitness cost of *rpoB* mutations was higher in laboratory-derived H526Y mutants than in S531L mutants, regardless of the strain background (84). In our study, H37Rv and the two *in vitro* mutants generated from this progenitor, as well as the two RIF-resistant clinical isolates appear to correspond with this finding. Interestingly, however, one of the three *rpoB*526 mutants exhibited a similar growth rate to its progenitor for the majority of the exponential phase. To further investigate these results, a competition assay could be done to determine the fitness cost in each of the strains used in this study.

Expression data for the drug-sensitive clinical isolate varied considerably from the *in vitro* mutants generated from this strain, and the same trend was not seen in the progenitor-mutant comparisons for H37Rv, therefore this dataset was excluded from subsequent analyses. As a

result, findings from the H37Rv-mutant comparison could not be confirmed in a second strain of *M. tuberculosis*. The differences that were observed could be due to additional mutations or in/dels acquired by the *in vitro* mutants during the Luria–Delbrück fluctuation assay. However, gene expression between the two *in vitro* mutants proved to be relatively similar and it is unlikely that the same mutational event occurred in both these samples. Furthermore, whole genome sequencing of the strains confirmed that no in/dels or non-synonymous SNPS were present when compared to the progenitor strain. Alternatively, it is possible that these transcriptional differences arose due to unknown technical problems during culture, RNA isolation or microarray sample preparation steps.

To investigate the influence of *rpoB* mutations on RNAp interactions, the affinity of WT and mutated RNAp for p/ppGpp and different σ factors can be evaluated as this will influence how *M. tuberculosis* adapts to external stress and internal stimuli. If RNAp with particular *rpoB* mutations favour binding of certain σ factors or have an increased sensitivity towards p/ppGpp, it could explain why these mutations lead to global changes in gene expression. Likewise, the relationship of WT and mutated RNAp to RNAp-binding sRNA could be investigated to determine whether there is a link between *rpoB* mutations and this aspect of transcriptional regulation.

As the WhiB7 regulon appeared to play a major role in driving differential expression in the *rpoB* mutants studied here, it would be worthwhile to investigate this further. Determining whether comparable differential expression of *whiB7* can be seen in other clinical isolates of RIF resistant *M. tuberculosis* would confirm the findings from this study. Future studies could include transcription profiling of isolates from different strain lineages of *M. tuberculosis* which appear to be driving the MDR-TB epidemic in South Africa, for example Latin

American-Mediterranean (LAM) and Low Copy Clade (LCC) lineages (117). Additionally, *rpoB*526 mutants in this study demonstrated upregulation of the *eis* gene; therefore we are interested to determine whether strains with this mutation would have decreased susceptibility to kanamycin as this would have an impact on the current second-line treatment regimen.

Recent studies have shown that synonymous polymorphisms should not be overlooked, therefore determining the role of the synonymous SNP we identified in TB16.3 would add to the findings in this study. However, as little is known concerning the hypothetical protein for which this gene encodes, investigating the effect of this SNP on mycobacterial physiology is challenging at this stage. Furthermore, different non-synonymous SNPs were found in the *in vitro* generated *rpoB* mutants, therefore we are not able to hypothesise whether either one of the polymorphisms play a role in a compensatory mechanism or drug resistance. Future studies should aim to determine whether the polymorphisms identified in this study are present in other RIF-resistant samples and the impact they have on mycobacterial physiology.

Lastly, as it is uncertain how the SNPs identified in this study may have influenced the transcriptome, differential gene expression observed when comparing progenitor and RIF-resistant mutants cannot be solely accounted for by the respective *rpoB* mutations. Additionally, regions absent from the H37Rv reference genome and repetitive elements such as PE/PPE genes were excluded from the whole genome sequencing analysis; therefore variants in these regions may have been overlooked. Although this hampers our aim to link *rpoB* mutations and the phenotype of *M. tuberculosis*, it is worthwhile to note that before the introduction of next generation sequencing technologies, researchers studied the effects of mutations unaware that additional polymorphisms could be playing a role. Since the dawn of the "omics" era, we will undoubtedly see more studies incorporating this aspect of validation.

Chapter 7

Conclusion

Investigating the transcriptome of RIF-resistant *M. tuberculosis* with different *rpoB* mutations has led to an unexpected finding. Strains with *rpoB* S531L and H526Y mutations exhibit downregulation of multiple metabolic and cellular pathways, which appears to be reminiscent of the transition from replicating to dormant populations of *M. tuberculosis*. This effect is amplified in S531L mutants, irrespective of the strain's genetic background. It is known that certain resistance-conferring mutations lead to a lower fitness cost in *M. tuberculosis*; however it may also be possible that certain RIF-resistance conferring mutations predispose the bacterium to the effect of cellular stress. Findings from this study suggest that *rpoB* S531L mutations could be advantageous to the bacterium in the event of cellular stress, as it is possible that bacilli which exhibit downregulation of numerous cellular pathways will be able to adapt more rapidly to the array of stress factors within the host. However, the mechanism through which these transcriptional changes are brought about remains unclear.

It is likely that the effects of *rpoB* mutation are multifactorial, involving altered RNAp interactions with termination signals, p/ppGpp, σ factors or transcriptional regulators such as Whib7. This indicates that although RIF resistance is a simple process, the aftermath of mutations which confer resistance are more complex than previously thought. It is also possible that differential expression of the Whib7 regulon leads to a plethora of downstream events which would account for the global downregulation we observe in the *rpoB* S531L mutants. Alternatively, there may be aspects of WhiB7 regulation that are currently unknown, which could involve transcriptional regulation of a number of different cellular pathways. Further elucidation of the role of WhiB7 in *rpoB* mutants will rely on findings from future studies.

In this study, results indicate that exposing RIF mono-resistant isolates to the critical concentration of RIF (2 μg/mL) is not sufficient to induce a transcriptional response. Therefore, we were unable to determine whether treatment with a drug to which the bacterium is already resistant would confer an advantage, for example activation of efflux pumps which would decrease the intracellular concentration of other anti-TB drugs.

Synonymous and non-synonymous SNPs that are currently unrelated to RIF-resistance were identified in 3 of the 4 *in vitro* mutants used in this study. This suggests that the mycobacterial genome is constantly evolving, perhaps not as slowly as previously thought. It is thought that the acquisition of polymorphisms is somewhat rare and isolated; however results from this study indicate that additional mutation events are taking place concurrently to those which confer drug resistance. As the *in vitro* mutants used in this study were generated using the Luria–Delbrück fluctuation assay, it can be assumed that mutational pressure would be even greater when the bacterium is in the human host, which would increase the likelihood of polymorphisms incorporated in the mycobacterial genome. Further studies will elucidate whether these seemingly unrelated polymorphisms play a role in the success of *M. tuberculosis* as a pathogen.

In conclusion, results from this study indicate that both the position of the *rpoB* mutation as well as the genetic background of the strain plays an important role in the gene expression profile of *rpoB* mutants. The combination of different drug resistance-conferring mutations and strain background may result in each strain of *M. tuberculosis* having unique characteristics, demanding ever-more sophisticated approaches to understanding the complex physiology of this Machiavellian pathogen.

