Characterisation of VapBC Toxin-Antitoxins from *Mycobacterium tuberculosis*

A thesis submitted in partial fulfilment

of the requirements for the degree

of

Master of Science in Biological Sciences

at

The University of Waikato

by

Abigail Victoria Sharrock

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2013

Abstract

Toxin-antitoxin (TA) systems were identified more than 20 years ago on the mini-F plasmid of *Escherichia coli* as plasmid stability elements; components responsible for purging bacterial cells that lack the plasmid from the population. More recent discovery of TA systems spanning a wide diversity of prokaryotic chromosomes, including that of *Mycobacterium tuberculosis* (*M. tb*), suggests a broader biochemical role. TA systems can be classified into a number of families, with the *vapBC* systems being by far the largest. The biochemical role of *vapBC* systems in *M. tb* remains unclear despite their abundance within the genome.

This thesis describes the biochemical and functional characterisation of two *vapBC* systems encoded by operons *Rv0065a/c* and *Rv0617a/c* in the *M. tb* genome. $VapC_{Rv0617}$ overexpression had a bacteriostatic effect on the growth of *Mycobacterium smegmatis* cultures. Therefore, the biochemistry underlying this phenotype was investigated, along with that of the *Rv0065a/c* system. $VapC_{Rv0065}$ and $VapC_{Rv0617}$ are Mg^{2+}-dependent, sequence-specific ribonucleases targeting GC-rich 4-mers. Ribonuclease assays with *in vitro* synthesised RNA transcripts suggested an additional layer of target recognition resides in RNA secondary structure, and revealed that both VapC proteins exhibit high activity against isolated *M. smegmatis* 16S and 23S rRNA. Electrophoretic mobility shift assays with purified VapBC protein and labelled DNA demonstrated an autoregulatory function for the $VapBC_{Rv0617a/c}$ complex and not the $VapBC_{Rv0065a/c}$ complex. $VapBC_{Rv0617a/c}$ bound specifically to a near-perfect inverted repeat in the *Rv0617a/c* promoter region that overlaps an annotated -10 *M. tb* promoter element. In contrast, the $VapBC_{Rv0065a/c}$ complex exhibited no DNA binding activity against a putative *Rv0065a/c* promoter region. Individualised *vapBC* transcriptional regulation mechanisms may help explain the persistence of such an expanded number of these systems in the genome.

Knowledge of the physiological role of *vapBC* systems in *M. tb* will enable a better understanding of how they contribute to the pathogenicity of this bacterium. This would serve as the basis for the design of drugs which interfere with *vapBC* system functioning and in turn the ability of *M. tb* cells to enter the persistent state.

Acknowledgements

First and foremost I would like to acknowledge Professor Vic Arcus for the guidance, support and opportunities that he has provided throughout the duration of my MSc degree. I valued his enthusiasm both in the toxin-antitoxin project and in my progression as a research scientist.

I'd like to thank everybody in the Proteins & Microbes lab for making my research year by far the most enjoyable year of my tertiary study. I was very fortunate to be working in a lab with such great people who are also such talented researchers.

A special mention needs to go out to all the mycobacterial girls, past and present, for introducing me to the toxin-antitoxin world and helping me along my way. This includes Jo M, Emily, Ali, Emma A and Chelsea. I'd also like to thank the remainder of those who have been a part of the Proteins & Microbes team whilst I completed my research; Judith, Jo H, Jo D, Mark, Emma S, Tifany, Joel, Kirsty, Vikas and Erica.

A big thank you to those who provided financial support throughout my degree; including the University of Waikato, the Brian Perry Charitable Trust, the Nancy Caiger Graduate Trust and the NZ Federation of Graduate Women.

I'd finally like to thank my family, non-scientific friends and Jono for their love and unwavering support throughout my studies, with a special thank you to Keith whose scientific knowledge and proof-reading skills were very much appreciated!

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Systeme International d'Unités (SI) abbreviations for units and standard notations for chemical formulae and elements are used throughout this thesis. Additional abbreviations used are listed below.

