Clinically relevant mutations contributing to drug resistance in *Mycobacterium tuberculosis*

---

by

Juanelle du Plessis

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

Promoter:

Professor Samantha Leigh Sampson

Co-promoters:

Professor Sivaramesh Wigneshweraraj

Professor Robin Mark Warren

December 2017

DECLARATION

By submitting this thesis 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.

Signature …… Date ……December 2017……

ABSTRACT

Single nucleotide variants are the underlying driver of drug resistance, strain fitness and adaptation in *Mycobacterium tuberculosis* and investigating the mechanistic and physiological aspects of these mutations is key to our understanding of the biology of this pathogen. Rifampicin, one of the most powerful first line drugs used to treat tuberculosis, inhibits transcription in *M. tuberculosis* by binding to the β subunit of RNA polymerase (RNAP). However, mycobacteria are able to evade binding of this drug by acquiring mutations in the *rpoB* gene. These resistance-conferring mutations are essential for the survival of *M. tuberculosis* in the presence of rifampicin, however they impart a fitness cost to the bacterium due to a presumed reduction in transcription efficiency and subsequent changes in gene expression. One of the mechanisms *M. tuberculosis* uses to buffer the effects of this fitness cost is to acquire compensatory mutations in *rpoC* and *rpoA*. As RNAP is at the core of all mechanisms of gene regulation in *M. tuberculosis*, it is not surprising that mutations within this enzyme led to pleiotropic effects. As this remains a poorly understood area of mycobacterial physiology, the current work encompasses a triad of studies which aims to better understand functional aspects of *M. tuberculosis* RNAP, and the role of *rpoB* and *rpoC* mutations in drug resistance.

First, the effect of a bacteriophage protein, Gp2, was investigated to determine whether it is able to inhibit RNAP in *M. tuberculosis*. As Gp2 is known to bind to the β' subunit of RNAP in *Escherichia coli*, positive findings from this work would provide a framework for the identification of novel compounds that inhibit transcription in the presence of *rpoB* mutations, affecting the β subunit. By way of *in vitro* and *in silico* analysis, it was found that Gp2 binds to and inhibits RNAP in *M. tuberculosis*, however to a much lesser degree than it does in its *de facto* host, *E. coli*. Nonetheless, future studies can build on our findings as *in silico* modification of Gp2 could identify a structure which allows for stronger binding affinity of the protein.

Secondly, the effect of *rpoB* and *rpoC* mutations on the function of mycobacterial RNAP was investigated using a series of *in vitro* transcription assays. Radioactivity-based assays were performed using purified wildtype and mutant versions of the RNAP complex, to assess enzyme activity and promoter affinity. Furthermore, the use of a fluorescence-based assay was trialled to develop a comparable method without the use of radiolabelled nucleotides.

Lastly, we undertook a study to understand the effect of *rpoC* mutations on the transcriptome of *M. tuberculosis.* For this purpose, serial clinical isolates were selected where the acquisition of an *rpoC* mutation was observed. These samples were used for whole genome sequencing and gene expression analysis, which revealed a potential link between the *rpoC* V483G mutation and upregulation of *Rv2416c* (*eis*) and *Rv1258c* (*tap*). Serendipitously, genomic data also revealed that an *ald* mutation was acquired alongside the *rpoC* mutation. Recently, *ald* has been described as a novel gene linked to D-cycloserine resistance in *M. tuberculosis*, however, to date, the mechanism of drug resistance has not been determined. Given the unique opportunity to study the effect of *ald* mutations on gene expression in our study, we investigated two genes which were found to be differentially expressed in a clinical isolate with an *ald* mutation. A turbidity-based microdilution assay revealed that upregulation of *Rv0577*, a putative glyoxalase, led to an increase in the minimum inhibitory concentration of D-cycloserine, a finding which provides novel insight into the mechanism of D-cycloserine resistance in *M. tuberculosis*. In summary, this body of work has contributed to existing knowledge surrounding drug resistance and compensatory adaptation in *M. tuberculosis*.

OPSOMMING

Enkel-nukleotiedvariante is die onderliggende dryfveer van middelweerstandigheid, fiksheid en aanpassing in *Mycobacterium tuberculosis* en die ondersoek van die meganistiese en fisiologiese aspekte van hierdie mutasies is die sleutel tot ons begrip omtrent die biologie van hierdie patogeen. Rifampisien, een van die mees kragtigste eerste linie medikasies wat gebruik word om tuberkulose te behandel, inhibeer transkripsie in *M. tuberculosis* deur te bind aan die β subeenheid van RNA polimerase (RNAP). Mikobakterieë kan egter die binding van hierdie geneesmiddel ontduik deur mutasies in die *rpoB* geen te ontwikkel. Hierdie weerstandigheidsmutasies is noodsaaklik vir die oorlewing van *M. tuberculosis* in die teenwoordigheid van rifampisien, maar as gevolg van ‘n vermoedelike vermindering in transkripsie doeltreffendheid en daaropvolgende veranderinge in geenuitdrukking, gee hierdie mutasies ‘n fiksheidskoste aan die bakterieë. Een van die meganismes wat *M. tuberculosis* gebruik om die effekte van hierdie fiksheidskoste te buffer, is om kompenserende mutasies in *rpoC* en *rpoA* te bekom. Aangesien RNAP die kern van alle meganismes van geen-regulering in *M. tuberculosis* is, is dit nie verbasend dat mutasies in hierdie ensiem tot pleiotropiese effekte lei nie. Aangesien dit 'n gebied van mikobakteriese fisiologie is wat steeds swak verstaan word, sluit die huidige werk 'n triade van studies in wat poog om die funksionele aspekte van *M. tuberculosis* RNAP en die rol van *rpoB* en *rpoC* mutasies in middelweerstandigheid beter te verstaan.

Eerstens is die effek van 'n bakteriofaag proteïen, Gp2, ondersoek om te bepaal of dit RNAP in *M. tuberculosi*s kan inhibeer. Dit is bekend dat Gp2 aan die β'-subeenheid van RNAP in *Escherichia coli* bind, daarom sal positiewe bevindings uit hierdie werk 'n raamwerk verskaf vir die identifisering van nuwe verbindings wat transkripsie inhibeer in die teenwoordigheid van *rpoB* mutasies, wat die β subeenheid beïnvloed. By wyse van *in vitro* en *in silico* analise is gevind dat Gp2 aan RNAP in *M. tuberculosis* bind en transkripsie inhibeer, maar tot 'n veel mindere mate as wat die geval is in sy *de facto* gasheer, *E. coli*. Nietemin kan toekomstige studies op ons bevindinge voortbou, aangesien *in silico* modifikasie van Gp2 'n struktuur kan identifiseer wat ’n sterker bindende affiniteit van die proteïen moontlik maak.

Tweedens is die effek van *rpoB* en *rpoC* mutasies op die funksie van mikobakteriese RNAP met behulp van 'n reeks *in vitro* transkripsie toetse ondersoek. Radioaktiwiteits-gebaseerde toetse is uitgevoer met behulp van gesuiwerde normale en mutante weergawes van die RNAP kompleks om ensiem aktiwiteit en promotor affiniteit te assesseer. Verder is fluoressensie-

gebaseerde toetse gebruik om 'n vergelykbare metode te ontwikkel sonder die gebruik van radioaktief-gemerkte nukleotiede.

Laastens het ons 'n studie onderneem om die effek van *rpoC* mutasies op die transkriptoom van *M. tuberculosis* te verstaan. Vir hierdie doel is ’n reeks opeenvolgende kliniese isolate gekies waar die ontwikkeling van die *rpoC* mutasie waargeneem is. Hierdie monsters is gebruik vir die heel genoom volgordebepaling en geenuitdrukking analise, wat 'n moontlike verband tussen die *rpoC* V483G mutasie en opregulering van *Rv2416c* (*eis*) en *Rv1258c* (*tap*) onthul het. Genomiese data het ook per toeval aan die lig gebring dat 'n *ald* mutasie saam met die *rpoC* mutasie ontwikkel het. *Ald* is onlangs beskryf as 'n nuutgevonde geen wat gekoppel is aan D-sikloserienweerstandigheid in *M. tuberculosis*, maar die meganisme van weerstand is egter tot op hede nog nie bepaal nie. Gegewe die unieke geleentheid om die effek van *ald* mutasies op geenuitdrukking in ons studie te bestudeer, het ons twee gene ondersoek wat differensiëel uitgedruk is in 'n kliniese isolaat met 'n *ald* mutasie. 'n Turbiditeits-gebaseerde mikroverdunningstoets het getoon dat opregulering van *Rv0577*, 'n vermeende glioksilase, gelei het tot 'n toename in die minimum inhibitiewe konsentrasie van D-sikloserien, 'n bevinding wat nuwe insig bied in die meganisme van D-sikloserienweerstandigheid in *M. tuberculosis*. In samevatting, hierdie werk het bygedra tot bestaande kennis rondom middelweerstandigheid en kompenserende aanpassing in *M. tuberculosis.*

AKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Professor Samantha Sampson – your continuous support, guidance and motivation has been a fundamental part of my success. I appreciate all you have done to contribute to my professional growth over the past number of years as well as the time and effort you put into every aspect of reviewing this thesis. To my UK-based co-supervisor, Professor Ramesh Wigneshweraraj – thank you for hosting me in your lab for 6 months, and to Professor Rob Warren – thank you for your insightful comments and suggestions and encouraging me to ask the types of questions that no one else does.

To all my friends and colleagues in the Division of Molecular Biology and Human Genetics – thank you for the laughs, essential coffee breaks and all your help and advice. Annika, Hanri, Tiaan and Jamie – your friendship has been one of the highlights of my time at the Tygerberg campus. Also, a big thanks to everyone in the Wigneshweraraj Lab, for all your help during my time at Imperial College and for making me feel at home from the day I arrived.

Completing this project would not have been possible without the funding I have received from The Harry Crossley Foundation, The Second Stella and Paul Lowenstein Charitable and Educational Trust, The Ethel and Eriksen Trust and the Stellenbosch Merit Bursary. I am also grateful for the financial support provided by the Royal Society Newton International Exchanges Award, The European Molecular Biology Organization, The WhiteSci Travel Award, and Boehringer Ingelheim Fonds for funding my research visits to Imperial College. Additionally, 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. Lastly, I would like to acknowledge the support of the South African Research Chairs Initiative award (UID 86539) and the South African Medical Research Council for providing funding for this study.

Most importantly, a special word of thanks to my parents, Jean and Estelle, who have encouraged me throughout my postgraduate studies, and to my sisters, Christelle and Rochelle, who have been my role models from a young age. Thank you for cheering me on when I needed it most, I wouldn't be where I am today if it weren't for you.

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

3D	:	three-dimensional
°C	:	degree Celsius
α	:	alpha
β	:	beta
β’	:	beta prime
σ	:	sigma
ω	:	omega
µg	:	microgram
µL	:	microlitre
µM	:	micromolar
µm	:	micrometre
Å	:	angstrom
A	:	alanine
ADC	:	albumin dextrose catalase
ADS	;	albumin dextrose saline
Ald	:	L-alanine dehydrogenase
ANOVA	:	analysis of variance
ATP	:	adenosine triphosphate
BCA	:	bicinchoninic acid
BDQ	:	bedaquiline
bp	:	base pair
BSA	:	bovine serum albumin
BSL-3	:	Biosafety Level 3
B. subtilis	:	*Bacillus subtilis*
C	:	cysteine
C. difficile	:	*Clostridium difficile*
CTP	:	cytidine triphosphate
CV	:	column volumes
D	;	aspartic acid
DCS	:	D-cycloserine
DNA	:	deoxyribonucleic acid
DOPE	:	discrete optimised protein energy
DST	:	drug susceptibility testing

DTT	:	dithiothreitol
E	:	glutamic acid
E. coli	:	*Escherichia coli*
EDTA	:	ethylenediaminetetraacetic acid
E. faecalis	:	*Enterococcus faecalis*
EMB	:	ethambutol
ETH	:	ethionamide
F	:	phenylalanine
FEL	:	free energy landscape
G	:	glycine
g	:	gram
gDNA	:	genomic DNA
GDP	:	guanosine diphosphate
GFP	:	green fluorescent protein
Gp	:	gene product
GTP	:	guanosine triphosphate
H	:	histidine
HCl	:	hydrochloric acid
His	:	histidine
hVISA	:	heterogeneously vancomycin-intermediate *S. aureus*
I	:	isoleucine
in/dels	:	insertions or deletions
INH	:	isoniazid
IPTG	:	isopropyl β-D-1-thiogalactopyranoside
IS	:	insertion sequence
K	:	lysine
KAN	:	kanamycin
KCl	:	potassium chloride
kDa	:	kilo-Dalton
L	:	leucine
LDSM	:	lowest DOPE score model
M	:	methionine
M	:	molar
MDR	:	multidrug-resistant

$MgCl_2$	:	magnesium chloride
MGIT	:	mycobacteria growth indicator tube
MIC	:	minimum inhibitory concentration
mL	:	millilitre
M. leprae	:	*Mycobacterium leprae*
mM	:	millimolar
MOX	:	moxifloxacin
mRNA	:	messenger RNA
M. smegmatis	:	*Mycobacterium smegmatis*
M. tuberculosis	:	*Mycobacterium tuberculosis*
N	:	asparagine
NaCl	:	sodium chloride
NAD	:	nicotinamide adenine dinucleotide
NADH	:	nicotinamide adenine dinucleotide hydride
NHLS	:	National Health Laboratory Service
Ni^{2+}	:	nickel ion
ng	:	nanogram
nM	:	nanomolar
nm	:	nanometre
NMR	:	nuclear magnetic resonance
OADC	:	oleic albumin dextrose catalase
OD	:	optical density
OFL	:	ofloxacin
P	:	proline
P. aeruginosa	:	*Pseudomonas aeruginosa*
PAGE	:	polyacrylamide gel electrophoresis
PCA	:	principal component analysis
PCR	:	polymerase chain reaction
PDIM	:	phthiocerol dimycocerosate
PE	:	proline-glutamate
PGRS	:	polymorphic GC-rich sequence
phage	:	bacteriophage
PPE	:	proline-proline-glutamate
ppGpp	:	guanosine tetraphosphate

pppGpp	:	guanosine pentaphosphate
PROVEAN	:	Protein Variation Effect Analyzer
PZA	:	pyrazinamide
Q	:	glutamine
R	:	arginine
RFLP	:	restriction fragment length polymorphism
RIF	:	rifampicin
RMSD	:	root mean square deviation
RMSF	:	root mean square fluctuation
RNA	:	ribonucleic acid
RNAP	:	RNA polymerase
RNasin	:	ribonuclease inhibitor
rNTP	:	ribonucleoside triphosphate
RPc	:	closed promoter complex
RPo	:	open promoter complex
RRDR	:	Rifampicin Resistance Determining Region
rRNA	:	ribosomal RNA
S	:	serine
SASA	:	solvent accessible surface area
S. aureus	:	*Staphylococcus aureus*
S. coelicolor	:	*Streptomyces coelicolor*
SDS	:	sodium dodecyl sulphate
S. lividans	:	*Streptomyces lividans*
S. mauvecolor	:	*Streptomyces mauvecolor*
SNP	:	single nucleotide polymorphism
sVISA	:	slow vancomycin-intermediate *S. aureus*
T	:	threonine
TB	:	tuberculosis
T_m	:	melting temperature
TR	:	transcriptional regulator
TRD	:	terizidone
tRNA	:	transfer RNA
T. thermophilus	:	*Thermus thermophilus*
UP	:	upstream promoter

UTP	:	uridine triphosphate
V	:	valine
VISA	:	vancomycin-intermediate *S. aureus*
VMD	:	Visual Molecular Dynamics
VOC	:	vicinal oxygen chelate
VSSA	:	vancomycin-susceptible *S. aureus*
W	:	tryptophan
WGS	:	whole genome sequencing
WT	:	wildtype
XDR	:	extensively drug-resistant
Y	:	tyrosine
Zn^{2+}	:	zinc ion

LIST OF TABLES AND FIGURES

Chapter 1

None

Chapter 2

Chapter 3

Chapter 5

Chapter 6

None

Chapter 1

General Introduction

1.1 Background

The incidence of tuberculosis (TB) in South Africa was estimated to be more than 834/100,000 in 2015, translating to a half a million cases in a population of approximately 50 million (1). According to data from the National Institute for Communicable Diseases, the frequency of multidrug-resistant (MDR) TB among new cases was 2.1% from 2012 to 2014 (2). MDR-TB is defined as *Mycobacterium tuberculosis* which has gained resistance to rifampicin (RIF) and isoniazid (INH), 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 (1). One such toxic drug is D-cycloserine, a second line antibiotic which has severe side effects, including paralysis, seizures, and memory loss.

The emergence of MDR-TB has had a severe impact on TB control programs in recent years and the lack of new drug candidates in the drug discovery pipeline is a cause for concern. A novel approach to finding solutions for treating bacterial infections is to investigate how bacteriophages (phages) subvert bacterial metabolic processes, such as DNA transcription, in order to facilitate phage genome transcription and development of phage progeny. Many lytic phages, which encode their own single subunit RNA polymerase (RNAP), only depend on the host bacterial RNAP for the transcription of early viral genes. Thereafter, the host RNAP is inhibited soon after the transcription of early viral genes and the transcription of middle and late viral genes is carried out by the phage-encoded RNAP. One of the ways in which phages inhibit the bacterial RNAP is through small proteins that specifically bind to and inactivate the enzymatic activity of the bacterial RNAP. A well-studied example of a phage-encoded bacterial RNAP inhibitor is T7 phage protein Gp2, which effectively inhibits RNAP activity in *Escherichia coli* (3–13).

Bacterial RNAP comprises of 5 subunits ($\alpha_2\beta\beta'\omega$) and one of several σ factors, which confers promoter specificity to the core enzyme. T7 Gp2 binds to a region called the jaw domain in the β’ subunit of RNAP (9). The jaw domain contributes to an important DNA binding channel in bacterial RNAP and functional analysis of Gp2 has indicated that it uses a multifaceted approach to inhibit the enzymatic activity of RNAP. Based on the findings from these studies, Wigneshweraraj and colleagues have identified a library of ~350 compounds (Gp2 mimetics) which can potentially inhibit bacterial RNAP in a similar mechanism as Gp2. In the current study, we aim to investigate whether Gp2 binds to mycobacterial RNAP and inhibits transcription, as this would allow for a new approach to using RNAP as a drug target.

Drug resistance in *M. tuberculosis* primarily arises as a result of chromosomal mutations in genes which encode drug targets. For example, RIF inhibits mycobacterial transcription by binding to the catalytic β subunit of DNA-dependent RNAP, encoded by the *rpoB* gene. Resistance to RIF is caused by mutations in *rpoB* which often leads to reduced affinity of RIF for RNAP. Clusters of these mutations fall within the 81 bp sequence called the "Rifampicin Resistance Determining Region" (RRDR) (14). As the majority of RIF-resistant clinical isolates are also resistant to INH, a mutation in the RRDR is seen as a surrogate marker for MDR-TB. Numerous studies have found that mutations at codons 526 and 531 are the most frequent in clinical isolates, however little is known concerning the effects of these mutations on RNAP function (15). Mutations in *rpoC*, which codes for the β' subunit of RNAP, have also been described in RIF-resistant strains and are thought to compensate for the fitness cost incurred by *rpoB* mutants, in particular those with a mutation in codon 531 (16–18). Subsequently, clinical strains of *M. tuberculosis* may also acquire mutations which allow for resistance to second line anti-TB drugs. An example of this includes mutations in *ald*, a gene implicated in D-cycloserine (DCS) resistance. In this way, the pathogen acquires multiple mutations within the host to adapt to cellular stress and antibiotic pressure. However, our understanding of the impact of these individual mutations remain elusive and present a considerable stumbling block in the treatment of drug resistant TB. One of the main aims of this study was to investigate the effect of *rpoB* and *rpoC* mutations on the function of RNAP in *M. tuberculosis* to better understand how compensatory mutations are able to ameliorate the fitness cost incurred by drug-resistant mutants. Furthermore, we aimed to study mutations in *M. tuberculosis* involved in RIF and DCS-resistance using a transcriptomics approach. This work will contribute to existing knowledge as it aims to provide new insight into the physiology of drug resistance and the mechanistic aspects of RNAP in *M. tuberculosis*.

1.2 Study 1

1.2.1 Hypothesis

Gp2 can be used as a framework for the generation of novel antibacterial compounds against *M. tuberculosis.*

1.2.2 Objectives

- Test the inhibitory effect of Gp2 on mycobacterial RNAP using *in vitro* assays.
- Compare binding of Gp2 to *E. coli* RNAP to that of *M. tuberculosis* RNAP, using *in silico* techniques.

1.3 Study 2

1.3.1 Hypothesis

Clinically relevant *rpoB* and *rpoC* mutations influence the function and activity of mycobacterial RNAP.

1.3.2 Objectives

- Generate *rpoB* and *rpoC* point mutations in genes encoding RNAP.
- Express and purify mutant and wildtype forms of mycobacterial RNAP.
- Investigate the effect of *rpoB* and *rpoC* mutations on transcription using previously described radioisotope-based assays.
- Develop a fluorescence-based *in vitro* transcription assay to investigate the effect of *rpoB and rpoC* mutations on mycobacterial transcription.

1.4 Study 3

1.4.1 Hypothesis

Compensatory *rpoC* mutations in RNAP alter the gene expression profile of *M. tuberculosis.*

1.4.2 Objectives

- Investigate the effect of compensatory *rpoC* mutations on the transcriptome of a clinical isolate of *M. tuberculosis*.
- Investigate the effect of *ald* mutations acquired alongside the compensatory *rpoC* mutation in serial clinical isolates of *M. tuberculosis.*

1.5 Structure of thesis

The chapters in this thesis consist of published and unpublished work. Chapters which have not been submitted for publication have been formatted according to the specifications of the Journal of Clinical Microbiology.

Chapter 1: General introduction to give a broad overview of the work contained in this thesis, providing a brief background for the study and underlining areas of research which are not well understood.

Chapter 2: Based on literature from a broad range of bacterial species, this review aims to provide an in-depth assessment of the structure and function of RNA polymerase and how mutations in the two largest subunits of this enzyme can influence bacterial physiology.

Chapter 3: A published chapter which consists of a study which used both *in vitro* and *in silico* methods to determine whether the bacteriophage protein, Gp2, is able to inhibit RNA polymerase in *M. tuberculosis*.

Chapter 4: The work described in this chapter includes *in vitro* assays to determine how drug resistance-conferring and compensatory mutations in mycobacterial RNA polymerase impact transcription.

Chapter 5: Chapter intended for publication which encompasses work done to investigate the mechanism of D-cycloserine resistance linked to *ald* mutations in *M. tuberculosis*, using whole genome sequencing and gene expression profiling. Furthermore, the addendum to this chapter describes additional analysis which was done using the same sample set of clinical isolates, to better understand the impact of *rpoC* mutations on the gene expression profile of *M. tuberculosis.*

Chapter 6: The final chapter summarises the findings from the individual studies in this thesis and highlights areas of research which would benefit from further studies.

1.6 References

1. World Health Organization. 2015. Global tuberculosis report 2015.
2. National Institute for Communicable Diseases. 2011. Communicable diseases surveillance bulletin.
3. 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.
4. Sheppard C, Cámara B, Shadrin A, Akulenko N, Liu M, Baldwin G, Severinov K, Cota E, Matthews S, Wigneshweraraj SR. 2011. Reprint of: inhibition of *Escherichia coli* RNAp by T7 Gp2 protein: role of negatively charged strip of amino acid residues in Gp2. J Mol Biol 412:832–841.
5. Mekler V, Minakhin L, Sheppard C, Wigneshweraraj S, Severinov K. 2011. Molecular mechanism of transcription inhibition by phage T7 Gp2 protein. Journal of Molecular Biology 413:1016–1027.
6. Shadrin A, Sheppard C, Severinov K, Matthews S, Wigneshweraraj S. 2012. Substitutions in the *Escherichia coli* RNA polymerase inhibitor T7 Gp2 that allow inhibition of transcription when the primary interaction interface between Gp2 and RNA polymerase becomes compromised. Microbiology (Reading, Engl) 158:2753–2764.
7. James E, Liu M, Sheppard C, Mekler V, Cámara B, Liu B, Simpson P, Cota E, Severinov K, Matthews S, Wigneshweraraj S. 2012. Structural and mechanistic basis for the inhibition of *Escherichia coli* RNA polymerase by T7 Gp2. Mol Cell 47:755–766.
8. Shadrin A, Sheppard C, Savalia D, Severinov K, Wigneshweraraj S. 2013. Overexpression of *Escherichia coli udk* mimics the absence of T7 Gp2 function and thereby abrogates successful infection by T7 phage. Microbiology (Reading, Engl) 159:269–274.
9. Nechaev S, Severinov K. 1999. Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein. Journal of Molecular Biology 289:815–826.
10. 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.
11. 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.

12. Hesselbach BA, Nakada D. 1977. I protein: bacteriophage T7-coded inhibitor of *Escherichia coli* RNA polymerase. J Virol 24:746–760.
13. 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.
14. 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.
15. Ramaswamy S. 1998. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tubercle & Lung Disease 79:3.
16. 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.
17. De Vos M, 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.
18. Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Molecular Microbiology 85:142–151.

Chapter 2

Literature Review

The phenotypic effect of *rpoB* and *rpoC* mutations

2.1 Abstract

RNA polymerase is a large, multi-subunit enzyme responsible for the transcription of DNA into RNA. Not surprisingly, inhibition of this essential enzyme is an established strategy for treating a broad range of bacterial infections. A primary example of this is rifampicin, an antibiotic which binds to a region near the active site of RNA polymerase and prevents elongation of the RNA transcript. Furthermore, viruses which infect bacteria often encode proteins which bind to bacterial RNA polymerase to regulate its function. Numerous studies have described polymorphisms in the *rpoB* gene, encoding the β subunit of RNA polymerase, which prevent binding of rifamycins and viral proteins. Additional mutations in the *rpoC*-encoded β' subunit, known as compensatory mutations, have been identified in numerous species, which allow the bacterium to overcome the fitness cost associated with drug resistance. These mutations are of particular importance to our understanding of drug resistant *Mycobacterium tuberculosis*, the causative agent of tuberculosis. However, current knowledge of the pleiotropic effects of compensatory and drug resistance-conferring mutations is relatively limited and requires further study to allow for a better understanding of bacterial physiology. This review aims to summarise available literature and evaluate findings from other organisms to advance our understanding of the functional and phenotypic effect of *rpoB* and *rpoC* mutations in RNA polymerase.

2.2 Introduction

As it is vital for prokaryotes to be able to modulate gene expression in response to external stress and internal stimuli, RNA polymerase (RNAP) plays an indispensable role in transcriptional regulation. The interplay between transcriptional activators, repressors, and their interaction with transcriptional machinery adds an extra layer of complexity to this regulation, making it an intriguing, yet challenging area of scientific study (Figure 2.1). Since rifampicin was introduced as part of a four-drug tuberculosis treatment regimen in the 1960's, numerous studies have described polymorphisms in the *rpoB* gene, which allows for bacterial resistance against this first-line drug (1–4). The identification of drug-resistance causing mutations has been useful in the development of novel diagnostic tools. Additionally, characterization of compensatory mutations, which offset the fitness cost of *rpoB* mutations, provide insight into how bacteria are able to adapt and respond to changes in their environment.

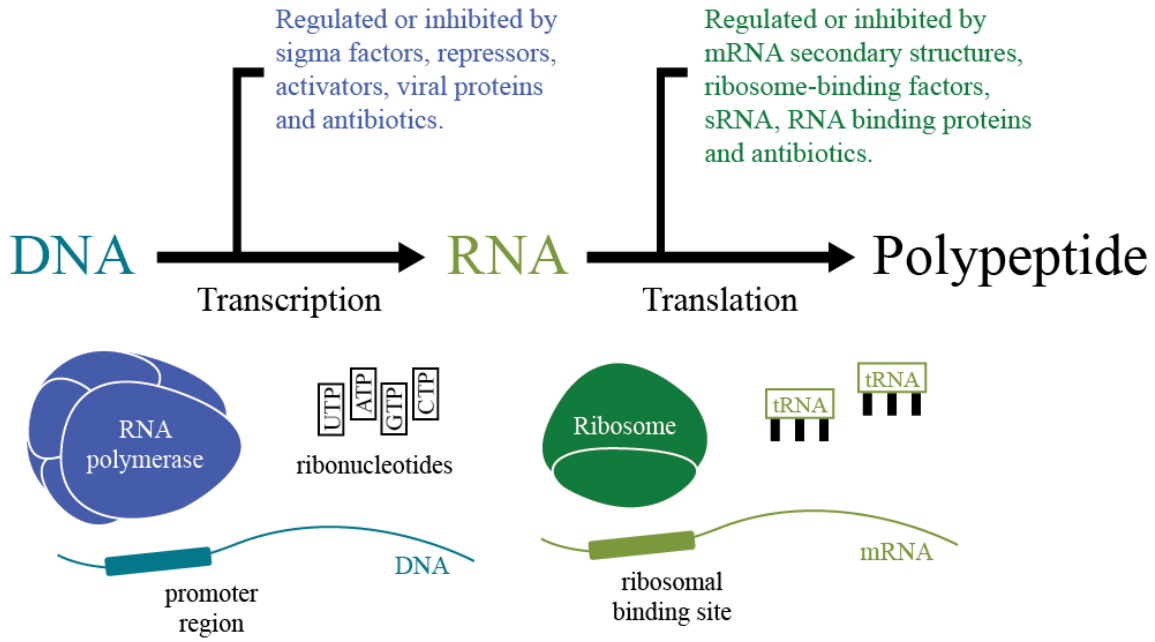


Figure 2.1 Regulatory factors which modulate gene expression and protein synthesis. RNA polymerase is the target of numerous transcription factors, viral proteins and antibiotics which can modify its function and have downstream effects on transcription and translation.

For *Mycobacterium tuberculosis*, understanding how these mutations influence transcription efficiency as well as regulation of gene expression is an important aspect of understanding how drug resistance influences features of bacterial physiology, including fitness, pathogenicity and

virulence. As only a limited number of studies have focused on investigating transcription in *M. tuberculosis*, the current review aims to expand our understanding on the subject by considering literature available for other prokaryotic species such as *Escherichia coli*, *Streptomyces coelicolor* and *Staphylococcus aureus*. A number of years ago, a comprehensive literature review by Koch *et al.* meticulously highlighted the impact of clinically relevant mutations on *M. tuberculosis* physiology. However, the authors noted that only a brief review of studies which investigated phenotypic consequences of *rpoB* mutations was provided, therefore we aimed to re-explore the topic and, wherever possible, include novel literature which provides insight into the functional and mechanistic effects of mutations in RNAP (5).

2.3 Bacterial transcription

Prokaryotic RNAP holoenzyme is a multi-subunit protein, consisting of the core enzyme ($\alpha_2\beta\beta'\omega$) and one of several sigma (σ) factors (6). During transcription initiation, a single σ factor binds to the RNAP core to form the holoenzyme. Although it does not play a direct role in RNA synthesis, the σ factor facilitates promoter recognition by decreasing the affinity of RNAP for non-specific DNA. Two upstream promoter regions, known as the −10 and −35 regions, play a key role in binding of RNAP through their interaction with the σ factor. Once the holoenzyme binds to a promoter region to form the closed promoter complex (RP_C), a series of conformational changes are initiated, which includes σ-mediated strand separation near the transcription start site. The resulting open promoter complex (RP_O) facilitates base pairing between the template strand and incoming ribonucleotides. Once a 9-10 nucleotide strand of RNA has been synthesized, the σ factor usually dissociates from the complex, allowing the core enzyme to continue with elongation of the RNA transcript (Figure 2.2).

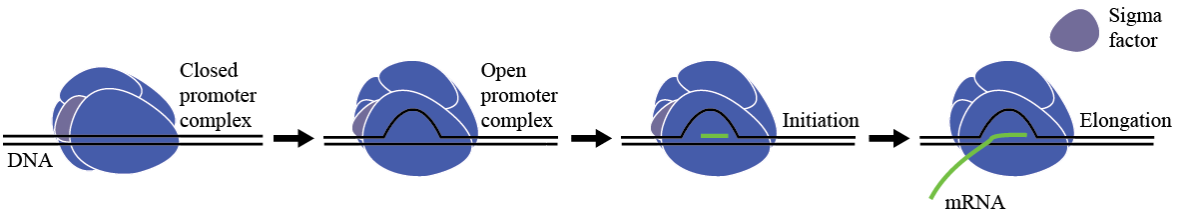


Figure 2.2 RNA polymerase undergoes conformational changes during transcription. Once DNA has melted, the open promoter complex (RP_O) allows for synthesis of the initial RNA transcript, which leads to elongation once the σ factor dissociates from the complex.

Termination of transcription can take place through either an intrinsic or factor-dependent mechanism. An example of the latter is termination which relies on the hexameric Rho protein. It is thought that Rho binds to a recognition site in the newly-transcribed strand of RNA, moves towards the RNAP:DNA complex driven by ATP hydrolysis and dissociates RNAP from the transcription bubble (7). However, intrinsic termination is mediated through the formation of RNA hairpin structures, encoded by termination sites in the DNA template. These sites consist of a GC-rich sequence, followed by a succession of 7-9 uridine residues, known as a U-tract (8). The hairpin structure causes the RNAP complex to pause and it is thought that the termination site in the DNA sequence then destabilizes the elongation complex.

2.4 Composition and Structure of RNA polymerase

The beta chains, β and β', encoded by *rpoB* and *rpoC*, respectively, are the largest subunits of the enzyme and resemble a crab pincer as they appear to clasp the DNA strand during transcription (Figure 2.3). The β' and β subunits are responsible for the formation of the DNA-binding channel, which includes the active-centre cleft necessary for ribonucleotide polymerization (9). Two identical alpha (α) chains are transcribed from *rpoA* and consist of an N-terminal and C-terminal domain connected by a flexible linker (10, 11). The C-terminal domain interacts with transcription factors and stabilizes the RNAP-promoter complex by binding to upstream promoter (UP) elements at some promoters.

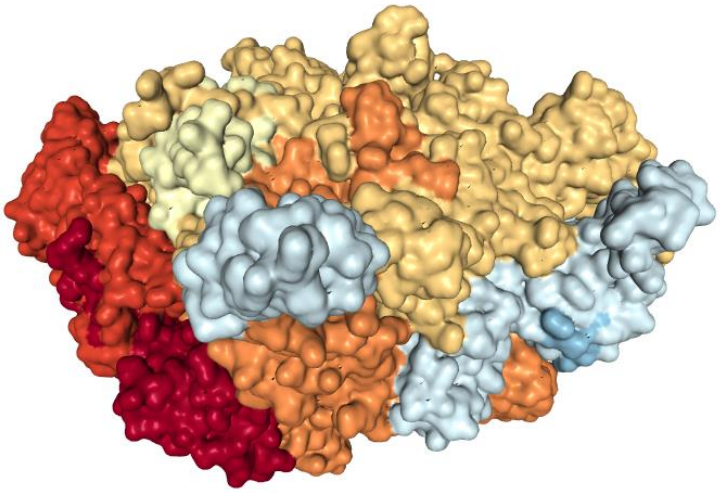


Figure 2.3 Crystal structure of *M. tuberculosis* RNA polymerase. Alpha (α) chains shown in dark red and light red, omega (ω) subunit shown in light green, dissociable sigma (σ) factor in blue and beta chains (β, β') in yellow and orange, respectively. Crystal structure solved by Lin *et al.* (12). Image generated using NGL Viewer (13, 14).

The omega (ω) subunit, encoded by *rpoZ*, is the smallest subunit in the holoenzyme, and is thought to facilitate recruitment of β′ during RNAP assembly (15). This small subunit is not essential in prokaryotes and was overlooked in early studies as it was only present in negligible amounts in crude extracts of RNAP. However, recent work has hinted at its importance in bacterial physiology. A *rpoZ* deletion strain of *S. coelicolor* exhibited a decrease in growth rate as well as reduced antibiotic production compared to the control strain (16). Similarly, a Δ*rpoZ* strain of *Mycobacterium smegmatis* presented a range of surface-related phenotypes such as a change in colony formation, a decrease in sliding motility and reduced biofilm formation (17). The authors associated these characteristics with a defect in the ability of mycobacteria to establish an extracellular matrix. Furthermore, Gunnelius *et al.* showed that recruitment of the housekeeping σ factor to RNAP core in cyanobacterium *Synechocystis* sp. PCC 6803 was less efficient in a strain where *rpoZ* was inactivated, and reported downregulation of usually highly expressed genes and upregulation of genes which are usually expressed at a low level (18). Together, these findings demonstrate some of the pleiotropic effects which can be observed when the transcriptional machinery is disrupted.

The association of σ factors with the core enzyme is required for initiation of transcription as it plays an important role in promoter recognition and strand separation. Two main families of σ factors have been described in prokaryotes, designated according to their molecular weight: σ^{54} and σ^{70}, which are 54 kDa and 70 kDa, respectively. Based on the presence or absence of four conserved domains, the σ^{70} family is further categorized into Group 1 and Groups 2-4, consisting of primary and alternative σ factors, respectively. A recent review by Paget (2015) provides an in-depth view on the structure and function of bacterial sigma factors (19). The primary σ factor, σ^{A}, belonging to Group 1, is responsible for the transcription of housekeeping genes. However alternative σ factors, belonging to Groups 2-4, can bind to RNAP to modify the affinity of the enzyme for specific promoter regions (20). Efficient modulation of bacterial gene expression is achieved through binding of different σ factors, which lead to diverse transcriptional responses through expression of particular sets of genes. The number of σ factors which a bacterial genome encodes is thought to reflect the degree of stress the species encounters in its immediate environment or host. For example, the symbiotic gut bacterium, *E. coli*, encodes 7 σ factors, however pathogenic *M. tuberculosis*, which is required to survive under anaerobic conditions within macrophages, encodes for 13 σ factors (21, 22).

2.5 Mutations in the β and β' subunits of RNA polymerase

2.5.1 Rifampicin resistance

Rifampicin (RIF) is a broad-spectrum antibiotic which binds to the β subunit of RNAP and inhibits mycobacterial transcription (2). The binding site of RIF is approximately 12 Å away from the active site of RNAP and prevents extension of RNA transcripts longer than 2-3 nucleotides (23). RIF-resistant isolates exhibit mutations in the *rpoB* gene which led to reduced binding of RIF to RNAP (24). The majority of these mutations fall within the 81 bp sequence called the "RIF Resistance Determining Region" (RRDR) (1). A systematic review by Ramaswamy and Musser in 1998, revealed that approximately 95% of RIF-resistant isolates have short insertions/deletions or point mutations in this region of *rpoB* (1). This has allowed for the development of diagnostic tools such as the Xpert® MTB/RIF which is able to rapidly identify strains of *M. tuberculosis* which are resistant to RIF without the need for phenotypic testing (25). As approximately 98% of RIF-resistant strains, are also resistant to the first-line drug isoniazid, a mutation within the RRDR of *rpoB* can be used as a surrogate marker for multidrug-resistant (MDR) TB (26). Interestingly, however, studies have described mutations within the RRDR which do not confer RIF resistance. Campbell and colleagues reported isolates harbouring H526L and H526N mutations which were RIF-susceptible (it is worthwhile to note that this refers to codon position 445 in *M. tuberculosis*, however, by convention, mutations are numbered as in *E. coli*) (27). Furthermore, Yoon *et al*. identified an *rpoB* H526C mutation in a RIF susceptible clinical isolate of *M. tuberculosis* and Lavania *et al.* reported that the *rpoB* Q517H substitution is not associated with RIF resistance in *Mycobacterium leprae* (28, 29). By virtue of their selection, these mutations bestow an advantage to the bacterium, possibly by conferring low level RIF resistance which would allow for growth in patients with low plasma concentrations of RIF. However, these findings could also be an indication that certain amino acid changes in the RRDR do not confer RIF resistance, suggesting an alternative role for *rpoB* mutations in clinical isolates of mycobacteria. One possibility could be that *rpoB* mutations allow for resistance to other anti-TB drugs, through the upregulation of efflux pumps. In possible support of this, Siddiqi *et al*. observed a 6- to 10-fold increase in the expression of *Rv1258c*, a gene encoding an integral membrane transport protein, when an MDR isolate of *M. tuberculosis*, harbouring an *rpoB* mutation, was grown in the presence of either ofloxacin or RIF (30).