Appendices

APPENDIX A

Preparation of growth media, reagents and buffers

CULTURE MEDIA:

- **Middlebrook 7H9 Broth**

  4.7 g Middlebrook 7H9

  900 mL ddH_2O

  2 mL glycerol

  500 µL Tween 80

  Autoclave

  Add 100 mL ADC

  Filter Sterilize and store at 4°C

- **Middlebrook 7H10 Agar**

  19 g Middlebrook 7H10

  900 mL ddH_2O

  5 mL glycerol

  Autoclave

  Add 100 mL OADC

  Pour into petri dishes and store at 4°C

- **Albumin-Dextrose-Catalase (ADC)**

  25 g bovine serum albumin

  10 g glucose

  0.75 mL catalase

  Add ddH_2O to a final volume of 500 mL

  Filter-sterilize and prepare 50 mL aliquots

  Store at 4°C

DRUG PREPARATION:

- **4000 µg/mL Rifampicin (light-sensitive)**

  Dissolve 40 mg of RIF in 1 mL DMSO

  Add sterile water to a final volume of 10 mL

  Prepare 500 µL aliquots in aluminium foil-wrapped tubes

  Store at -80°C

GEL ELECTROPHORESIS:

- **10x TBE Buffer (pH 8.3)**

  108 g 0.45 M Tris

  55g 0.44 M boric acid

  7.4g 10 mM EDTA

  Add ddH_2O to a final volume of 1 L

- **1.5% Agarose Gel**

  1.5 g agarose

  1.5 µL ethidium bromide

  100 mL 1x TBE buffer

- **Loading Dye**

  0.25 % xylene cyanol

  50 % glycerol

MICROARRAY HYBRIDIZATION:

- **Phosphate Buffer (pH 8.5-8.7)**

  9.5 mL 1 M K_2HPO_4

  0.5 mL 1 M KH_2PO_4

- **Phosphate Wash Buffer**

  0.5 mL 1 M KPO_4 (pH 8.5)

  15.25 mL MilliQ water

  84.25 mL 95% ethanol

- **Phosphate Elution Buffer**

  Dilute 1M KPO_4 (pH 8.5) to 4 mM with water

- **Aminoallyl-dNTP mix**

  5 µL 100 mM dGTP

  5 µL 100 mM dCTP

  5 µL 100 mM dATP

  3 µL 100 mM dTTP

  4 µL 100 mM aminoallyl-dUTP

  Add MilliQ water to a final volume of 40 µL

  Store at -20ºC

- **Sodium Carbonate Buffer (pH 9.0)**

  10.8 g Na_2CO_3

  80 mL MilliQ water

  Adjust pH to 9.0 with 12 M HCl

  Add MilliQ water to a final volume of 100 mL to give a 1 M solution

  Prepare a 0.1 M solution by diluting 1:10 with water

- **Cy-3 and Cy-5 dyes (light-sensitive)**

  Resuspend each tube of dried product in 73 µL DMSO

  Wrap tubes in aluminium foil

  Store aliquots at -80ºC

DNA/RNA ISOLATION:

- **75% Ethanol**

  150 mL absolute ethanol

  50 mL DEPC-treated water

  Refrigerate at -20°C for use during RNA isolation

- **Extraction Buffer**

  5 % sodium glutamate

  50 mM Tris-HCl (pH 7.4)

  25 mM EDTA

- **Proteinase K Buffer**

  5 % sodium dodecyl sulphate

  100 nM Tris-HCl (pH 7.8)

  50 mM EDTA

- **TE Buffer**

  10 mM Tris-HCl (pH 8.0)

  1 mM EDTA

APPENDIX B

Culturing of *M. tuberculosis* and the initial steps of RNA isolation was done in a BSL-3 laboratory. The following procedures are considered good laboratory practice with regard to the methods used in this study.

- **BIO-SAFETY CABINETS**

Before using the biosafety cabinet (BSC), all surfaces and equipment are disinfected with 10 % Incidin Plus (Antrix Hygiene Ltd, India) and 70 % ethanol. An autoclavable bottle containing 10 % Incidin Plus is used to discard cultures to give a final concentration of no less than 2 %. Pipette tips, inoculating loops, L-shaped spreaders, cuvettes and Eppendorf tubes are rinsed in 10 % Incidin Plus before discarding into a disposable screw-top container. All waste generated within the BSC is placed in an autoclave bag and sealed with autoclave tape. Before removing the autoclave bag, autoclavable container or any other item from the BSC, it is first disinfected with 10 % Incidin Plus and 70 % ethanol. Upon completing work in the BSC, all surfaces and equipment are cleaned as before. Aluminium centrifuge buckets are soaked overnight in a 2% solution of Incidin Plus.

- **INCUBATOR**

The temperature controlled warm room is set at 37ºC. Four minimum/maximum thermometers are attached to the shelving and daily temperature logs are kept to ensure proper functioning. All cultures to be incubated in the warm room are placed in an air-tight secondary container.

- **WASTE REMOVAL**

All waste, including autoclavable bottles, gowns, masks and overshoes, are autoclaved at 121°C for 30 minutes before removing from the BL3 laboratory and placing in designated waste bins for off-site disposal.

APPENDIX C

To prevent RNA degradation during the RNA isolation and purification steps, the following methods were used to maintain an RNase-free environment:

- Once samples were processed and removed from the BSL-3 laboratory, a designated RNase-free workspace; with laminar flow hood, heating block, centrifuge and pipettes; was used.
- RNase*Zap* (Ambion, Applied Biosystems, CA, USA) and 70% ethanol were used to clean all equipment and work surfaces before and after use.
- Powder-free nitrile gloves were used at all times and were changed frequently to prevent RNase contamination.
- RNase-free pipette tips, Eppendorf tubes and reagents were used and sample tubes were kept closed whenever possible.
- Glassware was treated with 1% SDS, rinsed with DEPC-treated water and oven-baked at 100ºC overnight to ensure that it was RNase-free.
- Water used to clean glassware and prepare 75% ethanol was treated with 0.1% DEPC, shaken overnight at 37°C and autoclaved at 121°C for 20 minutes to remove all traces of DEPC.

APPENDIX D

Concentrations of isolated RNA as determined by the Experion RNA StdSens analysis kit and Experion automated electrophoresis station.

		Concentration (ng/µL)	
Sample	**Triplicate number**	**Unexposed**	**RIF Exposed**
C526	1	464.34	530.23
	2	567.06	592.85
	3	761.48	517.88
C531	1	124.36	95.68
	2	332.69	468.73
	3	336.35	432.33
M526	1	692.04	880.87
	2	477.98	517.07
	3	253.22	367.37
M531	1	719.41	633.61
	2	665.59	614.39
	3	653.34	850.27
M0	1	417.53	-
	2	417.97	-
	3	512.73	-
H526	1	371.26	-
	2	869.19	-
	3	487.27	-
H531	1	292.79	-
	2	426.50	-
	3	513.83	-
H37Rv	1	408.21	-
	2	374.04	-
	3	586.79	-

APPENDIX E

Cellular pathways influenced by different *rpoB* mutations. Only pathways with at least 3 differentially expressed genes are listed here. The number of genes involved in each pathway is given in parentheses.