Ω	Ohms
µg	microgram (10^{-6} g)
µl	microliter (10^{-6} L)
µm	micrometer (10^{-6} m)
µM	micromolar (10^{-6} M)
3-HPA	3-hydroxypicolinic acid
A	adenine
ACN	acetonitrile
ADC	albumin dextrose catalase
ATP	adenosine triphosphate
APS	ammonium persulfate
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine
cDNA	complementary DNA
CFU	colony forming units
Da	Daltons
DEPC	diethyl pyruvate carbonate
DIG	digoxigenin
DNA	deoxyribose nucleic acids
DNase	deoxyribonuclease
ds	double-stranded
EDTA	ethylene diamine tetraacetic acid (disodium salt)
EMSA	electrophoretic mobility shift assay
FPLC	fast performance liquid chromatography
g	times the force of gravity
G	guanine
HdB	Hartmans-de Bont
HGT	horizontal gene transfer

His-tag	poly-histidine tag
HPLC	high performance liquid chromatography
IMAC	immobilising metal affinity chromatography
IR	inverted repeat
kb	kilobase
kDa	kilo Dalton
kV	kilo Volt
LB	Luria Bertani
LBT	Luria Bertani plus Tween-80
MALDI TOF	matrix assisted laser desorption ionisation time of flight
mAU	milli-absorbance units
MDT	multi-drug tolerant
MS	mass spectrometry
mg	milligram (10^{-3} g)
ml	millilitre (10^{-3} L)
mM	(10^{-3} M)
mRNA	messenger RNA
MS	mass spectrometry
MW	molecular weight
NPIR	near-perfect inverted repeat
NRP	non-replicating persistence
OD_{600}	optical density at 600 nm wavelength
ORF	open reading frame
PAGE	poly acrylamide gel electrophoresis
PCR	polymerase chain reaction
pI	isoelectric point
PIN	PilT N-terminal domain
PSK	post segregational killing
RHH	ribbon-helix-helix
R-M	restriction modification
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
SDS	sodium dodecyl sulphate

SOC	super optimal broth with catabolite repression
sRNA	small RNA
Ss	single stranded
T	thymine
TA	toxin-antitoxin
TAE	tris-acetate-EDTA
TB	tuberculosis
TE	tris EDTA buffer
TEMED	N, N, N, N,-tetramethylethylenediamine
T_m	melting temperature
TOF	time of flight
Tris	tris-hydroxymethyl aminomethane
tRNA	transfer ribonucleic acid
TSP	transcriptional start point
U	uracil
uF	micro Faraday
UTR	untranslated region
UV	ultra violet
Vap	virulence associated protein
$VapBC_{Rv0065a/c}$	VapBC encoded from the *Rv0065a/c* operon of *M. tb*
$VapBC_{Rv0617a/c}$	VapBC encoded from the *Rv0617a/c* operon of *M. tb*
$VapC_{Rv0065}$	VapC encoded from the *Rv0065* gene of *M. tb*
$VapC_{Rv0617}$	VapC encoded from the *Rv0617* gene of *M. tb*
$VapC_{MS1284}$	VapC from *M. smegmatis*
$VapB_{MS1283}$	VapB from *M. smegmatis*
v/v	volume per volume
WHO	World Health Organisation
w/v	weight per volume
w/w	weight per weight
XDR	extensively drug resistant

Chapter One: Introduction

1.1 Tuberculosis

The disease tuberculosis (TB), caused by *Mycobacterium tuberculosis* (*M. tb*), remains one of the most lethal infectious diseases despite the availability of chemotherapy and the BCG vaccine (Babajan et al. 2011). One-third of all humans world-wide are estimated to currently carry this microorganism, responsible for killing approximately 2 million people each year (Ramage et al. 2009). The World Health Organization reports that in 2011 an estimated 8.7 million new TB cases emerged (13% co-infected with HIV), with 1.4 million resulting in mortality. Geographically, the TB burden is greatest in Asia and Africa; India and China account for approximately 40% of the world's reported TB cases, while the African region, contributing 24%, exhibits the highest rate of cases and deaths per capita (WHO 2012).

M. tb is an intracellular bacterial pathogen, belonging to the genus *Mycobacterium* and the order Actinomycetales. Due to the unique structure of the mycobacterial cell wall, *M. tb* does not fall into the classic Gram negative/Gram positive classification. While still containing peptidoglycan, over 60% of the cell wall consists of complex lipids such as mycolic acids, cord factor and wax-D (Rajni et al. 2011). The unique cell wall structure plays an important role in virulence, contributing to resistance against antibiotics and lethal oxidants while assisting persistence within host cell environments such as within the macrophage (Rajni et al. 2011). The hardy cell wall also retards the passage of nutrients in and out of the cell, contributing to mycobacterial species' characteristically slow growth rate.