In another study, RIF-resistant *Enterococcus faecalis* was unexpectedly found to exhibit intrinsic resistance to cephalosporins, drugs which hinder peptidoglycan biosynthesis by

inhibiting penicillin-binding proteins (31). It was reported that the *rpoB* H486Y mutation, which corresponds to *rpoB* H526Y *in M. tuberculosis*, led to enhanced cephalosporin resistance without any effect on resistance towards drugs which target other aspects of cell wall synthesis or cellular processes. Conversely, other mutations at the same locus and at *rpoB*473, exhibited susceptibility to these drugs, which suggests that specific alleles in *rpoB* exert unique effects on *E. faecalis*. The authors proposed that the *rpoB* H526Y mutation could be linked to differential expression of genes that are critical for cellular adaptation to cephalosporin exposure.

Pang *et al.* observed that certain mutations in *M. tuberculosis rpoB* are associated with high-level RIF resistance and investigated the mechanistic effect of 6 different mutations (S531L, H526D, H526G, H526L, H526R, and L533P) using a three-dimensional *in silico* model of RNAP. Their findings revealed that mutations S531L and H526D disrupted the intermolecular hydrogen bonds between RNAP and RIF (32). Based on their analysis, the S531L mutation led to the formation of a structural bump whereas the H526D mutation introduced a negative charge, both of which would reduce the affinity of RNAP for RIF molecules. As a result, these amino acid changes appear to facilitate a mechanism which allows for high level RIF-resistance that has previously been associated with these substitutions (4). These underlying differences between mutations was further confirmed by Molodtsov *et al*., who determined the X-ray crystal structures of *E. coli* RNAP harbouring different *rpoB* mutations (33). Their study showed that S531L mutations confer little to no structural or functional impact on the enzyme when in the absence of RIF, however the H526Y mutation led to considerable steric conflicts and a change in the RIF binding pocket. Additional analysis by Singh *et al*. using a molecular model of *M. tuberculosis* RNAP with eight different amino acid substitutions at codon 526, further demonstrated the differences in structural mechanisms involved in RIF resistance (34).

Not surprisingly, the *rpoB* S531L mutation is one of the most common in RIF-resistant clinical isolates and *in vitro* generated mutants of *M. tuberculosis*. Besides allowing for high level RIF resistance, this could also be attributed to the low level of impact on mycobacterial fitness (3, 35). To investigate this, Gagneux *et al.* studied the fitness cost caused by *rpoB* mutations in RIF-resistant *in vitro* mutants and clinical isolates of *M. tuberculosis* (36). The fitness cost was determined according to the RIF-resistant strain's ability to compete for limited resources in the same environment as its RIF-sensitive progenitor. The authors reported a fitness cost in all *in vitro* generated mutants, yet no fitness cost was seen in most of the clinical isolates with a

S531L mutation. Other than this being an overall low-cost mutation, another possible explanation could be that strains with this mutation were able to employ a compensatory mechanism which mitigated the fitness cost associated with RIF resistance.

2.5.2 Compensatory mutations

Follow up studies describing compensatory mechanisms in RIF resistant clinical isolates of *M. tuberculosis* hypothesised that the acquisition of mutations in *rpoA* and *rpoC* led to the emergence of high fitness MDR strains (37, 38). These compensatory mutations occurred in a large number of clinical isolates, however the authors emphasised that the mechanism through which these mutations influence bacterial fitness is unknown. More recently, De Vos *et al.* explored the epidemiological significance of *rpoC* mutations in *M. tuberculosis* and reported that 23.5% of RIF-resistant clinical isolates included in their sample set harboured a non-synonymous polymorphism in *rpoC* (39). Additionally, these mutations were significantly associated with *rpoB* S531L mutations. The authors concluded that drug resistance-conferring mutations, *rpoC* mutations, as well as the genetic background of the strain all play a role in compensatory evolution of drug-resistant *M. tuberculosis*, however, the relationship between *rpoB* mutation and *rpoC* mutations requires further functional study in this organism.

Recently, Qi *et al.* explored the fitness cost of *rpoB* mutations and compensatory mechanisms in *Pseudomonas aeruginosa* (40). Firstly, a RIF sensitive strain carrying an integrated luciferase reporter construct was used as a parental strain from which several RIF-resistant mutants were generated (Q518L/R, D521G/N, H531L/R/Y, S536F, S579F). Each of the resistant strains, were placed in competition against a GFP-tagged control strain and the competitive fitness of the non-fluorescent strains were calculated relative to the proportion of the fluorescent control, with standardisation based on the RIF-sensitive progenitor. Results from this assay demonstrated that the *rpoB* mutations led to a growth fitness cost. In alignment with this, the transcriptional efficiency of the mutants were evaluated by comparing luciferase reporter activity to that of the wildtype progenitor. The results of this assay showed that all but one of the *rpoB* mutations had a direct impact on the activity of RNAP and reduced the transcriptional activity of the enzyme. This phenotype, as well as the fitness cost of the resistant strains, could be reversed in the presence of a second, compensatory mutation in *rpoB* (N573S, E528D). Interestingly, this correlates to findings from Meftahi *et al.*, where it was found that a secondary mutation in *rpoB* (V615M) compensated for the fitness cost in clinical isolates of *M. tuberculosis* with an *rpoB* S531L mutation (41). Further to this, the Qi *et al.* study reported that differential expression of

essential genes in the *rpoB* mutants did not lead to an additional fitness cost, but rather that the indirect consequence of these resistance-conferring mutations is linked to genome-wide changes in the transcriptional profiles of the mutants. Additionally, the authors speculated that upregulation of the *rpoBC* operon, which encodes the β and β' subunits, plays a role in the mechanisms used by the bacterium to buffer the fitness cost of RIF resistance.

2.5.3 Fidaxomicin (*lipiarmycin*) *resistance*

Mutations in *rpoB* have also been associated with decreased susceptibility to fidaxomicin (otherwise known as lipiarmycin), a macrocyclic antibiotic which is used to treat *Clostridium difficile* infections. Like RIF, fidaxomicin binds to the β subunit of RNAP and inhibits transcription by preventing synthesis of the RNA transcript (42). In a study by Leeds *et al*., several strains of *C. difficile* were subjected to serial passaging in the presence of increasing concentrations of fidaxomicin to generate spontaneous mutants (43). Subsequent whole genome sequencing of the mutants revealed an *rpoB* (Q1073R) mutation in a clone with reduced susceptibility to fidaxomicin. This finding was in line with a previous study which described strains of *C. difficile* with an increased minimum inhibitory concentration (MIC) of fidaxomicin as a result of a double mutation in *rpoC* (Q781R, D1127E) or a single mutation in *rpoB* (Q1074K) (44). Furthermore, a clinical isolate of *C. difficile* acquired an *rpoB* mutation (V1143G) after a patient was treated with fidaxomicin in a phase III clinical study (45, 46).

Interestingly, even though resistance to the compound is mediated through mutations in the β and β' subunit, lipiarmycin resistant mutants do not appear to be cross-resistant to RIF (47–49). Interestingly, even though resistance to the compound is mediated through mutations in the β and β' subunit, lipiarmycin resistant mutants do not appear to be cross-resistant to RIF (48). However, it is possible that clinical isolates with compensatory *rpoC* mutations would display cross-resistance to lipiarmycin. The *rpoC* mutations reported in this study include P326R and R412Q, which have not been described in *M. tuberculosis*, however these mutations are very close in proximity to G332R and V432A mutations, which have been reported previously (38, 39).

2.5.4 Vancomycin resistance

A clinical strain of vancomycin-intermediate *S. aureus* (VISA), Mu50, was found to harbour two mutations in genes involved in two-component regulatory systems, which led to the sequential acquisition of vancomycin resistance (50). The first mutation in *vraS* is responsible for the progression of vancomycin-susceptible *S. aureus* (VSSA) to a heterogeneously vancomycin-intermediate *S. aureus* (hVISA). Subsequently, a mutation in *graR* led to the conversion of this strain into VISA. As the introduction of the *graR* mutation into hVISA by gene replacement did not lead to the VISA phenotype, the group investigated the genome of the hVISA strain in comparison to Mu50 and identified a mutation in *rpoB* which was responsible for the increase in the level of vancomycin resistance necessary to achieve the VISA phenotype. Further investigation revealed that an *rpoB* mutation was present in 70% of their collection of clinical VISA strains and that this mutation could be linked to thickening of the cell wall and delayed growth rate. The mechanism by which this occurs could not be determined in this study, however the authors speculated that *rpoB* mutations change the global transcriptome of *S. aureus*, thereby altering the physiology of the cell and its susceptibility to the antibiotic. They further hypothesized that the *rpoB* mutation investigated in their study, H481Y, initiated the stringent response.

In line with this, a study by Matsuo *et al.* described an *rpoC* mutation (P440L) in *S. aureus* which led to a high MIC of vancomycin (51). The group hypothesised that the mutation causes upregulation of genes involved in peptidoglycan synthesis, while simultaneously repressing other metabolic processes in order to conserve energy. This leads to vancomycin-intermediate slow VISA, which differs from vancomycin susceptible hVISA in that it has a slow growth phenotype and a higher level of vancomycin resistance (Figure 2.4). However, a secondary mutation in *rpoC* (F562L) mitigated the negative effects of this resistance-conferring mutation, illustrating the same mechanism of compensatory evolution as seen in RIF resistant *M. tuberculosis*.

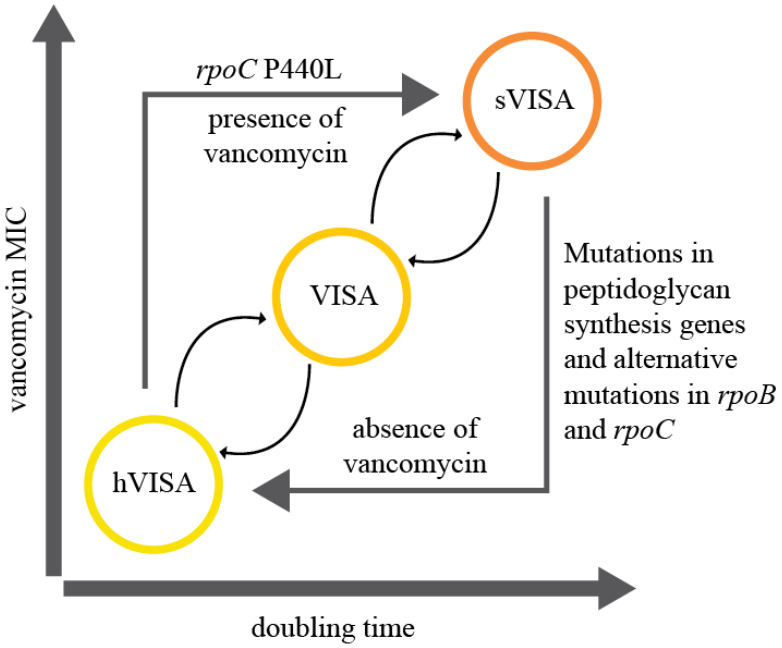


Figure 2.4 Illustration of the relationship between heterogeneously vancomycin-intermediate *S. aureus* (hVISA) and slow vancomycin-intermediate *S. aureus* (sVISA). Upon exposure to vancomycin, hVISA generates VISA and sVISA, which exhibits an increased MIC, yet a longer doubling time. This adaptation can be linked to the acquisition of the *rpoC*440 mutation. Subsequently, sVISA can revert back to hVISA in the absence of vancomycin, which is associated with additional mutations in *rpoB* and *rpoC* as well as mutations in genes responsible for peptidoglycan synthesis. Figure adapted from Matsuo *et al.* (51).

2.5.5 The stringent response

The stringent response is a global regulatory mechanism, which allows bacteria to adapt during unfavourable conditions by altering transcription of certain sets of genes. Stringent genes, which are negatively controlled during the stringent response, are expressed at high levels during ideal growth conditions. However, when amino acid availability is low, transcription of these genes discontinues in order to conserve cellular resources (Figure 2.5). Once the stringent response is triggered, increased activity of the enzyme RelA leads to accumulation of guanine nucleotides (ppGpp/pppGpp) which regulate transcription by binding to bacterial RNAP (52, 53). As a result, transcription from promoters which control expression of genes encoding stable RNA (rRNA and tRNA) are downregulated (54). In contrast, transcription from genes which encode amino acid biosynthesis and alternative σ factors are induced (55). Early studies confirmed that guanine nucleotides interact with bacterial RNAP at the N-terminal of the ω subunit and a double-psi β-barrel domain of the β' subunit, a site approximately 30 Å away

from the active site of RNAP (56–59). Recently, however, Ross *et al*. identified a second site for binding of guanine nucleotides, which is located at the β' subunit and the interface of DksA, a transcription factor which plays a key role in the stringent response (60). It is thought that DksA acts in an allosteric mechanism, whereby its coiled-coil domain interacts with the RNAP secondary channel, destabilizing RNAP-promoter complexes (61, 62). Dahl *et al*. reported that guanine nucleotides regulate the expression of 159 genes during starvation in *M. tuberculosis* (63). A number of these genes were found to be associated with persistence and the repression of metabolism, which allow for an increased chance of survival during unfavourable conditions. The role of guanine nucleotides, including the function of Rel in the context of *M. tuberculosis*, has been comprehensively reviewed by Hauryliuk *et al*. (64).

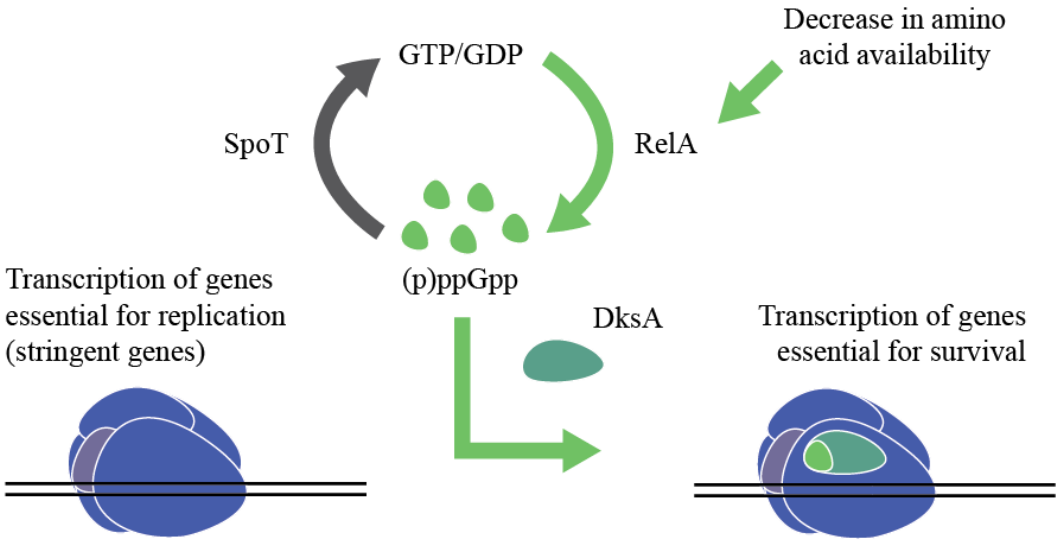


Figure 2.5 Stringent genes are transcribed during exponential growth, but repressed when nutrients are limited. RelA is responsible for the synthesis of (p)ppGpp, which regulates transcription by binding to RNAP together with DksA. Once suitable growth conditions return, the enzyme SpoT degrades guanine nucleotides, allowing for transcription of stringent genes.

Before it was confirmed that guanine nucleotides bind to RNAP, studies from as early as the 1980s showed a link between *rpoB* mutations and the stringent response (65–69). Little *et al.* isolated RIF-resistant strains of *E. coli* with mutations at codons 331 and 332 of *rpoB* and found that they exhibit a 20-fold increase in intracellular sensitivity to guanine nucleotides (65, 66). However, the levels of ppGpp in these *rpoB* mutants were approximately 10 times lower than their wildtype counterparts. This indicates that commonly-occurring *rpoB* mutations which

confer RIF resistance in prokaryotes, could affect guanine nucleotide metabolism and disturb the regulatory mechanism induced by their interaction with RNAP. In contrast to this, Nene and Glass isolated an *rpoB* mutant which displayed attenuation of the stringent response upon amino acid deprivation (67). It was later reported that purified RNAP from this mutant was found to exhibit increased resistance to guanine nucleotides *in vitro* (68). In 1992, Tedin and Bremer generated mutants by overexpressing RelA, which allowed for sublethal cytoplasmic levels of guanine nucleotides (70). The authors reported the isolation of *rpoB* mutants in which synthesis of RNA was only mildly affected by increased levels of guanine nucleotides, compared to wildtype *E. coli*, in which RNA accumulation (of which the majority is rRNA and tRNA) was reduced by 90%. These studies indicate that mutations in different regions of the β subunit may have distinct effects on the stringent response. In the latter case, it appears that binding of ppGpp/pppGpp to RNAP was altered, perhaps through a change in conformation or binding site, which thereby prevented the regulatory activity of guanine nucleotides.

It is also possible that the overarching effects of different mutations could be as a result of their influence on the mechanistic aspects of transcription in prokaryotes. Conrad *et al*. reported the discovery of *rpoB* and *rpoC* mutations upon adaptive evolution of *E. coli* in glycerol minimal medium (71, 72). Subsequent analysis of these mutants revealed that while they have a faster growth rate compared to the progenitor strain in minimal media, they exhibit slower growth in rich medium (72). Transcriptional pausing and longevity of open complexes at an rRNA promoter were evaluated and found to be decreased in the case of mutated RNAP, while the elongation rate increased. Furthermore, transcriptomics and genome-wide mapping of RNAP revealed drastic changes in global gene expression and confirmed that mutants exhibit a redistribution of polymerase from promoters of stringent genes to other regions of the genome. In a study by Zhuo *et al*., it was found that RNAP in *E. coli* is unstable at promoters of stringent genes and interchanges between stable and metastable states (73). Furthermore, *in vitro* transcription assays showed that RIF resistance-conferring mutations (*rpoB* S531F, Δ532A, L533P, T563P) further weaken the interaction between RNAP and stringently controlled promoters and *in vivo* results revealed that *E. coli* with these *rpoB* mutations mimic the stringent response. Mutated RNAP exhibited a 2- to 4-fold decrease in transcription from a promoter responsible for rRNA expression and a 2- to 3-fold increase in transcription from a non-stringent promoter. The authors proposed that the stability of initiation complexes plays an important role in the transcription of stringent genes.

RNAP with certain *rpoB* mutations has also been found to mimic stringent RNAP in *S. coelicolor* and *Streptomyces lividans* (which are members of the same order as *Mycobacterium sp.*) as they led to induction of genes involved in antibiotic biosynthesis, a mechanism usually attributed to p/ppGpp-bound RNAP (74–76). Strikingly, Hosaka *et al*. reported the biosynthesis of a novel class of antibacterials in the nonantibiotic-producing strain *Streptomyces mauvecolor* as a result of two *rpoB* mutations (H526D, H526L) (77). This was thought to be due to the increased affinity of mutated RNAP for promoters of silent genes. Furthermore, research by Inaoka *et al.* demonstrated that a RIF-resistant mutant of *Bacillus subtilis* produced neotrehalosadiamine, an aminosugar antibiotic which is not produced in the wild-type strain (78). The mutation found to be responsible for this change in cellular metabolism was the frequently reported *rpoB* S531L substitution, which led to upregulation of the *ntdABC* operon, confirmed to be responsible for neotrehalosadiamine biosynthesis. The authors proposed that this mutation enhanced the activity of RNAP at σ^A-regulated promoters, as is the case for the *ntdABC* operon, a phenomenon which was not seen with RNAP harbouring mutations at codon 526.

2.5.6 Physiology and adaptation

It is evident that mutations in RNAP have a profound impact on bacterial physiology, yet the numerous factors which contribute to these changes remain elusive. Studies which incorporate a systems biology approach are urgently needed to investigate the cellular effects in response to mutations in transcriptional machinery. A good example of this is a study by Utrilla *et al*., which used multi-scale computational and experimental analysis to study the effects of two *rpoB* mutations (E546V, E672K) in *E. coli* and demonstrated their role in resource allocation for cellular processes (79). The authors showed that the single point mutations led to reprogramming of cellular energy allocation towards growth, instead of mechanisms involved in stress functions such as survival during acid shock and growth on complex media. Interestingly, they proposed that the mutations, even though they are far apart, lie within the same structural community of ~250bp in *rpoB* (which refers to the sub-domain architecture of the protein) and therefore have the same pleiotropic effect. Given that the most prominent RIF-resistance causing mutations (H526Y, S531L) fall within the same region, it could be hypothesized that the same adaptive mechanism applies to mycobacterial RNAP with these mutations. Interestingly, molecular dynamic simulations revealed destabilization between the β and β’ subunits in the presence of particular mutations at codon 672, which correlated to the increased growth phenotype (79). This indicates that particular mutations could have distinct

effects on the intra-holoenzyme affinity of subunits which make up RNAP, another factor which may be driving the global changes in gene expression seen in *rpoB* mutants.

Furthermore, evidence suggests that mutations in *rpoB* could do more than just influence the phenotype of the bacteria, they could play an important role in host-pathogen interactions. Gao *et al*. reported global changes in the transcriptome of *S. aureus* harbouring an *rpoB* H481Y mutation (361 genes differentially expressed with a fold change cut-off of >1.5-fold and corrected p-value of < 0.05) (80). The authors reported that the *rpoB* mutant, although displaying attenuated virulence in a murine bacteremia model, showed enhanced survival in a whole-blood killing assay and found that this was due to upregulation of capsule production. Furthermore, the *rpoB* mutant demonstrated increased resistance to two human antimicrobial peptides, human neutrophil peptide 1 and human β-defensin 2. Interestingly, the study included a *relA* mutant which showed a very similar profile to the *rpoB* mutant with global changes in gene expression, including upregulation of capsule biosynthesis, and enhanced resistance against antimicrobial peptides and whole-blood killing.

The exact mechanisms which led to remodeling of mycobacterial physiology are unclear, however it could be linked to changes in the constituents of the cell wall. Bisson *et al*. investigated the effect of *rpoB* mutations on the proteome and metabolome of *M. tuberculosis* and reported upregulation of polyketide synthase genes (*ppsA-ppsE*, *drrA*) which are involved in the phthiocerol dimycocerosate (PDIM) biosynthetic pathway (81). The authors noted that PDIM is thought to play a role in the interaction between *M. tuberculosis* and the host macrophage and speculated that upregulation of the PDIM pathway could lead to remodeling of the cell wall. In agreement with the findings from this research, Lahiri *et al*. demonstrated that *rpoB* mutants (Q513E, H526Y, S531L) exhibit changes in the abundance of cell wall lipids, namely mycobactin siderophores and acetylated sulfoglycolipids (82). Similarly, a study using rifaximin-resistant bifidobacteria demonstrated an increase in saturated fatty acids, which indicates that this may be a conserved feature in *rpoB* mutants (83).

2.5.7 Transcription of bacteriophage genes

In nature, bacteria can be infected by viruses called bacteriophages (phages), which are often characterized by their life cycle. Lytic phages take over the host transcription machinery once they have infected the bacterial cell in order to transcribe viral genes needed for proliferation of the phage (84). Once new phage progeny have assembled, the host cell is promptly lysed to

release the phage particles. Lysogenic phages, on the other hand, integrate into the host genome or exist within their host as plasmids, where they can either remain for an indefinite period of time or switch to the lytic cycle upon an environmental trigger. Both these scenarios require the phage to manipulate bacterial RNAP at certain points in their life cycle. First to transcribe early viral proteins, and subsequently, to inhibit bacterial transcription once phage RNAP has been assembled. For this reason, phage proteins are known to interact with bacterial RNAP, either during transcription of viral genes or host shut-off.

Swapna *et al.* investigated the interaction between RNAP and transcription activator C from phage *Mu* (85). Unlike most activators in nature which interact with the α subunit and σ factor (86, 87), the study found that this particular activator binds to the β' subunit of RNAP holoenzyme in the presence of DNA containing the C binding sequence. Along with facilitating recruitment of RNAP to the phage gene, *mom*, the activator reduces abortive transcription and thereby improves promoter clearance. Interestingly, the group isolated an *rpoC* mutant (G524D) of *E. coli* which hampered clearance of the *mom* gene promoter within the phage genome, but did not compromise transcription at typical promoters in *E. coli*. This demonstrates that certain mutations in RNAP can have a distinct effect on a particular set of promoters, yet transcription activity at other locations in the genome will remain unaffected. As G524D does not fall within the binding site of the activator, the authors concluded that an allosteric mechanism of control may be involved.

Furthermore, a study by Georgopoulos *et al*. described the isolation of *E. coli* mutants which were able to block phage development (88). These mutants were termed *groN*, as they were able to prevent the interaction between RNAP and the product of gene *N* in phage λ, thus arresting growth of the phage. As the study was done prior to the widespread use of DNA sequencing, the authors were not able to identify the position of the mutation. However, as the mutated polymerase appeared to be more sensitive to rifamycin *in vitro*, it was assumed that the mutation was close to the RRDR in *E. coli*. Together, these studies illustrate that, just as bacteria acquire mutations to gain resistance to antibiotics, they are also able to overcome the activity of phage proteins which would otherwise lead to their demise.

2.6 Knowledge gaps

Research based largely on *E. coli*, but also other bacteria, has greatly improved our understanding of bacterial physiology; however, additional studies in *M. tuberculosis* are urgently needed to better our understanding of polymorphisms which affect the function and structure of mycobacterial RNAP. Many differences exist between transcription machinery in well-characterized *E. coli* and pathogenic bacteria such as *M. tuberculosis*, therefore inferences cannot always be made from one to better understand the other.

Primarily, a better understanding of the effect of *rpoB* and *rpoC* mutations on holoenzyme structure, transcription initiation, elongation and termination is beneficial for our understanding of how these mutations influence transcriptional machinery. Functional characterization of mycobacterial RNAP harbouring mutations which are often found in clinical isolates would greatly benefit this area of research. Furthermore, insight is needed on how drug-resistance and compensatory mutations influence the affinity of RNAP core enzyme for various sigma factors in bacteria. Alternatively, mutations in RNAP could alter the affinity of the holoenzyme for different promoter sequences or transcriptional regulators. These factors can have a major impact on global gene expression in bacteria, as highlighted in this review, and studies which are able to provide insight into the underlying mechanisms which are involved will improve our understanding of drug-resistant bacteria. Additionally, the recently solved crystal structures of *M. tuberculosis* and *M. smegmatis* RNAP will provide a considerable advantage to this field as these structures can be used to predict the roles of various mutations with greater accuracy as well as study how mycobacterial RNAP interacts with compounds and other proteins (12, 89). Lastly, studying the genomes of clinical isolates of *M. tuberculosis* could provide much needed insight into whether other polymorphisms are acquired alongside resistance-conferring mutations in RNAP, as is the case with compensatory *rpoC* mutations.

2.7 References

1. Ramaswamy S. 1998. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tuber Lung Dis 79:3.
2. 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.
3. Morlock GP, Plikaytis BB, Crawford JT. 2000. Characterization of spontaneous, *in vitro*-selected, rifampin-resistant mutants of *Mycobacterium tuberculosis* strain H37Rv. Antimicrob Agents Chemother 44:3298–3301.
4. 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.
5. Koch A, Mizrahi V, Warner DF. 2014. The impact of drug resistance on *Mycobacterium tuberculosis* physiology: what can we learn from rifampicin? Emerg Microbes Infect 3:e17.
6. A Darst S. 2001. Bacterial RNA polymerase. Curr Opin Struct Biol 11:155–162.
7. Richardson JP. 1990. Rho-dependent transcription termination. Biochim Biophys Acta BBA - Gene Struct Expr 1048:127–138.
8. Wilson KS, von Hippel PH. 1995. Transcription termination at intrinsic terminators: the role of the RNA hairpin. Proc Natl Acad Sci U S A 92:8793–8797.
9. 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.
10. Zhang G, Darst SA. 1998. Structure of the *Escherichia coli* RNA polymerase α subunit amino-terminal domain. Science 281:262–266.
11. Ebright RH, Busby S. 1995. The *Escherichia coli* RNA polymerase α subunit: structure and function. Curr Opin Genet Dev 5:197–203.
12. Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S, Feng Y, Zhang Y, Mandal S, Jiang Y, Liu S, Gigliotti M, Talaue M, Connell N, Das K, Arnold E, Ebright RH. 2017. Structural basis of *Mycobacterium tuberculosis* transcription and transcription inhibition. Mol Cell 66:169–179.e8.
13. Rose AS, Bradley AR, Valasatava Y, Duarte JM, Prlić A, Rose PW. 2016. Web-based molecular graphics for large complexes, p. 185–186. Proceedings of the 21st International Conference on Web3D Technology. ACM, New York, NY, USA.
14. Rose AS, Hildebrand PW. 2015. NGL Viewer: a web application for molecular visualization. Nucleic Acids Res 43:W576–W579.

15. Ghosh P, Ishihama A, Chatterji D. 2001. *Escherichia coli* RNA polymerase subunit ω and its N-terminal domain bind full-length β′ to facilitate incorporation into the α2β subassembly. Eur J Biochem 268:4621–4627.
16. Santos-Beneit F, Barriuso-Iglesias M, Fernández-Martínez LT, Martínez-Castro M, Sola-Landa A, Rodríguez-García A, Martín JF. 2011. The RNA polymerase omega factor RpoZ is regulated by PhoP and has an important role in antibiotic biosynthesis and morphological differentiation in *Streptomyces coelicolor*. Appl Environ Microbiol 77:7586–7594.
17. Mathew R, Mukherjee R, Balachandar R, Chatterji D. 2006. Deletion of the *rpoZ* gene, encoding the ω subunit of RNA polymerase, results in pleiotropic surface-related phenotypes in *Mycobacterium smegmatis*. Microbiology 152:1741–1750.
18. Gunnelius L, Hakkila K, Kurkela J, Wada H, Tyystjärvi E, Tyystjärvi T. 2014. The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. Nucleic Acids Res 42:4606–4614.
19. Paget MS. 2015. Bacterial sigma factors and anti-sigma factors: structure, function and distribution. Biomolecules 5:1245–1265.
20. Burgess RR, Travers AA. 1969. Factor stimulating transcription by RNA polymerase. Nature 221:43–46.
21. Sachdeva P, Misra R, Tyagi AK, Singh Y. 2010. The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. FEBS J 277:605–626.
22. Paget MSB, Helmann JD. 2003. The sigma70 family of sigma factors. Genome Biol 4:203.
23. 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.
24. Xu M, Zhou YN, Goldstein BP, Jin DJ. 2005. Cross-resistance of *Escherichia coli* RNA polymerases conferring rifampin resistance to different antibiotics. J Bacteriol 187:2783–2792.
25. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. 2014. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd.
26. Caws M, Duy PM, Tho DQ, Lan NTN, Hoa DV, Farrar J. 2006. Mutations prevalent among rifampin- and isoniazid-resistant *Mycobacterium tuberculosis* isolates from a hospital in Vietnam. J Clin Microbiol 44:2333–2337.

27. Campbell PJ, Morlock GP, Sikes RD, Dalton TL, Metchock B, Starks AM, Hooks DP, Cowan LS, Plikaytis BB, Posey JE. 2011. Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 55:2032–2041.
28. Yoon J-H, Nam J-S, Kim K-J, Choi Y, Lee H, Cho S-N, Ro Y-T. 2012. Molecular characterization of drug-resistant and -susceptible *Mycobacterium tuberculosis* isolated from patients with tuberculosis in Korea. Diagn Microbiol Infect Dis 72:52–61.
29. Lavania M, Hena A, Reja H, Nigam A, Biswas NK, Singh I, Turankar RP, Gupta U, Kumar S, Rewaria L, Patra PKR, Sengupta U, Bhattacharya B. 2016. Mutation at codon 442 in the *rpoB* gene of *Mycobacterium leprae* does not confer resistance to rifampicin. Lepr Rev 87:93–100.
30. Siddiqi N, Pathak N, Banerjee S, Ahmed N, Katoch VM, Hasnain SE, Das R. 2004. *Mycobacterium tuberculosis* isolate with a distinct genomic identity overexpresses a tap-like efflux pump. Infection 32:109–111.
31. Kristich CJ, Little JL. 2012. Mutations in the β subunit of RNA polymerase alter intrinsic cephalosporin resistance in enterococci. Antimicrob Agents Chemother 56:2022–2027.
32. 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.
33. Molodtsov V, Scharf NT, Stefan MA, Garcia GA, Murakami KS. 2017. Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important *rpoB* mutations found in *Mycobacterium tuberculosis*. Mol Microbiol 103:1034–1045.
34. Singh A, Grover S, Sinha S, Das M, Somvanshi P, Grover A. 2017. Mechanistic principles behind molecular mechanism of rifampicin resistance in mutant RNA polymerase beta subunit of *Mycobacterium tuberculosis*. J Cell Biochem.
35. Musser JM. 1995. Antimicrobial agent resistance in mycobacteria: molecular genetic insights. Clin Microbiol Rev 8:496–514.
36. 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.
37. Brandis G, Pietsch F, Alemayehu R, Hughes D. 2015. Comprehensive phenotypic characterization of rifampicin resistance mutations in Salmonella provides insight into the evolution of resistance in *Mycobacterium tuberculosis*. J Antimicrob Chemother 70:680–685.

38. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. 2012. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44:106–110.
39. De Vos M, 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.
40. Qi Q, Preston GM, MacLean RC. 2014. Linking system-wide impacts of RNA polymerase mutations to the fitness cost of rifampin resistance in *Pseudomonas aeruginosa*. mBio 5:e01562-14.
41. Meftahi N, Namouchi A, Mhenni B, Brandis G, Hughes D, Mardassi H. 2016. Evidence for the critical role of a secondary site *rpoB* mutation in the compensatory evolution and successful transmission of an MDR tuberculosis outbreak strain. J Antimicrob Chemother 71:324–332.
42. Sonenshein AL, Alexander HB. 1979. Initiation of transcription *in vitro* is inhibited by lipiarmycin. J Mol Biol 127:55–72.
43. Leeds JA, Sachdeva M, Mullin S, Barnes SW, Ruzin A. 2014. *In vitro* selection, via serial passage, of *Clostridium difficile* mutants with reduced susceptibility to fidaxomicin or vancomycin. J Antimicrob Chemother 69:41–44.
44. Babakhani F, Gomez A, Robert N, Sears P. 2011. Killing kinetics of fidaxomicin and its major metabolite, OP-1118, against *Clostridium difficile*. J Med Microbiol 60:1213–1217.
45. Goldstein EJC, Citron DM, Sears P, Babakhani F, Sambol SP, Gerding DN. 2011. Comparative susceptibilities to fidaxomicin (OPT-80) of isolates collected at baseline, recurrence, and failure from patients in two phase III trials of fidaxomicin against *Clostridium difficile* infection. Antimicrob Agents Chemother 55:5194–5199.
46. Goldstein EJC, Babakhani F, Citron DM. 2012. Antimicrobial activities of fidaxomicin. Clin Infect Dis Off Publ Infect Dis Soc Am 55:S143–S148.
47. Gualtieri M, Villain-Guillot P, Latouche J, Leonetti J-P, Bastide L. 2006. Mutation in the *Bacillus subtilis* RNA polymerase β′ subunit confers resistance to lipiarmycin. Antimicrob Agents Chemother 50:401–402.

48. 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.
49. Gualtieri M, Tupin A, Brodolin K, Leonetti J-P. 2009. Frequency and characterisation of spontaneous lipiarmycin-resistant *Enterococcus faecalis* mutants selected *in vitro*. Int J Antimicrob Agents 34:605–606.
50. Matsuo M, Hishinuma T, Katayama Y, Cui L, Kapi M, Hiramatsu K. 2011. Mutation of RNA polymerase β subunit (*rpoB*) promotes hVISA-to-VISA phenotypic conversion of strain Mu3. Antimicrob Agents Chemother 55:4188–4195.
51. Matsuo M, Hishinuma T, Katayama Y, Hiramatsu K. 2015. A mutation of RNA polymerase β' subunit (RpoC) converts heterogeneously vancomycin-intermediate *Staphylococcus aureus* (hVISA) into "slow VISA." Antimicrob Agents Chemother 59:4215–4225.
52. Cashel M, Gallant J. 1969. Two compounds implicated in the function of the RC gene of *Escherichia coli*. Nature 221:838–841.
53. 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.
54. Ryals J, Little R, Bremer H. 1982. Control of rRNA and tRNA syntheses in *Escherichia coli* by guanosine tetraphosphate. J Bacteriol 151:1261–1268.
55. 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.
56. 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.
57. Ross W, Vrentas CE, Sanchez-Vazquez P, Gaal T, Gourse RL. 2013. The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. Mol Cell 50:420–429.
58. Zuo Y, Wang Y, Steitz TA. 2013. The mechanism of *E. coli* RNA polymerase regulation by ppGpp is suggested by the structure of their complex. Mol Cell 50:430–436.
59. Mechold U, Potrykus K, Murphy H, Murakami KS, Cashel M. 2013. Differential regulation by ppGpp versus pppGpp in *Escherichia coli*. Nucleic Acids Res 41:6175–6189.

60. Ross W, Sanchez-Vazquez P, Chen AY, Lee J-H, Burgos HL, Gourse RL. 2016. ppGpp binding to a site at the RNAP-DksA interface accounts for its dramatic effects on transcription initiation during the stringent response. Mol Cell 62:811–823.
61. Rutherford ST, Villers CL, Lee J-H, Ross W, Gourse RL. 2009. Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA. Genes Dev 23:236–248.
62. Lennon CW, Ross W, Martin-Tumasz S, Toulokhonov I, Vrentas CE, Rutherford ST, Lee J-H, Butcher SE, Gourse RL. 2012. Direct interactions between the coiled-coil tip of DksA and the trigger loop of RNA polymerase mediate transcriptional regulation. Genes Dev 26:2634–2646.
63. Dahl JL, Kraus CN, Boshoff HIM, Doan B, Foley K, Avarbock D, Kaplan G, Mizrahi V, Rubin H, Barry CE. 2003. The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of *Mycobacterium tuberculosis* in mice. Proc Natl Acad Sci 100:10026–10031.
64. Hauryliuk V, Atkinson GC, Murakami KS, Tenson T, Gerdes K. 2015. Recent functional insights into the role of (p)ppGpp in bacterial physiology. Nat Rev Microbiol 13:298–309.
65. 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.
66. 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.
67. Nene V, Glass RE. 1983. Relaxed mutants of *Escherichia coli* RNA polymerase. FEBS Lett 153:307–310.
68. Glass RE, Jones ST, Nene V, Nomura T, Fujita N, Ishihama A. 1986. Genetic studies on the β subunit of *Escherichia coli* RNA polymerase. Mol Gen Genet MGG 203:487–491.
69. Baracchini E, Glass R, Bremer H. 1988. Studies *in vivo* on *Escherichia coli* RNA polymerase mutants altered in the stringent response. Mol Gen Genet MGG 213:379–387.
70. Tedin K, Bremer H. 1992. Toxic effects of high levels of ppGpp in *Escherichia coli* are relieved by *rpoB* mutations. J Biol Chem 267:2337–2344.
71. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BØ. 2006. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet 38:1406–1412.