DOWNREGULATED IN H526

Pantothenate and CoA biosynthesis (5)
Glycine, serine and threonine metabolism (4)
Geraniol degradation (4)
Aminoacyl-tRNA biosynthesis (4)
ABC transporters (4)
β-Alanine metabolism (4)
Glutathione metabolism (3)
Cysteine and methionine metabolism (3)
Purine metabolism (3)

UPREGULATED IN H526

Butanoate metabolism (3)

DOWNREGULATED IN H531

Limonene and pinene degradation (15)
Propanoate metabolism (13)
Butanoate metabolism (13)
Geraniol degradation (11)
Valine, leucine and isoleucine degradation (11)
Fatty acid metabolism (10)
Tryptophan metabolism (10)
Lysine degradation (9)
Benzoate degradation (9)
β-Alanine metabolism (9)
Aminobenzoate degradation (9)
Caprolactam degradation (7)
Purine metabolism (7)
Phenylalanine metabolism (7)

UPREGULATED IN H531

Histidine metabolism (4)
Aminoacyl-tRNA biosynthesis (3)
ABC transporters (3)

DOWNREGULATED IN H531

Histidine metabolism (7)

Arginine and proline metabolism (6)

Polycyclic aromatic hydrocarbon degradation (6)

Pyruvate metabolism (6)

Base excision repair (5)

Two-component system (5)

Biotin metabolism (5)

Citrate cycle (5)

Amino sugar and nucleotide sugar metabolism (4)

Glyoxylate and dicarboxylate metabolism (4)

Tyrosine metabolism (4)

Pantothenate and CoA biosynthesis (4)

Bisphenol degradation (4)

Glycolysis/Gluconeogenesis (4)

Glycerolipid metabolism (3)

Oxidative phosphorylation (3)

Nucleotide excision repair (3)

Ether lipid metabolism (3)

Sulfur metabolism (3)

Valine, leucine and isoleucine biosynthesis (3)

Homologous recombination (3)

Pentose and glucuronate interconversions (3)

Ribosome (3)

ABC transporters (3)

Toluene degradation (3)

APPENDIX F

FastQC reports used to assess quality of whole genome sequence data

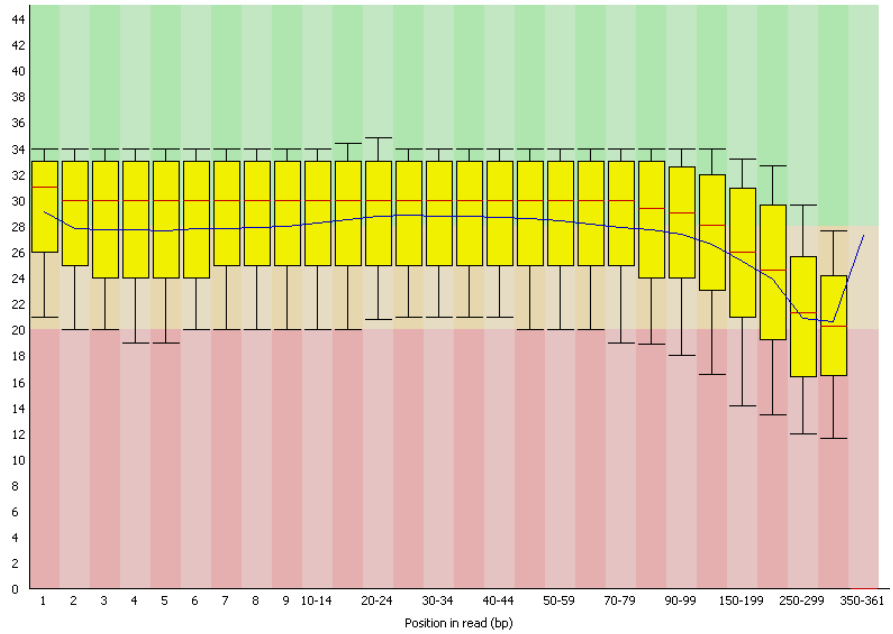

Figure F1. The distribution of quality scores across all bases. The blue line represents the mean quality, the red lines represent median values and the green and red regions of the plot indicate high and low quality scores, respectively.

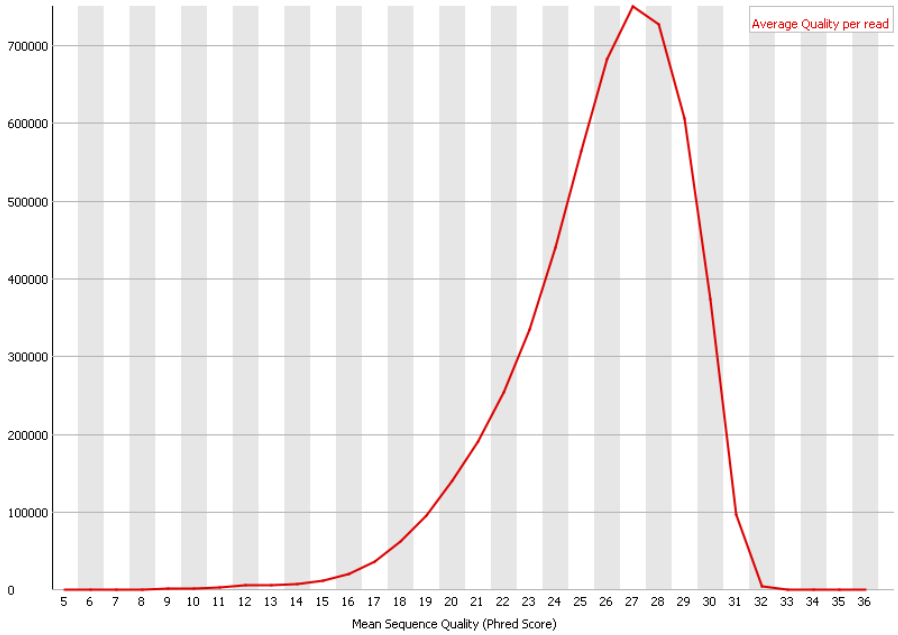

Figure F2. The distribution of quality across all sequences. Score used to assess whether a proportion of the sequences have universally low quality scores.

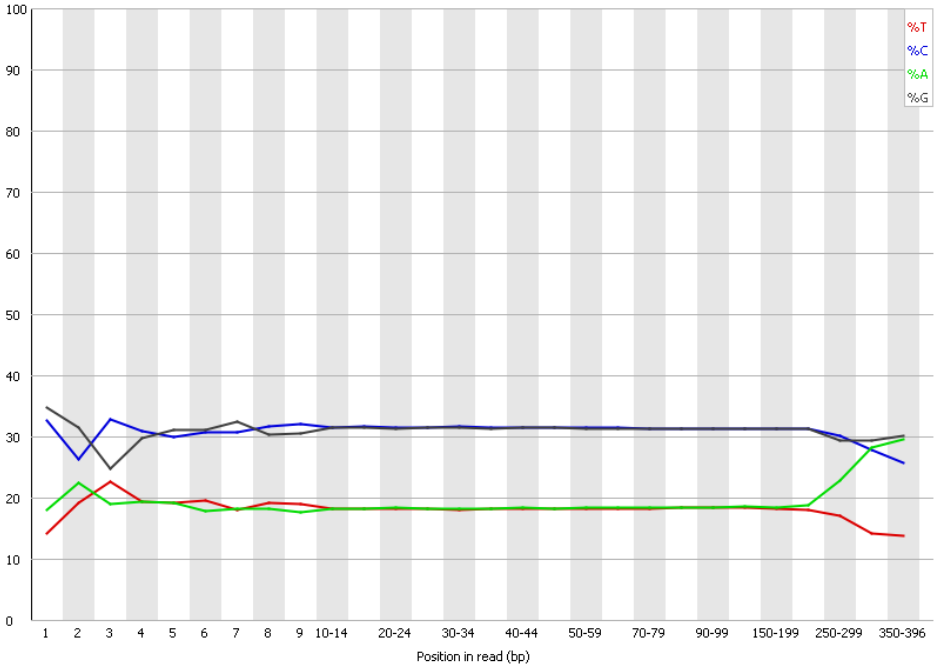

Figure F3. Sequence content across all bases. As *M. tuberculosis* has a GC-rich genome, the percentage GC content is expected to be >60%.