Persistence presents a significant hurdle for the successful treatment of *M. tb* infections, due to the ability of *M. tb* cells to enter into a 'dormant' state of growth. *In vivo*, dormancy is characterised by near zero growth, a state referred to as nonreplicating persistence (NRP) (Gill et al. 2009). Adopting a dormant phase of marginal growth renders cells refractory to antibiotic treatments that target rapidly dividing cells, while preserving energy and nutrients for survival in stressful environments. Subpopulations of the bacterial infection, termed multidrug tolerant (MDT) persisters, may exhibit resistance to a number of antimicrobial therapies. Alarmingly, 20% of previously treated cases and 3.7 % of

new cases are estimated to exhibit such resistance (WHO 2012). Equipped with the ability to selectively downregulate metabolism to the dormant level when required, *M. tb* is known to thrive in both nutrient rich and poor environments. Persistence within nutrient poor host macrophages (Pieters 2008) renders the immune system blind to the infection, allowing TB cases to be asymptomatic until cells emerge from dormancy, sometimes decades later (Murphy & Brown 2007). Out of the approximately 1.8 billion people currently infected with *M. tb*, the vast majority of cases display such clinical latency (Gill et al. 2009). To date, the physiology of bacteria in the dormant state, and the factors that establish and maintain this metabolic plasticity, are not well understood.

As the number of MDT TB cases increases, finding weaknesses in the pathogen that can be exploited in new treatments has become the main research thrust. Ultimately, studies aim to fully understand mechanisms by which *M. tb* is able to selectively and reversibly regulate growth when exposed to a variety of stresses. Characterizing the components involved in these mechanisms may lead to new targets for drug therapy, as interference with these pathways that permit *M. tb* cells to enter the dormant state may increase the effectiveness of existing drug treatments.

Entry into dormancy when faced with environmental stresses is a widely observed trait among pathogenic bacteria. Through adaptive evolution, multiple systems to ensure survival through dormancy have been developed, such as the toxin-antitoxin (TA) systems (Blower et al. 2012). The expansion and multiplicity of these systems in the *M. tb* genome implies they have significant a role in survival, making them of particular interest in the study of *M. tb* pathology.

1.2 Toxin-Antitoxin Systems

1.2.1 First Identification

TA systems residing on bacterial plasmids were first noted more than 20 years ago on the mini-F plasmid of *Escherichia coli* and were denoted as elements that protect low copy number plasmids in bacteria from loss during segregation (Ogura & Hiraga 1983). As replication of F and *oriC* plasmids is not coupled to cell division, a small population of cells would lack a plasmid after each divisional round. In order to minimize the production of plasmid-free segregants,

a genetic system encoded on the mini-F plasmid (later termed the *ccdAB* TA system) is employed to temporarily block cell division. Interplay between the two genes within this system (*ccdB* and *ccdA*) was found to be essential for maintaining the balance of such inhibition. The *ccdB* gene inhibits cell division by encoding a 'toxin' protein component while the *ccdA* gene lifts the inhibition through production of an 'antitoxin' protein component (Ogura & Hiraga 1983). If the plasmid is lost through segregation during cell division, the stoichiometric balance of toxin : antitoxin is thrown off due to a shorter antitoxin half-life. This results in bacteriostasis or cell death, purging plasmid-less cells from the population (Rankin et al. 2012). This strategy of plasmid-maintenance, outlined in Figure 1.1, is referred to as 'post-segregational killing' (PSK). As such, the underlying TA systems are often referred to as 'addiction modules', based on the observation that they function to 'addict' cell lines to the TA complex and loss of the TA system may result in PSK (Gerdes et al. 1986; Engelberg-Kulka & Glaser 1999). The interplay of a toxic component and a cognate toxicity quencher, as initially observed in the *ccdAB* system, is now known to be characteristic of all classified TA systems (Rankin et al. 2012).