72. Conrad TM, Frazier M, Joyce AR, Cho B-K, Knight EM, Lewis NE, Landick R, Palsson BØ. 2010. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. Proc Natl Acad Sci 107:20500–20505.
73. 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.
74. Xu J, Tozawa Y, Lai C, Hayashi H, Ochi K. 2002. A rifampicin resistance mutation in the *rpoB* gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). Mol Genet Genomics MGG 268:179–189.
75. Hu H, Zhang Q, Ochi K. 2002. Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. J Bacteriol 184:3984–3991.
76. Lai C, Xu J, Tozawa Y, Okamoto-Hosoya Y, Yao X, Ochi K. 2002. Genetic and physiological characterization of *rpoB* mutations that activate antibiotic production in *Streptomyces lividans*. Microbiol Read Engl 148:3365–3373.
77. Hosaka T, Ohnishi-Kameyama M, Muramatsu H, Murakami K, Tsurumi Y, Kodani S, Yoshida M, Fujie A, Ochi K. 2009. Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. Nat Biotechnol 27:462–464.
78. Inaoka T, Takahashi K, Yada H, Yoshida M, Ochi K. 2004. RNA polymerase mutation activates the production of a dormant antibiotic 3,3'-neotrehalosadiamine via an autoinduction mechanism in *Bacillus subtilis*. J Biol Chem 279:3885–3892.
79. Utrilla J, O'Brien EJ, Chen K, McCloskey D, Cheung J, Wang H, Armenta-Medina D, Feist AM, Palsson BO. 2016. Global rebalancing of cellular resources by pleiotropic point mutations illustrates a multi-scale mechanism of adaptive evolution. Cell Syst 2:260–271.
80. Gao W, Cameron DR, Davies JK, Kostoulias X, Stepnell J, Tuck KL, Yeaman MR, Peleg AY, Stinear TP, Howden BP. 2013. The *rpoB* H481Y rifampicin resistance mutation and an active stringent response reduce virulence and increase resistance to innate immune responses in *Staphylococcus aureus*. J Infect Dis 207:929–939.
81. 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.

82. Lahiri N, Shah RR, Layre E, Young D, Ford C, Murray MB, Fortune SM, Moody DB. 2016. Rifampin resistance mutations are associated with broad chemical remodeling of *Mycobacterium tuberculosis*. J Biol Chem 291:14248–14256.
83. Vitali B, Turroni S, Serina S, Sosio M, Vannini L, Candela M, Guerzoni ME, Brigidi P. 2008. Molecular and phenotypic traits of *in-vitro*-selected mutants of Bifidobacterium resistant to rifaximin. Int J Antimicrob Agents 31:555–560.
84. Young R, Wang I-N, Roof WD. 2000. Phages will out: strategies of host cell lysis. Trends Microbiol 8:120–128.
85. Swapna G, Chakraborty A, Kumari V, Sen R, Nagaraja V. 2011. Mutations in β' subunit of *Escherichia coli* RNA polymerase perturb the activator polymerase functional interaction required for promoter clearance. Mol Microbiol 80:1169–1185.
86. Hochschild A, Dove SL. 1998. Protein–protein contacts that activate and repress prokaryotic transcription. Cell 92:597–600.
87. Busby S, Ebright RH. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell 79:743–746.
88. Georgopoulos CP. 1971. Bacterial mutants in which the gene N function of bacteriophage lambda is blocked have an altered RNA polymerase. Proc Natl Acad Sci U S A 68:2977–2981.
89. Hubin EA, Fay A, Xu C, Bean JM, Saecker RM, Glickman MS, Darst SA, Campbell EA. 2017. Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. eLife 6:e22520.

Chapter 3

Study 1

Exploring the potential of T7 bacteriophage protein Gp2 as a novel inhibitor of mycobacterial RNA polymerase

Du Plessis J[1,a], Cloete R[2,a], Burchell L[3], Sarkar P[3], Warren RM[1], Christoffels A[2], Wigneshweraraj S[3], Sampson SL[1]

[1]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research / SA MRC Centre for TB Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa.

[2]South African National Bioinformatics Institute (SANBI), SA Medical Research Council Bioinformatics Unit, University of the Western Cape, South Africa

[3]MRC Centre for Molecular Bacteriology and Infection, Faculty of Medicine, South Kensington Campus, Imperial College, United Kingdom

[a]Authors contributed equally to this study

The following chapter consists of a manuscript which has been accepted for publication in the journal *Tuberculosis*. My contribution to this chapter is as follows:

Literature search
Study design
Experimental work (*in silico* analysis done by RC)
Data interpretation
Writing and editing of manuscript
Addressing reviewers' comments

3.1 Abstract

Over the past six decades, there has been a decline in novel therapies to treat tuberculosis, while the causative agent of this disease has become increasingly resistant to current treatment regimens. Bacteriophages (phages) are able to kill bacterial cells and understanding this process could lead to novel insights for the treatment of mycobacterial infections. Phages inhibit bacterial gene transcription through phage-encoded proteins which bind to RNA polymerase (RNAP), thereby preventing bacterial transcription. Gp2, a T7 phage protein which binds to the beta prime (β') subunit of RNAP in *Escherichia coli*, has been well characterized in this regard. Here, we aimed to determine whether Gp2 is able to inhibit RNAP in *Mycobacterium tuberculosis* as this may provide new possibilities for inhibiting the growth of this deadly pathogen. Results from an electrophoretic mobility shift assay and *in vitro* transcription assay revealed that Gp2 binds to mycobacterial RNAP and inhibits transcription; however, to a much lesser degree than in *E. coli*. To further understand the molecular basis of these results, a series of *in silico* techniques were used to assess the interaction between mycobacterial RNAP and Gp2, providing valuable insight into the characteristics of this protein-protein interaction.

3.2 Introduction

Mycobacterium tuberculosis is the causative agent of tuberculosis (TB), an infectious disease which is one of the leading causes of morbidity and mortality in developing countries (1). The emergence of drug resistant TB has negatively affected control programs worldwide and the absence of a large repertoire of anti-TB drugs exacerbates this problem. A new approach to finding a sustainable solution for treating bacterial infections such as TB is to determine how bacteriophages (phages) are able to kill bacterial cells. One of the ways in which phages prevent bacterial gene transcription is through small phage-encoded proteins which bind to and inhibit RNA polymerase (RNAP), the sole enzyme responsible for transcription in bacteria. A well-studied example of this interaction is that of T7 phage protein, Gp2, which binds to the beta prime (β') subunit of RNAP in *Escherichia coli*, encoded by the *rpoC* gene, thereby preventing the productive engagement of RNAP with the promoter (2–8). Bacterial RNAP consists of 5 subunits (α2ββ'ω) which make up the core enzyme, and one of several sigma (σ) factors, which confer promoter specificity to the resulting holoenzyme. Gp2 binds to the region comprised of amino acids 1145-1198, referred to as the 'jaw domain' of the β' subunit, a region which contributes to the DNA binding channel of RNAP. More specifically, residues E1158 and E1188 have previously been found to be critical for binding of Gp2 to RNAP in *E. coli* (7). Functional and structural analysis of Gp2 has revealed that Gp2 uses a multipronged mechanism to inhibit the enzymatic activity of bacterial RNAP: by preventing functionally obligatory conformational changes in the σ factor; by preventing functionally obligatory changes in RNAP; by sterically preventing the binding of DNA in the DNA binding channel, and through antagonising the interaction between DNA and the catalytic site of RNAP.

Similarly, RNAP is the target of rifampicin (RIF), one of the most powerful drugs currently used to treat TB. RIF binds to the beta (β) subunit of RNAP, encoded by the *rpoB* gene, and inhibits transcription by preventing initialization of the nascent RNA transcript (9). However, many clinical isolates harbour mutations in *rpoB*, which alters the binding site of RIF and confers resistance to the drug. In the current study, we used both wet bench and *in silico* techniques to better understand the interaction between the β' subunit of mycobacterial RNAP and Gp2. It is envisaged that studying the binding of Gp2 and RNAP in *M. tuberculosis* could pave the way for novel approaches to RNAP inhibition as a means of TB treatment, especially for strains of *M. tuberculosis* which harbour *rpoB* mutations, as this may prove to be a solution for the current drug resistance epidemic.

3.3 Methods

3.3.1 Protein purification

N-terminal 6x His-tagged Gp2 was expressed and purified using the pSW33::Gp2^{WT} vector as previously described (7, 8). Dual expression vectors pACYCDuet *rpoA-sigA* and pETDuet *rpoB-rpoC* were used for the purpose of RNAP purification, both a kind gift from the Mukhopadhyay laboratory (10). Chemically competent *E. coli* BL21(DE3) was co-transformed with the expression vectors and selected on Luria Bertani agar containing ampicillin and chloramphenicol at final concentrations of 100 µg/mL and 34 µg/mL, respectively. Single colonies were picked and inoculated into 7 mL of 2xYT media and incubated overnight at 37°C. Stationary phase primary cultures were used to inoculate secondary cultures of 500 mL which were incubated at 37°C to an OD_{600} of 0.5. Thereafter, cultures were induced by the addition of IPTG (Calbiochem) to a final concentration of 0.5 mM and incubated at 16°C for 16 hours. Cells were harvested at 7000g for 20 minutes at 4°C and resuspended in lysis buffer [10 mM Tris-HCl pH 7.9, 50 mM KCl, 10 mM β-mercaptoethanol, 15 mM $MgCl_2$, 5 mM EDTA, 1x cOmplete protease inhibitor tablet (Sigma-Aldrich)] followed by sonication with a probe sonicator (5-second pulses, 2 x 10-minute cycles).

Sonicated cells were centrifuged at 23,500g for 30 minutes at 4°C, thereafter 10% polyethyleneimine (Sigma-Aldrich) was added to the supernatant and allowed to mix for an additional 30 minutes at 4°C. The solution was centrifuged at 23,500g for 20 minutes at 4°C and the pellet resuspended in wash buffer [10 mM Tris HCl pH 7.9, 6 mM β-mercaptoethanol, 5% glycerol, 0.1 mM EDTA, 500 mM NaCl]. Following centrifugation, the pellet was resuspended in extraction buffer [10 mM Tris HCl pH 7.9, 6 mM β-mercaptoethanol, 5% glycerol, 0.1 mM EDTA, 1 M NaCl] and allowed to incubate at 4^0C for 45 minutes with stirring. The solution was centrifuged at 23,500g for 20 minutes at 4°C and the supernatant precipitated with saturated ammonium sulphate (Sigma-Aldrich). After a 30-minute incubation step at 4°C with stirring, the solution was centrifuged and the resulting pellet resuspended in Buffer A [25 mM NaH_2PO_4 pH 7, 0.5 M NaCl, 5% glycerol].

In vivo reconstituted RNAP, incorporating an N-terminally 10x His-tagged α subunit, was purified using a HisTrap HP column (GE Healthcare Life Sciences) pre-equilibrated with Buffer A. After addition of the protein solution, the column was washed with 4 column volumes (CV) Buffer A, followed by 4 CV of 3% Buffer B [25 mM NaH_2PO_4 pH 7, 0.5 M NaCl, 5% glycerol, 1 M imidazole]. Protein was eluted with Buffer B over a linear gradient of 8 CV and

collected in 1 mL fractions. Based on the elution profile, fractions containing the proteins of interest were pooled and dialysed in dialysis buffer [10 mM Tris HCl pH 7.9, 5% glycerol, 50 mM NaCl, 1 mM DTT, 0.1 mM EDTA].

In a second round of purification, protein was injected onto a HiTrap Heparin HP column (GE Healthcare Life Sciences) pre-equilibrated with Buffer C [10 mM Tris HCl pH 7.9, 5% glycerol, 0.1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$, 50 mM NaCl]. The column was washed with 8 CV Buffer C, followed by elution with Buffer D [10 mM Tris HCl pH 7.9, 5% glycerol, 0.1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$, 1 M NaCl] over a linear gradient of 20 CV. Resulting elution fractions were run on a 10% SDS-PA gel and those containing the desired proteins which were free of any contaminating proteins were pooled and dialysed overnight in storage buffer [10 mM Tris HCl pH 7.9, 50 mM NaCl, 50% glycerol, 0.1 mM EDTA, 1 mM DTT] using a dialysis membrane (Spectra/Por) with molecular weight cut off of 6-8000 kDa.

3.3.2 Electrophoretic mobility shift assay

An electrophoretic mobility shift assay was done to determine if there is a direct interaction between Gp2 and *M. tuberculosis* RNAP. Reactions containing 100 nM of Gp2 labelled with [γ^{32}P]-ATP, were incubated with increasing concentrations of RNAP (0-800 nM) in reaction buffer [18mM Tris-HCl pH 8, 10mM NaCl, 10mM $MgCl_2$, 8mM β-mercaptoethanol] at 37°C. The reaction was stopped after 10 minutes and subjected to 4.5% native polyacrylamide gel electrophoresis (PAGE). The dried gel was exposed overnight and visualized using a Typhoon FLA 7000 (GE Healthcare Life Sciences). ImageJ v1.49 software was used to quantify the bands and compare the level of intensity between lanes.

3.3.3 Transcription assay

A 10 µL reaction containing 400 nM mycobacterial RNAP, transcription buffer [18 mM Tris-HCl pH 8, 10 mM NaCl, 10 mM $MgCl_2$, 8 mM β-mercaptoethanol] and 20 U of RNasin was prepared to assess the inhibitory activity of Gp2 on transcription. The mycobacterial promoter region of the *rrs* gene, which was amplified using *M. tuberculosis* DNA and primers 5'-TCGTGGAGAACCTGGTGAGTCTCGG-3' and 5'-ACGCCGCCAGCGTTCGTC-3', was used as a template for the purpose of *in vitro* transcription. The concentration of Gp2 in the reactions ranged from 0 to 3200 nM, correlating to a 1:0 to 1:8 fold titration of RNAP:Gp2. Upon addition of rNTPs to a final concentration of 0.01 mM and 0.005 mM [α^{32}P]-UTP, the reaction was incubated for 30 minutes at 37°C to allow transcription to take place. Transcription

was terminated with the addition of stop buffer [80% formamide, 10 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol]. Thereafter, reactions were heated at 95°C for 5 minutes, centrifuged at 13,000g for 1 minute and resolved by urea-PAGE. The wet gel was exposed for 3 hours and phosphorimaging was done using the Typhoon FLA 7000 (GE Healthcare Life Sciences). As before, ImageJ v1.49 software was used to quantify the intensity of the bands.

The following controls were included in the abovementioned assay. Identical reactions were prepared using 100 nM of *E. coli* RNAP and the T7A1 promoter, to demonstrate inhibition of transcription at a ratio of 1:2, as previously described (7, 11). Furthermore, reactions with Gp2 harbouring an R56E mutation were included to demonstrate that the inhibition of transcription is Gp2-specific, as the mutated peptide is unable to bind to *E. coli* RNAP (8).

3.3.4 Homology modelling

MODELLER was used to predict the holoenzyme structure of RNAP in *M. tuberculosis* by including all the subunit sequences (RpoA, RpoB, RpoC, RpoZ and SigA) within the alignment using the backslash option (12). A total of 50 models were constructed and the lowest DOPE score model (LDSM) was used for comparison to the holoenzyme structure of *E. coli* RNAP in complex with Gp2 (PDB ID: 4LLG). The Gp2 peptide was extracted from the *E. coli* template using the align function of the PyMOL Molecular Graphics System to obtain a complete holoenzyme structure for *M. tuberculosis* RNAP. We also predicted the three-dimensional (3D) structure for *M. tuberculosis* RNAP (RpoC-SigA-Gp2 complex) using MODELLER. The amino acid sequence for the β' subunit of *M. tuberculosis* RNAP (RpoC) was retrieved from Uniprot and used as query to search for homologous protein structures in the Protein Databank (PDB) using the profile.build module from MODELLER. The crystal structure for *Thermus thermophilus* RNAP holoenzyme in complex with streptolydigin (PDB ID: 2A6H) and *E. coli* RNAP in complex with Gp2 (PDB ID: 4LLG) were used as homologous template structures (13). A multiple sequence to structure alignment (MSSA) was constructed between amino acid sequences for templates *T. thermophilus* (RpoC) and *E. coli* (RpoC and RpoD) using salign and align2d routines from MODELLER. RpoD (σ^{70}) is the housekeeping σ factor for RNAP in *E. coli* whereas SigA (σ^{A}) is the housekeeping σ factor in *M. tuberculosis*. The alignment was manually edited to include sequences for Gp2 and *M. tuberculosis* SigA using the backslash option. The final optimised alignment was used as input to build 50 protein models using model_mult module from MODELLER. MODELLER builds 3D structures for proteins based

on the satisfaction of spatial restraints derived from the alignment between the target and templates. Based on the good agreement between the models of the holoenzyme and the subunits, subsequent steps only considered RpoC, housekeeping σ factors (RpoD and SigA) and Gp2 for *E. coli* and *M. tuberculosis*, as these units are crucial for the binding of Gp2 to RNAP in *E. coli*. Recently, the closely related crystal structure of RNAP for *M. smegmatis* was solved experimentally, therefore this template (PDB ID: 5TW1) was used to predict the structure of σ^A and the β' subunit of *M. tuberculosis* RNAP using the SWISS-MODEL webserver (14, 15).

3.3.5 Model evaluation

The model with the lowest discrete optimised protein energy (DOPE) score was selected for further analysis as it is considered to be closest to its native form. The quality of the LDSM was assessed using the GA341 composite score. Stereochemical analysis was done using PROCHECK to determine if residues of the LDSM were located in allowed regions of the Ramachandran plot (16). The ProSA II online web-tool was used to assess the reliability of regions within the modelled target and to calculate a normalised Z-score to determine the total energy deviation of the structure using random conformations (17). The normalised Z-score was used for comparison to the template structures and to determine whether the protein falls within the range of experimentally solved proteins of similar size. The root mean square deviations (RMSD) between the LDSM and the two templates were assessed by aligning all atoms to one another calculated in PyMOL.

3.3.6 Molecular dynamic simulations

All the molecular dynamic simulations were performed using GROMACS 4.6.5 with the AMBER SB99 force field (18, 19). Five systems in total were prepared; two wildtype (WT) systems were prepared for simulation of the LDSM of *M. tuberculosis* (RpoC-SigA-Gp2) and apo RpoC-SigA, two for *E. coli* template (RpoC-RpoD-Gp2) and apo (RpoC-RpoD) and one mutant system of *M. tuberculosis* RpoC-SigA-Gp2. All the systems were immersed in a 16 or 18 nm cubic box of TIP3P water molecules placed between the protein and the edges of the box which was sufficient for our systems. The WT and mutant *M. tuberculosis* Rpoc-SigA-Gp2 and the apo RpoC-SigA were neutralized by replacing 54 and 49 sodium ions with water molecules respectively, while the *E. coli* RpoC-RpoD-Gp2 and apo RpoC-RpoD systems were neutralized by replacing 16 and 11 sodium ions with water molecules, respectively. The five systems were initially energy-minimized by steepest descents for 50,000 steps to equilibrate the systems.

Thereafter each system was subjected to equilibration steps which included canonical ensemble (NVT) for 100ps and position-restraint dynamics simulation (NPT) for 500ps at 300K. Finally, the five systems were subjected to full molecular dynamic simulation of 50ns each and 1 bar pressure with a coupling time constant of 2ps with coordinates of the trajectory being saved every 20ps. The particle mesh Ewald method was used to treat long range Coulombic interactions and the V-rescale temperature coupling method was implemented (20). The molecular dynamic simulations were repeated using a random seed number without changing any of the parameters to validate the reproducibility of the results, verifying that all observables, including the trajectory, converge to reach their equilibrium values.

3.3.7 Principal component analysis and free energy landscape

For principal component analysis (PCA), the translational and rotational movements were removed from the systems using g_covar from GROMACS to construct a covariance matrix. Subsequently, eigenvectors and eigenvalues were calculated by diagonalising the matrix. The eigenvectors which correspond to the largest eigenvalues are called principal components, as they represent the largest-amplitude collective motions. The original trajectory was filtered and the section along the most important eigenvectors, namely vector 1 and 2, was projected out using g_anaeig from GROMACS utilities. Furthermore, the free energy landscape (FEL) was calculated using the GROMACS utility g_sham along the first two eigenvectors.

3.3.8 Analysis of molecular dynamics trajectory

The total and potential energy terms were calculated using Gromacs tool g_energy and the backbone atoms RMSD and C-alpha atoms root mean square fluctuation (RMSF) values were calculated using Gromacs utilities g_rms over the whole trajectory and g_rmsf over the last 40ns of the trajectory. RMSF is a measure of deviation between the position of a particle and some reference position averaged over time. The solvent accessible surface area (SASA) was calculated for the protein using g_sas. For visual inspection of the protein dynamics, short simulation movies were generated using Visual Molecular Dynamics (VMD) (21) of the five systems (files provided in supplementary CD).

3.4 Results

3.4.1 The effect of Gp2 on mycobacterial RNA polymerase

To determine whether Gp2 binds to mycobacterial RNA polymerase (RNAP), radiolabelled Gp2 was incubated with increasing concentrations of purified mycobacterial RNAP in an electrophoretic mobility shift assay (Figure 3.1A). This revealed that a Gp2:RNAP ratio of up to 1:8 did not allow for complete binding of Gp2, whereas the control reaction displays complete binding of Gp2 to *E. coli* RNAP at a ratio of 1:2. Furthermore, *in vitro* transcription assays were done to test whether Gp2 is able to inhibit the functioning of mycobacterial RNAP (Figure 3.1B). Transcription from the mycobacterial *rrs* promoter proved successful, with increasing levels of inhibition observed at higher concentrations of Gp2. However, when compared to *E. coli* reactions (Figure 3.1C), it is evident that the same level of inhibition of mycobacterial RNAP is not obtained. As would be predicted from these results, overexpression of Gp2 in *M. smegmatis* did not have a long-term impact on growth in a whole cell assay (data not shown). Together, these results indicate that Gp2 does not bind to mycobacterial RNAP with the same degree of affinity as in *E. coli*.

3.4.2 Model building

To further investigate the mechanistic basis for the weak affinity of Gp2 for mycobacterial RNAP, *in silico* approaches were utilised to understand structural features important for Gp2 binding. To identify the most suitable template for constructing a model of the *M. tuberculosis* RNAP holoenzyme, we assessed sequence identity between RpoC of *M. tuberculosis*, *E. coli* and *T. thermophilus*. *M. tuberculosis* showed the highest sequence identity to *M. smegmatis*, with 92%, for both RpoA and RpoB, while RpoC, RpoZ and SigA subunits showed 91%, 82% and 96%, respectively. When *M. tuberculosis* was compared to *T. thermophilus*, there was 53%, 51%, 41%, 18% and 51% identity for the RpoA, RpoB, RpoC, RpoZ and SigA subunits, respectively. Compared to *E. coli*, there was 57%, 56%, 48%, 53% and 34% identity for the RpoA, RpoB, RpoC, RpoZ and SigA subunits, respectively.

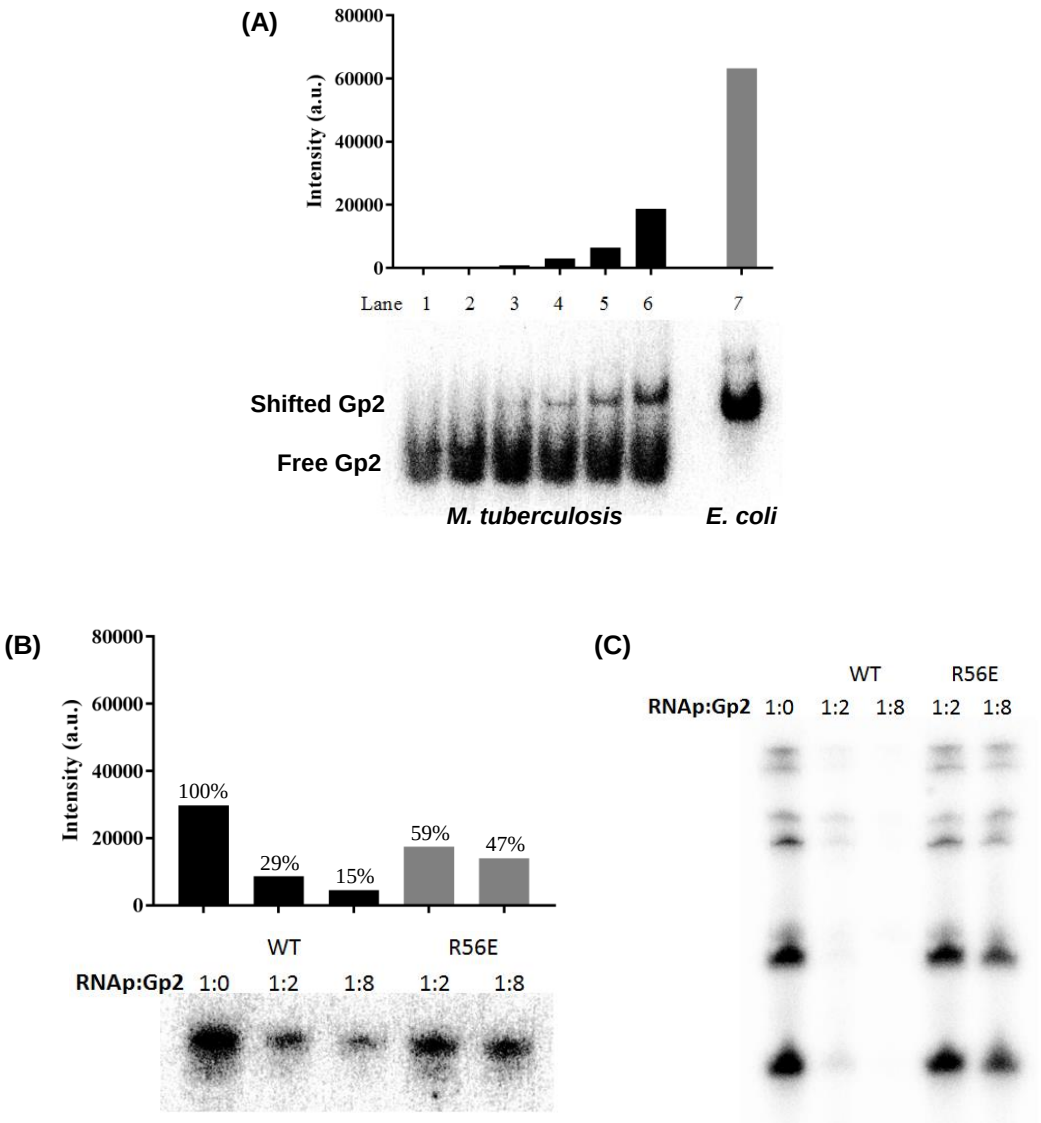


Figure 3.1 Assessment of Gp2 binding and RNAP inhibition.

(A) Electrophoretic mobility shift assay using radiolabelled Gp2 (100 nM) and increasing concentrations of mycobacterial RNA polymerase (Lane 1 = 0 nM; Lane 2 = 50 nM; Lane 3 = 100 nM; Lane 4 = 200 nM; Lane 5 = 400 nM; Lane 6 = 800 nM; Lane 7 illustrates the same reaction with 200 nM *E. coli* RNA polymerase). (B) *In vitro* transcription with mycobacterial RNA polymerase from the *rrs* promoter. RNAP:Gp2 ratios specified above each lane. Reactions in lanes 4 and 5 were done using Gp2 harbouring an R56E substitution, which has been shown to prevent binding to *E. coli* RNA polymerase. (C) *In vitro* transcription with *E. coli* RNA polymerase from the T7A1 promoter. Lanes correspond to those in (B).

To further assess the degree of conservation of the RpoC subunit in *M. tuberculosis*, *T. thermophilus* and *E. coli*, a multiple sequence to structure alignment (MSSA) was performed. This analysis indicated that the binding site residues of Gp2 in *E. coli* (E1158 and E1188) are not conserved in RpoC of *M. tuberculosis*. Figure 3.2A shows the multiple sequence alignment of RpoC in *M. tuberculosis*, *T. thermophilus*, *E. coli*, and closely related *M. smegmatis* which was generated using ClustalW. The multiple sequence alignment between *M. tuberculosis* SigA and three homologous templates (*T. thermophilus*, *E. coli* and *M. smegmatis*) is shown in Figure 3.2B. The multiple sequence alignments illustrate that *M. tuberculosis* RpoC and SigA subunits are more closely related to *M. smegmatis* than *T. thermophilus* and *E. coli*.

A total of 50 protein models were constructed for both the *M. tuberculosis* RNAP holoenzyme and the RpoC-SigA-Gp2 complex using the two homologous templates (PDB ID: 4LLG (*E. coli*) and 2A6H (*T. thermophilus*)). The crystal structure for the *E. coli* holoenzyme and the lowest DOPE score models (LDSM) for *M. tuberculosis* holoenzyme and the RpoC-SigA-Gp2 complex have been visually represented using PyMOL (Figures 3.3A-C). The secondary structural arrangement of LDSM for the *M. tuberculosis* RpoC-SigA-Gp2 complex includes 45 alpha helices, 29 beta strands and 73 turns, while *E. coli* RpoC-RpoD-Gp2 consists of 38 alpha helices, 49 beta sheets and 83 coil regions (Figure 3.3A and 3.3B). This indicates key structural differences in the number of beta sheets which contribute to the stronger affinity of Gp2 to *E. coli* RNAP. To compare structural similarity, the models consisting of subunits RpoC-SigA-Gp2 built with SWISS-MODEL using template 5TW1 (*M. smegmatis*) was superimposed on the model (subunit RpoC-SigA-Gp2) built using templates 4LLG (*E. coli*) and 2A6H (*T. thermophilus*). This revealed a 3.233Å difference between the carbon alpha residues, suggesting that there are structural differences within some domains although these domains are not crucial for Gp2 binding between the two models (Figure 3.4). For future studies, the *M. tuberculosis* model built using template 5TW1 from *M. smegmatis* might be beneficial due to its high sequence identity and structural overlap. For the purpose of this study, further assessment was done on the models generated using both *E. coli* and *T. thermophilus* templates and the one generated using *M. smegmatis* was selected for comparison.

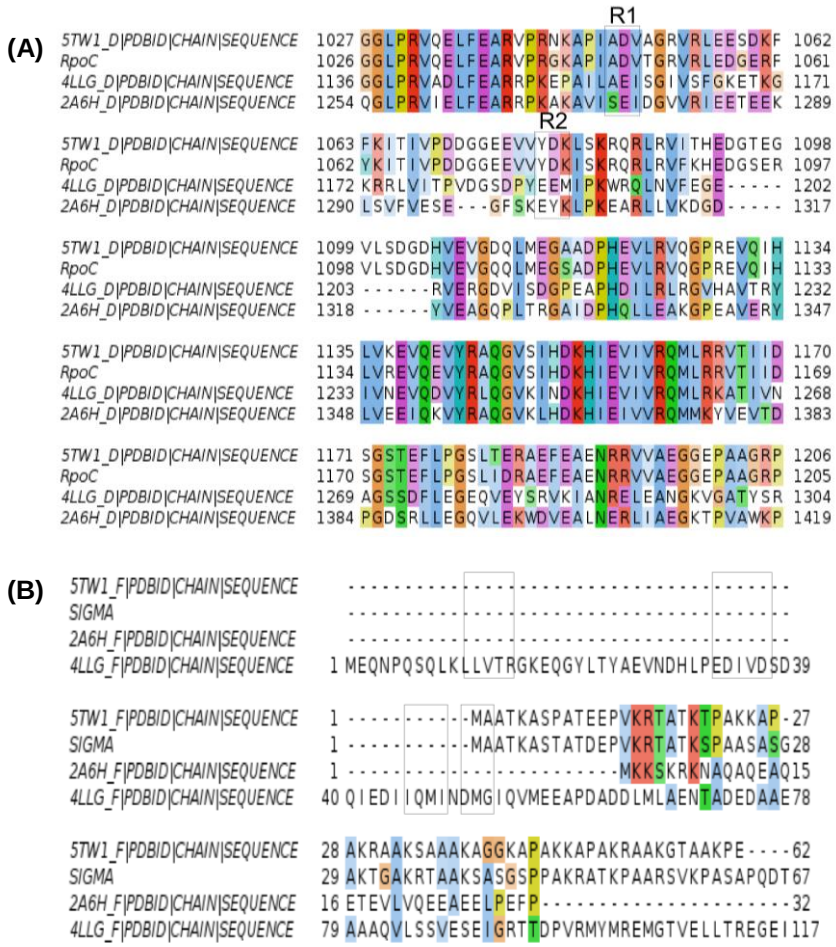


Figure 3.2 Multiple sequence alignments of *M. tuberculosis* RpoC and SigA with three homologous templates.

(A) Snapshot of the multiple sequence alignment between *M. tuberculosis* RpoC and template *M. smegmatis* (5TW1_D), *E. coli* (4LLG chain_D) and *T. thermophilus* (2A6H_D). The boxed regions 1 and 2 show the non-conserved E1158 and E1188 residues between templates 4LLG and *M. tuberculosis* RpoC. R1 and R2 incorporate the conserved residues E1158 and E1188. (B) Snapshot of the multiple sequence alignment between *M. tuberculosis* SigA (SIGMA) and template *M. smegmatis* (5TW1_F), *E. coli* (4LLG_F) and *T. thermophilus* (2A6H_F). The boxed regions show the known 4LLG Sig70 Gp2-interacting residues which are absent in *M. tuberculosis* SigA indicated by dashes except for the last boxed region.

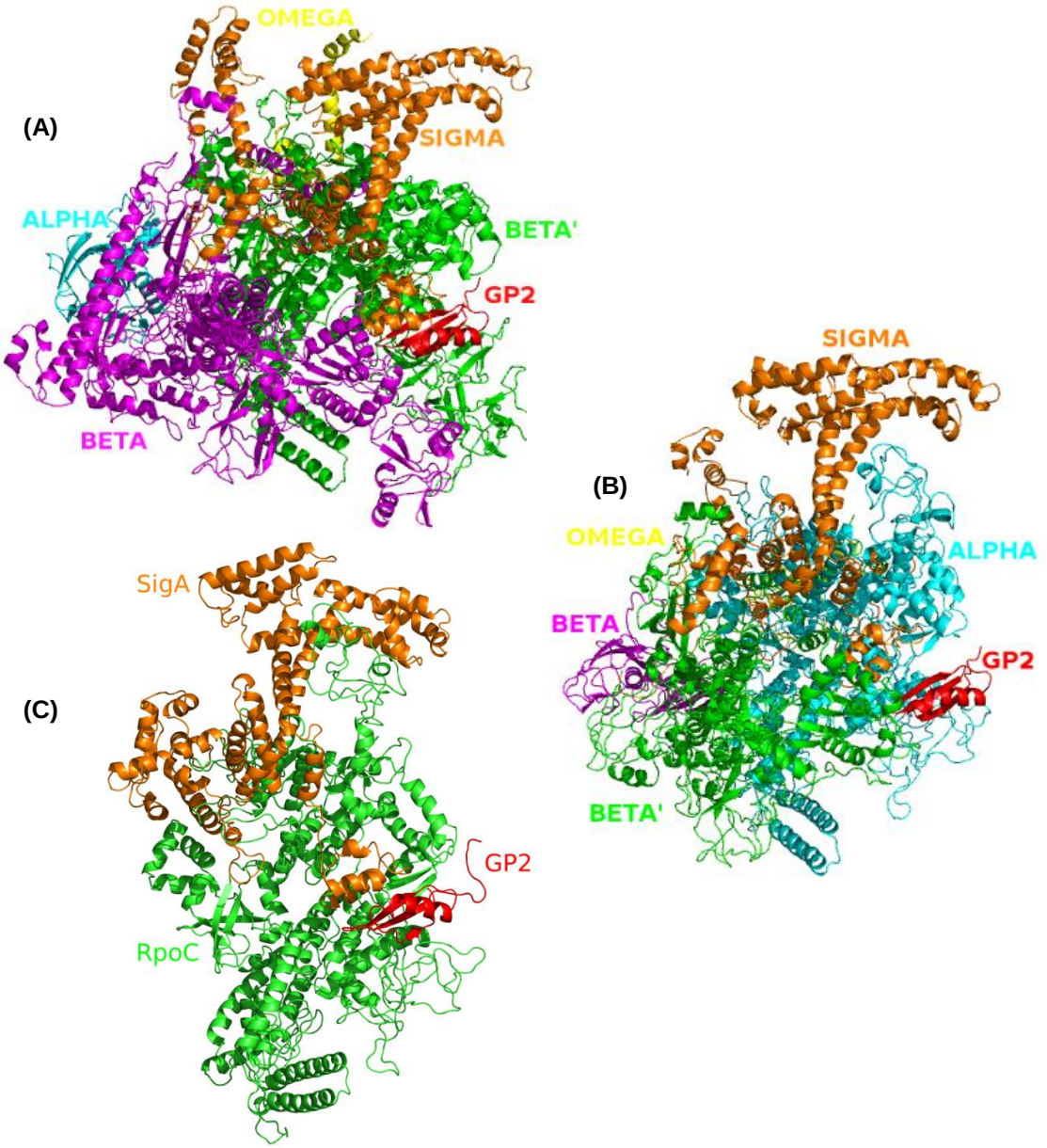


Figure 3.3 Crystal structure of the *E. coli* 4LLG RNAP holoenzyme and the *M. tuberculosis* lowest DOPE score models for the RNAP holoenzyme and the RpoC-SigA-Gp2 complex.

(A) Crystal structure of the *E. coli* holoenzyme in complex with Gp2. (B) The lowest DOPE score model for the holoenzyme of *M. tuberculosis* RNAP. (C) Lowest DOPE score model (multiple templates *E. coli* 4LLG and *T. thermophilus* 2A6H) of *M. tuberculosis* RpoC-SigA-Gp2 complex.