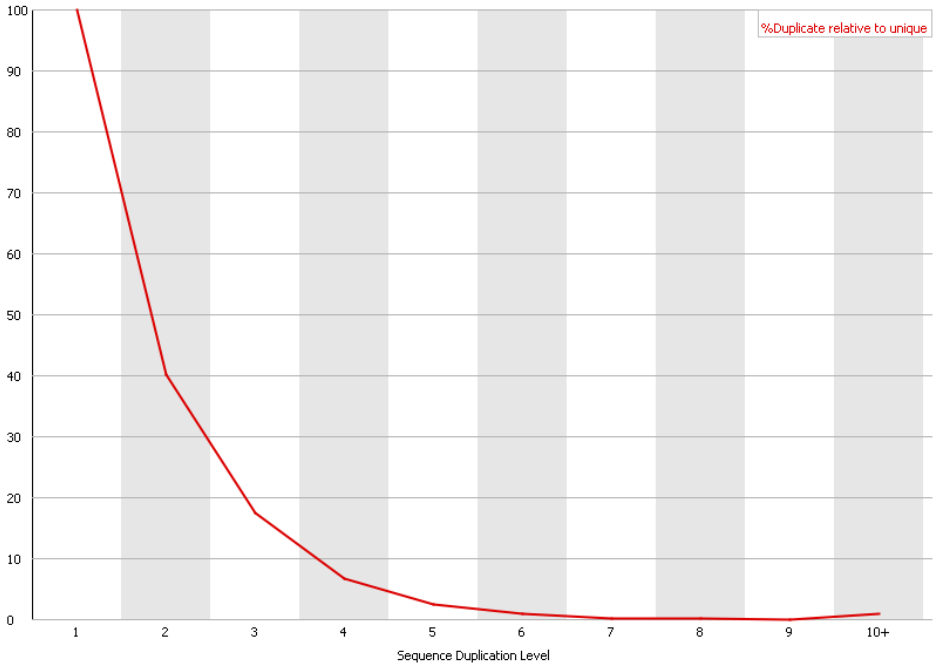

Figure F4. Level of sequence duplication. A high level of duplication could suggest an enrichment bias such as PCR over-amplification.

APPENDIX G

Do *rpoB* Mutations Induce a Stringent Response?

One of the most important interactions RNAp undergoes to fulfil its enzymatic function is during transcription initiation, when the holoenzyme recognizes and binds to certain sequences within the promoter region. In *E. coli*, it has been found that particular *rpoB* mutations (S531F, D532A, L533P, T563P) lead to downregulation of the *rpoD* operon, which is negatively regulated by the stringent response (118). Likewise, induction of genes that are positively controlled by the stringent response was observed in these mutants (119, 120). The same four RIF-resistance conferring mutations were found to facilitate the growth of *relA*/*spoT* double-deletion mutants in minimal media, whereas strains without an *rpoB* mutation require supplementation with amino acids. Zhuo and Jin speculated that these findings indicate that mutated RNAp mimics WT RNAp during the stringent response. In the study which aimed to investigate this hypothesis, the authors concluded that mutated RNAp causes unstable initiation complexes at stringent promoters in *E. coli* (121). As a result, transcription of stringent genes is decreased, which allows more RNAp molecules to be available for transcription from promoters which are positively regulated by the stringent response.

Guanosine nucleotides (p/ppGpp) and the transcriptional regulator, CarD, initiate the stringent response in prokaryotes when amino acid availability is low. As ppGpp/pppGpp is able to regulate transcription by binding directly to RNAp at a site near the C-terminus of the β subunit, it is possible that an *rpoB* mutation could also influence this interaction. Strains of *E. coli* harbouring an *rpoB* mutation have been found to be more sensitive to ppGpp/pppGpp which would lead to the activation of the stringent response without the necessary increase in ppGpp/pppGpp (40). Furthermore, it has been found that CarD is essential for resistance to

oxidative stress and that weakening the interaction between CarD and RNAp leads to increased sensitivity of *M. tuberculosis* to RIF (42, 122).

Dahl and colleagues demonstrated that ppGpp/pppGpp regulates the expression of 159 genes in *M. tuberculosis* during starvation (38). Results from this study were highly comparable to a study by Betts *et al*, which investigated the transcriptomic response of *M. tuberculosis* to nutrient starvation (21). This verifies the role that ppGpp/pppGpp plays in orchestrating the stringent response and subsequent shutdown of prokaryotic metabolism. Genes which were downregulated after 24 hours of starvation were compared to the present study to determine whether a similar trend could be observed, however the same extent of differential expression was not seen. Genes which were greatly downregulated in the study by Betts *et al* were only marginally downregulated in the present study. Furthermore, a number of genes which were downregulated in the starvation study were upregulated in either of the *rpoB* mutants studied here. This indicates that the complete stringent response is not triggered in *rpoB* mutants; however, it is a possibility that the downregulation of metabolic pathways that we observed in this study demonstrates a diminished stringent response.

APPENDIX H

Do *rpoB* Mutations Induce a Stress Response?

To determine whether *rpoB* mutations elicit a constitutive stress response, Bergval and colleagues investigated the expression of *recA* and *dnaE2* in six *in vitro* mutants with different *rpoB* mutations (123). No differential expression of *recA* was seen, however a 2-5 fold induction of *dnaE2* was seen in 4 of the 6 mutants used in the study. Interestingly, *rpoB*531 and *rpoB*526 mutants were the two which did not demonstrate upregulation of either *recA* or *dnaE2*. Expression data from the present study agrees with these findings, therefore we can conclude that the *rpoB* mutations studied here do not facilitate *recA* and *dnaE2*-mediated adaptation in *M. tuberculosis*. However, it is possible that a stress response is initiated by the binding of alternative σ factors to the RNAp core enzyme.

If *rpoB* mutations lead to a change in the affinity of RNAp for particular σ factors, a transcriptional response could be initiated by directing RNAp to different promoters. The basis of such a response could be due to an alteration in the interaction between the σ factor and β-subunit of RNAp. As different σ factors bind to RNAp to increase the affinity of the enzyme for specific promoter regions, we can hypothesize that the presence of an *rpoB* mutation would interfere with the usual behaviour of these two subunits, and may lead to increased affinity of alternate σ factors. This would explain the difference in gene expression we observe when comparing WT and *rpoB* mutants of *M. tuberculosis*.