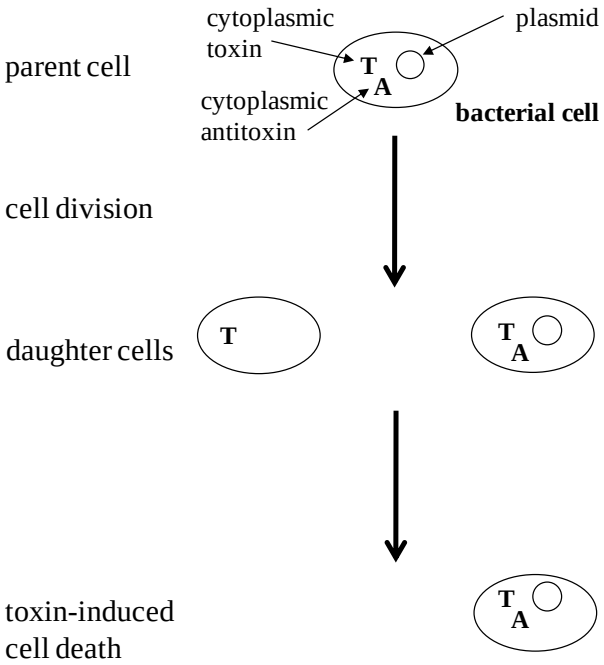


Figure 1.1 Schematic diagram of plasmid addiction showing presence (or absence) of the toxin (T) and antitoxin (A) within the cytoplasm. If cell division results in loss of the plasmid, the antitoxin will be degraded more quickly leaving an unmasked toxin. As a result the daughter cell is killed. Figure adapted from Rankin et al (2012).

Genetic systems exhibiting similar interactions to the *ccdAB* system were later identified on the plasmids of a variety of other bacterial species (Gerdes et al. 1986), and the discovery of chromosomally-encoded TA systems in disparate prokaryotic and archaic taxa suggested they play a broader role in bacterial biology. Interestingly, a number of pathogenic bacterial species exhibit large expansions in the number of TA systems encoded in their genomes, for example *E. coli* 0157 and CFT073, *Streptococcus pneumoniae, Vibrio cholera* and *M. tb* (Pandey & Gerdes 2005).

1.2.2 Toxin-Antitoxin System Classification

TA systems are bicistronic operons identified by their characteristic protein interactions (Anantharaman & Aravind 2003). One gene in the operon encodes a 'toxic' component that targets an essential cellular function such as replication or translation. Ectopic overexpression of the toxin may result in cell death or stasis. The other, generally smaller, gene encodes a cognate 'antitoxin' component that quenches the toxin's deleterious effect. This interaction is depicted in Figure 1.2. A balance of toxicity is achieved through differential stability of the stable toxin and the more unstable antitoxin (Fineran et al. 2009). Greater lability of the antitoxin requires it to be produced at a higher rate in order to maintain constant suppression of the toxin's effects (McKenzie 2011).

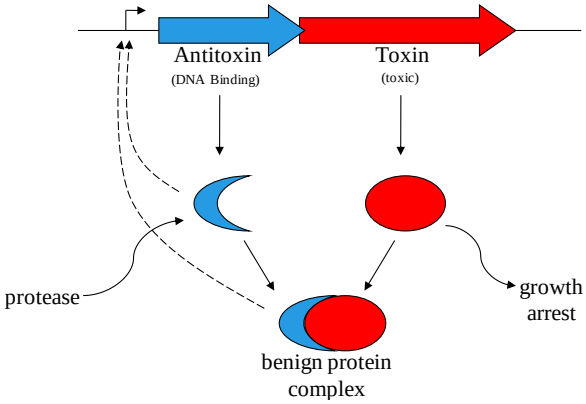


Figure 1.2 Schematic diagram of interactions in a generalised toxin-antitoxin system. Toxin and antitoxin genes are encoded in a two-gene operon. The toxin protein exhibits deleterious effects such as growth arrest when not bound in a benign complex with the antitoxin. The antitoxin alone, and in complex with the toxin, binds to the promoter region of the TA operon. The antitoxin has greater lability and is more susceptible to proteolytic degradation. Figure adapted from McKenzie (2011).

TA systems have been currently classified into five types based on the nature of the interaction between the toxin and antitoxin components (Brantl 2012).

Type I systems have been identified in both Gram-negative and Gram-positive bacteria and are characterised by the antitoxin component acting as a small RNA (sRNA). The antitoxin and toxin gene-pair is either overlapping and convergently transcribed, where the antitoxin is an antisense RNA, or divergently transcribed, where the antitoxin is a trans-encoded sRNA. The antitoxin RNA specifically binds to and inhibits translation of the toxic protein mRNA transcript. Complementary base-pairing between the antitoxin and toxin mRNA forms a double-stranded (ds) RNA region across the ribosomal binding site, which blocks translation initiation and recruits endoribonuclease III for toxin RNA degradation (Brantl 2012). Examples of type I TA systems include the *ldr/rdl* and *tisB/istR-1* systems of *E. coli* (Kawano 2012; Wagner & Unoson 2012).

Type II TA systems are characterised by a smaller, unstable antitoxin protein that sequesters the toxin protein in a benign heterodimeric protein complex. A type II operon normally consists of two genes, where the upstream gene encodes the antitoxin (Van Melderen & Saavedra De Bast 2009). The *vapBC* and *mazEF* systems fall into the class of type II TA systems.