3.4.3 Qualitative analysis of LDSM

To assess the quality of the LDSM structure compared to other protein models, MODELLER was used to calculate the discrete optimized protein energy (DOPE) for each protein model generated using *E. coli* and *T. thermophilus* templates. The lowest DOPE score was equal to -118799.52 for protein model 12 (Figure 3.3C) which corresponds to a particular native state of the protein. Furthermore, the DOPE score energy profile of the *M. tuberculosis* RpoC-SigA-Gp2 LDSM (model 12) was plotted relative to the homologous templates and it was found to be comparable to that of the homologous templates in *T. thermophilus* and *E. coli*, with no distinct regions of extremely high energy (Figure S1), suggesting that no problematic loop regions were encountered during modelling. This indicates that the generated model is close to its native conformation. The LDSM for *M. tuberculosis* RpoC-SigA-Gp2 satisfies stereo-chemical restraints according to PROCHECK. The Ramachandran plot indicates that 88.4% of residues are in most favoured regions and 0.8% are located in disallowed regions, suggesting that most residues satisfied phi and psi dihedral angle distributions within the protein structure. For the LDSM of *M. tuberculosis* RpoC-SigA-Gp2, the GA341 score was equal to 1 and the ProSA normalised Z-score was -9.86, comparable to experimental structures of similar size and shape. This suggests that the correct fold has been assigned to the predicted 3D model of *M. tuberculosis* RpoC-SigA-Gp2, while the structure might be less energetically stable than its homologous template structures. The RMSD values for the superimposition of the LDSM onto the two templates was 1.358Å and 4.166Å for *T. thermophilus* and *E. coli*, respectively, which suggests that the LDSM is structurally more similar to *T. thermophilus*. The global quality estimation scores of the 3D structures predicted for *M. tuberculosis* RpoC and SigA using template 5TW1 on the SWISS-MODEL workspace were 0.59 and 0.95, respectively. As these values are larger than 0 and close to 1, this suggests that each subunit was reliably built using the alignments provided. Superimposition of the model built using SWISS-MODEL (chains D and F) onto *M. smegmatis* chains, indicated RMSD values of 0.183Å and 0.068Å, respectively. This suggests that there is very little deviation in carbon backbone atoms and structural similarity between RNAP in *M. tuberculosis* and *M. smegmatis*. Further, this validates the use of the selected model for further analysis of Gp2 binding.

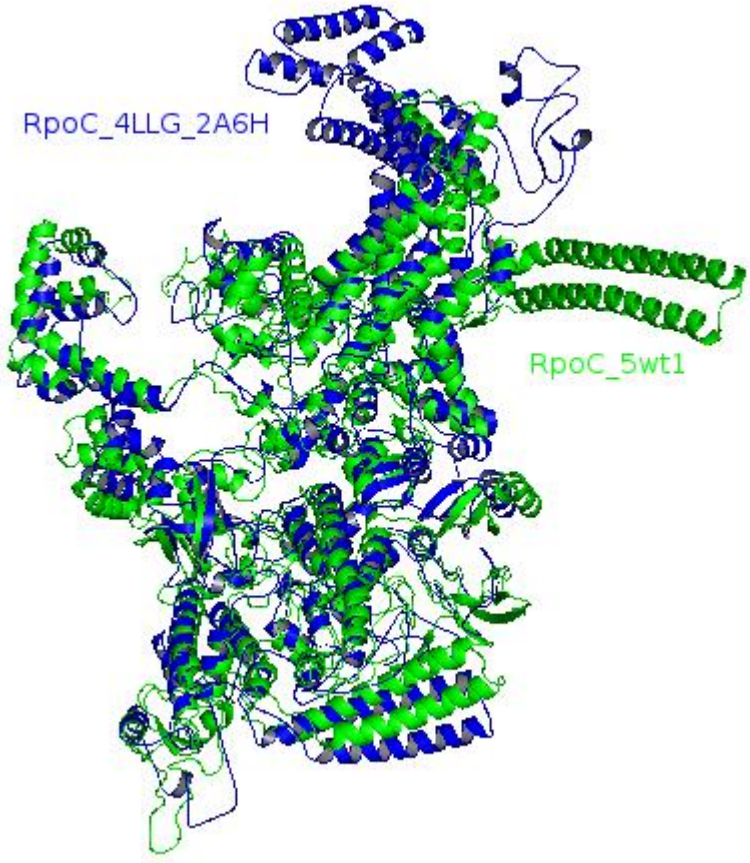


Figure 3.4 Juxtaposition of the LDSM in blue (templates *E. coli* 4LLG and *T. thermophilus* 2A6H) on top of the model built using template 5TW1 in green (RMSD = 3.233A).

3.4.4 Molecular dynamic simulation analysis

In order to identify large scale structural changes due to Gp2 binding, we simulated the movement of atoms over time for both the *M. tuberculosis* and *E. coli* RNAP systems. Analysis of the average RMSD values for the backbone atoms in the wildtype (WT) *M. tuberculosis* RpoC-SigA-Gp2 revealed the largest mean values of 1.32 ±0.22 nm, compared to values of 1.00 ±0.13 nm and 1.24 ±0.19 nm, respectively for the mutant *M. tuberculosis* RpoC-SigA and apo RpoC-SigA (Figure S2A), suggesting that the mutant is more stable. Also, the *E. coli* RpoC-RpoD-Gp2 showed higher RMSD values, 0.82 ±0.14 nm, compared to apo *E. coli* RpoC-RpoD of 0.64 ±0.09 nm (Figure S2A). The root mean square fluctuation (RMSF) for the WT *M. tuberculosis* RpoC-SigA-Gp2 model averaged approximately 0.36 ±0.24 nm during the 50ns simulation period, while the mutant *M. tuberculosis* Rpo-SigA-Gp2 showed the least amount of flexibility (0.309±0.20 nm) in comparison to the apo *M. tuberculosis* RpoC-SigA of 0.326±0.23 nm. Again, the apo *E. coli* RpoC-RpoD showed the lowest RMSF value

(0.279±0.14 nm) compared to the RpoC-RpoD-Gp2 of 0.308±0.23 nm (Figure S2B). To assess the magnitude of binding-induced conformational changes, the solvent accessible surface area (SASA) was determined. The WT *M. tuberculosis* RpoC-SigA-Gp2 displayed a slightly higher total SASA values (1235±27.09 nm^2) compared to the mutant *M. tuberculosis* RpoC-SigA-Gp2 value (1232±25.84 nm^2), while the apo *M. tuberculosis* RpoC-SigA displayed the lowest total SASA values of 1194±24.41 nm^2 (Figure S2C). The *E. coli* RpoC-RpoD-Gp2 had higher SASA values (1115±12.25 nm^2) compared to the apo *E. coli* RpoC-RpoD which showed considerably low SASA values (1094±18.05 nm^2). This suggests high stability within the hydrophobic core of the apo *E. coli* RNAP structure (Figure S2C). Secondary structure analysis of the jaw domain of the β' subunit (region spanning residues 1048 to 1078) indicates a gain of a 3-helix and a bend within the mutant *M. tuberculosis* RpoC-SigA-Gp2, while the WT *M. tuberculosis* RpoC-SigA-Gp2 loses a 3-helix, gains a turn and maintains a coil throughout the 50ns simulation supporting the stability of the jaw domain when mutated residues E1048 and E1078 are present. Analysis of the *E. coli* RpoC-RpoD-Gp2 indicated a loss of a coil arrangement for a bend within the region spanning 1158 to 1188, while the apo *E. coli* RpoC-RpoD remained stable with no changes within the same region (Figures S3A–E, see simulation movies SS_4LLG, SS_4LLG_noGp2, SS_Mtb_RpoC_WT, SS_Mtb_RpoC_MUT and SS_Mtb_noGp2 provided in supplementary CD). The observed interaction between Gp2 and *M. tuberculosis* RNAP is different to that with *E. coli* RNAP due to the lack of critical conserved residues. This is further supported by interaction analysis between WT *M. tuberculosis* RpoC (5TW1) and Gp2, which lacks polar contacts between aspartate residues (1048 and 1078), while mutated *M. tuberculosis* RpoC shows interaction for both glutamate residues (1048 and 1078) (Figures S4A and S4B). The mutated *M. tuberculosis* RpoC showed a higher number of interactions compared to WT *M. tuberculosis* RpoC, suggesting a stronger interaction with Gp2, which might result in a more stabilized structure. Furthermore, we propose that the four interactions (R282, R1041, K1043, E1120) found between residues of the *M. tuberculosis* WT complex and Gp2 (Figure S4) may account for the weak interaction observed in wet bench results.

Principal component analysis (PCA) was carried out to describe predominant movement of the protein structures for each of the two systems as a measure of protein-peptide stability. PCA is a statistical technique which reduces the complexity of a dataset in order to extract biologically relevant movements of protein domains from irrelevant localised motions of atoms. For each of the model structures, the first two principal components explained more

than 80% of the motion in the system. The free energy landscape (FEL) shows the local energy minima conformations or folded states for each system (Figures S5A-E). The FEL represents all possible conformations a protein can adopt during a simulation and is typically reported as Gibbs free energy. The dark blue dots indicate the local energy minima wells of the protein in the folding landscape. The *E. coli* RpoC-RpoD-Gp2 showed approximately 14 local energy minima conformations despite having higher RMSD, RMSF and SASA values compared to the more stable *E. coli* RpoC-SigA without Gp2, which displayed only one local energy minima conformation (Figures S5A and B). The WT model structure of the *M. tuberculosis* RpoC-SigA-Gp2 complex displayed fewer collective movements, with three local minima conformations in the essential subspace compared to the five states observed for mutant *M. tuberculosis* RpoC-SigA-Gp2, while the WT *M. tuberculosis* RpoC-SigA without Gp2 displayed only two local energy minima (Figures S5C – E). Interestingly, the opposite is seen in *E. coli* when Gp2 is bound to *E. coli* RNAP, where it becomes less stable and gains more flexibility, thereby preventing the enzyme from performing its function. The concerted movement for the mutant *M. tuberculosis* RpoC-SigA-Gp2 is further supported by the decrease in the covariance matrix value after diagonalisation from 1037 nm^2 for WT *M. tuberculosis* RpoC-SigA-Gp2 to 761 nm^2 for mutant *M. tuberculosis* RpoC-SigA-Gp2. Coincidently, the *M. tuberculosis* RpoC-SigA without Gp2 displayed less concerted movements with a covariance matrix value of 862 nm^2 compared to the mutant *M. tuberculosis* RpoC-SigA-Gp2. This indicates that mutant *M. tuberculosis* RpoC-SigA-Gp2 is more thermodynamically stable than the WT and apo structures and that Gp2 might have a stronger interaction with the mutant structure resulting in stronger inhibition of RNAP in *M. tuberculosis*. The protein instability observed in the WT *M. tuberculosis* RpoC-SigA-Gp2 could be due to the difference in topology of the jaw domain in the β' subunit, specifically the lack of the two critical glutamate residues (E1048 and E1078), which introduces large differences in movement along the principal components of WT *M. tuberculosis* RpoC-SigA-Gp2, leading to a weaker interaction with Gp2. Analysis of the energy and average temperature terms for the five systems indicated that each system reached convergence fluctuating around stable average values (Figures S6A-B, S7A-B, S8A-B, S9A-B and S10A-B).

3.5 Discussion

Bacteriophage T7 Gp2 is a 7 kDa peptide which binds to the β' subunit of RNA polymerase (RNAP) in *E. coli* and inhibits the transcription of DNA, thereby preventing bacterial growth (3). Wigneshweraraj and colleagues have identified compounds which mimic the inhibitory effect of Gp2 in *E. coli*, a discovery which could lead to the generation of novel antibacterial drugs. In the current study, we aimed to determine whether Gp2 is able to inhibit RNAP in *M. tuberculosis* as this may provide new possibilities for inhibiting the growth of this deadly pathogen.

Results from the electrophoretic mobility shift assay demonstrated that Gp2 does indeed bind to mycobacterial RNAP; however, a much higher concentration of Gp2 was required in order to correlate with results seen in the reaction with *E. coli* RNAP. Furthermore, *in vitro* transcription assays confirmed that Gp2 was able to inhibit transcription of mycobacterial RNAP, albeit at a much lesser degree than with *E. coli* RNAP. These findings hint at a reduced interaction between Gp2 and mycobacterial RNAP, however a series of *in silico* techniques were used in order to fully understand the molecular basis of these results. In summary, initial modelling was carried out using templates from *E. coli* and *T. thermophilus* as described below. Subsequently, these findings were independently verified using a model generated from *M. smegmatis* when this structure became available. Since submitting the current manuscript for publication, however, the crystal structure for *M. tuberculosis* transcription initiation complex was resolved by Lin and colleagues (PDBID: 5UH5) (22). Upon comparing this structure to the predicted holo structure generated using the *M. smegmatis* template, an RMSD value of 0.881Å was found between the structures (data not shown), suggesting a good overall agreement. We will consider the use of the *M. tuberculosis* crystal structure in future studies.

RNAP holoenzyme from *M. tuberculosis* as well as σ^A and the β' subunit in complex with Gp2 were modelled using two homologous templates, *T. thermophilus* and *E. coli*. The models were then used to compare the *de facto* interaction of Gp2 and RNAP in *E. coli*, to the modelled interaction in *M. tuberculosis*. Results from the multiple sequence alignment indicate that the binding site residues of Gp2 in *E. coli* RNAP, namely E1158 and E1188, are not conserved in the β' subunit of *M. tuberculosis*. These residues are crucial for structural stability of Gp2 within the binding cleft of RNAP in order to elicit its inhibitory activities on *E. coli* RNAP. This was further supported by interaction analysis after remodelling the WT *M. tuberculosis* RpoC-SigA-Gp2 complex using the closely related homolog *M. smegmatis*, which showed no interaction

between Gp2 and the aspartate residues at positions 1048 and 1078 (corresponding to E1158 and E1188 in *E. coli*). Furthermore, the WT *M. tuberculosis* RpoC-SigA-Gp2 displayed less stability within the hydrophobic core of the complex and the jaw domain does not possess a 3-helical topological conformation, but a turn at position 1048. These findings provide an explanation for results obtained in the wet-lab experiments, suggesting that Gp2 would not inhibit the growth of *M. tuberculosis* as it does not interact with mycobacterial RNAP in the same manner as in *E. coli*. However, despite the apparent low affinity of Gp2 binding to the four residues (R282, R1041, K1043, E1120) of *M. tuberculosis* RNAP, this molecule, together with models generated and tested here, could be used as framework for the *in silico* design of peptides that have a stronger affinity for RNAP in *M. tuberculosis*. Furthermore, it is possible that inhibitory peptides similar to Gp2 exist in mycobacteriophages, which would provide promising candidates for novel anti-TB drugs. *In silico* modelling approaches such as those described here may aid the identification of such peptides. Another strategy could entail introduction of mutations into Gp2 and performing docking studies to identify high affinity binders which could then be pursued within the laboratory.

3.6 Conclusion

In the present study, *in silico* investigation of the interaction between an inhibitory peptide, Gp2, and RNAP in *M. tuberculosis* gave an accurate account of the structural features important for Gp2 binding. Findings from molecular modelling and simulations aligned seamlessly with results obtained from wet-lab experiments, providing valuable insight into the mechanism of Gp2 binding to *M. tuberculosis* RNAP, an otherwise challenging task when relying on traditional molecular techniques. Furthermore, the results from this study lend support to the notion that performing *in silico* validation, prior to costly wet-lab experiments, is a reliable means of investigating protein-protein interactions.

3.7 Acknowledgements

The authors would like to thank the South African Medical Research Council (MRC), the National Research Foundation (NRF) and the Harry Crossley Foundation for financial support of this work. JdP would like to acknowledge travel funding from the Boehringer Ingelheim Fonds, the Newton International Exchanges scheme and the NRF. SW is a recipient of a Wellcome Trust Investigator Award. SLS is funded by the South African Research Chairs Initiative of the Department of Science and Technology and NRF of South Africa, award number UID 86539. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NRF.

3.8 References

1. Organization WH. 2015. Global tuberculosis report 2015. World Health Organization.
2. Sheppard C, Cámara B, Shadrin A, Akulenko N, Liu M, Baldwin G, Severinov K, Cota E, Matthews S, Wigneshweraraj SR. 2011. Reprint of: inhibition of *Escherichia coli* RNAp by T7 Gp2 protein: role of negatively charged strip of amino acid residues in Gp2. J Mol Biol 412:832–841.
3. Mekler V, Minakhin L, Sheppard C, Wigneshweraraj S, Severinov K. 2011. Molecular mechanism of transcription inhibition by phage T7 Gp2 protein. Journal of Molecular Biology 413:1016–1027.
4. Shadrin A, Sheppard C, Severinov K, Matthews S, Wigneshweraraj S. 2012. Substitutions in the *Escherichia coli* RNA polymerase inhibitor T7 Gp2 that allow inhibition of transcription when the primary interaction interface between Gp2 and RNA polymerase becomes compromised. Microbiology (Reading, Engl) 158:2753–2764.
5. James E, Liu M, Sheppard C, Mekler V, Cámara B, Liu B, Simpson P, Cota E, Severinov K, Matthews S, Wigneshweraraj S. 2012. Structural and mechanistic basis for the inhibition of *Escherichia coli* RNA polymerase by T7 Gp2. Mol Cell 47:755–766.
6. Shadrin A, Sheppard C, Savalia D, Severinov K, Wigneshweraraj S. 2013. Overexpression of *Escherichia coli udk* mimics the absence of T7 Gp2 function and thereby abrogates successful infection by T7 phage. Microbiology (Reading, Engl) 159:269–274.
7. Nechaev S, Severinov K. 1999. Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein. Journal of Molecular Biology 289:815–826.
8. 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.
9. McClure WR, Cech CL. 1978. On the mechanism of rifampicin inhibition of RNA synthesis. J Biol Chem 253:8949–8956.
10. Banerjee R, Rudra P, Prajapati RK, Sengupta S, Mukhopadhyay J. 2014. Optimization of recombinant *Mycobacterium tuberculosis* RNA polymerase expression and purification. Tuberculosis 94:397–404.
11. 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.

12. Eswar N, Eramian D, Webb B, Shen M-Y, Sali A. 2008. Protein structure modelling with MODELLER. Methods Mol Biol 426:145–159.
13. Deshpande N, Addess KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, Knezevich C, Xie L, Chen L, Feng Z, Green RK, Flippen-Anderson JL, Westbrook J, Berman HM, Bourne PE. 2005. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. Nucleic Acids Res 33:D233-237.
14. Hubin EA, Fay A, Xu C, Bean JM, Saecker RM, Glickman MS, Darst SA, Campbell EA. 2017. Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. eLife 6:e22520.
15. 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.
16. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26:283–291.
17. Wiederstein M, Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407-410.
18. Hess B, Kutzner C, van der Spoel D, Lindahl E. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447.
19. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic Molecules. J Am Chem Soc 117:5179–5197.
20. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. The Journal of Chemical Physics 103:8577–8593.
21. Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. J Mol Graph 14:33–38, 27–28.
22. Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S, Feng Y, Zhang Y, Mandal S, Jiang Y, Liu S, Gigliotti M, Talaue M, Connell N, Das K, Arnold E, Ebright RH. 2017. Structural basis of *Mycobacterium tuberculosis* transcription and transcription inhibition. Mol Cell 66:169–179.e8.

3.9 Supplementary material

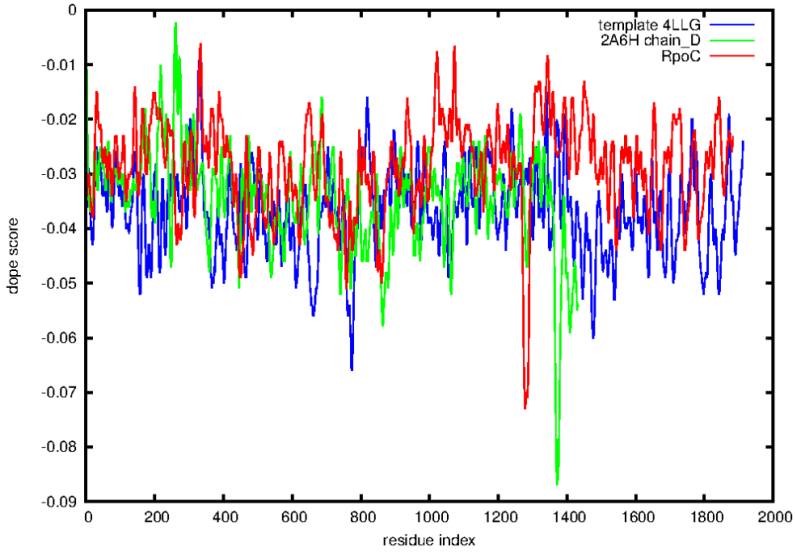


Figure S1. Discrete optimised protein energy score profiles for the LDSM of *M. tuberculosis* RpoC-SigA-Gp2 (red), *E. coli* 4LLG-Beta'-SigA-Gp2 (blue) and *T. thermophilus* 2A6H (green) templates.

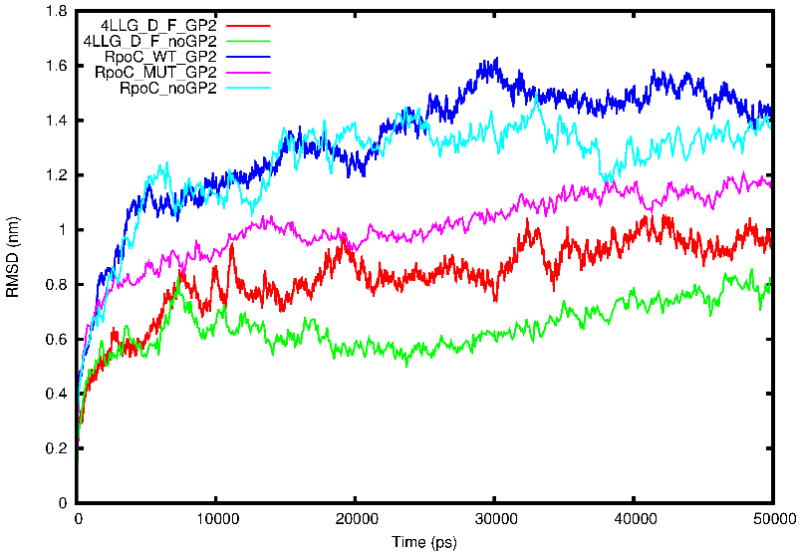


Figure S2. Trajectory analysis of the RMSD, RMSF and SASA parameters for *E. coli* 4LLG-RpoC-SigA-Gp2, apo 4LLG-RpOC-SigA and WT *M. tuberculosis* RpoC-SigA-Gp2, mutant RpoC-SigA-Gp2 complexes and apo RpoC-SigA.

(A) RMDS of the backbone atoms for 4LLG-RpoC-SigA-Gp2, 4LLG-RpoC-SigA-noGp2 *M. tuberculosis* RpoC-SigA-Gp2 WT, *M. tuberculosis* RpoC-SigA-Gp2 mutant complexes and *M. tuberculosis* RpoC-SigA-noGp2. Red line corresponds to 4LLG, green to 4LLG without Gp2, blue to *M. tuberculosis* RpoC WT structure with Gp2, magenta to *M. tuberculosis* mutant structure with Gp2 and cyan to *M. tuberculosis* RpoC without Gp2.

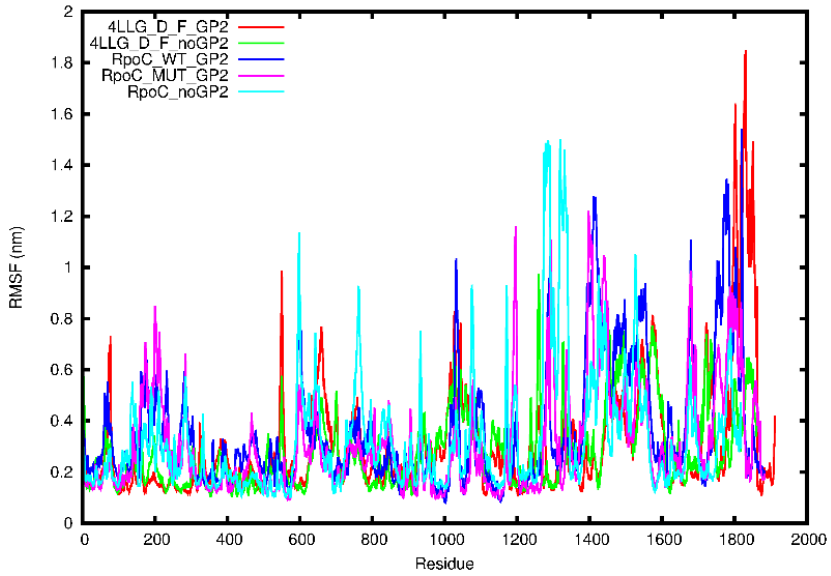


Figure S2. (B) Root mean square fluctuation of the C-alpha atoms for 4LLG-RpoC-SigA-Gp2, 4LLG-RpoC-SigA-noGp2 *M. tuberculosis* RpoC-SigA-Gp2 WT, *M. tuberculosis* RpoC-SigA-Gp2 mutant complexes and *M. tuberculosis* RpoC-SigA-noGp2.Red line corresponds to 4LLG, green to 4LLG without Gp2, blue to *M. tuberculosis* RpoC WT structure with Gp2, magenta to *M. tuberculosis* mutant structure with Gp2 and cyan to *M. tuberculosis* RpoC without Gp2.

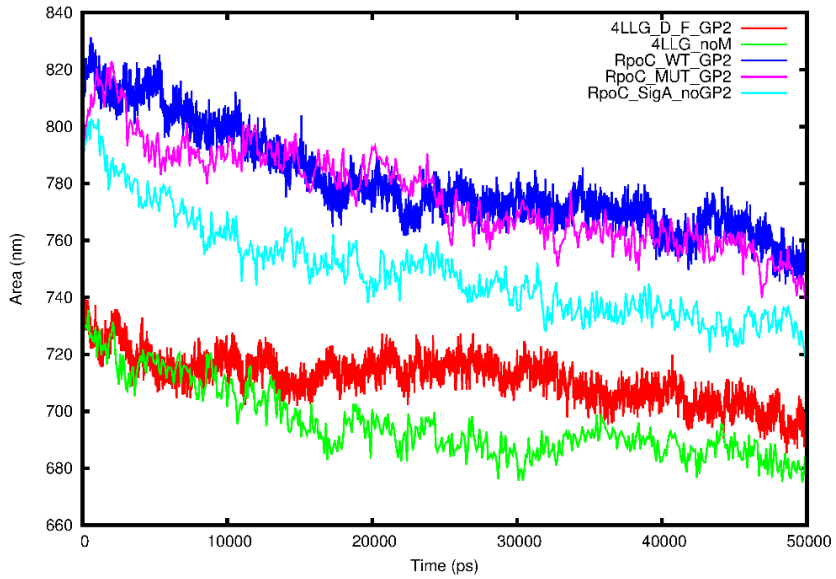


Figure S2. (C) Total solvent accessible surface area of the protein residues for 4LLG-RpoC-SigA-Gp2, 4LLG-RpoC-SigA-noGp2 and *M. tuberculosis* RpoC-SigA-Gp2 WT, *M. tuberculosis* RpoC-SigA-Gp2 mutant complexes and *M. tuberculosis* RpoC-SigA-noGp2. Red line corresponds to 4LLG with Gp2, green to 4LLG without Gp2, blue to *M. tuberculosis* RpoC WT structure with Gp2, magenta to *M. tuberculosis* mutant structure with Gp2 and cyan to *M. tuberculosis* RpoC without Gp2.

Figure S3. Fluctuation of secondary structure elements for *E. coli* 4LLG-RpoC-SigA-Gp2, apo 4LLG-RpoC-SigA and WT *M. tuberculosis* RpoC-SigA-Gp2, mutant RpoC-SigA-Gp2 complexes and apo RpoC-SigA over 50ns simulation.

Files too large to add to document – please refer to supplementary CD or links below.

(A) Secondary structure analysis of the protein residues for the *E. coli* 4LLG-RpoC-SigA-Gp2 complex system. [goo.gl/ZnIecm]
(B) Secondary structure analysis of the protein residues for the apo *E. coli* 4LLG-RpoC-SigA system. [goo.gl/Yh1cNy]
(C) Secondary structure analysis of the protein residues for the WT *M. tuberculosis* RpoC-SigA-Gp2 complex system. [goo.gl/b12HLs]
(D) Secondary structure analysis of the protein residues for the mutant *M. tuberculosis* RpoC-SigA-Gp2 complex system. [goo.gl/k9nHHc]
(E) Secondary structure analysis of the protein residues for the apo *M. tuberculosis* RpoC-SigA system. [goo.gl/QguiAu]

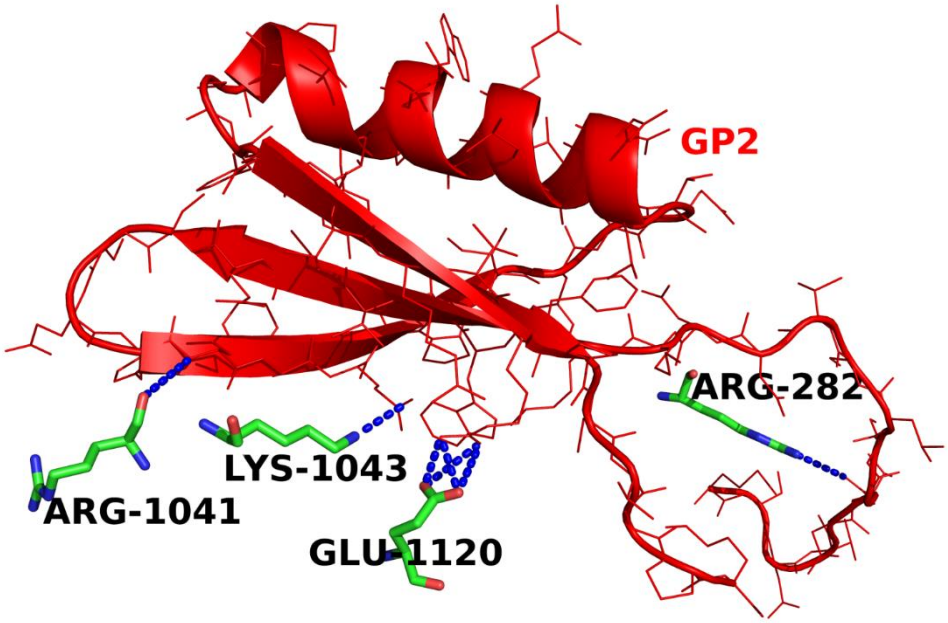
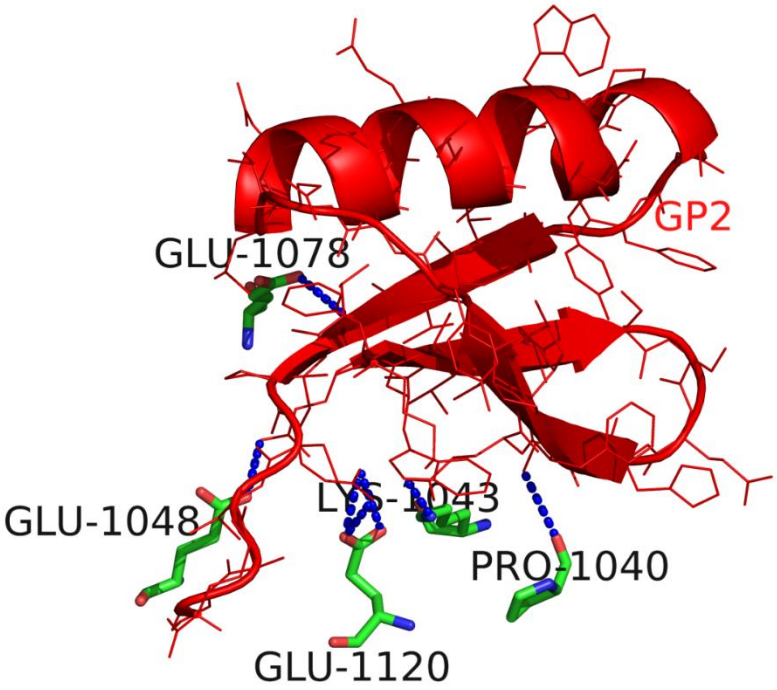


Figure S4. Polar contacts between the model built with template 5TW1 and Gp2 for the WT and mutated protein residues.

(A) Interactions of the model built with template 5TW1 with Gp2 for the WT *M. tuberculosis* protein. Blue dashes represent polar contacts; interacting residues are labelled and shown as sticks while Gp2 is shown in a red cartoon diagram.



(B) Interactions of the model built with template 5TW1 with Gp2 for the mutated *M. tuberculosis* protein (1048E and 1078E). Blue dashes represent polar contacts; interacting residues are labelled and shown as sticks while Gp2 is shown in a red cartoon diagram.

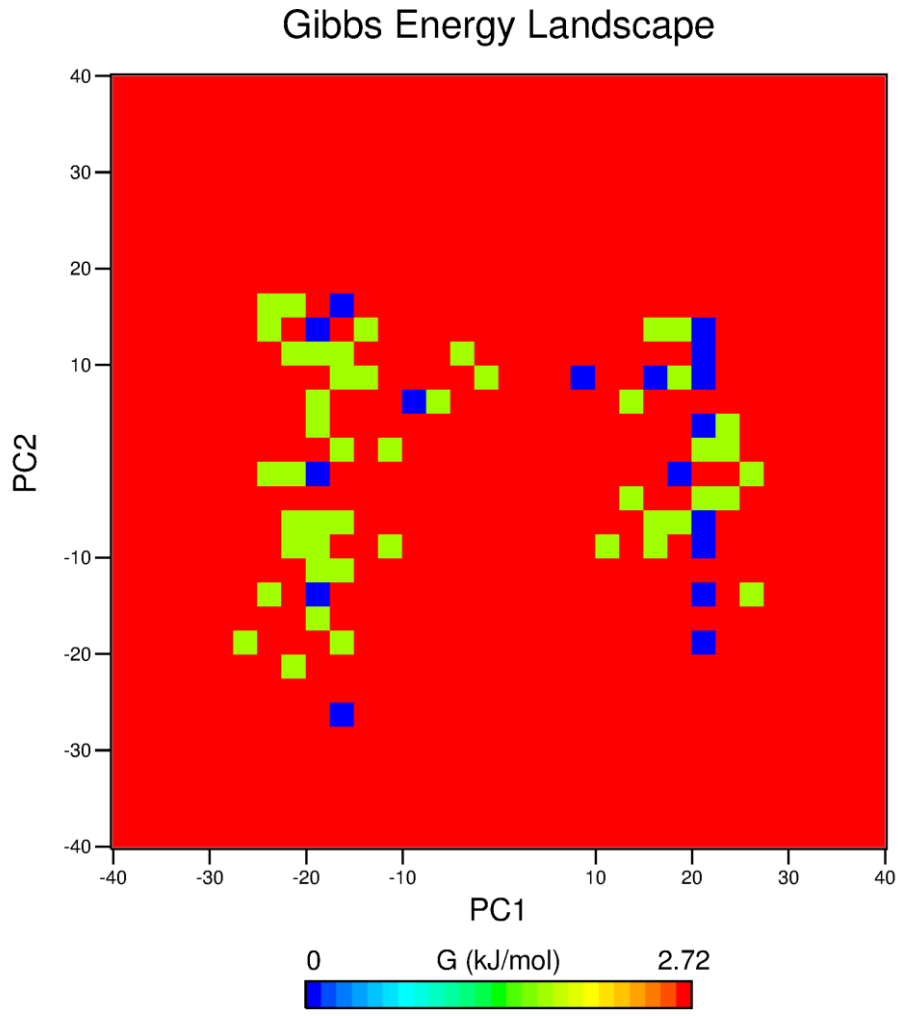


Figure S5. Local energy minima conformations for *E. coli* 4LLG-RpoC-SigA-Gp2, 4LLG-RpoC-SigA-noGp2, WT *M. tuberculosis* RpoC-SigA-Gp2, mutant *M. tuberculosis* RpoC-SigA-Gp2 complexes and apo *M. tuberculosis* RpoC-SigA system.

(A) Gibbs free energy landscape for backbone atoms of the *E. coli* RpoC-RpoD-Gp2 complex system. Dark blue dots represent local energy minima conformations (n = 14).

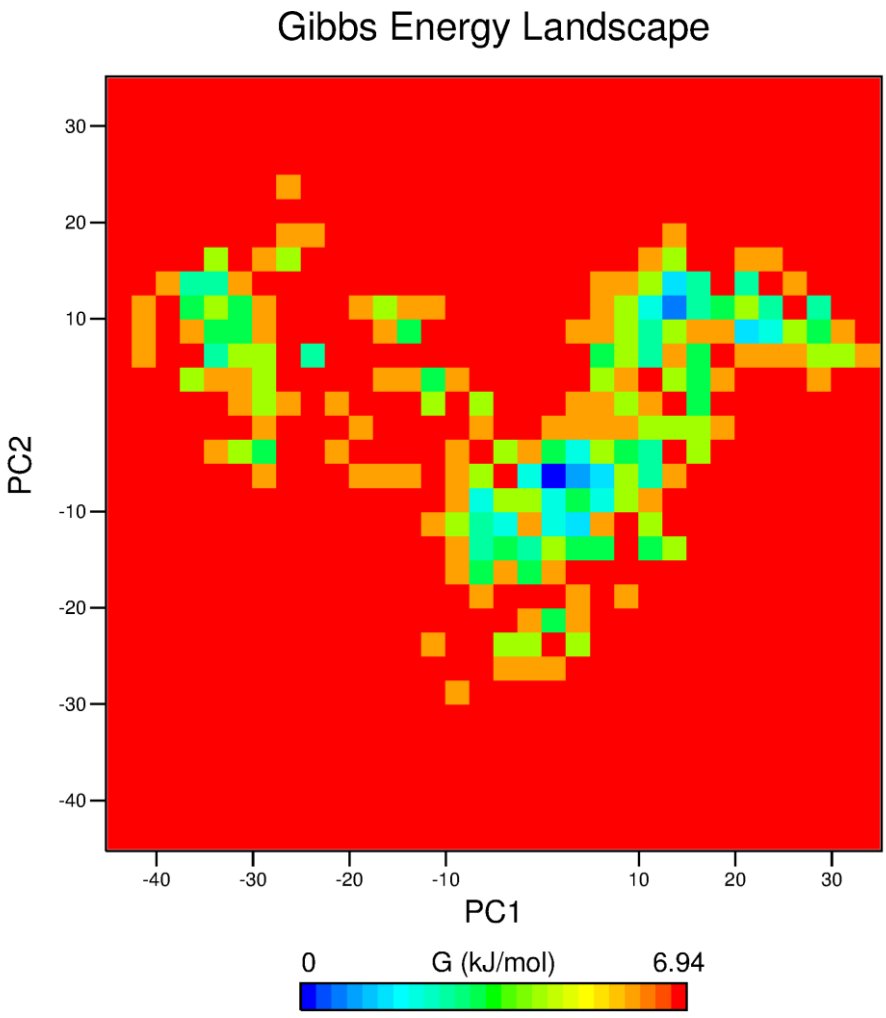


Figure S5. (B) Gibbs free energy landscape for backbone atoms of the apo *E. coli* RpoC-RpoD complex system. Dark blue dots represent local energy minima conformations (n = 1).