Numerous studies have been done to investigate which operons are under the control of different σ factors found in *M. tuberculosis*; a review by Sachdeva *et al* summarizes this information comprehensively (124). We aimed to investigate whether the genes which were differentially expressed in each of the *rpoB* mutants studied here could be associated with a

particular σ factor. Surprisingly, only a small group of genes that are linked to σ^A appeared to match this criterion, therefore the influence of *rpoB* mutations on the affinity of RNAp for alternative σ factors remains unknown.

References

1. **Koch R**. 1942. The aetiology of tuberculosis (translation of Die Aetiologie der Tuberculose [1882]). Source book of medical history. Dover Publications, Inc., New York 392–406.
2. **World Health Organization**. 2012. Global tuberculosis report 2012. Geneva: The Organization.
3. **World Health Organization**. 2008. Global tuberculosis control: surveillance planning financing. Geneva: The Organization
4. **National Institute for Communicable Diseases**. 2011. Communicable Diseases Surveillance Bulletin.
5. **Ramaswamy S**. 1998. Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. TUBERCLE & LUNG DISEASE **79**:3.
6. **Louw GE, Warren RM, Gey van Pittius NC, Leon R, Jimenez A, Hernandez-Pando R, McEvoy CRE, Grobbelaar M, Murray M, van Helden PD, Victor TC**. 2011. Rifampicin Reduces Susceptibility to Ofloxacin in Rifampicin-resistant Mycobacterium tuberculosis through Efflux. American Journal of Respiratory and Critical Care Medicine **184**:269–276.
7. **Bisson GP, Mehaffy C, Broeckling C, Prenni J, Rifat D, Lun DS, Burgos M, Weissman D, Karakousis PC, Dobos K**. 2012. Upregulation of the Phthiocerol Dimycocerosate Biosynthetic Pathway by Rifampin-Resistant, rpoB Mutant Mycobacterium tuberculosis. J. Bacteriol. **194**:6441–6452.
8. **Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, 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 MA, 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.
9. **Zhang G, Darst SA**. 1998. Structure of the Escherichia coli RNA Polymerase α Subunit Amino-Terminal Domain. Science **281**:262–266.
10. **Ebright RH, Busby S**. 1995. The Escherichia coli RNA polymerase α subunit: structure and function. Current Opinion in Genetics & Development **5**:197–203.
11. **Darst SA, Polyakov A, Richter C, Zhang G**. 1998. Structural Studies of Escherichia coli RNA Polymerase. Cold Spring Harb Symp Quant Biol **63**:269–276.

12. **Burgess RR, Travers AA**. 1969. Factor Stimulating Transcription by RNA Polymerase. Nature **221**:43–46.
13. **Murakami KS**. 2013. X-ray Crystal Structure of Escherichia coli RNA Polymerase σ70 Holoenzyme. J. Biol. Chem. **288**:9126–9134.
14. **Madej T, Addess KJ, Fong JH, Geer LY, Geer RC, Lanczycki CJ, Liu C, Lu S, Marchler-Bauer A, Panchenko AR, Chen J, Thiessen PA, Wang Y, Zhang D, Bryant SH**. 2012. MMDB: 3D structures and macromolecular interactions. Nucleic Acids Res. **40**:D461–464.
15. **Saecker RM, Tsodikov OV, McQuade KL, Schlax Jr PE, Capp MW, Thomas Record Jr M**. 2002. Kinetic Studies and Structural Models of the Association of E. coli σ70 RNA Polymerase with the λPR Promoter: Large Scale Conformational Changes in Forming the Kinetically Significant Intermediates. Journal of Molecular Biology **319**:649–671.
16. **Helmann JD, Chamberlin MJ**. 1988. Structure and Function of Bacterial Sigma Factors. Annual Review of Biochemistry **57**:839–872.
17. **Missiakas D, Raina S**. 1998. The extracytoplasmic function sigma factors: role and regulation. Molecular Microbiology **28**:1059–1066.
18. **Volpe E, Cappelli G, Grassi M, Martino A, Serafino A, Colizzi V, Sanarico N, Mariani F**. 2006. Gene expression profiling of human macrophages at late time of infection with Mycobacterium tuberculosis. Immunology **118**:449–460.
19. **Manganelli R, Dubnau E, Tyagi S, Kramer FR, Smith I**. 1999. Differential expression of 10 sigma factor genes in Mycobacterium tuberculosis. Molecular Microbiology **31**:715–724.
20. **Hu Y, Coates ARM**. 1999. Transcription of Two Sigma 70 Homologue Genes,sigA and sigB, in Stationary-PhaseMycobacterium tuberculosis. J. Bacteriol. **181**:469–476.
21. **Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K**. 2002. Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Molecular Microbiology **43**:717–731.
22. **Voskuil MI, Visconti KC, Schoolnik GK**. 2004. Mycobacterium tuberculosis gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis **84**:218–227.
23. **Graham JE, Clark-Curtiss JE**. 1999. Identification of Mycobacterium tuberculosis RNAs synthesized in response to phagocytosis by human macrophages by selective capture of transcribed sequences (SCOTS). PNAS **96**:11554–11559.

24. **Michele TM, Ko C, Bishai WR**. 1999. Exposure to Antibiotics Induces Expression of theMycobacterium tuberculosis sigF Gene: Implications for Chemotherapy against Mycobacterial Persistors. Antimicrob. Agents Chemother. **43**:218–225.

25. **Fernandes ND, Wu Q, Kong D, Puyang X, Garg S, Husson RN**. 1999. A Mycobacterial Extracytoplasmic Sigma Factor Involved in Survival following Heat Shock and Oxidative Stress. J. Bacteriol. **181**:4266–4274.

26. **Hu Y, Coates ARM**. 2001. Increased levels of sigJ mRNA in late stationary phase cultures of Mycobacterium tuberculosis detected by DNA array hybridisation. FEMS Microbiology Letters **202**:59–65.

27. **Cappelli G, Volpe E, Grassi M, Liseo B, Colizzi V, Mariani F**. 2006. Profiling of Mycobacterium tuberculosis gene expression during human macrophage infection: Upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. Research in Microbiology **157**:445–455.

28. **Hu Y, Kendall S, Stoker NG, Coates ARM**. 2004. The Mycobacterium tuberculosis sigJ gene controls sensitivity of the bacterium to hydrogen peroxide. FEMS Microbiology Letters **237**:415–423.

29. **Avarbock D, Salem J, Li L, Wang Z, Rubin H**. 1999. Cloning and characterization of a bifunctional RelA/SpoT homologue from Mycobacterium tuberculosis. Gene **233**:261–269.

30. **Xiao H, Kalman M, Ikehara K, Zemel S, Glaser G, Cashel M**. 1991. Residual guanosine 3',5'-bispyrophosphate synthetic activity of relA null mutants can be eliminated by spoT null mutations. J. Biol. Chem. **266**:5980–5990.

31. **Avarbock A**. 2005. Functional regulation of the opposing (p) ppGpp synthetase/hydrolase activities of RelMtb from Mycobacterium tuberculosis. BIOCHEMISTRY-US **44**:9913.

32. **Cashel M, Gallant J**. 1969. Two Compounds implicated in the Function of the RC Gene of Escherichia coli. Nature **221**:838–841.

33. **Lazzarini RA, Cashel M, Gallant J**. 1971. On the Regulation of Guanosine Tetraphosphate Levels in Stringent and Relaxed Strains of Escherichia coli. J. Biol. Chem. **246**:4381–4385.