While type I and II TA systems were discovered on plasmids in the 1980s, type III systems weren't identified until 2009 (Fineran et al. 2009). In type III systems, the antitoxin is an RNA containing a pseudoknot motif that directly binds to and inhibits the toxin protein. The antitoxin is encoded by a number of short tandem repeat sequences upstream of the toxin gene. These repeats serve as a transcriptional terminator to regulate the ratio of antitoxic to toxic RNA transcripts (Fineran et al. 2009; Blower et al. 2012). The *toxIN* locus in phytopathogen *Pectobacterium atrosepticum* encodes the first identified type III TA system. Here, the toxin protein, ToxN, is encoded downstream of a 36 bp ToxI antitoxin region encompassing an array of 5.5 tandem repeats (Fineran et al. 2009).

Recently proposed type IV TA systems involve the antitoxin protein quenching the toxin's activity through interfering with the toxin's ability to bind its target. This is most recently observed in the *yeeU-yeeV* system of *E. coli* (Masuda et al. 2012).

The antitoxin component of type V TA systems acts as a ribonuclease (RNase) to cleave the toxin-encoding mRNA. To date, only the *ghoS/ghoT* system falls into this class, and provides the only example of an antitoxin which functions by cleaving toxin mRNA (Wang et al. 2012).

Other possible classes of TA systems include restriction modification (R-M) systems. These systems similarly act as selfish genetic elements and consist of two genes encoding a restriction endonuclease and a DNA methyltransferase. The endonuclease acts as the toxic component, introducing double-strand breaks at short recognition sequences in the DNA provided that these are devoid of methylation. The methyltransferase acts as the cognate antitoxin, methylating and in turn protecting regions of the genome targeted by the endonuclease (Naito et al. 1995). As seen with classical TA systems, there is a degree of multiplicity in the R-M systems encoded in various genomes (Magnuson 2007). For example, the *Helicobacter pylori* J99 genome encodes approximately two dozen R-M loci (Kong et al. 2000).

1.2.3 Cellular Targets and Activity of Toxin Proteins

At present, nine TA families have been classified based on amino acid sequence homology of the toxin component (Van Melderen & Saavedra De Bast 2009). The toxins have a range of targets within the cell, including DNA gyrases (*ccdAB* and *parDE* loci), the translating ribosome *(relBE* loci*)* and RNA (*mazEF* and *vapBC* loci*)* (Bahassi et al. 1999). Thus far, the most common means of toxicity is observed to be non-specific or site-specific mRNA cleavage which, in turn, results in translational inhibition of target transcripts (Van Melderen & Saavedra De Bast 2009). Several toxins displaying RNase activity contain PIN domains, a motif thought to be associated with such a function (Clissold & Ponting 2000). Selective activation of mRNases may allow the bacteria to adapt to stress by altering the transcriptome, allowing rapid changes in metabolic processes (Ramage 2010). Toxin-mediated inhibition of protein synthesis through interaction with RNA has also been observed via an mRNA-cleavage independent route. Here, selective, stable binding of the toxin to the transcript is sufficient for translational inhibition (Sharp et al. 2012). Table 1.1 provides a summary of annotated TA families and the proposed activity of their toxin within the cell.

Table 1.1 The nine toxin-antitoxin families

TA Family	Toxin Target	Activity	Cellular Process
CcdAB	DNA Gyrase	Generates dsDNA breaks	Replication
RelBE	Translating Ribosome	Induces mRNA cleavage	Translation
MazEF	mRNAs	Endoribonuclease	Translation
ParDE	DNA Gyrase	Generates double-stranded DNA breaks	Replication
Phd/doc	Translating Ribosome	Induces mRNA cleavage	Translation
VapBC	mRNAs	Endoribonuclease	Translation
HipBA	EF-Tu	Protein Kinase	Translation
HicRA	ND	mRNA cleavage	Translation
ω-ε-ζ	ND	Phosphotransferase	ND

Abbreviations: ND, not determined; Ccd, coupled cell division; Par, partitioning; doc, death on curing; Vap, virulence associated protein; Hip, high persistence. Table adapted from Van Melderen & Saavedra De Bast (2009).