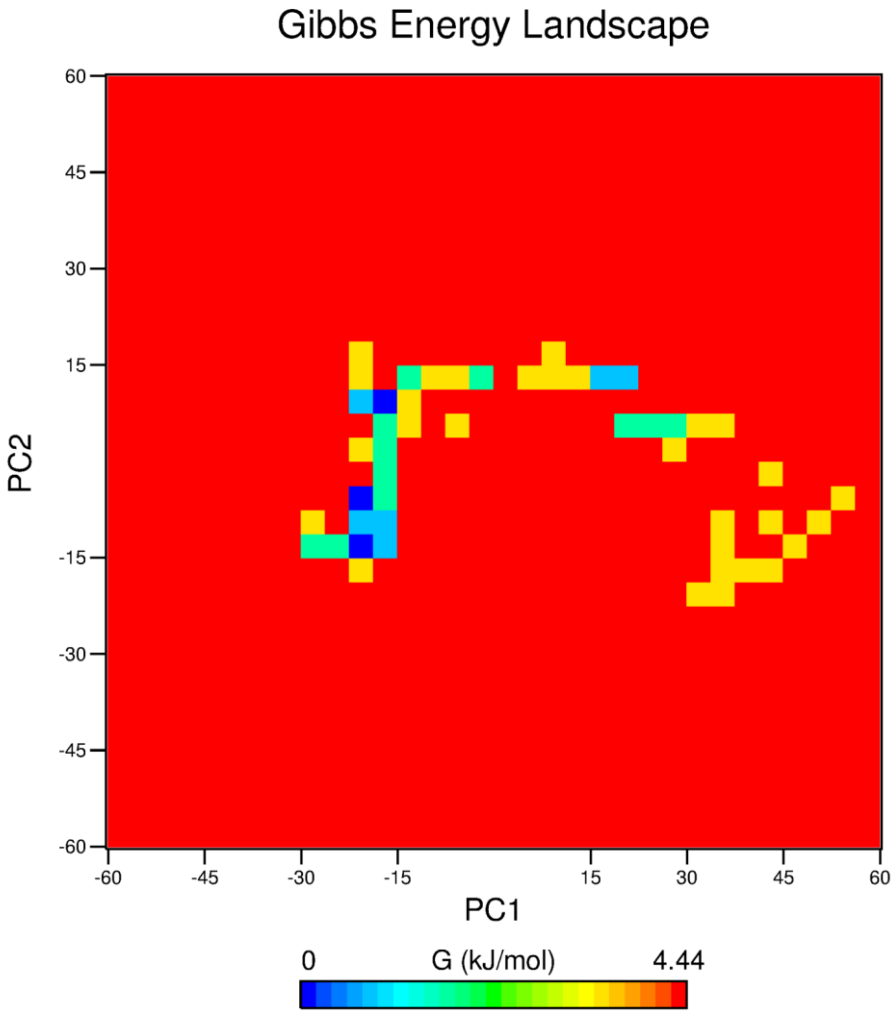


Figure S5. (C) Gibbs free energy landscape for backbone atoms of the WT *M. tuberculosis* RpoC-SigA-Gp2 complex system. Dark blue dots represent local energy minima conformations (n = 3).

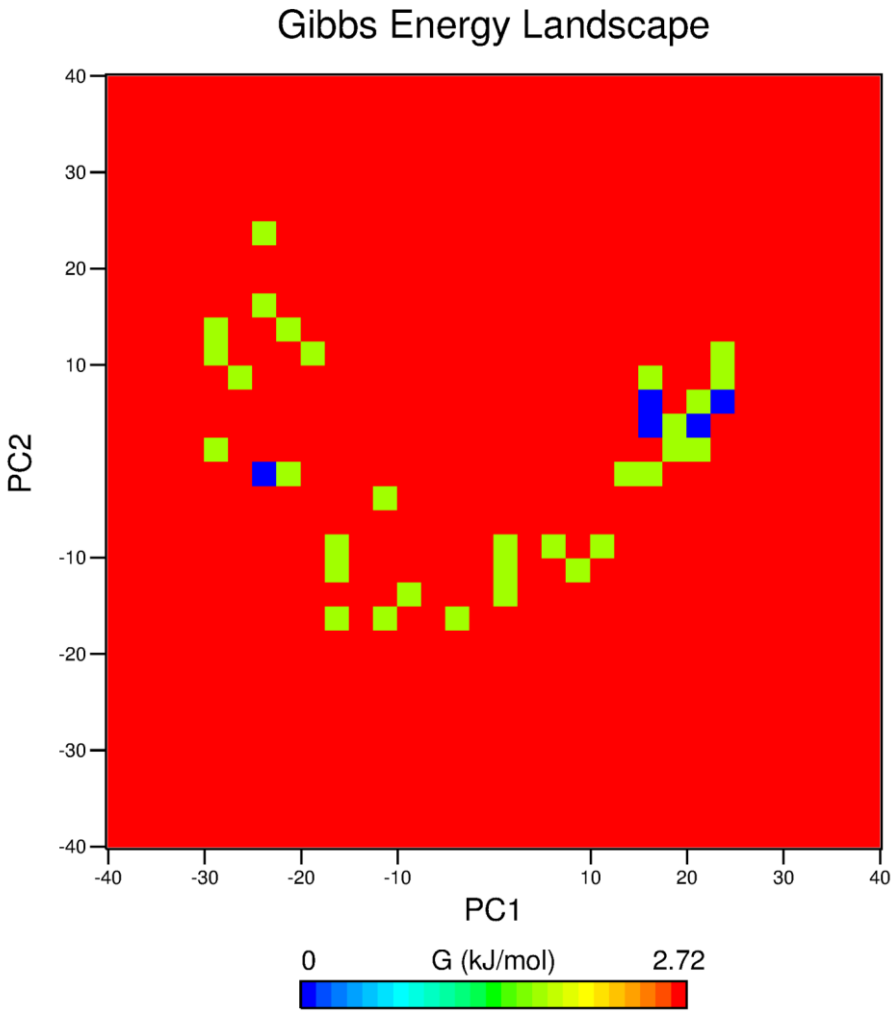


Figure S5. (D) Gibbs free energy landscape for backbone atoms of the mutant *M. tuberculosis* RpoC-SigA-Gp2 complex system. Dark blue dots represent local energy minima conformations (n = 5).

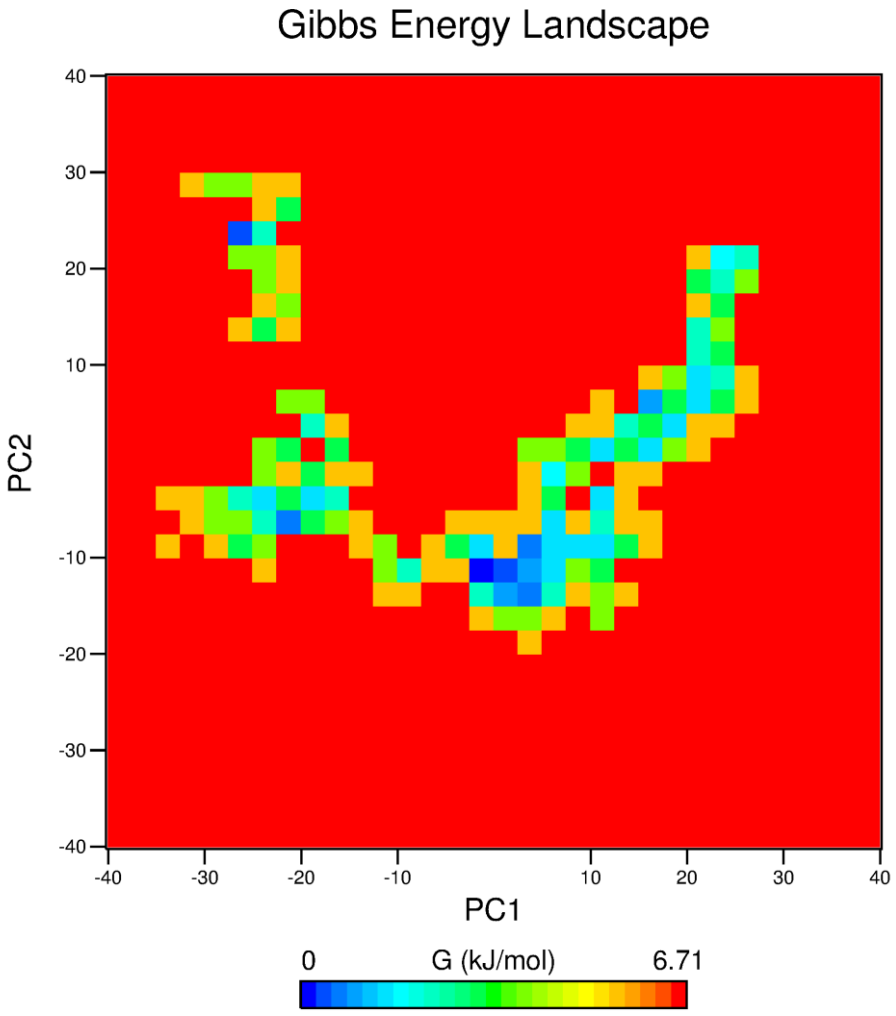


Figure S5. (E) Gibbs free energy landscape for backbone atoms of the apo *M. tuberculosis* RpoC-SigA system. Dark blue dots represent local energy minima conformations (n = 2).

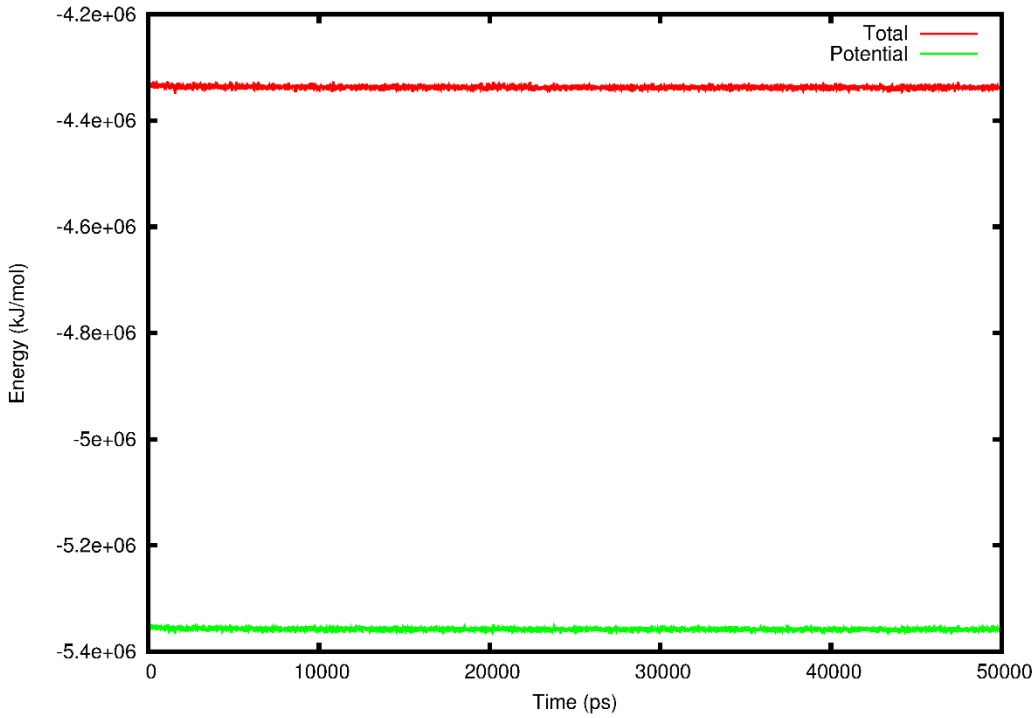
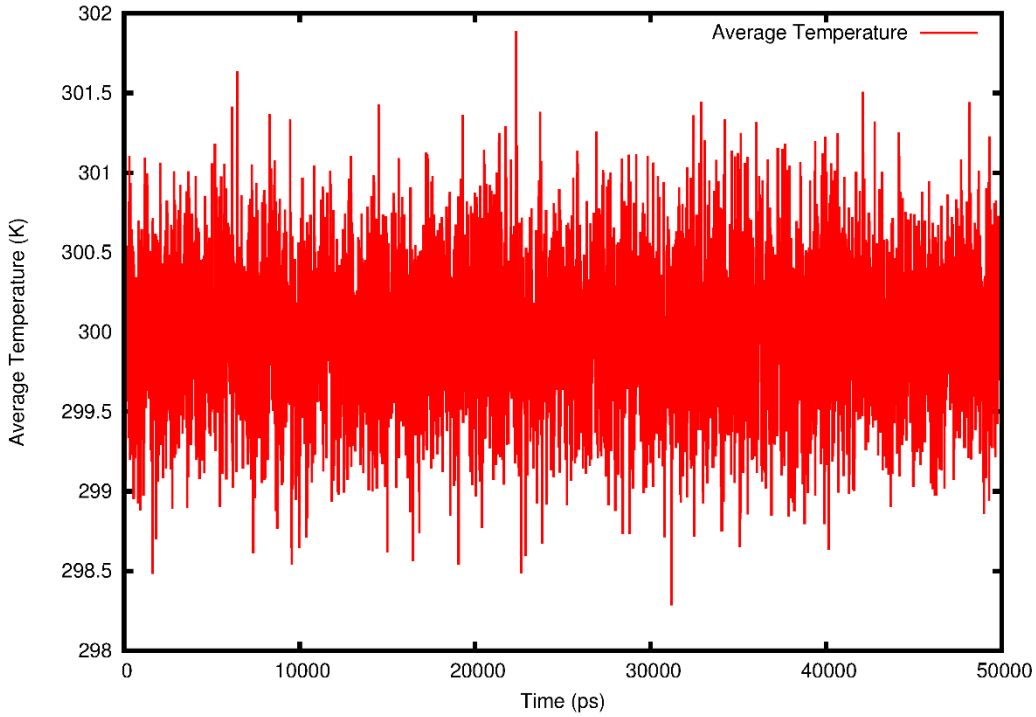


Figure S6. Energy and average temperature parameters for *E. coli* 4LLG-RpoC-SigA-Gp2 plotted over 50ns.

(A) Total energy (red line) and potential energy (green line) for the *E. coli* 4LLG-RpoC-SigA-Gp2 complex system.



(B) Average temperature fluctuation for *E. coli* 4LLG-RpoC-SigA-Gp2 complex system.

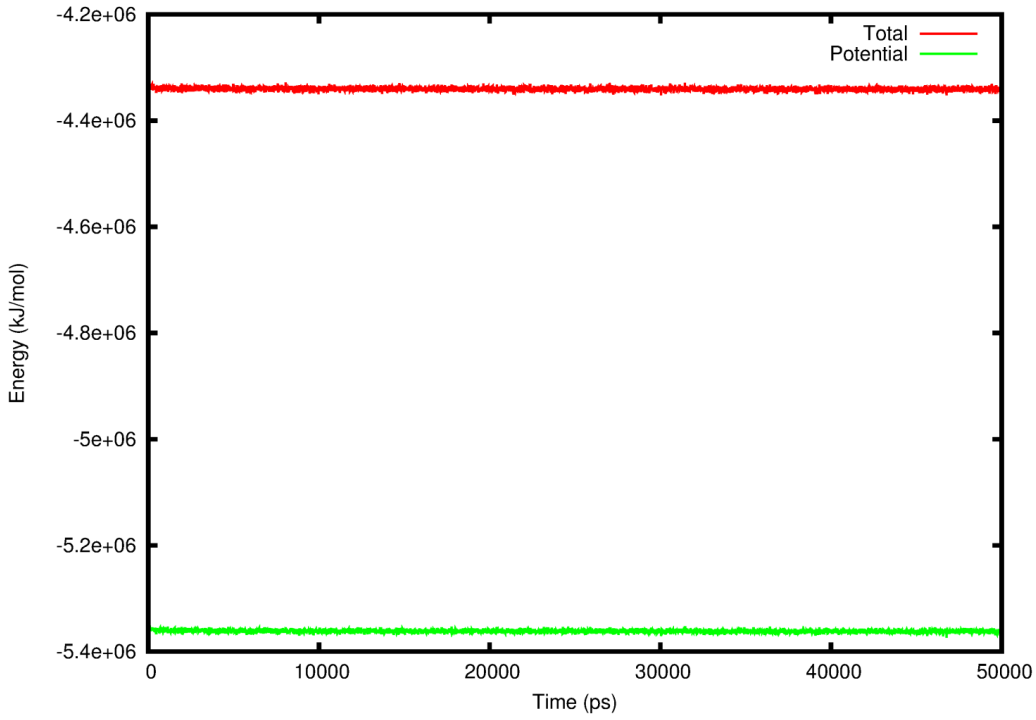
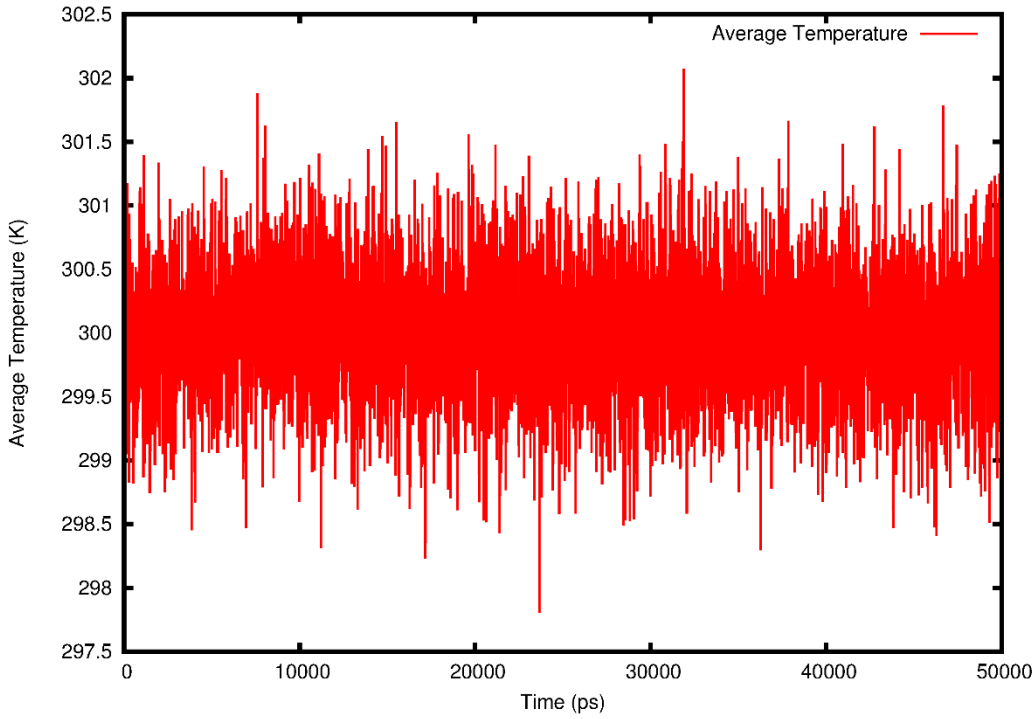


Figure S7. Energy and average temperature parameters for the apo *E. coli* 4LLG-RpoC-SigA structure plotted over 50ns.

(A) Total energy (red line) and potential energy (green line) for the apo *E. coli* 4LLG-RpoC-SigA system.



(B) Average temperature fluctuation for apo *E. coli* 4LLG-RpoC-SigA system.

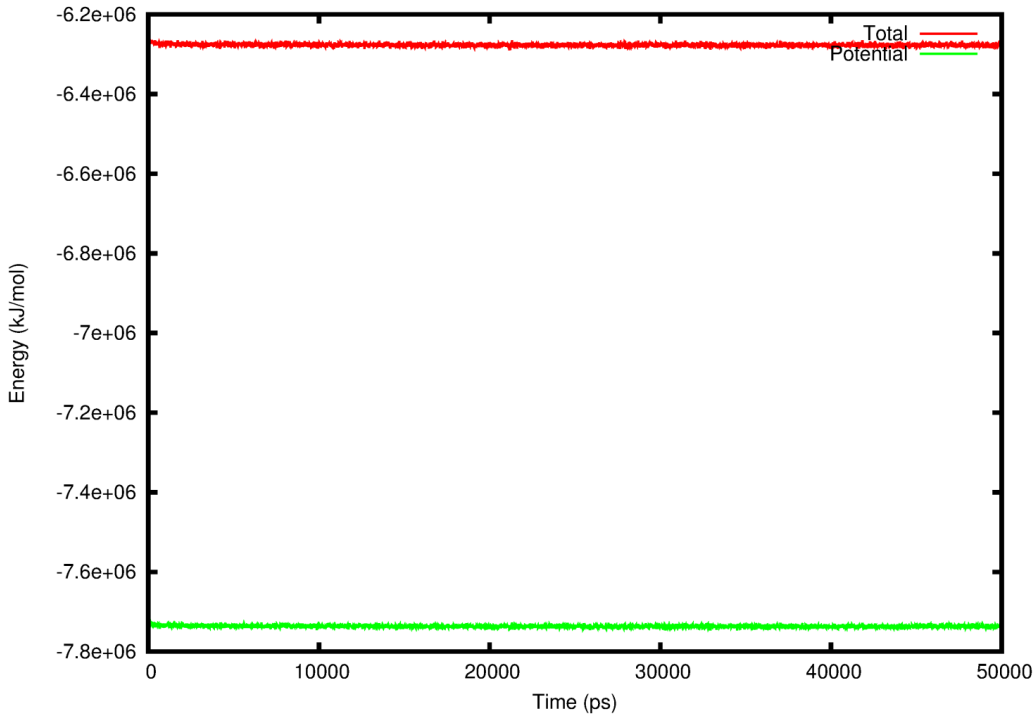
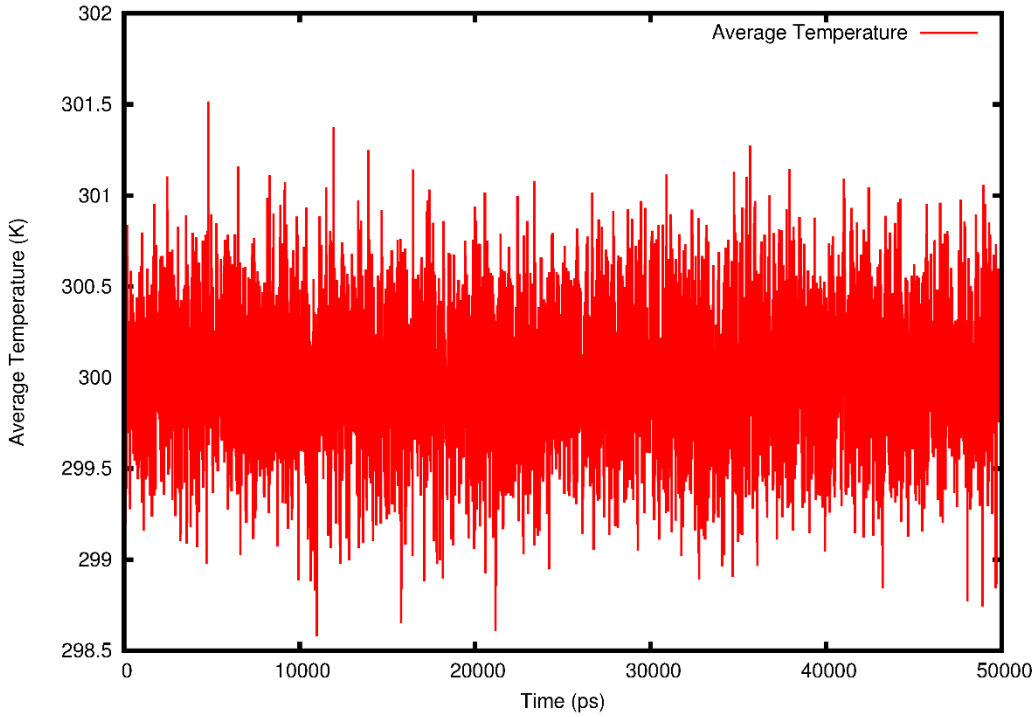


Figure S8. Energy and average temperature parameters for the WT *M. tuberculosis* RpoC-SigA-Gp2 complex system plotted over 50ns.

(A) Total energy (red line) and potential energy (green line) for the WT *M. tuberculosis* RpoC-SigA-Gp2 complex.



(B) Average temperature fluctuation for WT *M. tuberculosis* RpoC-SigA-Gp2 complex system.

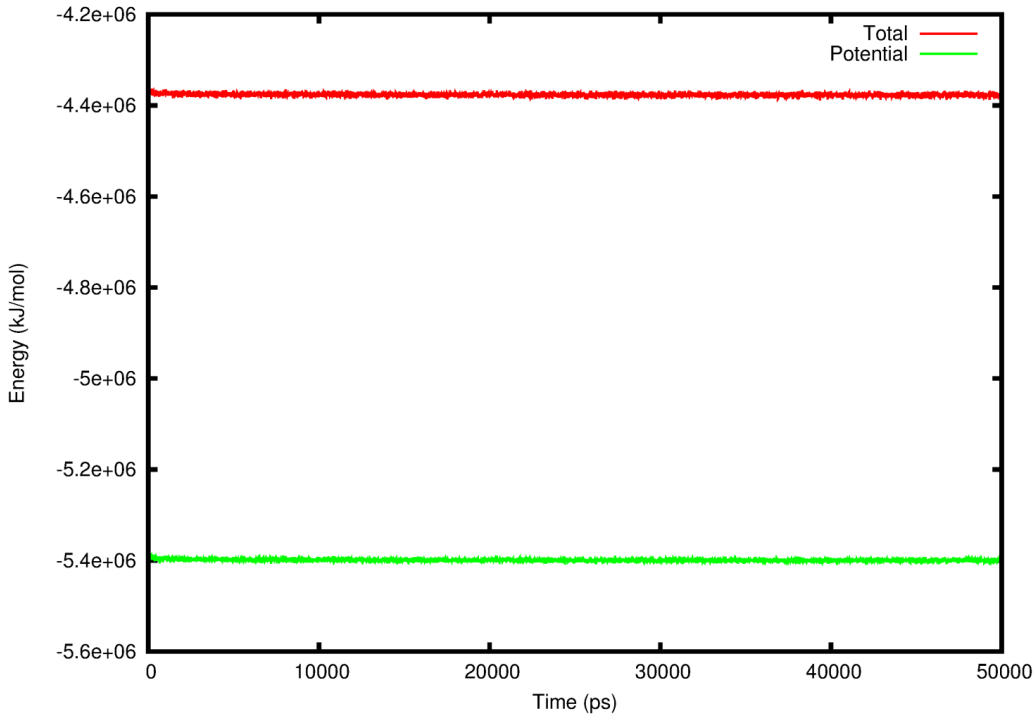
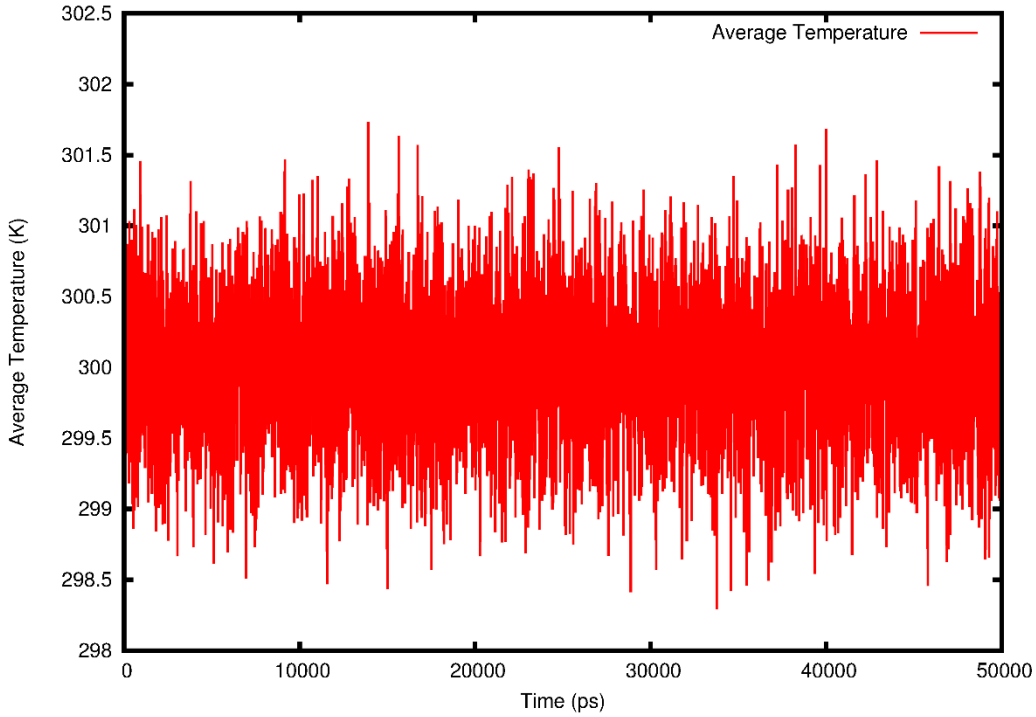


Figure S9. Energy and average temperature parameters for the mutant *M. tuberculosis* RpoC-SigA-Gp2 complex system plotted over 50ns.

(A) Total energy (red line) and potential energy (green line) for the mutant *M. tuberculosis* RpoC-SigA-Gp2 complex system.



(B) Average temperature fluctuation for the mutant *M. tuberculosis* RpoC-SigA-Gp2 complex system.

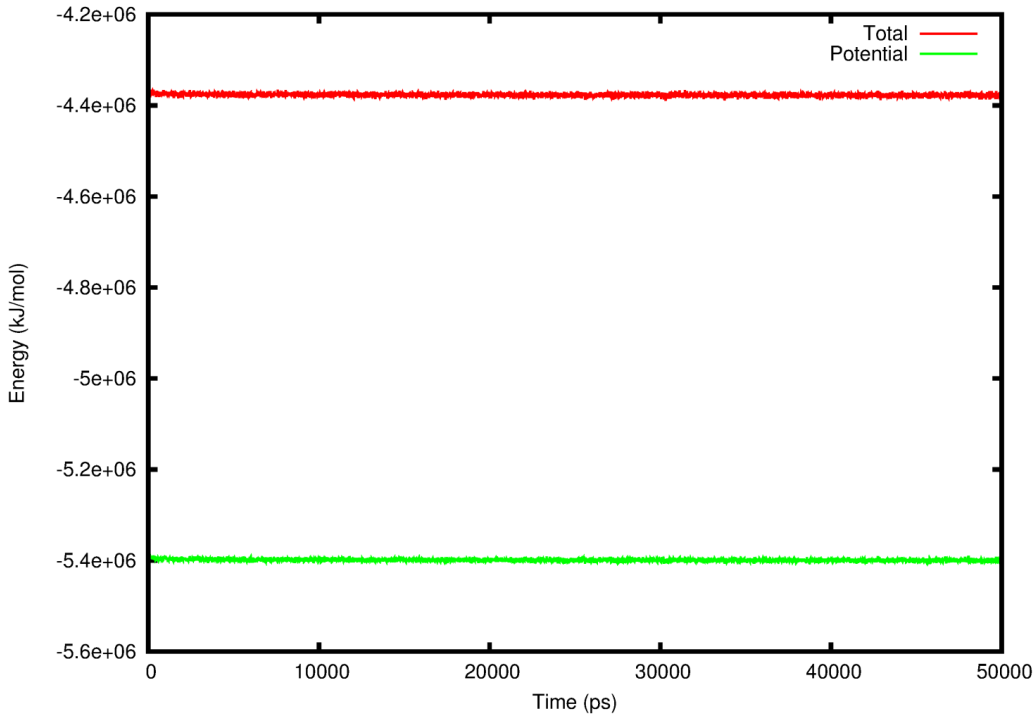
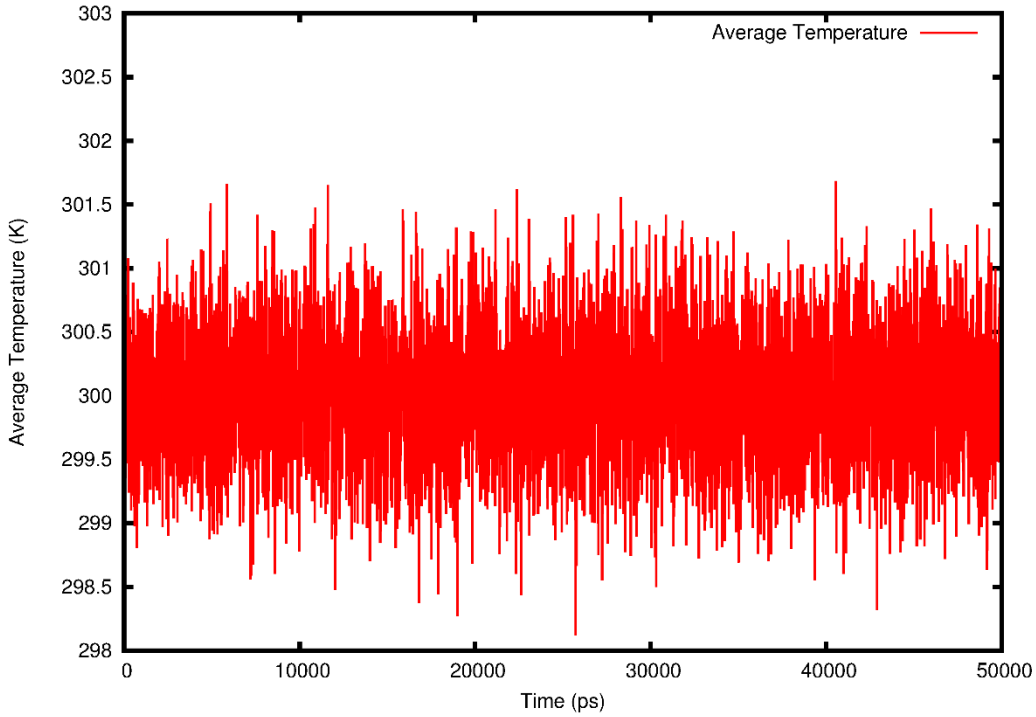


Figure S10. Energy and average temperature parameters for the apo *M. tuberculosis* RpoC-SigA system plotted over 50ns.

(A) Total energy (red line) and potential energy (green line) for the apo *M. tuberculosis* RpoC-SigA system.



(B) Average temperature fluctuation for the apo *M. tuberculosis* RpoC-SigA system.

Chapter 4

Study 2

Investigating the influence of *rpoB* and *rpoC* mutations on the function of RNA polymerase in *Mycobacterium tuberculosis*

The following chapter consists of unpublished work. My contribution to this chapter is as follows:

Literature search

Study design

Experimental work

Interpretation of data and results

Writing and editing of chapter

4.1 Abstract

Bacterial transcription is a complex process catalysed by the multi-subunit enzyme RNA polymerase (RNAP). The various steps involved in transcription initiation and elongation in *Escherichia coli* have been the focus of many studies over the years and the use of mutant RNAP has aided these studies to better understand the structure and function of this essential enzyme. However, RNAP in *Mycobacterium tuberculosis* has not been well characterized in this regard and the effects of clinically relevant mutations on enzyme function are poorly understood. As RNAP is a highly conserved enzyme, it could be assumed that mechanistic aspects of RNAP function would remain unchanged across bacterial species. However, evidence suggests that there is a great deal of variation in enzyme dynamics and promoter binding affinity. For this reason, this study aimed to characterise wildtype and mutant RNAP from *M. tuberculosis* using a mycobacterial promoter (*sigA*) and promoter-independent minimum scaffold. *In vitro* transcription from the *sigA* promoter indicated that variation exists between RNAP with different mutations. A significant difference was seen between RNAP with the *rpoB*526 mutation compared to RNAP with the *rpoB*531 and *rpoC*483 mutations. Clinically, both these enzymes would exhibit resistance to the first-line drug rifampicin. However, the former exhibits drastically increased levels of transcripts initiated from the *sigA* promoter. Results from the minimum scaffold assay indicated that the *rpoB* S531L mutation may affect specific activity of the enzyme, as a significant increase in signal was observed for RNAP with this mutation compared to the wildtype. Additionally, clinically relevant *rpoB* and *rpoC* mutations studied here appear to play a role in the relationship between the subunits which make up the RNAP holoenzyme. Our findings highlight the importance of further studies which focus on understanding rifampicin resistance in *M. tuberculosis*.

4.2 Introduction

RNA polymerase (RNAP) holoenzyme in *Mycobacterium tuberculosis* consists of 5 subunits (2αββ'ω) and one of 13 sigma (σ) factors (Figure 4.1) which drive transcription from certain sets of promoters. Clinically relevant mutations in RNAP include *rpoB* mutations, which confer resistance to the first-line drug rifampicin, and mutations in *rpoA* and *rpoC*, which compensate for the fitness cost conferred by *rpoB* mutations (1–4). As RNAP is the sole enzyme responsible for transcription in bacteria, its function and regulation play a pivotal role in gene expression. Understanding how these mutations influence enzyme function is of utmost importance to understanding the physiology of drug resistant *M. tuberculosis*.

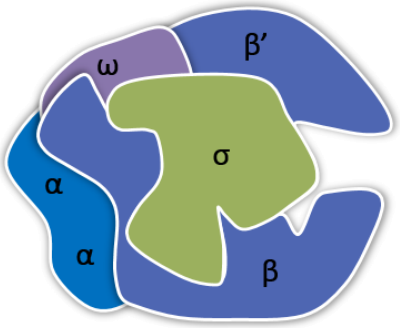


Figure 4.1 RNA polymerase holoenzyme. Multi-subunit RNA polymerase consists of β, β', ω, two identical α subunits and one of 13 σ factors in *M. tuberculosis*.

A previous study by our group investigated the effect of two of the most common *rpoB* mutations, H526Y and S531L, on the gene expression profile of *M. tuberculosis* (5). Findings from this study revealed that strains with these mutations exhibit downregulation of manifold genes involved in metabolic and cellular pathways. Furthermore, this effect appeared to be amplified in S531L mutants, irrespective of the genetic background of the strain. These substantial changes in transcription could be due to various reasons. Mutations in *rpoB* could influence the affinity and binding of RNAP to transcriptional regulators such as guanine nucleotides (p/ppGpp) or important transcriptional machinery such as sigma factors, or other RNAP-associated proteins such as CarD and RbpA. Likewise, *rpoB* mutations could have an impact on the binding or affinity of RNAP for certain promoters in the genome. Alternatively, transcriptional processes such as initiation, elongation or termination could be affected due to a change in the enzyme's conformation or stability. In support of the latter, Yanofsky *et al.* revealed that the *rpoB* H526Y mutation in *Escherichia coli* decreases transcription termination at the tryptophan operon attenuator (6). As termination efficiency was not reduced at operons

which do not form transcript termination structures, the authors suggested that particular *rpoB* mutations affect the ability of RNAP to recognize these termination signals. Furthermore, a study which investigated the fitness cost of RIF-resistant *M. tuberculosis*, reported that transcription efficiency of strains harbouring the S531L mutation was half that of the wildtype (4).

To determine the impact of *rpoB* and *rpoC* mutations on the function of mycobacterial RNAP, the current study aimed to investigate their role in transcription initiation and elongation using several *in vitro* transcription assays. Radioactivity-based assays were used to assess catalytic activity and promoter affinity. Furthermore, the use of a fluorescence-based assay was trialled to develop a comparable method without the use of radiolabelled nucleotides. Both the *rpoB* H526Y and S531L mutations were investigated to assess their impact on transcription. Additionally, two *rpoC* mutations were included in the study, as these are often acquired subsequent to *rpoB*531 mutations in *M. tuberculosis*. Previous studies which have described compensatory mutations have shown that the V483G substitution is the most commonly found *rpoC* mutation in clinical isolates of *M. tuberculosis* (1, 2). In contrast to this, Comas *et al*. also reported the D485N substitution, which is very close in proximity to the *rpoC*483 mutation, however does not occur as frequently (1). By selecting these mutations, we aimed to determine the mechanism through which the *rpoC*483 mutation is able to ameliorate the fitness cost conferred by *rpoB* mutations.