34. **Chatterji D, Fujita N, Ishihama A**. 1998. The mediator for stringent control, ppGpp, binds to the β-subunit of Escherichia coli RNA polymerase. Genes to Cells **3**:279–287.

35. **Ryals J, Little R, Bremer H**. 1982. Control of rRNA and tRNA syntheses in Escherichia coli by guanosine tetraphosphate. J. Bacteriol. **151**:1261–1268.

36. **Gentry DR, Hernandez VJ, Nguyen LH, Jensen DB, Cashel M**. 1993. Synthesis of the stationary-phase sigma factor sigma s is positively regulated by ppGpp. J. Bacteriol. **175**:7982–7989.

37. **Jishage M, Kvint K, Shingler V, Nyström T**. 2002. Regulation of ς factor competition by the alarmone ppGpp. Genes Dev. **16**:1260–1270.

38. **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. PNAS **100**:10026–10031.

39. **Toulokhonov II, Shulgina I, Hernandez VJ**. 2001. Binding of the Transcription Effector ppGpp to Escherichia coli RNA Polymerase Is Allosteric, Modular, and Occurs Near the N Terminus of the β′-Subunit. J. Biol. Chem. **276**:1220–1225.

40. **Little R, Ryals J, Bremer H**. 1983. Physiological characterization of Escherichia coli rpoB mutants with abnormal control of ribosome synthesis. J. Bacteriol. **155**:1162–1170.

41. **Little R, Ryals J, Bremer H**. 1983. rpoB mutation in Escherichia coli alters control of ribosome synthesis by guanosine tetraphosphate. J Bacteriol **154**:787–792.

42. **Stallings CL, Stephanou NC, Chu L, Hochschild A, Nickels BE, Glickman MS**. 2009. CarD Is an Essential Regulator of rRNA Transcription Required for Mycobacterium tuberculosis Persistence. Cell **138**:146–159.

43. **Magnusson LU, Gummesson B, Joksimović P, Farewell A, Nyström T**. 2007. Identical, Independent, and Opposing Roles of ppGpp and DksA in Escherichia coli. J. Bacteriol. **189**:5193–5202.

44. **Wassarman KM, Storz G**. 2000. 6S RNA Regulates E. coli RNA Polymerase Activity. Cell **101**:613–623.

45. **Trotochaud AE, Wassarman KM**. 2005. A highly conserved 6S RNA structure is required for regulation of transcription. Nat Struct Mol Biol **12**:313–319.

46. **Gildehaus N, Neußer T, Wurm R, Wagner R**. 2007. Studies on the function of the riboregulator 6S RNA from E. coli: RNA polymerase binding, inhibition of *in vitro* transcription and synthesis of RNA-directed de novo transcripts. Nucl. Acids Res. **35**:1885–1896.

47. **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**:e1002342.

48. **Keren I, Minami S, Rubin E, Lewis K**. 2011. Characterization and Transcriptome Analysis of Mycobacterium tuberculosis Persisters. mBio **2**:e00100–11.

49. **Windbichler N, von Pelchrzim F, Mayer O, Csaszar E, Schroeder R**. 2008. Isolation of small RNA-binding proteins from E. coli: evidence for frequent interaction of RNAs with RNA polymerase. RNA Biol **5**:30–40.

50. **Fischer D, Teich A, Neubauer P, Hengge-Aronis R**. 1998. The General Stress Sigma Factor ςS ofEscherichia coli Is Induced during Diauxic Shift from Glucose to Lactose. J. Bacteriol. **180**:6203–6206.

51. **Hengge-Aronis R**. 1999. Interplay of global regulators and cell physiology in the general stress response of Escherichia coli. Current Opinion in Microbiology **2**:148–152.

52. **Majdalani N, Cunning C, Sledjeski D, Elliott T, Gottesman S**. 1998. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Proc Natl Acad Sci U S A **95**:12462–12467.

53. **Sledjeski DD, Whitman C, Zhang A**. 2001. Hfq Is Necessary for Regulation by the Untranslated RNA DsrA. J. Bacteriol. **183**:1997–2005.

54. **Majdalani N, Chen S, Murrow J, St John K, Gottesman S**. 2001. Regulation of RpoS by a novel small RNA: the characterization of RprA. Molecular Microbiology **39**:1382–1394.

55. **Soper TJ, Woodson SA**. 2008. The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. RNA **14**:1907–1917.

56. **Sledjeski D, Gottesman S**. 1995. A small RNA acts as an antisilencer of the H-NS-silenced rcsA gene of Escherichia coli. Proc Natl Acad Sci U S A **92**:2003–2007.

57. **Miotto P, Forti F, Ambrosi A, Pellin D, Veiga DF, Balazsi G, Gennaro ML, Di Serio C, Ghisotti D, Cirillo DM**. 2012. Genome-Wide Discovery of Small RNAs in Mycobacterium tuberculosis. PLoS ONE **7**:e51950.

58. **Pellin D, Miotto P, Ambrosi A, Cirillo DM, Di Serio C**. 2012. A Genome-Wide Identification Analysis of Small Regulatory RNAs in Mycobacterium tuberculosis by RNA-Seq and Conservation Analysis. PLoS ONE **7**:e32723.

59. **Hesselbach BA, Nakada D**. 1977. "Host shutoff" function of bacteriophage T7: involvement of T7 gene 2 and gene 0.7 in the inactivation of Escherichia coli RNA polymerase. J. Virol. **24**:736–745.

60. **Hesselbach BA, Nakada D**. 1975. Inactive complex formation between E. coli RNA polymerase and inhibitor protein purified from T7 phage infected cells. Nature **258**:354–357.

61. **Hesselbach BA, Nakada D**. 1977. I protein: bacteriophage T7-coded inhibitor of Escherichia coli RNA polymerase. J. Virol. **24**:746–760.

62. **Hesselbach BA, Yamada Y, Nakada D**. 1974. Isolation of an inhibitor protein of E. coli RNA polymerase from T7 phage infected cell. Nature **252**:71–74.

63. **Nechaev S, Severinov K**. 1999. Inhibition of Escherichia coli RNA Polymerase by Bacteriophage T7 Gene 2 Protein. Journal of Molecular Biology **289**:815–826.

64. **Cámara B, Liu M, Reynolds J, Shadrin A, Liu B, Kwok K, Simpson P, Weinzierl R, Severinov K, Cota E, Matthews S, Wigneshweraraj SR**. 2010. T7 phage protein Gp2 inhibits the Escherichia coli RNA polymerase by antagonizing stable DNA strand separation near the transcription start site. PNAS **107**:2247–2252.

65. **Sheppard C, James E, Barton G, Matthews S, Severinov K, Wigneshweraraj S**. 2013. A non-bacterial transcription factor inhibits bacterial transcription by a multipronged mechanism. RNA Biology **10**:495–501.

66. **Stevens A**. 1972. New Small Polypeptides Associated with DNA-Dependent RNA Polymerase of Escherichia coli after Infection with Bacteriophage T4. Proc Natl Acad Sci U S A **69**:603–607.

67. **Stevens A**. 1973. An inhibitor of host sigma-stimulated core enzyme activity that purifies with DNA-dependent RNA polymerase of E. coli following T4 phage infection. Biochem. Biophys. Res. Commun. **54**:488–493.