4.3 Methods

4.3.1 Vectors

Dual expression vectors containing gene sequences for mycobacterial RNA polymerase (RNAP) were used for the purpose of protein expression and purification in this study (Figure 4.2). The pAcYc Duet vector contains gene sequences for 10x His-tagged α-subunit (*rpoA*) and sigma factor σ^A (*sigA*). Likewise, pET Duet contains gene sequences for the β (*rpoB*) and β' (*rpoC*) subunits. These vectors, a kind gift from Professor Mukhopadhyay's laboratory at the Bose Institute in India, allow for the purification of high yields of *in vivo* assembled *M. tuberculosis* RNAP (7).

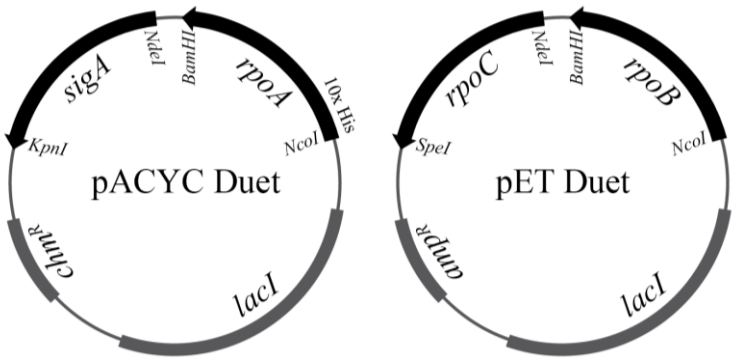


Figure 4.2 Vectors used for protein expression. Dual expression vectors containing gene sequences for *sigA*, *rpoA*, *rpoC* and *rpoB* were used for expression and purification of *M. tuberculosis* RNA polymerase using chloramphenicol (chm) and ampicillin (amp) for plasmid selection. Image adapted from Banerjee *et al*. (7).

4.3.2 Site-directed mutagenesis

As this study aimed to investigate the influence of clinically relevant mutations on the function of RNAP in *M. tuberculosis*, primers were designed to generate base changes correlating to *rpoB* (H526Y, S531L) and *rpoC* (V483G, D485N) mutations by site-directed mutagenesis (Table 4.1). Mutations were generated using the Q5 Site-Directed Mutagenesis kit (New England BioLabs) and confirmed by targeted DNA sequencing using the ABI 3500XL sequencer (Applied Biosystems).

Table 4.1 Primers used for site-directed mutagenesis. The following primer sets were used to generate mutations (denoted by lowercase bold letters) in the dual expression vector containing gene sequences for *rpoB* and *rpoC*.

Mutation	Primer Sequence	T_m (°C)
rpoB H526Y	5' GGGGTTGACC**t**ACAAGCGCCG 3'	69
	5' GACAGCGGGTTGTTCTGGTC 3'	
rpoB S531L	5' CGCCGACTGT**t**GGCGCTGGGG 3'	72
	5' CTTGTGGGTCAACCCCGACAGC 3'	
rpoC V483G	5' CGCCCCCAAG**g**GTGGGATGTG 3'	72
	5' CTGGCGCTCCACCATGCG 3'	
rpoC D485N	5' CCAAGTGTGG**a**ATGTGCTCGAAGAGGTCATCG 3'	72
	5' GGGCGCTGGCGCTCCACC 3'	

4.3.3 Protein expression and extraction

RNAP holoenzyme derivatives ($\alpha_2\beta\beta'\sigma^A$) were expressed as fusion proteins with a polyhistidine tag as previously described (7). In summary, chemically competent *E. coli* BL21(DE3) was co-transformed with dual expression vectors pACYCDuet *rpoA-sigA* and pETDuet *rpoB-rpoC* and selected on Luria Bertani agar containing ampicillin and chloramphenicol at final concentrations of 100 µg/mL and 34 µg/mL, respectively. Single colonies were picked and inoculated into 7 mL of 2xYT media and incubated overnight at 37°C. Saturated primary cultures were used to inoculate secondary cultures of 500 mL which were incubated at 37°C to an OD_{600} of 0.5. Thereafter, cultures were induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (Calbiochem) to a final concentration of 0.5 mM and incubated at 16°C for 16 hours. Cells were harvested at 7000*g* for 20 minutes at 4°C and resuspended in lysis buffer [10 mM Tris-HCl pH 7.9, 50 mM KCl, 10 mM β-mercaptoethanol, 15 mM $MgCl_2$, 5 mM EDTA, cOmplete ULTRA protease inhibitor (Roche) prepared as per manufacturer's instruction], followed by sonication with a probe sonicator (40% amplitude, 5-second pulses, 2 x 10-minute cycles).

Sonicated cells were centrifuged at 23,500*g* for 30 minutes at 4°C and 10% polyethyleneimine solution (Sigma-Aldrich) was added to the supernatant, thereafter the solution was allowed to mix for an additional 30 minutes at 4°C. The solution was centrifuged at 23,500*g* for 20 minutes at 4°C and the pellet resuspended in wash buffer [10 mM Tris HCl pH 7.9, 6 mM β-mercaptoethanol, 5% glycerol, 0.1 mM EDTA, 500 mM NaCl]. Following centrifugation, the pellet was resuspended in extraction buffer [10 mM Tris HCl pH 7.9; 6 mM β-mercaptoethanol, 5% glycerol, 0.1 mM EDTA, 1 M NaCl] and allowed to incubate at 4^0C for 45 minutes with

stirring. The solution was centrifuged at 23,500*g* for 20 minutes at 4°C and the supernatant precipitated with saturated ammonium sulphate solution (Sigma-Aldrich). After a 30-minute incubation step at 4°C with stirring, the solution was centrifuged and the resulting pellet resuspended in Buffer A [25 mM NaH_2PO_4 pH 7, 0.5 M NaCl, 5% glycerol].

4.3.4 Protein purification

Recombinant protein complexes were purified using the ÄKTA pure chromatography system with UNICORN 6.4 software (GE Healthcare Life Sciences). For the purpose of affinity purification, the protein solution was injected into a 5 mL HisTrap HP column (GE Healthcare Life Sciences) pre-equilibrated with Buffer A. The column was washed with 4 column volumes (CV) Buffer A, followed by 4 CV of 3% Buffer B [25 mM NaH_2PO_4 pH 7, 0.5 M NaCl, 5% glycerol, 1 M imidazole]. Protein was eluted with Buffer B over a linear gradient of 8 CV and collected in 1 mL fractions. Based on the elution profile, fractions containing the proteins of interest were pooled and precipitated with saturated ammonium sulphate solution. The solution was centrifuged at 23,500*g* for 20 minutes at 4°C and the pellet resuspended in Buffer C [10 mM Tris HCl pH 7.9, 5% glycerol, 0.1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$, 50 mM NaCl].

In a second round of purification, protein was injected into a 5 mL HiTrap Heparin HP column (GE Healthcare Life Sciences) pre-equilibrated with Buffer C. The column was washed with 8 CV Buffer C, followed by elution with Buffer D **[**10 mM Tris HCl pH 7.9, 5% glycerol, 0.1 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$, 1 M NaCl] over a linear gradient of 20 CV. Resulting elution fractions were run on a 10% SDS-PA gel and those containing the desired proteins which were free of contaminating proteins were pooled and dialysed overnight in storage buffer [10 mM Tris HCl pH 7.9, 50 mM NaCl, 50% glycerol, 0.1 mM EDTA, 1 mM DTT] using a dialysis membrane (Spectra/Por) with molecular weight cut off of 6-8000 kDa.

4.3.5 Protein quantification

Purified proteins were separated on a 10% SDS-PA gel and quantified densitometrically using the Chemi Doc System (Bio-Rad). Thereafter, the Pierce BCA Protein Assay kit (Thermo Scientific) was used to determine the concentration of the purified proteins. For long term storage, samples were stored in small aliquots at -80°C. Working stocks were diluted to a concentration of 1 µM and stored at -20°C for up to one week.

4.3.6 Minimal scaffold assay

A minimal scaffold assay was used to determine the activity of RNAP as previously described (8). The assay measures the ability of RNAP to add a single nucleotide onto the end of a pre-made RNA strand. The use of a DNA/RNA hybrid in the assay mimics the elongation complex and thereby allows for the study of promoter-independent catalytic activity (Figure 4.3).

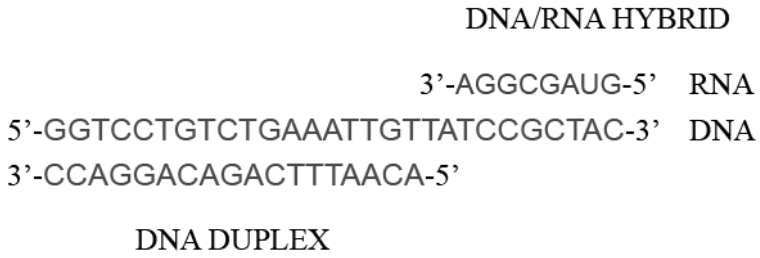


Figure 4.3 DNA/RNA hybrid used in the minimal nucleic acid scaffold assay. Adapted from Kulbachinskiy *et al*. (8).

PAGE-purified DNA and RNA (Sigma-Aldrich) were diluted to prepare a 10 µM working stock of each. Subsequently, a 500 nM stock of minimal scaffold was prepared in nuclease-free water (10 mM Tris HCl pH10, 1 mM $MgCl_2$, 0.5 µM template DNA, 0.5 µM non-template DNA, 0.5 µM RNA). The mixture was heated at 95°C for 5 minutes and allowed to cool slowly overnight. The minimal scaffold assay was set up in technical triplicates using 400 nM of purified mycobacterial RNAP, 20 U of RNasin (Promega) and reaction buffer (40 mM Tris pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT). Samples were incubated for 5 minutes at 37°C, supplemented with 100 nM minimal nucleic acid scaffold and thereafter incubated for a further 10 minutes before adding [α^{32}P]-UTP. The reaction was stopped after 10 minutes and the radiolabelled 9-nt RNA product was quantified by running the reaction on a 20% urea gel. Subsequent phosphorimaging was performed using the Typhoon FLA 7000 (GE Healthcare Life Sciences).

4.3.7 Radioactivity-based transcription assay

The *sigA* promoter, previously described by Banerjee *et al.* 2014 (7), was used for the purpose of radioactivity-based transcription assays. The template was prepared by PCR amplification of the respective promoter region using *M. tuberculosis* H37Rv genomic DNA as template (Table 4.2), with subsequent purification of the PCR reaction using the Wizard SV Gel and PCR Clean-Up System (Promega).

Table 4.2 Primers used to prepare templates for *in vitro* transcription assays.

Gene	Primer Sequence	Product Size (bp)	T_m (°C)
sigA	5' AGAACACCGCCGGAATTGTCGGTG 3'	318	66
	5' GACGCGGAAGCCGCGGGC 3'		

A 10 µL reaction containing 800 nM RNAP, transcription buffer [40 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM] and 20 U of RNasin was set up for the purpose of the *in vitro* transcription assay using radiolabelled GTP (Figure 4.4). After the addition of the rNTP mix to a final concentration of 0.01mM, the reaction was incubated for 30 minutes at 37°C to allow transcription to take place. Only CTP and [α^{32}P]-GTP were added to the reaction to allow for transcription of a 17-nucleotide product from the DNA template.

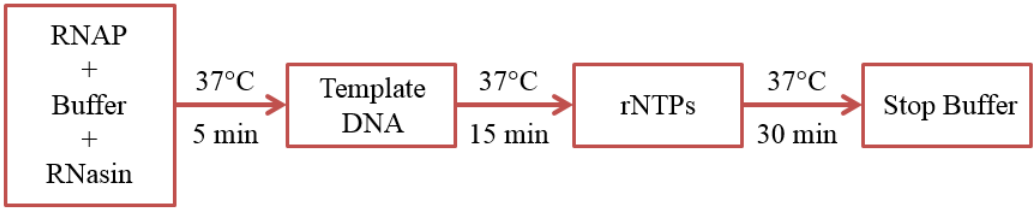


Figure 4.4 Outline of *in vitro* transcription assay using radiolabelled GTP to detect transcribed RNA.

Transcription was terminated with the addition of stop buffer [80% formamide, 10 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol]. Thereafter, reactions were heated at 95°C for 5 minutes, centrifuged at 13,000*g* for 1 minute and resolved with a 20% urea gel run at 200 volts for 60 minutes. The wet gel was exposed overnight and phosphorimaging was performed using the Typhoon FLA 7000 (GE Healthcare Life Sciences). Band intensities were quantified using ImageJ software and a one-way ANOVA was used to determine statistical difference between reactions from wildtype and mutant RNAP, using the Bonferroni correction to adjust the confidence interval.

4.3.8 Fluorescence-based transcription assay

A fluorescence-based *in vitro* transcription assay was used to determine the activity of the wildtype and mutant polymerases. The pUC19 *sinP3* plasmid, provided by the Mukhopadhyay lab (7), was used as a template for RNA synthesis from the *sinP3* promoter in a 20 µL reaction containing 50 nM RNAP in transcription buffer [250 mM Tris-HCl pH 8.0, 500 mM KCl, 50 mM $MgCl_2$, 25% glycerol, 5 mM DTT, 50 µg/mL BSA] (Figure 4.5).

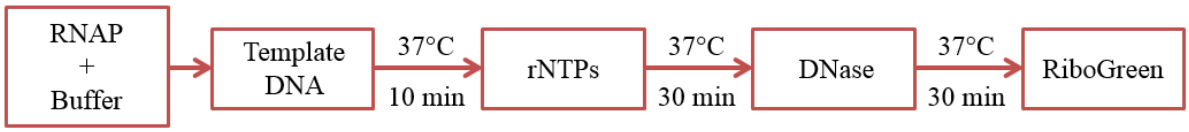


Figure 4.5 Outline of *in vitro* transcription assay using RiboGreen reagent to measure RNA synthesis.

After the addition of rNTPs to a final concentration of 0.1 mM, the reaction was incubated for 30 minutes at 37°C to allow transcription to take place. Once treated with DNaseI (Agilent) to remove the DNA template, RNA transcripts were quantified using Quant-iT RiboGreen reagent (Thermo Scientific) which fluoresces when bound to nucleic acid. Reactions were set up in triplicate to ensure reproducibility and negative controls, containing storage buffer, DNA template and rNTPs, were included in each experiment to allow for an accurate baseline fluorescence reading. Fluorescence intensity was measured with an excitation wavelength of 485 nm and an emission wavelength of 520 nm using the POLARstar Omega plate reader (BMG Labtech). As before, a one-way ANOVA was used to examine differences between reactions from wildtype and mutant RNAP, using a Bonferroni-adjusted p-value to determine statistical significance.

4.4 Results

4.4.1 Protein quantification

Wildtype and mutated RNA polymerase (RNAP) was successfully purified using affinity chromatography. Purified protein was visualised by SDS-PAGE which revealed differing ratios of subunits across wildtype and mutant RNAP preparations (Figure 4.6). Subunit abundances of RNAP with the *rpoB*531 mutation, with and without the additional *rpoC*483 mutation, appeared to be most similar to wildtype RNAP, where much of the preparation consisted of β and β' subunits (Table 4.3). The relative proportion of 10x His-tagged α subunit remained relatively constant across all preparations, with the exception of RNAP with the *rpoC*485 mutation, where it was marginally higher. Interestingly, the highest proportion of σ^A was seen in RNAP with the *rpoB*526 mutation. We considered normalising the signals obtained from *in vitro* transcription assays based on these protein quantification values, however concluded that doing so could potentially skew the results. Therefore, downstream assays were performed with equimolar amounts of RNAP using concentrations determined by the BCA Protein Assay. The implications of this are discussed in the limitations section of this chapter.

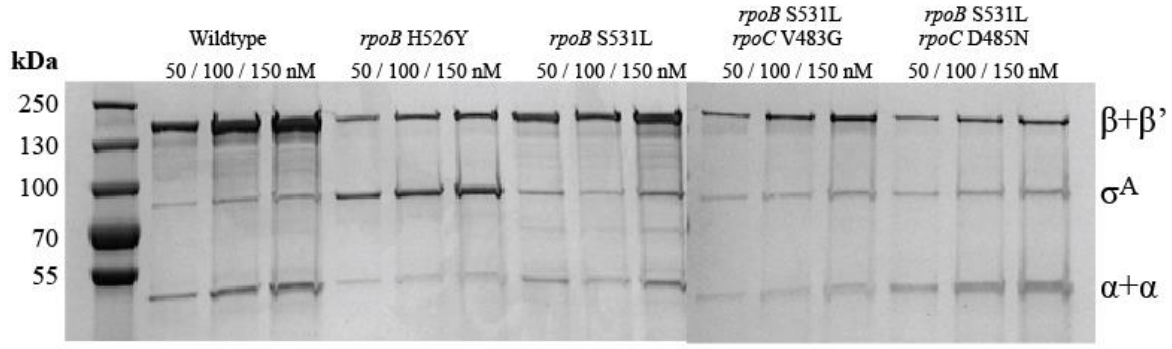


Figure 4.6 Increasing concentrations (50-150 nM) of purified RNAP preparations following fractionation with SDS-PAGE.

Table 4.3 Quantification of 50 nM protein bands. Values represent percentage intensity relative to bands in the same lane.

	WT	***rpoB* H526Y**	***rpoB* S531L**	***rpoB* S531L *rpoC* V483G**	***rpoB* S531L *rpoC* D485N**
β+β'	75.4	42.3	67.1	68.5	39.9
σ^A	9.2	47.9	15.4	17.3	19.4
α	15.5	9.8	17.5	14.2	40.7

4.4.2 Radioactivity-based assays

Purified RNAP was used in a minimal scaffold assay to measure the ability of RNAP to add a nucleotide to the end of a pre-made RNA strand (Figure 4.7). Surprisingly, wildtype RNAP displayed the lowest activity in this promoter-independent assay. A comparable level of activity was seen in all the mutant forms of the enzyme. Furthermore, RNAP harbouring *rpoB*531 (P = 0.002) and *rpoB*531 with *rpoC*483 (P = 0.001) mutations, were significantly higher than wildtype. The *in vitro* transcription assay which included the *sigA* promoter as a DNA template, revealed a very different set of results. Firstly, wildtype and mutant forms harbouring the *rpoB*531 and *rpoC*483 mutations showed similar levels of transcription (Figure 4.8). However, RNAP with the *rpoB*526 mutation yielded significantly increased levels of transcript compared to RNAP harbouring the *rpoB*531 and *rpoC*485 mutations (P = 0.002).

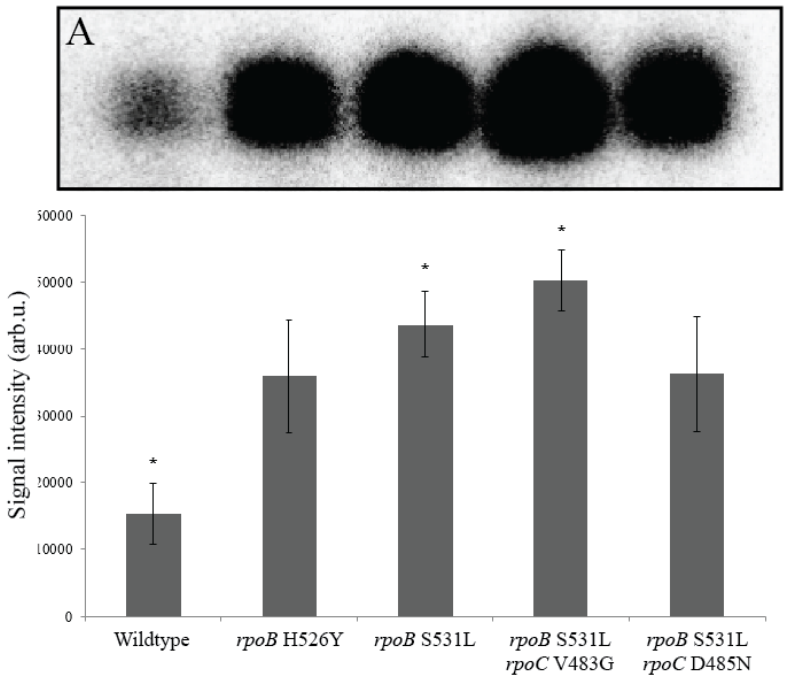


Figure 4.7 Minimum scaffold assay. Bars represent the mean obtained from technical triplicates with error bars depicting standard deviation. *Bonferroni post hoc test: $p < 0.005$.

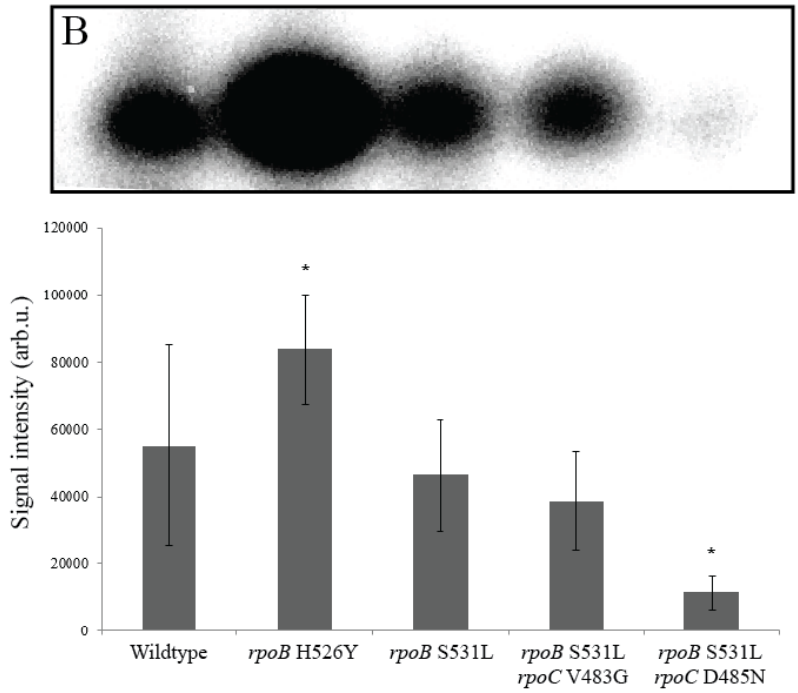


Figure 4.8 *In vitro* transcription from the *sigA* promoter. Each bar represents the mean value obtained from technical quadruplicates with error bars depicting standard deviation. *Bonferroni post hoc test: $p < 0.005$.

4.4.3 Fluorescence-based transcription assay

Radioactivity-based assays are considered to be the gold-standard to study transcription *in vitro*. However, a fluorescence-based assay was established as it would eliminate the use of harmful radioisotopes in the experimental set-up. Results from this assay demonstrated that RNAP with the *rpoB*526 and *rpoB*531 mutation are 20-26% less active compared to wildtype, however statistical analysis indicated that the differences seen between these reactions are not significant (Figure 4.9).

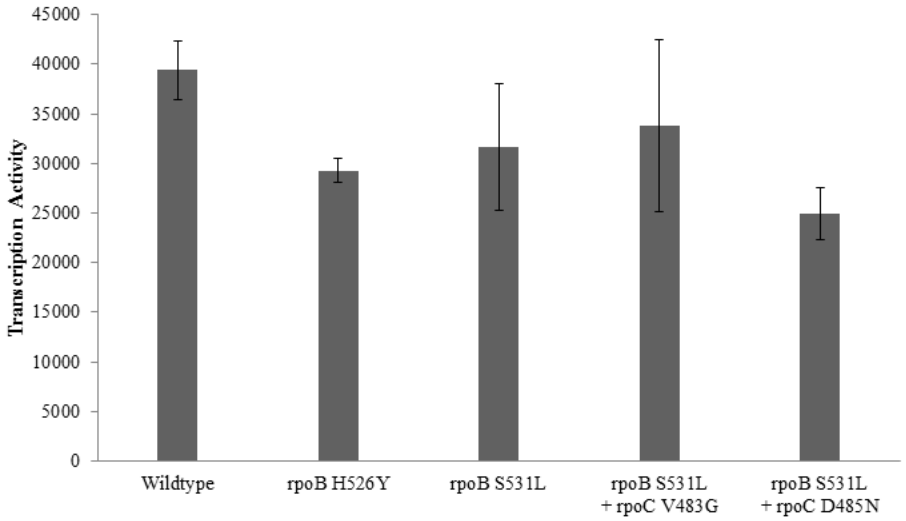


Figure 4.9 Fluorescence-based transcription assay. Fluorescence values obtained from the *in vitro* transcription assay where a nucleic acid stain is used to measure transcript abundance. Each bar represents the mean obtained from technical triplicates with error bars depicting standard deviation.

4.5 Discussion

In the current study, we assessed the impact of *rpoB* and *rpoC* mutations on the function of RNA polymerase (RNAP). As a previous study from our group indicated that different *rpoB* mutations led to distinct gene expression profiles, we were interested to investigate the cause of this variation at a molecular level. Furthermore, studies have found that *rpoC* mutations are able to restore the fitness cost incurred by *rpoB* mutants, yet the mechanism through which this occurs remains largely unknown (1, 2). The present study undertook to investigate the effect of previously described *rpoC* mutations on *in vitro* transcription, a rudimentary yet essential step to understanding the compensatory mechanism of these mutations. Additionally, positive outcomes for the current works include the generation of new tools to study RNAP physiology. Ours is the first study to incorporate the use of purified mycobacterial RNAP which harbours clinically relevant *rpoB* and *rpoC* mutations for the purpose of *in vitro* transcription assays.

4.5.1 Radioactivity-based assays

As the minimum scaffold assay is promoter nonspecific, results from this assay gave an indication of the enzyme's catalytic activity. This assay revealed that RNAP with the *rpoB*531, with or without the additional *rpoC*483 mutation, has an enhanced ability to bind to the nucleic acid template. We propose that this might be due to a conformational change in RNAP, which promotes nucleic acid entry into the DNA binding channel. Binding of the σ factor to the RNAP core is known to prevent the interaction between RNAP and the minimal scaffold (8). Therefore, results from this assay suggest that the wildtype enzyme exhibits a higher affinity for σ^A compared to RNAP with an *rpoB*531 mutation, thereby displaying a lower level of transcription activity.

Of note to these results, a study by Jin and Turnbough demonstrated abortive transcription for *E. coli* RNAP with an R529C mutation (9). We could speculate that although the *rpoB*531 mutants studied here led to increased signal in the minimum scaffold assay, most likely due to increased leniency towards the open complex formation (RP_O), this might not correlate to increased transcript *in vivo*. Results from the minimum scaffold assay are based on the addition of a single radiolabelled nucleotide to an RNA template. Mutants which display higher activity in this assay may thus also display a higher level of abortive transcription, which would relate to an overall decrease in successful transcription elongation *in vivo*. This hypothesis is supported by previous work from our group which indicated that *M. tuberculosis* with the *rpoB*531 mutation exhibited downregulation of genes which are involved in metabolic and

cellular pathways (5). Additional studies are needed to assess the degree of abortive transcription and how it influences enzyme efficiency at different mycobacterial promoters.

The *in vitro* transcription assay revealed a contrasting set of results. For this assay, the abundance of σ^A in the protein preparation could play a considerable role as it is necessary for recognition of the *sigA* promoter in the DNA template. As a result, the higher ratio of σ^A in the *rpoB*526 mutant could be linked to increased levels of transcription as seen in this assay. However, statistical analysis indicated that the only significant difference was that of RNAP with the *rpoB*526 mutation compared to RNAP with the *rpoB*531 and *rpoC*483 mutations. These results align with findings from a previous study which revealed upregulation of the WhiB7 regulon in an *rpoB*526 mutant (5). Based on the promoter sequence upstream of the gene, it is thought that expression of *whiB7* is transcribed by the housekeeping sigma factor in *M. tuberculosis* and once expressed, can bind to σ^A as a transcriptional activator (10). Expression of *whiB7* has been found to be induced upon exposure to antibiotics and has been linked to intrinsic drug resistance; therefore, upregulation of this gene would be advantageous to the bacterium (11). Furthermore, it is interesting to note that from a clinical perspective, both RNAP with the *rpoB*526 mutation and RNAP with the *rpoB*531 and *rpoC*483 mutations would exhibit rifampicin resistance phenotypically. However, our results reveal that the former exhibits increased levels of transcript at the *sigA* promoter, which could alter the phenotype of the bacterium in aspects beyond drug resistance. We suggest that further studies are needed to determine whether RNAP with the *rpoB*526 mutation exhibits increased affinity for promoters recognized by σ^A.

The low transcript abundance in the reaction with the *rpoB*531/*rpoC*485 double mutant indicates that this *rpoC* mutation severely hampers RNAP function. This may be the reason this particular substitution is not seen as frequently as the *rpoC*483 mutation in clinical isolates of *M. tuberculosis* (1, 2). In fact, a recent study by Li and colleagues which included the use of 332 RIF resistant clinical isolates did not report at mutation at codon 483 of *rpoC* (12). We speculate that the compensatory mechanism evoked by this particular mutation may lie beyond the scope of the assays which were employed in this study.

4.5.2 Fluorescence-based transcription assay

There are numerous drawbacks to the fluorescence-based transcription assay which was tested in this study. The main concern is that the transcription product is only quantified in relative fluorescence units, and cannot be visually inspected. As a result, shorter transcripts, such as those produced during abortive transcription, would contribute to the fluorescence reading. We propose that this may be the reason for the discrepancy seen when comparing these results to that of the radioactivity-based assay. An additional concern is that RiboGreen reagent is not RNA-specific and binds to all nucleic acid, therefore the presence of residual DNA template could erroneously amplify the fluorescence reading in this assay. Furthermore, the technique is very sensitive to pipetting error due to the additional DNase treatment step and readings from triplicate reactions varied noticeably among samples. This substantiates the continued use of radioactivity-based assays as the gold-standard for investigating transcription *in vitro*.

4.5.3 Limitations

One of the limitations of the current study is the absence of the ω subunit in the purified RNAP preparations. The role of this small and seemingly insignificant subunit in *M. tuberculosis* is not yet fully understood, however recent work indicates that it plays a role in enzyme stability and transcriptional specificity in *Staphylococcus aureus* (13). Studies have found that the presence of the ω subunit does not play a role in yield or activity of RNAP during *in vitro* transcription, however this may not be the case for mutated RNAP, where the structural stability of the enzyme is most likely altered (7, 14). Similarly, the absence of RNAP-binding factors such as CarD and RbpA in transcription reactions could lead to results which do not accurately represent how the enzyme functions *in vivo*. A study by Srivastava *et al.* suggests that CarD is a component of the transcriptional machinery necessary during transcription initiation and highlighted its importance during open complex (RP_O) formation in *M. tuberculosis* (15). Recent findings from Davis *et al.* revealed weak activity of mycobacterial RNAP at the *rrnA*-P3 promoter in the absence of CarD (16). The authors reported that mycobacterial RP_O complexes are extremely unstable compared to *E. coli* RNAP, and that CarD is responsible for stabilising mycobacterial RPo by preventing collapse of the transcription bubble. Likewise, RbpA is an essential transcription factor in *M. tuberculosis*, which has been found to stabilize RNAP-promoter complexes and stimulate transcription from σ^A-dependant promoters (17). It is possible that the affinity of RNAP for these factors may be altered in the event of *rpoB* or *rpoC* mutations, a finding which could be investigated in future studies.

Lastly, as differing ratios of RNAP subunits were seen across wildtype and mutant preparations, accurately measuring the concentration of RNAP holoenzyme was a major constraint. Furthermore, normalising the signal from transcription assays using the SDS-PAGE results would not correlate to intact RNAP holoenzyme, neither would it give an indication of the number of transcriptionally active complexes. There are also multiple technical considerations which could pose a problem, such as signal saturation, the dynamic range and sensitivity of the stain as well as inconsistencies in signal strength for different subunits, especially across more than one gel. Our results indicate that *rpoB* and *rpoC* mutations alter intra-complex affinity, a limitation which would be difficult to overcome unless subunits are prepared individually and reconstituted *in vitro* using equivalent ratios. In this case, holoenzyme formation would still be dependent on subunit affinity, however ratios of subunits would not be further influenced by subjection to affinity chromatography. We suggest that future studies involving the use of different forms of mutated RNAP employ an accurate method of normalising *in vitro* assay results to the amount of transcriptionally-active RNAP holoenzyme.

4.5.4 Future studies

As no evidence for the compensatory effect of *rpoC* mutations was seen in the assays used in this study, alternative aspects of RNAP structure and function could be investigated to better understand their role in bacterial fitness. Future studies could include the use of *in vivo* pull-downs to ascertain whether different RNAP-binding factors are co-purified in the presence of different *rpoB* and *rpoC* mutations. The use of *in silico* modelling of RNAP would also greatly support our understanding of the holoenzyme stability when mutations are present. In aid of this, Lin *at al* recently solved a 3.91 Å-resolution crystal structure of the transcription initiation complex of *M. tuberculosis* (18).

As mentioned previously, the absence of CarD and RbpA in the *in vitro* transcription assays were a limitation in the current study, therefore future work should incorporate the use of these RNAP-binding factors, as well as different promoter templates from *M. tuberculosis*. Furthermore, the addition of rifampicin (RIF) to the transcription assays may led to interesting results as it would give an indication of how this anti-tuberculosis drug influences transcriptional activity. Our study aimed to investigate the inhibitory effect of RIF on RNAP in the fluorescence-based assay described here, however it was found that the absorption spectrum of RIF overlaps with the emission spectrum of RiboGreen reagent, therefore the signal was quenched and accurate results could not be obtained (results not shown). Similar to our study,

Rodrigues *et al.* reported fluorescence resonance energy transfer between fluorescent probes and RIF (19). For this reason, it is recommended that radioactivity-based assays be used for future studies involving the use of this compound.

Another alternative to investigating transcriptional efficiency *in vivo* is the approach taken by Qi *et al.*, where a susceptible strain of *Pseudomonas aeruginosa* containing an integrated luciferase reporter construct was used to generate rifampicin resistant mutants (20). Luciferase reporter activity was then used to compare the effects of different *rpoB* mutations on RNAP activity, as well as the restorative effect of secondary compensatory mutations in *rpoB*. Although individual aspects of transcription, such as initiation and elongation, cannot be investigated using this approach, it circumvents the need for purified RNAP and gives a good representation of transcriptional activity within the bacterium as the necessary transcription factors are present within the cell. Furthermore, previous work has shown that bioluminescent reporters have been successfully used in *M. tuberculosis*, making this a worthwhile option for future studies (21).

4.6 Conclusion

Our results indicate that the *rpoB* S531L mutation in RNA polymerase (RNAP) may affect specific activity of the enzyme in a promoter-independent assay, yet this did not correlate to increased transcript from a mycobacterial promoter. We speculate that the reason for this may be linked to an increased level of abortive transcription in rifampicin resistant *M. tuberculosis* harbouring this mutation, however additional studies will be needed to confirm this hypothesis. Nonetheless, the techniques and assays described here provide a good basis for future work using mutated mycobacterial RNAP since we have been able to highlight and discuss limitations and potential problems which could arise. Furthermore, clinically relevant mutations studied here appear to play a role in the relationship between the subunits which make up the RNAP holoenzyme. Although not studied here, our results can be extrapolated to hypothesise that the enzyme's affinity for RNAP-binding proteins such as CarD and RbpA could be altered in a similar fashion. These factors play a critical role in the structure and function of RNAP and could be linked to the global changes in gene expression seen in rifampicin resistant clinical isolates of *M. tuberculosis*.

4.7 References

1. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. 2012. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 44:106–110.
2. De Vos M, Müller B, Borrell S, Black PA, van Helden PD, 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.
3. Brandis G, Wrande M, Liljas L, Hughes D. 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. Mol Microbiol 85:142–151.
4. Song T, Park Y, Shamputa IC, Seo S, Lee SY, Jeon H-S, Choi H, Lee M, Glynne RJ, Barnes SW, Walker JR, Batalov S, Yusim K, Feng S, Tung C-S, Theiler J, Via LE, Boshoff HIM, Murakami KS, Korber B, Barry CE, Cho S-N. 2014. Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the β′ subunit of RNA polymerase. Mol Microbiol 91:1106–1119.
5. Du Plessis J. 2014. Deciphering the impact of *rpoB* mutations on the gene expression profile of *Mycobacterium tuberculosis*. Thesis: Stellenbosch University.
6. 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.
7. Banerjee R, Rudra P, Prajapati RK, Sengupta S, Mukhopadhyay J. 2014. Optimization of recombinant *Mycobacterium tuberculosis* RNA polymerase expression and purification. Tuberculosis 94:397–404.
8. Kulbachinskiy A, Feklistov A, Krasheninnikov I, Goldfarb A, Nikiforov V. 2004. Aptamers to *Escherichia coli* core RNA polymerase that sense its interaction with rifampicin, σ-subunit and GreB. Eur J Biochem 271:4921–4931.
9. Jin DJ, Turnbough Jr CL. 1994. An *Escherichia coli* RNAP defective in transcription due to its overproduction of abortive initiation products. J Mol Biol 236:72–80.
10. Burian J, Ramón-García S, Howes CG, Thompson CJ. 2012. WhiB7, a transcriptional activator that coordinates physiology with intrinsic drug resistance in *Mycobacterium tuberculosis*. Expert Rev Anti Infect Ther 10:1037–1047.

11. 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.
12. Li Q-J, Jiao W-W, Yin Q-Q, Xu F, Li J-Q, Sun L, Xiao J, Li Y-J, Mokrousov I, Huang H-R, Shen A-D. 2016. Compensatory mutations of rifampin resistance are associated with transmission of multidrug-resistant *Mycobacterium tuberculosis* Beijing genotype strains in China. Antimicrob Agents Chemother 60:2807–2812.
13. Weiss A, Moore BD, Tremblay MHJ, Chaput D, Kremer A, Shaw LN. 2017. The ω subunit governs RNA polymerase stability and transcriptional specificity in *Staphylococcus aureus*. J Bacteriol 199:e00459-16.
14. Jacques J-F, Rodrigue S, Brzezinski R, Gaudreau L. 2006. A recombinant *Mycobacterium tuberculosis in vitro* transcription system. FEMS Microbiol Lett 255:140–147.
15. Srivastava DB, Leon K, Osmundson J, Garner AL, Weiss LA, Westblade LF, Glickman MS, Landick R, Darst SA, Stallings CL, Campbell EA. 2013. Structure and function of CarD, an essential mycobacterial transcription factor. Proc Natl Acad Sci U S A 110:12619.
16. Davis E, Chen J, Leon K, Darst SA, Campbell EA. 2015. Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. Nucleic Acids Res 43:433–445.
17. Hu Y, Morichaud Z, Chen S, Leonetti J-P, Brodolin K. 2012. *Mycobacterium tuberculosis* RbpA protein is a new type of transcriptional activator that stabilizes the σ^A-containing RNA polymerase holoenzyme. Nucleic Acids Res 40:6547–6557.
18. Lin W, Mandal S, Degen D, Liu Y, Ebright YW, Li S, Feng Y, Zhang Y, Mandal S, Jiang Y, Liu S, Gigliotti M, Talaue M, Connell N, Das K, Arnold E, Ebright RH. 2017. Structural basis of *Mycobacterium tuberculosis* transcription and transcription inhibition. Mol Cell 66:169–179.e8.
19. Rodrigues C, Gameiro P, Prieto M, de Castro B. 2003. Interaction of rifampicin and isoniazid with large unilamellar liposomes: spectroscopic location studies. Biochim Biophys Acta BBA - Gen Subj 1620:151–159.
20. Qi Q, Preston GM, MacLean RC. 2014. Linking system-wide impacts of RNA polymerase mutations to the fitness cost of rifampin resistance in *Pseudomonas aeruginosa*. mBio 5:e01562-14.
21. Andreu N, Zelmer A, Fletcher T, Elkington PT, Ward TH, Ripoll J, Parish T, Bancroft GJ, Schaible U, Robertson BD, Wiles S. 2010. Optimisation of bioluminescent reporters for use with mycobacteria. PloS One 5:e10777.