68. **Hinton DM, March-Amegadzie R, Gerber JS, Sharma M**. 1996. Bacteriophage T4 middle transcription system: T4-modified RNA polymerase; AsiA, a σ70 binding protein; and transcriptional activator MotA, p. 43–57. *In* Sankar Adhya (ed.), Methods in Enzymology. Academic Press.

69. **Geiduschek EP**. 1991. Regulation of Expression of the Late Genes of Bacteriophage T4. Annual Review of Genetics **25**:437–460.

70. **Orsini G, Ouhammouch M, Caer JPL, Brody EN**. 1993. The asiA gene of bacteriophage T4 codes for the anti-sigma 70 protein. J. Bacteriol. **175**:85–93.

71. **Telenti A, Imboden P, Marchesi F, Matter L, Schopfer K, Bodmer T, Lowrie D, Colston M., Cole S**. 1993. Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. The Lancet **341**:647–651.

72. **Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA**. 2001. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. Cell **104**:901–912.

73. **Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB**. 2009. Tuberculosis Drug Resistance Mutation Database. PLoS Med **6**:e1000002.

74. **Kapur V, Li LL, Iordanescu S, Hamrick MR, Wanger A, Kreiswirth BN, Musser JM**. 1994. Characterization by automated DNA sequencing of mutations in the gene (rpoB) encoding the RNA polymerase beta subunit in rifampin-resistant Mycobacterium tuberculosis strains from New York City and Texas. J. Clin. Microbiol. **32**:1095–1098.

75. **Bodmer T, Zürcher G, Imboden P, Telenti A**. 1995. Mutation position and type of substitution in the beta-subunit of the RNA polymerase influence in-vitro activity of rifamycins in rifampicin-resistant Mycobacterium tuberculosis. J. Antimicrob. Chemother. **35**:345–348.

76. **Ramaswamy SV, Dou S-J, Rendon A, Yang Z, Cave MD, Graviss EA**. 2004. Genotypic analysis of multidrug-resistant Mycobacterium tuberculosis isolates from Monterrey, Mexico. J. Med. Microbiol. **53**:107–113.

77. **Kim BJ, Kim SY, Park BH, Lyu MA, Park IK, Bai GH, Kim SJ, Cha CY, Kook YH**. 1997. Mutations in the rpoB gene of Mycobacterium tuberculosis that interfere with PCR-single-strand conformation polymorphism analysis for rifampin susceptibility testing. J. Clin. Microbiol. **35**:492–494.

78. **Donnabella V, Martiniuk F, Kinney D, Bacerdo M, Bonk S, Hanna B, Rom WN**. 1994. Isolation of the gene for the beta subunit of RNA polymerase from rifampicin-resistant Mycobacterium tuberculosis and identification of new mutations. Am. J. Respir. Cell Mol. Biol. **11**:639–643.

79. **Moghazeh SL, Pan X, Arain T, Stover CK, Musser JM, Kreiswirth BN**. 1996. Comparative antimycobacterial activities of rifampin, rifapentine, and KRM-1648 against a collection of rifampin-resistant Mycobacterium tuberculosis isolates with known rpoB mutations. Antimicrob. Agents Chemother. **40**:2655–2657.

80. **Pang Y, Lu J, Wang Y, Song Y, Wang S, Zhao Y**. 2013. Study of the Rifampin Monoresistance Mechanism in Mycobacterium tuberculosis. Antimicrob. Agents Chemother. **57**:893–900.

81. **Huitric E, Werngren J, Juréen P, Hoffner S**. 2006. Resistance Levels and rpoB Gene Mutations among *In Vitro*-Selected Rifampin-Resistant Mycobacterium tuberculosis Mutants. Antimicrob. Agents Chemother. **50**:2860–2862.

82. **Musser JM**. 1995. Antimicrobial agent resistance in mycobacteria: molecular genetic insights. Clin. Microbiol. Rev. **8**:496–514.

83. **Morlock GP, Plikaytis BB, Crawford JT**. 2000. Characterization of Spontaneous, *In Vitro*-Selected, Rifampin-Resistant Mutants of Mycobacterium tuberculosisStrain H37Rv. Antimicrob. Agents Chemother. **44**:3298–3301.

84. **Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM**. 2006. The Competitive Cost of Antibiotic Resistance in Mycobacterium tuberculosis. Science **312**:1944–1946.

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

86. **Brandis G, Wrande M, Liljas L, Hughes D**. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Molecular Microbiology **85**:142–151.

87. **Vos M de, Müller B, Borrell S, Black PA, Helden PD van, Warren RM, Gagneux S, Victor TC**. 2013. Putative Compensatory Mutations in the rpoC Gene of Rifampin-Resistant Mycobacterium tuberculosis Are Associated with Ongoing Transmission. Antimicrob. Agents Chemother. **57**:827–832.

88. **Traynor K**. 2011. Fidaxomicin approved for C. difficile infections. Am J Health Syst Pharm **68**:1276.

89. **Coronelli C, White RJ, Lancini GC, Parenti F**. 1975. Lipiarmycin, a new antibiotic from Actinoplanes. II. Isolation, chemical, biological and biochemical characterization. J. Antibiot. **28**:253–259.

90. **Kurabachew M, Lu SHJ, Krastel P, Schmitt EK, Suresh BL, Goh A, Knox JE, Ma NL, Jiricek J, Beer D, Cynamon M, Petersen F, Dartois V, Keller T, Dick T, Sambandamurthy VK**. 2008. Lipiarmycin targets RNA polymerase and has good activity against multidrug-resistant strains of Mycobacterium tuberculosis. J. Antimicrob. Chemother. **62**:713–719.

91. **Tupin A, Gualtieri M, Leonetti J-P, Brodolin K**. 2010. The transcription inhibitor lipiarmycin blocks DNA fitting into the RNA polymerase catalytic site. EMBO J **29**:2527–2537.

92. **Srivastava A, Talaue M, Liu S, Degen D, Ebright RY, Sineva E, Chakraborty A, Druzhinin SY, Chatterjee S, Mukhopadhyay J, Ebright YW, Zozula A, Shen J, Sengupta S, Niedfeldt RR, Xin C, Kaneko T, Irschik H, Jansen R, Donadio S, Connell N, Ebright RH**. 2011. New target for inhibition of bacterial RNA polymerase: "switch region". Current Opinion in Microbiology **14**:532–543.

93. **Pretorius GS, Helden V, D P, Sirgel F, Eisenach KD, Victor TC**. 1995. Mutations in katG Gene Sequences in Isoniazid-Resistant Clinical Isolates of Mycobacterium Tuberculosis Are Rare. Antimicrob. Agents Chemother. **39**:2276–2281.

94. **Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J**. 2000. A concise guide to cDNA microarray analysis. BioTechniques **29**:548–550, 552–554, 556 passim.

95. **Warren RM, van Helden PD, Gey van Pittius NC**. 2009. Insertion element IS6110-based restriction fragment length polymorphism genotyping of Mycobacterium tuberculosis. Methods Mol. Biol. **465**:353–370.

96. **Arnold K, Bordoli L, Kopp J, Schwede T**. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics **22**:195–201.

97. **Benkert P, Biasini M, Schwede T**. 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics **27**:343–350.

98. **Ng PC, Henikoff S**. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. **31**:3812–3814.