Chapter 5

Study 3

Gaining a better understanding of D-cycloserine resistance in *Mycobacterium tuberculosis* through next-generation technologies

Du Plessis J[1], De Vos M[1], Pethe K[2], Ab-Rahman NB[2], Burns S[3], Posey JE[3], Theron D[4], van Helden PD[1], Warren RM[1], Sampson SL[1]

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

[2]Lee Kong Chian School of Medicine and School of Biological Sciences, Nanyang Technological University, 59 Nanyang Drive, Singapore 636 921, Singapore.

[3]Laboratory Branch, Division of Tuberculosis Elimination, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia 30329, United States.

[4]Brewelskloof Hospital, Department of Health, Province of Western Cape, South Africa.

The following chapter consists of a manuscript which is intended for publication in a peer reviewed journal. My contribution to this chapter, as well as the addendum which follows directly thereafter, includes: Literature search
Study design
Experimental work
Writing and editing of manuscript
Interpretation of data and results
Analysis and interpretation of commercially generated data and data provided by co-authors (genomic data generated by SB and JEP, susceptibility data provided by KP and NBA).

5.1 Abstract

This study focuses on understanding the mechanism of resistance of *Mycobacterium tuberculosis* to D-cycloserine, a poorly understood second-line drug currently used for the treatment of multidrug-resistant tuberculosis. Our analysis includes the use of a unique set of serial clinical isolates, which were collected from a patient over a period of 21 months. The isolates underwent routine drug susceptibly testing at the time of sample collection and were further analysed by whole genome sequencing and RNAseq for the purpose of this study. Whole genome sequencing revealed that the isolates had multiple mutations conferring resistance to anti-tuberculosis drugs. Of interest to this study, the strain acquired mutations in *ald*, a gene that has been associated with resistance to D-cycloserine. Transcriptomic analysis of this isolate revealed upregulation of *Rv0576* and *Rv0577*; the latter is a putative glyoxalase which could be involved in metabolic detoxification. *Rv0576* and *Rv0577* were cloned and transformed into the reference strain *M. tuberculosis* H37Rv and a clinical isolate of *M. tuberculosis* to determine whether over-expression of these two genes could be linked to the mechanism of D-cycloserine resistance. We assessed the susceptibility of parental and transformed strains to a panel of anti-tuberculosis drugs including D-cycloserine and demonstrated that overexpression of *Rv0577* led to an increase in the minimum inhibitory concentration of D-cycloserine. These findings provide novel insight into the mechanism of D-cycloserine resistance in *M. tuberculosis*.

5.2 Introduction

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB) is a global threat to human health. The World Health Organization estimated that there were 10.4 million new cases of active TB disease in 2015, with the African Region displaying the most acute burden relative to the population: 275 cases per 100,000 people (1). Treatment of drug resistant TB raises further concern, as it requires the use of highly toxic drugs and has poor outcomes, with only 52% of patients with multidrug-resistant (MDR) TB being treated successfully (1). Despite considerable progress in TB research and control in recent years, a key factor which hinders treatment of drug resistant infections is the lack of knowledge concerning a number of drug-resistance mechanisms. The current study focuses on D-cycloserine (DCS), a second-line drug with severe side-effects, currently used to treat MDR-TB. To date, drug susceptibility testing (DST) for DCS has not been standardized, the breakpoint concentration which defines resistance is yet to be described, the mechanism of drug resistance is poorly understood, and the effect of resistance-conferring mutations on the physiology of mycobacteria remains largely unknown. Furthermore, the frequency of DCS resistance among drug-resistant clinical isolates of *M. tuberculosis* has not been determined.

Genes that have previously been linked to DCS resistance in mycobacteria include *Rv3423c* (*alr*) and *Rv2981c* (*ddlA*), both of which play a key role in peptidoglycan synthesis (Figure 5.1) (2–4). DCS resistance in *M. tuberculosis* has been linked to an S22L polymorphism in the *alr* gene (5), as well as to overexpression of *alr* and *ddl* in *M. smegmatis* (2, 3). DCS, a cyclic analogue of D-alanine, has been found to competitively inhibit both alanine racemase and D-alanine:D-alanine ligase, which has led to much ambiguity over which of these enzymes is the target of DCS in *M. tuberculosis* (2–4, 6–8). Recently, however, Prosser and Carvalho used stable isotope metabolomic profiling to confirm that DdlA is the *in vivo* lethal target of DCS, a finding that was further confirmed by Halouska *et al.* using NMR metabolomics (9, 10).

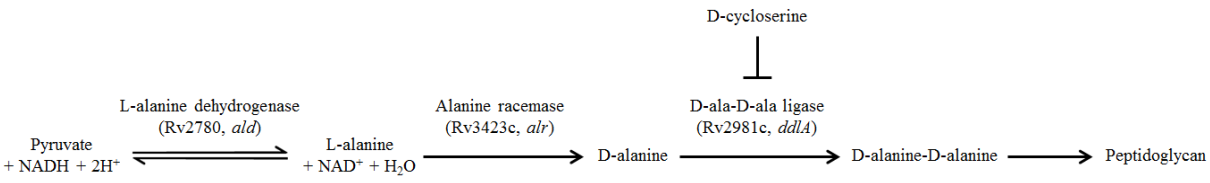


Figure 5.1 Ald and the biosynthesis of peptidoglycan in *M. tuberculosis*. D-cycloserine is a cyclic analogue of D-alanine and competitively inhibits D-alanine:D-alanine ligase *in vivo* (9).

Besides *alr* and *ddlA*, Desjardins *et al.* ingeniously revealed a link between DCS resistance and *Rv2780* (*ald*), which encodes L-alanine dehydrogenase (11). The authors reported multiple loss-of-function mutations in the gene, which led to an increase in the minimum inhibitory concentration (MIC) of clinical isolates of *M. tuberculosis* to DCS. Ald is a 40 kDa multifunctional enzyme, which catalyses the reductive amination of pyruvate to L-alanine in the peptidoglycan biosynthesis pathway, as well as the reverse reaction, oxidising L-alanine to pyruvate (12). Furthermore, it was recently revealed that Ald exhibits glyoxylate aminase activity, a function that converts glyoxylate to glycine (13). Further investigation confirmed that Ald is involved in redox homeostasis during recovery from hypoxia, by maintaining optimal ratios of NADH/NAD (14).

The mechanism by which *ald* mutations lead to DCS resistance remains elusive, however Desjardins *et al.* hypothesised that strains harbouring loss-of-function mutations in *ald* would be unable to oxidise L-alanine to pyruvate. As a result, the pool of available L-alanine would increase, thereby overriding the effect of DCS on Alr and/or DdlA, and driving continued biosynthesis of peptidoglycan. In the current study, we exploited a unique collection of serial clinical isolates from a patient with MDR-TB who additionally developed resistance to terizidone (a bi-cyclic derivative of DCS) to investigate the link between *ald* mutations and the mechanism of drug resistance. The set of six serial clinical isolates, obtained from June 2005 to April 2007, were characterized by drug susceptibility testing, whole genome sequencing (WGS) and RNAseq to reveal a novel link between *ald* mutations and the mechanism of DCS resistance in *M. tuberculosis.*

5.3 Methods

5.3.1 Sample selection

Six serial clinical isolates were selected from a reference database linked to a culture bank of drug resistant *M. tuberculosis* isolates, maintained at Stellenbosch University, South Africa. The selected isolates were characterized by IS*6110*-based restriction fragment length polymorphism (RFLP), which confirmed that they belong to the X-family (15, 16). Furthermore, routine drug susceptibility testing (DST) results for these clinical samples were collected from the National Health Laboratory Service (NHLS) in South Africa.

5.3.2 Collection of clinical data

A folder review was done at the hospital where the patient was treated to determine which drugs were used during TB therapy, thereby giving an indication of the selective pressure the strain encountered during this timeframe. The necessary ethical approval was obtained from the Health Research Ethics Committee at Stellenbosch University (Ethics protocol number N10/04/126) and all patient identifiers were removed to safeguard patient confidentiality.

5.3.3 Bacterial culture

Clinical isolates were cultured on Löwenstein-Jensen slants in a Biosafety Level 3 (BSL-3) laboratory for 4 to 6 weeks. Subsequently, a scrape of the culture was inoculated into 5 mL of Middlebrook 7H9 broth (Becton, Dickinson and Company) enriched with 10% albumin dextrose catalase (ADC), 0.05% Tween 80 (Merck) and 0.2% glycerol. 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 ($OD_{600} \approx 0.8$) was reached. Cultures were inspected for contamination by inoculating blood agar plates and through Ziehl-Neelsen differential staining. Glycerol stocks of these cultures were stored at -80°C in 1 mL aliquots for downstream experiments. All six samples were subjected to whole genome sequencing, whereas Isolates 3 and 4 were selected for RNAseq analysis to study the effect of *ald* mutations on the gene expression profile of the strain.

5.3.4 Whole genome sequencing

Each isolate was cultured on OADC-enriched Middlebrook 7H10 media and DNA was extracted as previously described for the purpose of whole genome sequencing (WGS) (17). DNA quality and quantity were determined spectrophotometrically using the NanoDrop-1000 (Thermo Scientific). DNA was normalized so that total input for library preparation was 1.0-

1.5 µg per sample. Samples were fragmented to an average length of 600 bp using a focused-ultrasonicator (Covaris LE220). Following fragmentation, samples underwent size selection via 0.6X Ampure bead purification. Library preparation was completed, per the manufacturer's recommendation using the NEBNext Ultra DNA library prep kit for Illumina (New England BioLabs) with dual indices. Samples were pooled to a final loading concentration of 1 pM. Sequencing was performed on a NextSeq 550 using the 500/550 mid-output v2 (300 cycle) kit.

FASTQ files for each of the genomes were generated and FastQC was used to assess the quality of the raw sequence reads (18). Thereafter, an in-house automated pipeline was used to map the raw sequence reads to the *M. tuberculosis* H37Rv reference genome (Genbank: AL123456) and identify single nucleotide polymorphisms and short in/dels as previously described (19). Variants in *pe/ppe* gene families and other repetitive regions were not included in the analysis. To allow for the identification of underlying sub-populations, variants were not filtered for a specific frequency. Subsequently, the effect of novel variants linked to drug resistance in *M. tuberculosis* was assessed using PROVEAN (Protein Variation Effect Analyzer) (20).

5.3.5 RNA isolation and transcriptomic analysis

Isolates 3 and 4 were selected for RNAseq analysis to study the effect of *ald* mutations on the gene expression profile of the strain. Firstly, the growth characteristics of each isolate were determined by culture in ADC-enriched Middlebrook 7H9 broth, inoculated with a primary culture to an OD_{600} of 0.1. Thereafter, 200 µL of the culture was added in quadruplicate to a flat bottom 96-well plate and incubated at 37°C for a period of 25 days. Growth was monitored by measuring the optical density of the cultures using the POLARstar Omega plate reader (BMG Labtech).

For the isolation of RNA, 50 mL cultures were inoculated in biological duplicate with a primary culture to an OD_{600} of 0.1 and incubated at 37°C until mid-log phase was reached. RNA isolation was done using the FastRNA® Blue kit (MP Biomedicals LLC) according to the protocol provided by the manufacturer. Thereafter, samples were treated with RQ1 DNase (Promega) to remove residual genomic DNA (gDNA) and further purified by phase separation with phenol:chloroform. RNA quality and quantity were determining using the Experion RNA StdSens analysis kit and automated electrophoresis station (Bio-Rad Laboratories Inc.) according to the manufacturer's instructions. Furthermore, samples were tested for gDNA

contamination by polymerase chain reaction (PCR) amplification of the *katG* gene with published primer set RTB59/RTB38 using HotStarTaq DNA polymerase (Qiagen) (21).

RNA samples were sequenced at the Agricultural Research Centre (ARC) in South Africa, using the HiSeq 2500 sequencing system (Illumina). FastQC was used to assess the quality of the raw reads (18), thereafter Chipster was used for analysis of the sequencing data (22). In summary, sequencing adapters were trimmed using Trimmomatic (23) before aligning reads to the reference genome *M. tuberculosis* H37Rv using TopHat2 (24). Reads were counted using HTSeq (25) and differential expression was determined using EdgeR (26). Furthermore, a linear correlation test was done to confirm satisfactory R^2 values between sample replicates. Following Benjamini Hochberg multiple testing correction, a p-value of <0.05 was used to determine statistical significance.

5.3.6 Construction of plasmids

Two genes which were found to be differentially expressed in the isolates used for RNAseq analysis, namely *Rv0577* and *Rv0576*, were included in downstream analysis. The gene sequence of *Rv0577* was cloned with and without *Rv0576* into the episomal vector pMV262 under control of the constitutive *hsp60* promoter to generate pMV262_*Rv0576-Rv0577* and pMV262_*Rv0577*, respectively. Both constructs were transformed into the reference strain H37Rv and a pan-susceptible clinical isolate of *M. tuberculosis*, N0145, to determine their role in D-cycloserine (DCS) resistance.

5.3.7 Drug susceptibility testing

A turbidity-based microdilution assay was used to determine inhibitory concentrations, which was defined as the lowest concentration of drug which allowed for 50% growth inhibition (MIC_{50}). Of note, MIC_{50} was only determined when a complete typical sigmoidal curve was obtained. Moxifloxacin (MOX) and isoniazid (INH) were used as reference compounds; PA-824 and bedaquiline (BDQ) were included to determine whether the transformed strains would display cross-resistance to next-generation drugs, and ethambutol (EMB) was included as a less potent control targeting the cell-wall. For this experiment, DCS was prepared as a stock of 200 mM in dH_20, whereas all other drugs were dissolved in 90% DMSO to a concentration of 20 mM. Two-fold serial dilutions of the drugs were added to 96-well flat bottom plates. Filter-sterilised albumin dextrose saline (ADS) enriched Middlebrook 7H9 broth without glycerol was used to culture the transformed strains of *M. tuberculosis*. Inocula were then

prepared by adjusting the bacterial density to an OD_{600} of 0.005 with a final volume of 200 µL per well. Cultures were incubated for 5 days at 37°C and optical density readings were taken using the BioTeK CYTATION 3 multimode reader. The experiment was performed twice, with each concentration tested in triplicate. MIC_{50} values were calculated using GraphPad Prism 6.

5.4 Results

5.4.1 Mapping clinical data and whole genome sequencing data

Six serial clinical isolates, collected from a patient over a period of 21 months, were subjected to whole genome sequencing (WGS) with a mean depth coverage of 113 to identify single nucleotide variants and small insertions and deletions which were acquired by the strain while the patient was undergoing tuberculosis (TB) treatment. To gain a better understanding of the selective pressure that the bacterium experienced during this timeframe, clinical data was collected from the hospital where the patient received treatment. Figure 5.2 illustrates the anti-TB drugs which were administered, results from drug susceptibility testing (DST), and which sequence variants were identified during the patient's treatment.

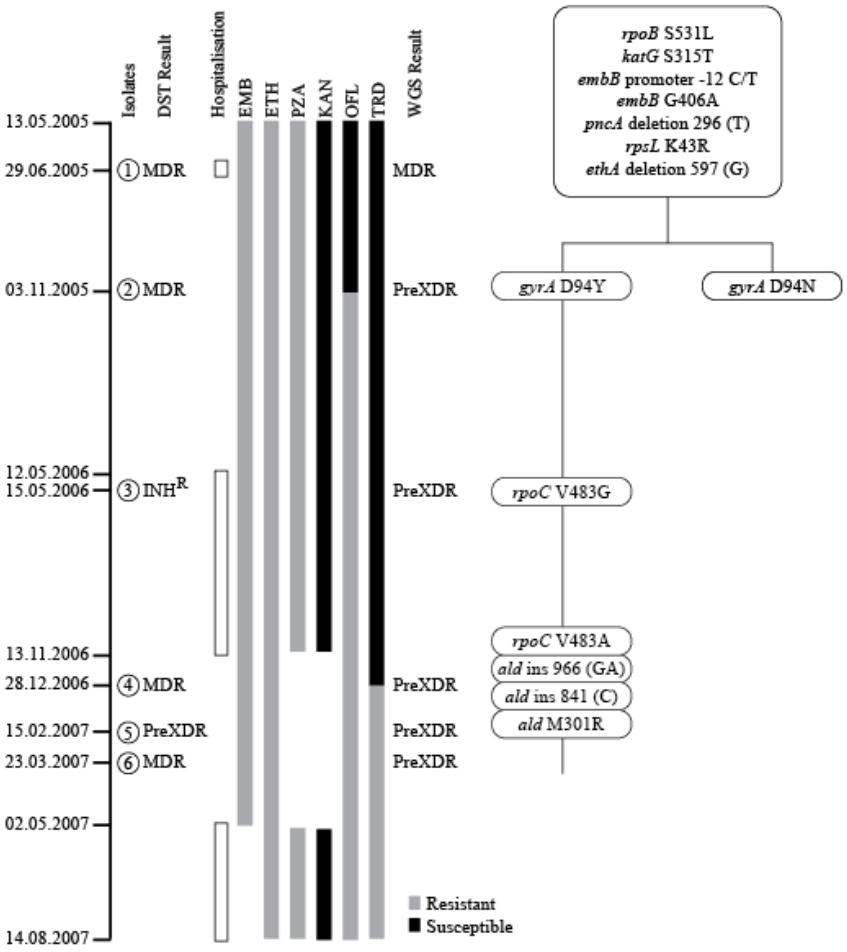


Figure 5.2 Patient timeline. Chronology of isolate collection illustrating routine drug susceptibility testing (DST) results, patient treatment information, acquired genomic variants linked to drug resistance and overall drug susceptibility determined by whole genome sequencing (WGS). Drug abbreviations: ethambutol (EMB), ethionamide (ETH), pyrazinamide (PZA), kanamycin (KAN), ofloxacin (OFL), terizidone (TRD).

The patient was diagnosed with TB in May of 2005, and was admitted to hospital once DST results indicated that it was a case of MDR-TB. At that time, routine DST included testing for rifampicin (RIF) and isoniazid (INH) resistance only, thus the patient was initiated on a standardised MDR-TB regimen consisting of a combination of ethambutol (EMB), ethionamide (ETH), pyrazinamide (PZA), kanamycin (KAN), ofloxacin (OFL) and terizidone (TRD), which is a derivative of D-cycloserine (DCS). Retrospective WGS analysis of Isolate 1 identified the following mutations to be fixed in the population (>95% of the bacilli harboured these mutations): *rpoB* S531L, *katG* S315T, *embB* promoter -12C/T, *embB* G406A, *rpsL* K43R, *pncA* deletion 296 (T) and *ethA* deletion 597 (G). In summary, in addition to resistance to RIF and INH, these mutations confer resistance to three of the administered second-line drugs (EMB, ETH, PZA). WGS of Isolate 2, collected four months later, revealed the acquisition of two *gyrA* mutations, D94Y and D94N, at a variant frequency of 55% and 45%, respectively. Once these *gyrA* mutations were acquired, conferring resistance to OFL, the only remaining effective drugs in the treatment regimen were KAN and TRD.

WGS of Isolate 3, collected in mid-2006, showed the loss of the D94N *gyrA* mutation and acquisition of a V483G substitution in *rpoC*, a compensatory mutation which ameliorates the fitness cost associated with *rpoB* mutations (27, 28). By the end of 2006, treatment with KAN and PZA were stopped, most likely due to adverse side effects, and the only remaining effective drug in the treatment regimen was terizidone. Not surprisingly, the strain acquired resistance to this drug within a very short period of time by acquiring a mutation and two insertions in *ald* (M301R, 1bp insertion (C) at position 841, 2bp insertion (GA) at position 966), a gene which has previously been linked to DCS resistance. WGS of Isolate 4 also revealed the presence of a second polymorphism at the same position of *rpoC* (V483A) at a variant frequency of 26%, while the frequency of the initial *rpoC* V483G polymorphism decreased from >99% to 74%.

By February of 2007, the panel of drugs used for routine DST in South Africa was expanded to include OFL, EMB, ETH and amikacin. The isolate was found to be resistant to the former two drugs and was classified as pre-XDR, an unofficial term used to describe cases of TB which are resistant to INH, RIF and either a fluoroquinolone or a second-line injectable drug (29). Once the DST results were available, treatment with EMB was replaced with KAN and PZA once more, but unfortunately, despite another extended period of hospitalisation, the patient succumbed to the disease shortly thereafter. Sequencing of the final samples, Isolate 5 and Isolate 6, revealed that all but one of the resistance-associated variants that were acquired during

this timeframe remained present to varying degrees within the bacterial population, with *gyrA* D94N being the only mutation that was lost (Table 5.1).

Of interest in this study, was the acquisitioning of mutations in *ald* due to inadvertent monotherapy with terizidone. The 2bp insertion at position 966 has previously been described as a loss-of-function mutation by Desjardin *et al*. Furthermore, the *ald* M301R substitution was found to be deleterious when the functional effect of this mutation was assessed using PROVEAN (Protein Variation Effect Analyzer) (30). Similarly, the 1bp insertion at position 841 of *ald* is thought to be deleterious as it causes a frameshift mutation at codon 281, which leads to the alteration of the stop codon.

Table 5.1 Frequency of variants (%) identified by WGS analysis.

Variant	Isolate					
	1	**2**	**3**	**4**	**5**	**6**
pckA 680 GGC/GAC	54	0	0	0	0	0
[a]*pyrG* 191 TAC/TCC		44	0	0	0	0
[a]2954957 G/A		16	0	0	0	0
[a]*Rv2681* 660 A del		22	0	0	0	0
[a]*leuC* 110 ATC/GTC		35	0	0	0	0
[a]*glpk* 368 CTG/CCG		31	0	0	0	0
[b]*Rv3747* 117 CTG/CCG		46	100	100	100	100
[b]4243221 C/T		55	100	100	100	100
[b]*mycp2* 105 G ins		55	100	100	100	100
gyrA 94 GAC/AAC		45	0	0	0	0
gyrA 94 GAC/TAC		55	100	100	100	100
rpoC 483 GTG/GGG			*100	74	48	55
rpoC 483 GTG/GCG				26	51	45
ald 841 C ins				20	33	34
ald 301 ATG/AGG				60	44	52
ald 966 GA ins				10	0	5
cspA 22 GAC/GAG					15	23

[a]Probable hitchhiking variants associated with *gyrA* 94GAC/AAC
[b]Probable hitchhiking variants associated with *gyrA* 94GAC/TAC

5.4.2 Transcriptomic analysis of clinical isolates

Isolate 3 (with wildtype *ald*) and Isolate 4 (harbouring mutations in *ald*) were used for transcriptomic analysis to determine the effect of *ald* polymorphisms on the gene expression profile of *M. tuberculosis*. Determination of the growth profiles for these two samples confirmed no difference in growth (Figure 5.3). Differential gene expression analysis using EdgeR revealed that the isolate with mutations in *ald* exhibited a 3.11 and 2.64-fold upregulation of *Rv0576* and *Rv0577*, respectively (Table 5.2). *Rv0577* encodes a protein annotated as Cfp32 on Tuberculist and *Rv0576* is thought to be a transcriptional regulator which lies directly upstream (31). Additionally, *ald* was upregulated 1.37-fold in the strain containing variants in this gene. Upregulation of *ald* can be expected if loss-of-function variants in the protein signal for increased transcription of the gene through a regulatory feedback mechanism. *Rv2526* (VapB17) was downregulated 3.07–fold in Isolate 4, and two other genes, *Rv3508* (PE-PGRS54) and *Rv3512* (PE-PGRS56), were downregulated 1.64 and 2.47-fold, respectively.

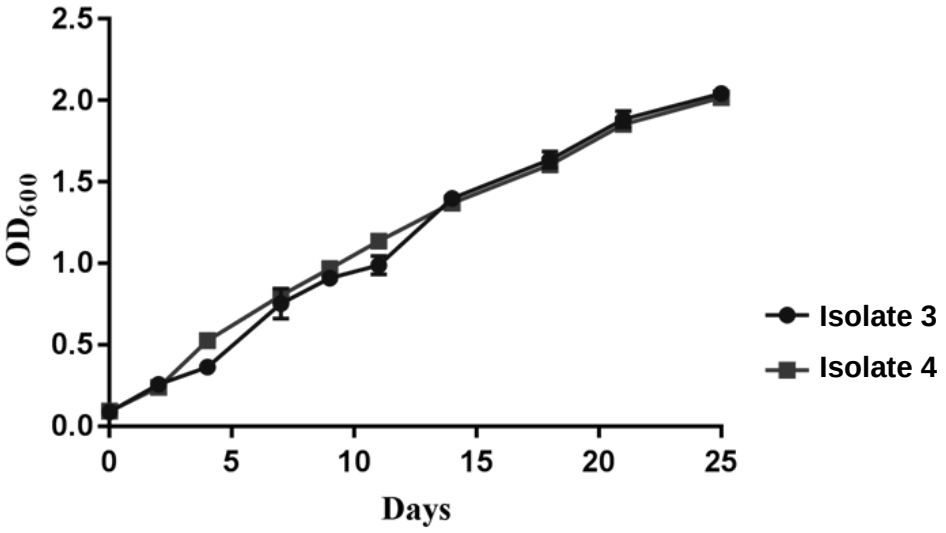


Figure 5.3 Growth profiles. No difference in growth was observed between clinical isolates selected for RNAseq. Absorbance values represent the mean of technical quadruplicates with error bars depicting standard deviation.

Table 5.2 Differential expression of six genes in Isolate 4, harbouring two insertions and a point mutation in *ald*, relative to Isolate 3, lacking *ald* mutations.

Gene	Gene Product	Fold Change	Adjusted p-value
Rv0576	Transcriptional regulator	3.11 up	2.90E-43
Rv0577	Conserved protein (Cfp32)	2.64 up	1.38E-41
Rv2526	Possible antitoxin VapB17	3.07 down	0.006986
Rv2780	L-alanine dehydrogenase, Ald (TB43)	1.37 up	0.018802
Rv3508	PE-PGRS54	1.64 down	0.014622
Rv3512	PE-PGRS56	2.47 down	7.84E-05

5.4.3 Overexpression of Rv0576 and Rv0577

Since upregulation of *Rv0576* and *Rv0577* was observed in Isolate 4 which harboured *ald* mutations, further experiments were performed to determine whether overexpression of these genes could be linked to DCS resistance in *M. tuberculosis*. Two over-expression constructs were prepared (one with only *Rv0577* and one with both *Rv0576* and *Rv0577*) and transformed into the reference strain *M. tuberculosis* H37Rv, and the clinical isolate N0145 (32). Drug susceptibility of the resulting strains was determined using an established turbidity based assay (33, 34). This revealed that overexpression of *Rv0577* led to a 2-fold increase in the MIC_{50} of *M. tuberculosis* to DCS, whereas no effect was seen for any of the other drugs included in the assay (Figure 5.4). Overexpression of *Rv0576* did not appear to contribute towards the level of DCS-resistance in either the clinical isolate or reference strain when co-expressed with *Rv0577*.

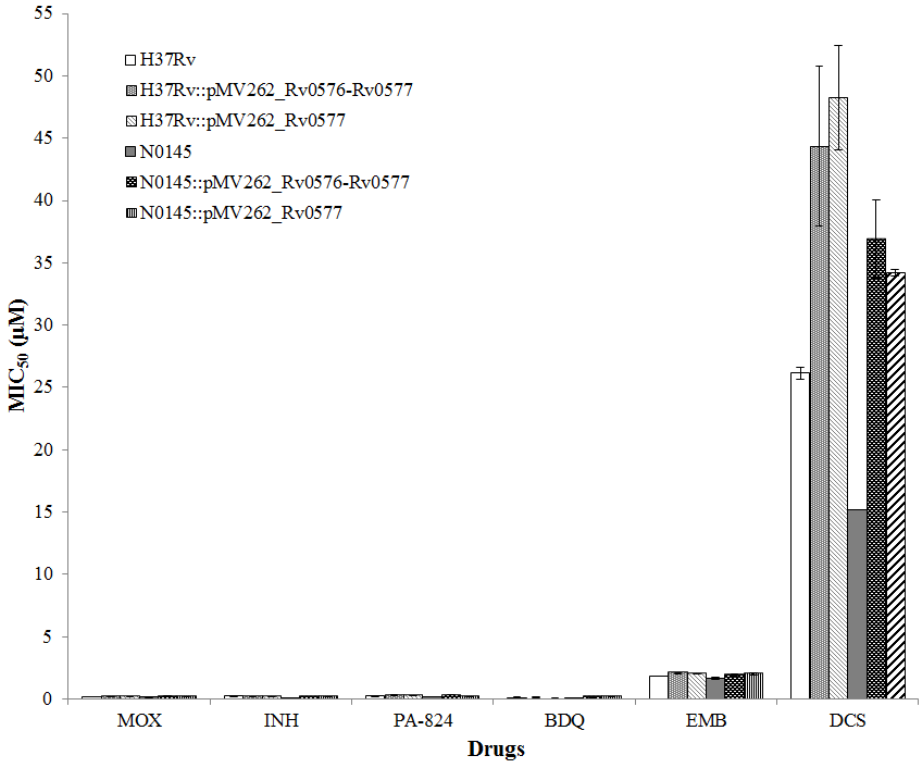


Figure 5.4 MIC_{50} values for transformed strains of *M. tuberculosis* against six anti-TB drugs. Drugs included moxifloxacin (MOX), isoniazid (INH), PA-824 and bedaquiline (BDQ), ethambutol (EMB) and D-cycloserine (DCS). Values represent the mean of technical duplicates with error bars depicting standard deviation.

5.5 Discussion

The current study shows the emergence of D-cycloserine (DCS) resistance during treatment of multidrug resistant tuberculosis (MDR-TB). Here we aimed to understand the mechanism of resistance to DCS associated with *ald* mutations by using a set of serial clinical isolates collected from a patient over a period of 2 years. The isolates underwent routine drug susceptibly testing at the time of sample collection and were further analysed by whole genome sequencing (WGS) and RNAseq. Furthermore, differentially expressed genes identified through transcriptomic analysis were investigated using a turbidity based susceptibility assay.

5.5.1 In vivo acquired polymorphisms identified by WGS

Whole genome sequencing of serial clinical isolates from a patient who was inadvertently placed on a weakened MDR treatment regimen of three anti-tuberculosis drugs - kanamycin, ofloxacin and terizidone (a DCS-derivative) - revealed the initial acquisition of ofloxacin resistance, followed by mutations in *rpoC* and *ald*. The latter gene has recently been linked to DCS-resistance (11). Desjardins *et al.* described 11 loss-of-function mutations conferring resistance to DCS, of which one (*ald* 966 ins) was also identified in our study. We also report two additional mutations in *ald*, M301R and a 1bp insertion at position 841, which both appear to be deleterious to protein function.

As much remains to be understood about the function of L-alanine dehydrogenase (Ald) in *M. tuberculosis*, it is unclear how mutations in *ald* influence mycobacterial physiology. Andersen *et al.* reported the presence of Ald in 4-day-old culture filtrate and proposed that it is secreted across the cell membrane to assist in synthesis of the peptidoglycan layer (12). Raynaud *et al.* showed that Ald could not be detected in culture filtrate of non-pathogenic *M. smegmatis*, yet was present in *M. tuberculosis* culture filtrate (35). Accordingly, it could be hypothesised that secreted Ald may fulfil another role in *M. tuberculosis*, perhaps one involved in virulence or pathogenesis. This is supported by the observation that an *ald* knockout of *Pseudomonas aeruginosa* exhibited reduced virulence in a rat lung model of infection (36). Another study found that *ald* is upregulated in response to hypoxia, implicating its involvement in adaptation to anaerobic conditions (13). Further studies investigating the effect of *ald* mutations would greatly benefit our understanding of this protein and its role in mycobacterial physiology.

5.5.2 Transcriptomic effect of ald mutations

Transcriptomic analysis of Isolates 3 and 4 was conducted to understand the effect of *ald* mutations on the gene expression profile of *M. tuberculosis*. Results from the analysis revealed moderate upregulation of *Rv0576* and *Rv0577* in the clinical isolate harbouring multiple variants in *ald*. To further investigate these findings, the genes were cloned into an expression vector and transformed in *M. tuberculosis* H37Rv and a pan-susceptible clinical isolate. A susceptibility assay revealed that overexpression of *Rv0577* led to a 2-fold increase in the MIC_{50} of DCS for both strains, with no effect seen for any of the other drugs tested.

Rv0576 is part of the ArsR subfamily of transcriptional regulators as it possesses a domain which is found in the Arsenical Resistance Operon Repressor protein. This family of repressors is known to facilitate the response to metal ion stress by acting as metal sensors (37). Binding of surplus metal ions to ArsR regulators, modulate the protein's ability to bind to DNA, thereby regulating genes involved in metal efflux and detoxification (38). Despite a 3.11-fold upregulation of *Rv0576* in the clinical isolate with an *ald* mutation, overexpression of *Rv0576* did not directly contribute to DCS-resistance in the subsequent susceptibility assay. Andreu *et al.* reported that *Rv0576* and *Rv0577* are co-transcribed from an upstream promoter and that the product of *Rv0576* is a transcriptional regulator that represses the expression of the operon (39). *Rv0576* contains putative binding sites for Zn^{2+} and a hydrophobic ligand (40). Interestingly, *Rv2034*, another ArsR-like regulator, also binds Zn^{2+}, has an N-terminus which is highly similar to that of *Rv0576* (40.2% identity; 57.3% similarity), and is also capable of auto-regulation (41). Gao *et al.* demonstrated that *Rv2034* is involved in regulating lipid metabolism and hypoxic response by positively regulating the expression of *phoP*, *dosR*, and *groEL2* (41, 42), therefore *Rv0576* could play an equally important role in regulating transcription.

In the current study, we speculate that loss-of-function *ald* mutations may be involved in the level of intracellular Zn^{2+}, of which an excess would release the transcriptional regulator *Rv0576* from its operon. Kim *et al.* reported that Zn^{2+} is an allosteric cofactor of Ald in *Bacillus subtilis* at pH < 9.0, which binds at the same site as its substrate (43). Ald in *M. tuberculosis* shares 51% identity and 67% similarity with the amino acid sequence of Ald in *B. subtilis*, therefore the same allosteric interaction could be proposed for Mycobacteria, linking both *ald* and *Rv0576* to regulation by Zn^{2+}.

The second gene of interest, *Rv0577*, encodes for Cfp32, a 27 kDa antigen which was been identified in the culture filtrate of *M. tuberculosis* (44). The protein consists of two domains which are both part of the vicinal oxygen chelate (VOC) superfamily. The VOC superfamily encompasses a range of structurally related proteins, including glyoxalase I, which is a metalloenzyme involved in the methylglyoxal detoxification pathway, requiring Zn^{2+} or Ni^{2+} ions for activity (45). Several antibiotic resistance proteins are also part of the VOC family and use various mechanisms to inhibit the function of antibiotics. Bleomycin resistance protein is able to inactivate bleomycin by binding to it, whereas the fosfomycin resistance proteins, (FosA, FosB, FosX) are known to open the oxirane ring of fosfomycin, thus rendering the drug ineffective (46).

Desjardins *et al.* hypothesised that loss of function mutations in *ald* led to an increase in the pool of L-alanine by preventing its conversion to pyruvate. As a result, an increase in L-alanine would overcome the competitive inhibition of Alr and DdlA by DCS, thereby reducing the strain's susceptibility to this drug. Here we report that DCS resistance-conferring mutations in Ald led to upregulation of the *Rv0576-Rv0577* operon. As Cfp32, encoded by *Rv0577*, is a putative glyoxalase, this could allow for increased conversion of methylglyoxal into lactic acid, a precursor for pyruvate (33). More importantly however, we propose that Cfp32 possesses the capability of rendering DCS ineffective in a mechanism similar to fosfomycin and bleomycin resistance proteins (Figure 5.5). Further studies are required to investigate the two VOC domains of Cfp32 and confirm whether it is able to cleave the ring structure of DCS.

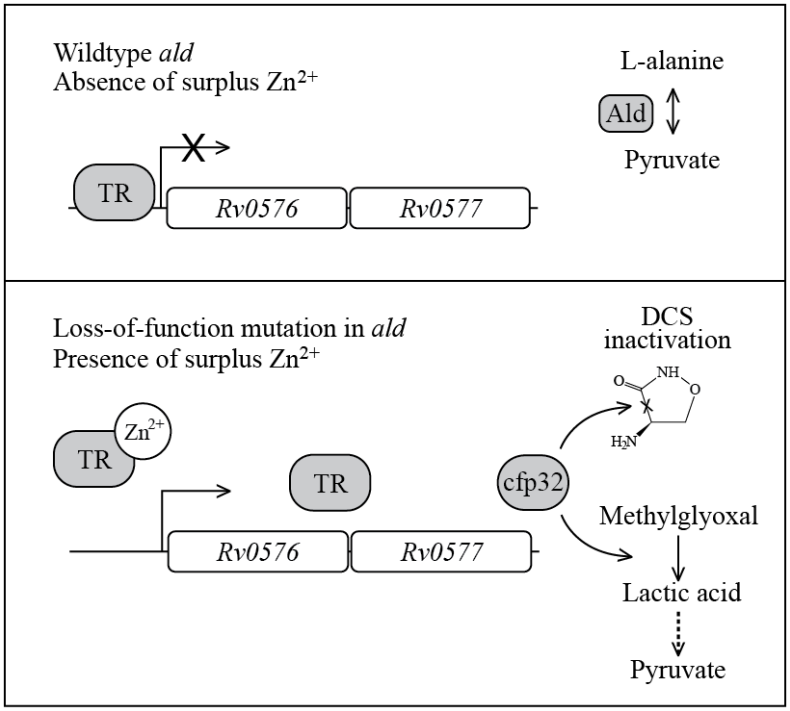


Figure 5.5 Model proposed for the mechanism of D-cycloserine (DCS) resistance associated with loss-of-function *ald* mutations. Expression of *Rv0577* (encoding Cfp32) is induced in response to a surplus of Zn^{2+} ions which binds to the transcriptional repressor (TR) of the operon. Once expressed, Cfp32 is thought to act as an antibiotic resistance protein which cleaves the ring structure of DCS. Furthermore, Cfp32, a putative glyoxalase, could be responsible for an increase in the amount of lactic acid, which is a precursor for pyruvate.

5.6 Conclusion

Here we describe novel *ald* mutations implicated in D-cycloserine (DCS) resistance in *M. tuberculosis*, which will support the development of a molecular diagnostic test for resistance against this second-line anti-tuberculosis drug. Transcriptional profiling of a clinical isolate of *M. tuberculosis* harbouring polymorphisms in *ald* showed upregulation of *Rv0576* and *Rv0577*. Subsequent susceptibility testing of strains overexpressing these genes, revealed a 2-fold increase in the inhibitory concentration of DCS. Our findings show for the first time that overexpression of *Rv0577*, encoding a putative glyoxalase (Cfp32), decreases the strain's susceptibility to DCS and propose that this enzyme is involved in the molecular mechanism of resistance, possibly through increased conversion of methylglyoxal to lactic acid, a precursor for pyruvate. Furthermore, we hypothesise that Cfp32 harbours a domain which cleaves the ring structure of DCS, rendering the drug ineffective. This is the first study to suggest a mechanism whereby Cfp32 could be involved in DCS resistance in *M. tuberculosis*. Additionally, our findings will contribute to further studies on the peptidoglycan biosynthesis pathway as a drug target for the treatment of tuberculosis.

5.7 Acknowledgements

The authors would like to thank the South African National Research Foundation and the Harry Crossley Foundation for funding. Furthermore, research reported in this article was supported by the South African Medical Research Council. SLS is funded by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation (NRF) of South Africa, award number UID 86539. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NRF. This research was supported in part by a Lee Kong Chian School of Medicine, Nanyang Technological University Start-Up Grant to KP. The authors gratefully acknowledge the assistance of Dr KR Jacobson for the retrieval of clinical data, and Mrs T Dolby for drug susceptibility testing data.

5.8 References

1. WHO | Global tuberculosis report 2016. WHO.
2. Cáceres NE, Harris NB, Wellehan JF, Feng Z, Kapur V, Barletta RG. 1997. Overexpression of the D-alanine racemase gene confers resistance to D-cycloserine in *Mycobacterium smegmatis*. J Bacteriol 179:5046–5055.
3. Feng Z, Barletta RG. 2003. Roles of *Mycobacterium smegmatis* D-alanine:D-alanine ligase and D-alanine racemase in the mechanisms of action of and resistance to the peptidoglycan inhibitor D-cycloserine. Antimicrob Agents Chemother 47:283–291.
4. David HL. 1971. Resistance to D-cycloserine in the tubercle bacilli: mutation rate and transport of alanine in parental cells and drug-resistant mutants. Appl Microbiol 21:888–892.
5. Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rüsch-Gerdes S, Fattorini L, Oggioni MR, Cox H, Varaine F, Niemann S. 2013. Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. PLOS ONE 8:e82551.
6. Chacon O, Feng Z, Harris NB, Cáceres NE, Adams LG, Barletta RG. 2002. *Mycobacterium smegmatis* D-alanine racemase mutants are not dependent on d-alanine for growth. Antimicrob Agents Chemother 46:47–54.
7. Milligan DL, Tran SL, Strych U, Cook GM, Krause KL. 2007. The alanine racemase of *Mycobacterium smegmatis* is essential for growth in the absence of D-alanine. J Bacteriol 189:8381–8386.
8. Awasthy D, Bharath S, Subbulakshmi V, Sharma U. 2012. Alanine racemase mutants of *Mycobacterium tuberculosis* require D-alanine for growth and are defective for survival in macrophages and mice. Microbiology (Reading, Engl) 158:319–327.
9. Prosser GA, de Carvalho LPS. 2013. Metabolomics reveal D-alanine:D-alanine ligase as the target of D-cycloserine in *Mycobacterium tuberculosis*. ACS Med Chem Lett 4:1233–1237.
10. Halouska S, Fenton RJ, Zinniel DK, Marshall DD, Barletta RG, Powers R. 2014. Metabolomics analysis identifies D-alanine-D-alanine ligase as the primary lethal target of D-cycloserine in Mycobacteria. Journal of proteome research 13:1065.
11. Desjardins CA, Cohen KA, Munsamy V, Abeel T, Maharaj K, Walker BJ, Shea TP, Almeida DV, Manson AL, Salazar A, Padayatchi N, O'Donnell MR, Mlisana KP, Wortman J, Birren BW, Grosset J, Earl AM, Pym AS. 2016. Genomic and functional analyses of

Mycobacterium tuberculosis strains implicate *ald* in D-cycloserine resistance. Nat Genet 48:544–551.

12. Andersen AB, Andersen P, Ljungqvist L. 1992. Structure and function of a 40,000-molecular-weight protein antigen of *Mycobacterium tuberculosis*. Infect Immun 60:2317–2323.
13. Giffin MM, Modesti L, Raab RW, Wayne LG, Sohaskey CD. 2012. *ald* of *Mycobacterium tuberculosis* encodes both the alanine dehydrogenase and the putative glycine dehydrogenase. J Bacteriol 194:1045–1054.
14. Giffin MM, Shi L, Gennaro ML, Sohaskey CD. 2016. Role of alanine dehydrogenase of *Mycobacterium tuberculosis* during recovery from hypoxic nonreplicating persistence. PLOS ONE 11:e0155522.
15. Embden JD van, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, Al E. 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. Journal of Clinical Microbiology 31:406.
16. Streicher EM, Victor TC, Spuy G van der, Sola C, Rastogi N, Helden PD van, Warren RM. 2007. Spoligotype signatures in the *Mycobacterium tuberculosis* complex. J Clin Microbiol 45:237–240.
17. 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.
18. Babraham Bioinformatics - FastQC a quality control tool for high throughput sequence data.
19. Black P, De Vos M, Louw G, van der Merwe R, Dippenaar A, Streicher E, Abdallah A, Sampson S, Victor T, Dolby T, Simpson J, van Helden P, Warren R, Pain A. 2015. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. BMC Genomics 16:857.
20. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. PLOS ONE 7:e46688.
21. 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.

22. Kallio MA, Tuimala JT, Hupponen T, Klemelä P, Gentile M, Scheinin I, Koski M, Käki J, Korpelainen EI. 2011. Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics 12:507.
23. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics btu170.
24. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14:R36.
25. Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.
26. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.
27. De Vos M, 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.
28. 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.
29. Mlambo CK, Warren RM, Poswa X, Victor TC, Duse AG, Marais E. 2008. Genotypic diversity of extensively drug-resistant tuberculosis (XDR-TB) in South Africa. Int J Tuberc Lung Dis 12:99–104.
30. Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics btv195.
31. Lew JM, Kapopoulou A, Jones LM, Cole ST. 2011. TubercuList – 10 years after. Tuberculosis 91:1–7.
32. Rose G, Cortes T, Comas I, Coscolla M, Gagneux S, Young DB. 2013. Mapping of genotype-phenotype diversity among clinical isolates of *Mycobacterium tuberculosis* by sequence-based transcriptional profiling. Genome Biol Evol 5:1849–1862.
33. Pethe K, Sequeira PC, Agarwalla S, Rhee K, Kuhen K, Phong WY, Patel V, Beer D, Walker JR, Duraiswamy J, Jiricek J, Keller TH, Chatterjee A, Tan MP, Ujjini M, Rao SPS, Camacho L, Bifani P, Mak PA, Ma I, Barnes SW, Chen Z, Plouffe D, Thayalan P, Ng SH, Au M, Lee BH, Tan BH, Ravindran S, Nanjundappa M, Lin X, Goh A, Lakshminarayana

SB, Shoen C, Cynamon M, Kreiswirth B, Dartois V, Peters EC, Glynne R, Brenner S, Dick T. 2010. A chemical genetic screen in *Mycobacterium tuberculosis* identifies carbon-source-dependent growth inhibitors devoid of *in vivo* efficacy. Nature Communications 1:57.

34. Pethe K, Bifani P, Jang J, Kang S, Park S, Ahn S, Jiricek J, Jung J, Jeon HK, Cechetto J, Christophe T, Lee H, Kempf M, Jackson M, Lenaerts AJ, Pham H, Jones V, Seo MJ, Kim YM, Seo M, Seo JJ, Park D, Ko Y, Choi I, Kim R, Kim SY, Lim S, Yim S-A, Nam J, Kang H, Kwon H, Oh C-T, Cho Y, Jang Y, Kim J, Chua A, Tan BH, Nanjundappa MB, Rao SPS, Barnes WS, Wintjens R, Walker JR, Alonso S, Lee S, Kim J, Oh S, Oh T, Nehrbass U, Han S-J, No Z, Lee J, Brodin P, Cho S-N, Nam K, Kim J. 2013. Discovery of Q203, a potent clinical candidate for the treatment of tuberculosis. Nat Med 19:1157–1160.
35. Raynaud C, Etienne G, Peyron P, Lanéelle M-A, Daffé M. 1998. Extracellular enzyme activities potentially involved in the pathogenicity of *Mycobacterium tuberculosis*. Microbiology 144:577–587.
36. Boulette ML, Baynham PJ, Jorth PA, Kukavica-Ibrulj I, Longoria A, Barrera K, Levesque RC, Whiteley M. 2009. Characterization of alanine catabolism in *Pseudomonas aeruginosa* and its importance for proliferation *in vivo*. J Bacteriol 191:6329–6334.
37. Summers AO. 2009. Damage control: regulating defences against toxic metals and metalloids. Curr Opin Microbiol 12:138–144.
38. Arunkumar AI, Campanello GC, Giedroc DP. 2009. Solution structure of a paradigm ArsR family zinc sensor in the DNA-bound state. PNAS 106:18177–18182.
39. Andreu N, Soto CY, Roca I, Martín C, Gibert I. 2004. *Mycobacterium smegmatis* displays the *Mycobacterium tuberculosis* virulence-related neutral red character when expressing the *Rv0577* gene. FEMS Microbiol Lett 231:283–289.
40. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43:D222-226.
41. Gao C, Yang M, He Z-G. 2012. Characterization of a novel ArsR-like regulator encoded by *Rv2034* in *Mycobacterium tuberculosis*. PLoS One 7.
42. Gao C-H, Yang M, He Z-G. 2011. An ArsR-like transcriptional factor recognizes a conserved sequence motif and positively regulates the expression of *phoP* in mycobacteria. Biochemical and Biophysical Research Communications 411:726–731.

43. Soo-Ja Kim, Yu Jin Kim, Mi Ran Seo, Bong Sook Jhun. 2000. Regulatory mechanism of L-alanine dehydrogenase from *Bacillus subtilis*. Bull Korean Chem Soc 21:1217–1221.
44. Rosenkrands I, Weldingh K, Jacobsen S, Veggerby Hansen C, Florio W, Gianetri I, Andersen P. 2000. Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection. Electrophoresis 21:935–948.
45. Thornalley PJ. 2003. Glyoxalase I – structure, function and a critical role in the enzymatic defence against glycation. Biochemical Society Transactions 31:1343–1348.
46. Rigsby RE, Fillgrove KL, Beihoffer LA, Armstrong RN. 2005. Fosfomycin resistance proteins: a nexus of glutathione transferases and epoxide hydrolases in a metalloenzyme superfamily. p. 367–379. *In* Packer, HS and L (ed.), Methods in Enzymology. Academic Press.

Addendum to: Gaining a better understanding of D-cycloserine resistance in *Mycobacterium tuberculosis* through next-generation technologies

Note: The strains described in the preceding chapter originally drew our attention due to the mutations observed in the *rpoB* and *rpoC* genes. Although the manuscript to be submitted focuses on clinically relevant mutations associated with D-cycloserine resistance, we also investigated the impact of *rpoC* mutations on gene expression in the same strains. The results from this analysis are presented here.

A5.1 Introduction

Various mutations in *rpoC*, which code for the β' subunit RNA polymerase, have previously been described in rifampicin resistant strains of *Mycobacterium tuberculosis* (1, 2). Resistance to rifampicin is caused by mutations in *rpoB* which often leads to reduced affinity of RNA polymerase for the drug. Mutations in *rpoC* are thought to compensate for the fitness cost incurred by *rpoB* mutants, in particular those with a S531L mutation, however little is known concerning the effects of *rpoC* mutations on the gene expression profile of *M. tuberculosis* (1, 2). Using multidrug resistant clinical isolates of *M. tuberculosis*, we undertook to investigate the compensatory mechanism which benefit strains harbouring *rpoC* mutations. A better understanding of the effect of these mutations on gene expression could allow for novel insight into mycobacterial physiology and complement current literature surrounding compensatory mutations in RNA polymerase.

A5.2 Methods

A5.2.1 Sample selection, clinical data and bacterial culture

Mycobacterium tuberculosis clinical samples were collected and cultured as described in the preceding chapter. Since Isolate 4 was a heterogeneous population exhibiting two different non-synonymous polymorphisms at codon 483 of *rpoC*, single colony forming units were prepared and cultured, to generate Samples 4.c1 and 4.c2 (Table A1). These samples were included in the transcriptomic analysis to investigate the individual subpopulations of Isolate 4. Briefly, Isolate 4 was cultured to mid-log phase and a 1 mL aliquot of the culture was sonicated in an ultrasonic bath (UC-1D, Zeus Automation) at 37 kHz for 12 minutes to disperse clumps. Serial dilutions were prepared and grown on Middlebrook 7H10 media enriched with 10% oleic albumin dextrose catalase (OADC) and 0.5% glycerol for 21 days. Single colonies were picked and inoculated into 5 mL of ADC-enriched Middlebrook 7H9 media. Glycerol stocks of these cultures were prepared and stored as in Chapter 5. Thereafter, PCR amplification (with primers in Table A2) and Sanger sequencing with the ABI 3500XL sequencer (Applied Biosystems) were used to identify the mutation that was present at codon 483 of *rpoC*. Based on screening results, single colonies exhibiting either the V483A polymorphism (Sample 4.c1) or the V483G polymorphism (Sample 4.c2) were selected for downstream analyses. Additionally, variants were called using RNAseq data to identify which variants were present in each population.

Table A1. Heterogeneous clinical isolate. The multidrug resistant clinical isolate and subsequent single colonies isolated from this heterogeneous population.

Sample	Sample Type	*rpoC* Mutation	Amino Acid Change
Isolate 4	Clinical isolate	V483A/G	GTG → GSG
Sample 4.c1	Colony forming unit	V483A	GTG → GCG
Sample 4.c2	Colony forming unit	V483G	GTG → GGG

Table A2. Primers used to amplify region of *rpoC* flanking codon 483.

Name	Sequence	Product Size (bp)	Tm (°C)
CMR_F	5’ CAACCATGCGCAGAACATCA 3’	763	62
CMR_R	5’ CTTGAGCTTGTCGACGGTCT 3’		

A5.2.2 Whole genome sequencing

Bacterial culture, DNA isolation, subsequent whole genome sequencing and data analysis were performed as described in Chapter 5.

A5.2.3 RNA isolation and transcriptomic analysis

Single colony forming units from Isolate 4, namely 4.c1 and 4.c2, were selected for transcriptomic analysis to study the effect of different *rpoC* mutations on the gene expression profile of the strain. Firstly, the growth characteristics of each colony forming unit was determined as described in the Chapter 5. Subsequently, RNA was isolated, sequenced, and RNAseq data was analysed as previously described in Chapter 5.

A5.3 Results

A5.3.1 Identification of variants

Detailed analysis of all mutations identified in 6 serial clinical isolates is provided in the preceding chapter. Here, we focus on *rpoC* mutations identified in single colonies (Sample 4.c1 and Sample 4.c2). Whole genome sequencing (WGS) of Isolate 3, collected in mid-2006, showed the acquisition of a V483G substitution in *rpoC*, a compensatory mutation which ameliorates the fitness cost associated with *rpoB* mutations (1, 2) (Table A3).

Table A3. Variant identification. Frequency of variants (%) which were acquired subsequent to the *rpoC* V483G mutation, as identified by whole genome sequencing.

Variant	Isolate					
	1	**2**	**3**	**4**	**5**	**6**
[b]*rpoC* 483 GTG/GGG			100	74	48	55
[a]*rpoC* 483 GTG/GCG				26	51	45
[a]*ald* 841 C ins				20	33	34
[b]*ald* 301 ATG/AGG				60	44	52
ald 966 GA ins				10	0	5
[a]*cspA* 22 GAC/GAG					15	23

[a]Variants present in Sample 4.c1

[b]Variants present in Sample 4.c2

WGS of Isolate 4 revealed the presence of a second polymorphism at the same position of *rpoC* (V483A) at a variant frequency of 26%, while the frequency of the initial *rpoC* V483G polymorphism decreased from >99% to 74%. Although all reads exhibited the *rpoC* V483G substitution in Sample 3, we propose that a population with wildtype *rpoC* was still present in the patient, from which the *rpoC*483 GTG/GCG variant arose. To support our hypothesis, Sanger sequencing of *rpoC* using crude DNA from Isolate 3, revealed a wildtype allele (GTG) at position 48. However, as purified DNA which was used for whole genome sequencing was isolated from a culture which had undergone additional passages (Figure A1), this indicates that cells harbouring the *rpoC* mutation experienced a fitness advantage during *in vitro* culturing, and thereby outcompeted the population with wildtype *rpoC*.

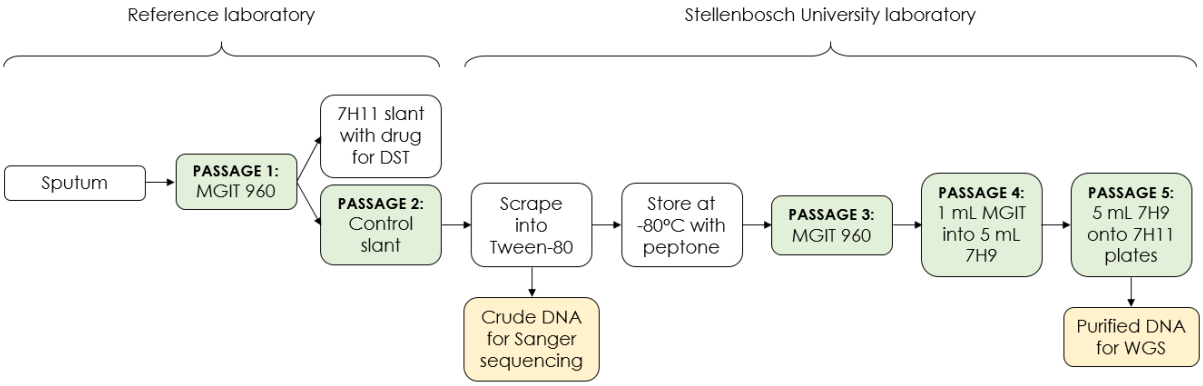


Figure A1. Culturing of clinical isolates. Sanger sequencing was done using crude DNA, whereas whole genome sequencing was performed using purified DNA which had undergone three additional passages.

Sequencing of the final samples, Isolate 5 and Isolate 6, revealed that the two *rpoC* mutations remained present at relatively equal proportions in the population. Furthermore, a D22E substitution was identified in *Rv3648c* which encodes a probable cold shock protein, CspA. The *cspA* polymorphism was associated with the population harbouring the *rpoC* V483A mutation and was present in 15% and 23% of reads in Isolates 5 and 6, respectively. Analysis using PROVEAN (Protein Variation Effect Analyzer) indicated that this polymorphism is deleterious to protein function (3).

A5.3.2 Investigating variants identified through RNAseq

To determine whether the variants in *ald* were linked to a particular *rpoC* mutation, variants were called using RNAseq data obtained from the single colonies isolated from Isolate 4 (Samples 4.c1 and 4.c2). Subsequent analysis revealed that the *ald* M301R mutation was associated with the *rpoC* V483G substitution, whereas the 1bp insertion at position 841 was acquired alongside *rpoC* V483A (Table A3 and Figure A2).

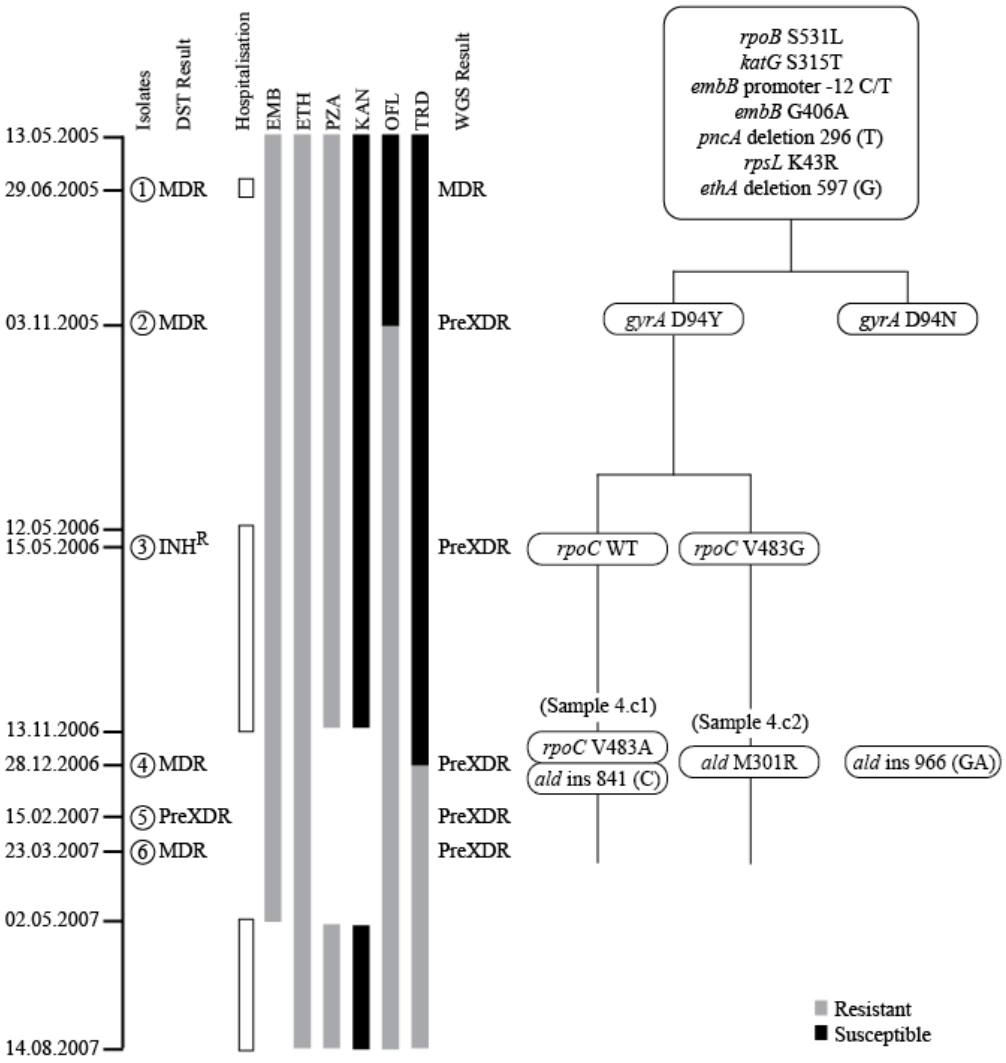


Figure A2. Patient timeline. Chronology of isolate collection illustrating routine drug susceptibility testing (DST) results, patient treatment information and acquired genomic variants linked to drug resistance. Additional analysis using data from Samples 4.c1 and 4.c2 revealed that the *ald* M301R mutation was associated with the *rpoC* V483G substitution, whereas the 1bp insertion at position 841 was acquired alongside *rpoC* V483A.

A5.3.3 Transcriptomic analysis of single colonies

Samples 4.c1 and 4.c2 were selected for transcriptomic analysis to determine the impact of different mutations in *rpoC* on gene expression (Table A4). Determination of the growth profiles for these two samples confirmed no difference in growth (Figure A3). Differential gene expression analysis using EdgeR revealed 23 genes which were upregulated in Sample 4.c2, which harbours the *rpoC* V483G substitution, relative to Sample 4.c1. Interestingly, a 2.49-fold upregulation of *Rv2416c* (*eis*) was observed in the clone with the *rpoC* V483G polymorphism, as well as a 2.85-fold upregulation of *Rv1258c* (*tap*), both of which are part of the Whib7

regulon. These findings aligned with gene expression data for Isolate 3 where the *rpoC* V483G polymorphism was present at a frequency of >99%, relative to Sample 4.c1 harbouring the V483A mutation, which revealed that 8 of these genes followed the same trend. We acknowledge that expression analysis between a single colony forming unit and a clinical isolate is not an ideal comparison seeing as much of the sensitivity will be lost, however this finding further demonstrates that upregulation of these genes can be attributed to the *rpoC* V483G substitution.

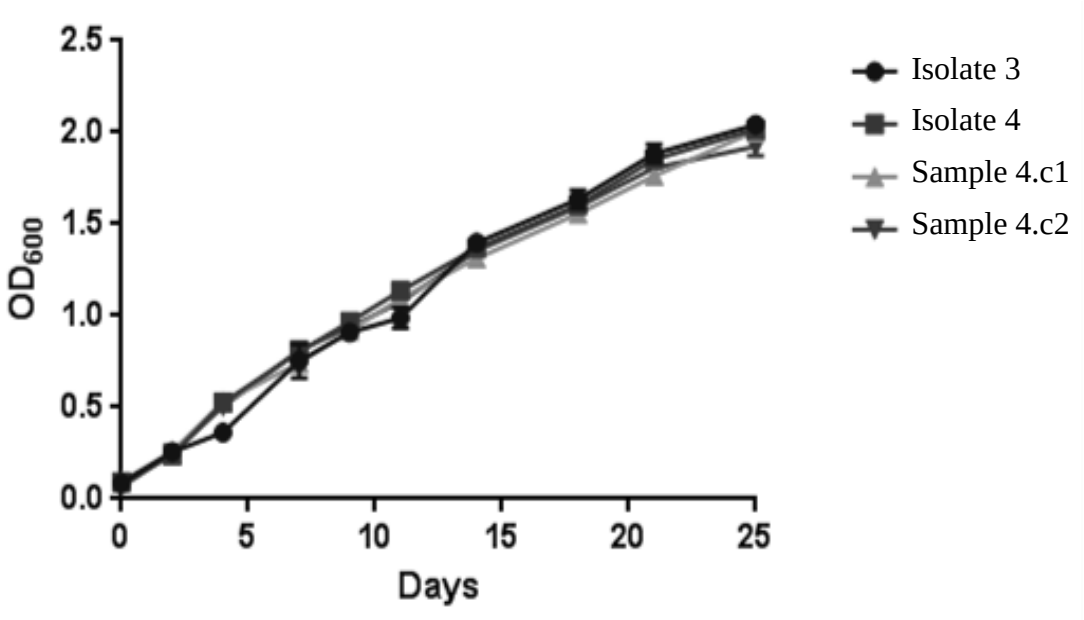


Figure A3. Growth profiles. No difference in growth was observed between the clinical isolates and single colonies selected for RNAseq. Absorbance values represent the mean of technical quadruplicates with error bars depicting standard deviation.

Table A4. Differential expression. Genes which were differentially expressed (fold change >2-fold, adjusted p-value < 0.05) in Sample 4.c2, harbouring a V483G polymorphism in *rpoC*, compared to Sample 4.c1, harbouring a V483A polymorphism in *rpoC*. Genes in bold were also found to be upregulated in Isolate 3, harbouring the *rpoC* V483G polymorphism, relative to Sample 4.c1 harbouring the V483A mutation.

Gene	Gene Product	Fold Change	Adjusted p-value
Rv0279c	**PE-PGRS family protein PE_PGRS4**	3.00 up	6.38E-07
Rv0544c	Possible conserved transmembrane protein	2.32 up	0.012888
Rv0649	Possible malonyl CoA-acyl carrier protein transacylase FabD2 (MCT)	3.23 up	0.001125
Rv0747	PE-PGRS family protein PE_PGRS10	2.17 up	3.6E-05
Rv1067c	PE-PGRS family protein PE_PGRS19	2.12 up	0.01303
Rv1087	PE-PGRS family protein PE_PGRS21	2.07 up	0.00153
Rv1091	PE-PGRS family protein PE_PGRS22	2.05 up	0.000114
Rv1130	**Possible methylcitrate dehydratase PrpD**	3.31 up	9.48E-40
Rv1131	**Probable methylcitrate synthase PrpC**	2.68 up	9.05E-28
Rv1257c	**Probable oxidoreductase**	2.10 up	7.75E-07
Rv1258c	**Probable conserved integral membrane transport protein (Tap)**	2.85 up	1.83E-12
Rv2270	Probable lipoprotein LppN	2.64 down	0.017121
Rv2416c	**Enhanced intracellular survival protein Eis**	2.49 up	7.89E-13
Rv2526	Possible antitoxin VapB17	2.74 up	0.001566
Rv2541	Hypothetical alanine rich protein	4.40 up	0.001719
Rv2725c	**Probable GTP-binding protein HflX**	2.05 up	8.82E-07
Rv3344c	PE-PGRS family protein PE_PGRS49	3.31 up	0.000354
Rv3345c	PE-PGRS family protein PE_PGRS50	2.03 up	1.23E-05
Rv3388	PE-PGRS family protein PE_PGRS52	2.47 up	0.000261
Rv3508	**PE-PGRS family protein PE_PGRS54**	5.84 up	6.46E-07
Rv3511	PE-PGRS family protein PE_PGRS55	2.10 up	0.000448
Rv3512	PE-PGRS family protein PE_PGRS56	4.87 up	0.00153
Rv3653	PE-PGRS family protein PE_PGRS61	2.08 up	0.03992
Rv3657c	Possible conserved alanine rich membrane protein	2.11 up	0.047196

A5.4 Discussion

To better understand the impact of compensatory mutations on the physiology of *M. tuberculosis*, this study used a set of clinical isolates to study the effect of *rpoC* mutations on gene expression. A drug resistant isolate harbouring two different mutations at the same codon of *rpoC* (V483G and V483A) was used to isolate single colony forming units. The separated populations were then subjected to RNAseq, to compare the effect of these *rpoC* mutations on gene expression. Although not described here, numerous attempts were made to isolate a population with wildtype *rpoC* from the preceding clinical isolate, as this would allow for a comparison between wildtype and *rpoC* mutant. However, as mutations in this gene are known to compensate for the fitness deficit experienced by *rpoB* mutants, it is thought that *in vitro* culturing selected for the mutant *rpoC* population. Nonetheless, we had at our disposal two *rpoC* mutants with near identical genetic backgrounds, which served as a useful tool to understand how *rpoC* mutations allow for compensation of fitness.

Upon investigating the transcriptomic effect of an *rpoC* V483G mutation relative to the *rpoC* V483A substitution, we observed a 2.49 and 2.85-fold upregulation of *Rv2416c* (*eis*) and *Rv1258c* (*tap*), respectively. Both genes have been found to be downregulated in clinical isolates of *M. tuberculosis* harbouring *rpoB* S531L mutations, compared to progenitors without this mutation (4), which may explain why V483G substitutions are the most frequent in clinical isolates of *M. tuberculosis* (1, 2). Furthermore, previous studies have found that upregulation of *eis* allows for low level resistance to kanamycin, and that upregulation of the efflux pump encoded by *tap* is associated with low-level streptomycin resistance (5, 6). Both of these factors could contribute to the previously described fitness advantage of isolates with the *rpoC* V483G mutation. Furthermore, as 12 genes which are part of the *pe-pgrs* family were found to be differentially expressed between the two *rpoC* mutants, it can be hypothesised that they may be linked to the compensatory mechanism of *rpoC* mutations. However, as the functions of these genes remain largely unknown, additional studies are needed to elucidate their role in the physiology of drug resistant *M. tuberculosis*.

Lastly, whole genome sequencing of Sample 4.c1 revealed the acquisition of a polymorphism in *cspA* alongside *rpoC* V483A and the 1bp insertion at position 841 of *ald*. The link between these genes is unknown, however studies in *Escherichia coli* have indicated that CspA is involved in transcription as an RNA chaperone. CspA is believed to bind to RNA to prevent the formation of secondary structures which cause termination of transcription at low

temperatures (7, 8). Fang *et al.* demonstrated that *cspA* is constitutively expressed at all temperatures, however translation is prevented at 37°C by destabilisation of its mRNA (9). As the polymorphism described here is thought to be deleterious to protein function and it is unlikely that the strain encountered temperatures low enough to enable its translation, it is probable that this is a hitchhiking polymorphism which does not play a direct role in strain fitness. Nevertheless, there is a possibility that this polymorphism, as well as the different *ald* variants present in each population, had an effect on the transcriptome. Furthermore, it is also possible that additional variants are present in unmapped regions of the clinical isolates, which may have influenced gene expression. For this reason, additional studies are needed to confirm the findings reported in this study.

A5.5 References

1. De Vos M, 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.
2. 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.
3. Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics btv195.
4. Du Plessis J. 2014. Deciphering the impact of *rpoB* mutations on the gene expression profile of *Mycobacterium tuberculosis*. Thesis: Stellenbosch University.
5. Zaunbrecher MA, Sikes RD, Metchock B, Shinnick TM, Posey JE. 2009. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase *eis* confers kanamycin resistance in *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA 106:20004–20009.
6. Reeves AZ, Campbell PJ, Sultana R, Malik S, Murray M, Plikaytis BB, Shinnick TM, Posey JE. 2013. Aminoglycoside cross-resistance in *Mycobacterium tuberculosis* due to mutations in the 5′ untranslated region of *whiB7*. Antimicrob Agents Chemother 57:1857–1865.
7. Bae W, Jones PG, Inouye M. 1997. CspA, the major cold shock protein of *Escherichia coli*, negatively regulates its own gene expression. J Bacteriol 179:7081–7088.
8. Bae W, Xia B, Inouye M, Severinov K. 2000. *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. PNAS 97:7784–7789.
9. Fang L, Jiang W, Bae W, Inouye M. 1997. Promoter-independent cold-shock induction of *cspA* and its derepression at 37°C by mRNA stabilization. Mol Microbiol 23:355–364.

Chapter 6

General Conclusion

Drug-resistant tuberculosis remains a major concern in countries around the globe, therefore the current study undertook to investigate multiple facets of drug resistance in *Mycobacterium tuberculosis*. Treatment of multidrug-resistant tuberculosis requires the use of drugs with severe negative side-effects and often leads to poor treatment outcomes. For this reason, novel drugs are urgently needed to improve the current repertoire of anti-tuberculosis drugs. As multidrug-resistant strains of *M. tuberculosis* harbour mutations in the β subunit of RNA polymerase (RNAP), the approach taken here was to investigate inhibition of transcription by binding of an inhibitory protein, Gp2, to the β' subunit of RNAP. Findings from *in vitro* assays, done in parallel with *in silico* modelling, revealed that Gp2 binds to mycobacterial RNAP, however a greater degree of affinity is required for this protein-protein interaction to effectively inhibit transcription. The analysis indicated that crucial binding site residues of Gp2 in *E. coli* RNAP are not conserved in the β' subunit of *M. tuberculosis*, however, our findings suggest that this molecule can be used for the *in silico* design of peptides that have a stronger affinity for mycobacterial RNAP. This could entail the development of Gp2 domains into stapled peptides to improve its longevity and provide a stronger interaction with *M. tuberculosis* RNAP. Another strategy could involve the identification of inhibitory proteins from bacteriophages which specifically infect mycobacteria; as these proteins would provide promising candidates for novel anti-tuberculosis drugs. Furthermore, the recent elucidation of the crystal structure of *M. tuberculosis* RNAP will greatly benefit studies which aim to investigate binding of compounds and proteins in the presence of point mutations in this enzyme.

In a second set of experiments, *rpoB* and *rpoC* mutations which are commonly identified in clinical isolates were investigated to determine their influence on mycobacterial transcription. As mutations at codons 531 and 526 of *rpoB* are the most frequent in clinical isolates, these were the mutations studied here to understand their effects on RNAP function. Mutations in *rpoC* have also been described in rifampicin-resistant strains and compensate for the fitness cost incurred by *rpoB* S531L mutants. Therefore, one of the aims of this study was to investigate how compensatory mutations allow for the amelioration of fitness in rifampicin-resistant mutants. For this, mutated forms of RNAP were generated by site-directed mutagenesis of cloned gene sequences which encode mycobacterial RNAP. Subsequent expression of these genes allowed for purification of wildtype and mutant RNAP, which were used in radioactivity-based *in vitro* transcription assays. Results from this work indicates that *rpoB* and *rpoC* mutations play a role in the relationship between the subunits which make up the RNAP holoenzyme. Furthermore, positive outcomes include the generation of RNAP harbouring

mutations which are frequently found in clinical isolates of *M. tuberculosis*, a new tool to study RNAP physiology using *in vitro* transcription. Additionally, a fluorescence-based *in vitro* transcription assay was investigated with the aim of developing a comparable method without the use of radiolabelled nucleotides. However, numerous drawbacks to this assay demonstrates the value of radioactivity-based assays as the gold-standard for investigating *in vitro* transcription. Future studies could incorporate the use of RNAP-binding factors, such CarD and RbpA, as well as different promoter templates from *M. tuberculosis*. The use of *in vivo* pull-downs could also be used to ascertain whether different RNAP-binding factors are co-purified in the presence of *rpoB* and *rpoC* mutations as this would provide a good foundation for our understanding of how mutated RNAP differs from that of wildtype.

The final aim proposed for this study was to investigate the effect of compensatory *rpoC* mutations on the gene expression profile of a clinical isolate of *M. tuberculosis*. However, *in vivo* evolution of strains is complex and analysis of the genomic data for the clinical isolates selected for this study revealed the acquisition of *ald* mutations alongside the *rpoC* mutation. Loss-of-function mutations in *ald* have been associated with D-cycloserine (DCS) resistance, therefore this unexpected finding allowed us to investigate the mechanism of resistance to this drug. Transcriptomic data revealed the upregulation of *Rv0577*, encoding a putative glyoxalase (Cfp32), which was cloned and overexpressed in the reference strain H37Rv and a pan-susceptible clinical isolate of *M. tuberculosis*. Susceptibility testing of these strains revealed an increase in the minimum inhibitory concentration of DCS, which suggests that the gene product of *Rv0577* plays a role in the mechanism of drug resistance. We hypothesise that strains harbouring mutations in *ald* experience increased conversion of methylglyoxal to lactic acid, which is a precursor for pyruvate. Additionally, our findings indicate that Cfp32 harbours a domain which may cleave the ring structure of DCS, rendering the drug ineffective. Additional studies are needed to validate this finding; however, this is the first report to illustrate a mechanism of DCS resistance linked to *ald* mutations in *M. tuberculosis*. Future studies could include the use of enzymatic assays to determine whether cleavage of DCS by Cfp32 is dependent on pH and whether metal co-factors such as Zn^{2+} are required for this reaction.

Lastly, using transcriptomic data generated from single colony forming units of the same clinical isolate, we investigated the effect of compensatory *rpoC* mutations on the transcriptome of *M. tuberculosis*. Here, we report a potential link between the *rpoC* V483G mutation and upregulation of *Rv2416c* (*eis*) and *Rv1258c* (*tap*), both of which have been found to be

downregulated in a clinical isolate harbouring an *rpoB* S531L mutation without an additional compensatory mutation. As upregulation of *eis* and *tap* have been associated with low level resistance to kanamycin and streptomycin, respectively, these findings may indicate that strains with *rpoC* mutations benefit from an additional advantage within a host being treated with these drugs. This is of importance in the broader clinical setting, as undetected cross resistance could undermine anti-tuberculosis treatment regimens. However, additional studies are necessary to confirm these findings.

Overall, the results from this body of work support and advance current understanding of the physiology of drug-resistant *M. tuberculosis* and point to many new avenues for future research. Furthermore, it is evident that multiscale studies which encompass both an omics approach and subsequent molecular validation are needed to gain a deeper understanding of the complex physiology of this Machiavellian pathogen.