99. **Burian J, Ramón-García S, Sweet G, Gómez-Velasco A, Av-Gay Y, Thompson CJ**. 2012. The Mycobacterial Transcriptional Regulator whiB7 Gene Links Redox Homeostasis and Intrinsic Antibiotic Resistance. J. Biol. Chem. **287**:299–310.

100. **Recchi C, Sclavi B, Rauzier J, Gicquel B, Reyrat J-M**. 2003. Mycobacterium tuberculosis Rv1395 Is a Class III Transcriptional Regulator of the AraC Family Involved in Cytochrome P450 Regulation. J. Biol. Chem. **278**:33763–33773.

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

102. **Singh A, Jain S, Gupta S, Das T, Tyagi AK**. 2003. mymA operon of Mycobacterium tuberculosis: its regulation and importance in the cell envelope. FEMS Microbiology Letters **227**:53–63.

103. **Morris RP, Nguyen L, Gatfield J, Visconti K, Nguyen K, Schnappinger D, Ehrt S, Liu Y, Heifets L, Pieters J, Schoolnik G, Thompson CJ**. 2005. Ancestral antibiotic resistance in Mycobacterium tuberculosis. PNAS **102**:12200–12205.

104. **Burian J, Yim G, Hsing M, Axerio-Cilies P, Cherkasov A, Spiegelman GB, Thompson CJ**. 2013. The mycobacterial antibiotic resistance determinant WhiB7 acts as a transcriptional activator by binding the primary sigma factor SigA (RpoV). Nucleic Acids Res.

105. **Yanofsky C, Horn V**. 1981. Rifampin resistance mutations that alter the efficiency of transcription termination at the tryptophan operon attenuator. J. Bacteriol. **145**:1334–1341.

106. **Awasthy D, Gaonkar S, Shandil RK, Yadav R, Bharath S, Marcel N, Subbulakshmi V, Sharma U**. 2009. Inactivation of the ilvB1 gene in Mycobacterium tuberculosis leads to branched-chain amino acid auxotrophy and attenuation of virulence in mice. Microbiology **155**:2978–2987.

107. **Slekar KH, Kosman DJ, Culotta VC**. 1996. The Yeast Copper/Zinc Superoxide Dismutase and the Pentose Phosphate Pathway Play Overlapping Roles in Oxidative Stress Protection. J. Biol. Chem. **271**:28831–28836.

108. **Nogae I, Johnston M**. 1990. Isolation and characterization of the ZWF1 gene of Saccharomyces cerevisiae, encoding glucose-6-phosphate dehydrogenase. Gene **96**:161–169.

109. **Gorsich SW, Dien BS, Nichols NN, Slininger PJ, Liu ZL, Skory CD**. 2006. Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in Saccharomyces cerevisiae. Appl Microbiol Biotechnol **71**:339–349.

110. **Peloquin CA**. 1997. Using therapeutic drug monitoring to dose the antimycobacterial drugs. CLIN CHEST MED **18**:79.

111. **Garnham JC, Taylor T, Turner P, Chasseaud LF**. 1976. Serum concentrations and bioavailability of rifampicin and isoniazid in combination. Br J Clin Pharmacol **3**:897–902.

112. **Singh G, Singh G, Jadeja D, Kaur J**. 2010. Lipid hydrolizing enzymes in virulence: Mycobacterium tuberculosis as a model system. Critical Reviews in Microbiology **36**:259–269.

113. **Sassetti CM, Boyd DH, Rubin EJ**. 2003. Genes required for mycobacterial growth defined by high density mutagenesis. Mol. Microbiol. **48**:77–84.

114. **Fu LM**. 2006. Exploring drug action on Mycobacterium tuberculosis using affymetrix oligonucleotide genechips. Tuberculosis **86**:134–143.

115. **Sawicki MP, Samara G, Hurwitz M, Passaro Jr. E**. 1993. Human Genome Project. The American Journal of Surgery **165**:258–264.

116. **Safi H, Lingaraju S, Amin A, Kim S, Jones M, Holmes M, McNeil M, Peterson SN, Chatterjee D, Fleischmann R, Alland D**. 2013. Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-β-D-arabinose biosynthetic and utilization pathway genes. Nat Genet **45**:1190–1197.

117. **Chihota VN, Müller B, Mlambo CK, Pillay M, Tait M, Streicher EM, Marais E, Spuy GD van der, Hanekom M, Coetzee G, Trollip A, Hayes C, Bosman ME, Pittius NCG van, Victor TC, Helden PD van, Warren RM**. 2012. Population Structure of Multi- and Extensively Drug-Resistant Mycobacterium tuberculosis Strains in South Africa. J. Clin. Microbiol. **50**:995–1002.

118. **Zhou YN, Jin DJ**. 1997. RNA polymerase beta mutations have reduced sigma70 synthesis leading to a hyper-temperature-sensitive phenotype of a sigma70 mutant. J. Bacteriol. **179**:4292–4298.

119. **Grossman AD, Taylor WE, Burton ZF, Burgess RR, Gross CA**. 1985. Stringent response in Escherichia coli induces expression of heat shock proteins. Journal of Molecular Biology **186**:357–365.

120. **Burton ZF, Gross CA, Watanabe KK, Burgess RR**. 1983. The operen that encodes the sigma subunit of RNA polymerase also encodes ribosomal protein S21 and DNA primase in E. coli K12. Cell **32**:335–349.

121. **Zhou YN, Jin DJ**. 1998. The rpoB mutants destabilizing initiation complexes at stringently controlled promoters behave like “stringent” RNA polymerases in Escherichia coli. Proc Natl Acad Sci U S A **95**:2908–2913.

122. **Weiss LA, Harrison PG, Nickels BE, Glickman MS, Campbell EA, Darst SA, Stallings CL**. 2012. Interaction of CarD with RNA Polymerase Mediates Mycobacterium tuberculosis Viability, Rifampin Resistance, and Pathogenesis. J Bacteriol **194**:5621–5631.

123. **Bergval I, Kwok B, Schuitema A, Kremer K, van Soolingen D, Klatser P, Anthony R**. 2012. Pre-Existing stance, but Not the Genotype of Mycobacterium Tuberculosis Drives Rifampicin Resistance Codon Preference *in Vitro*. PLoS ONE **7**:e29108.

124. **Sachdeva P, Misra R, Tyagi AK, Singh Y**. 2010. The sigma factors of Mycobacterium tuberculosis: regulation of the regulators. FEBS Journal **277**:605–626.

125. **Collins DM, Kawakami RP, Lisle GW de, Pascopella L, Bloom BR, Jacobs WR**. 1995. Mutation of the principal sigma factor causes loss of virulence in a strain of the Mycobacterium tuberculosis complex. PNAS **92**:8036–8040.

126. **Steyn AJC, Collins DM, Hondalus MK, Jacobs WR, Kawakami RP, Bloom BR**. 2002. Mycobacterium tuberculosis WhiB3 interacts with RpoV to affect host survival but is dispensable for in vivo growth. PNAS **99**:3147–3152.

127. **Rohde KH, Abramovitch RB, Russell DG**. 2007. Mycobacterium tuberculosis Invasion of Macrophages: Linking Bacterial Gene Expression to Environmental Cues. Cell Host & Microbe **2**:352–364.