1.2.4 Regulation of Toxin-Antitoxin systems

1.2.4.1 Operon Organisation

In order to orchestrate tight control of toxin activity within the cell, many levels of regulation are proposed to control the expression of TA operons. In bicistronic TA operons, organisation of the antitoxin upstream of the toxin may be an adaption resulting in a gene dosage effect, where the antitoxin is produced in a higher dose. In agreement with this, a study of the *ccdAB* type II TA system confirmed higher levels of the antitoxin compared to the toxin in *E. coli* (Afif et al. 2001). The authors noted, however, that this observation may have been partially attributed to a post-transcriptional regulation mechanism of 3'-degradation of the TA transcript, as is seen for *kid/kis* mRNA (Ruiz-Echevarría et al. 1995).

1.2.4.2 Antitoxin Auto-Regulation of Transcription

A characteristic feature of most TA systems examined is the interaction between the antitoxin/benign TA complex, and the TA operon promoter (Christensen et al. 2004). This interaction serves to sterically block access of RNA polymerase to the promoter, thereby down-regulating the production of TA operon transcripts and in turn the protein products. Two-component TA systems, such as the *mazEF* and *vapBC* systems, exhibit a characteristic autoregulation interaction where binding occurs between DNA-binding motifs at the antitoxin N-termini, and palindromic/repeat sequences in the promoter DNA.

1.2.4.3 Ratio of Toxin to Antitoxin

The ratio of toxin to antitoxin can play an important role in regulating TA operon transcription. While the antitoxin alone is usually sufficient for repression, in most cases the toxin acts as a co-repressor and enhances binding through strengthening the antitoxin-DNA interaction (Afif et al. 2001; Christensen et al. 2001; Garcia-Pino et al. 2010; Ramage 2010). For example, the FitA antitoxin binds to a target inverted repeat (IR) sequence in its promoter with relatively low affinity (K_d=178 nM), which increases by 38-fold when the FitB toxin forms a heteromeric complex with FitA (K_d=4.5 nM) (Wilbur et al. 2005). 'Conditional cooperativity' describes the phenomenon whereby TA complexes only repress transcription from their cognate promoter when the antitoxin and toxin are present in the correct stoichiometric amounts (Brown et al. 2013). This is the case for the *ccdAB* TA system, where the levels of CcdB toxin can either enhance or inhibit the DNA-binding activity of the antitoxin. When present in equal amount to the CcdA antitoxin, CcdB significantly enhances the affinity of CcdA for DNA. Surpassing this equimolar CcdB:CcdA ratio causes destabilision of the antitoxin-DNA complex (Afif et al. 2001). Conversely, it was recently revealed in the *mqsAR* TA system of *E. coli* that the toxin, MqsR, functions to destabilize the MsqA antitoxin-promoter DNA complex under all conditions and at all toxin:antitoxin stoichiometries. This is the first identified TA system where under no conditions does the toxin act as a transcriptional co-repressor, and instead, actively destabilizes the antitoxin-DNA interaction (Brown et al. 2013).

1.2.5 Proposed Physiological Roles of Toxin-Antitoxin Systems

While the mechanisms of bacterial TA systems have been extensively studied, the physiological roles they play are not universally understood. Widely suggested involvements include in bacterial persistence, programmed cell death and the cellular stress response (Amitai et al. 2004; Pandey & Gerdes 2005; McKenzie 2011). TA-mediated stabilisation of integrons (Christensen-Dalsgaard & Gerdes 2006) and contribution to the biofilm formation process (Garcia-Contreras et al. 2008) have also been proposed.

1.2.5.1 Persistence

It is observed that the majority of bacteria encoding TA operons exhibit a common metabolic state of dormancy or slow growth (Miallau et al. 2009). As

such, TA systems are suspected to play a role in the formation of MDT persister subpopulations. (Balaban et al. 2004; Keren et al. 2004). Studies conducted in *E. coli* demonstrated that deletion of a *hipBA* TA operon resulted in a decrease in MDT formation, while over-expression of the toxin HipA dramatically increased the persister population. Similarly, over-expression of the RelE toxin of the *relBE* TA operon resulted in a 10,000 fold increase in persisters in an *E. coli* bacterial culture post-antibiotic treatment (Keren et al. 2004). Studies of TA systems in *M. tb* have drawn comparable conclusions, for example the expression of an *M. tb* chromosomally encoded TA system (*Rv1102c - Rv1103c*) in *M. smegmatis* was seen to induce growth stasis and led to a significant increase in the persister cell population (Han et al. 2010).

Persistence may result from TA systems conferring a competitive advantage to the genetic element they are located on, as Cooper & Heinemann (2000) observed that plasmids containing a TA system outcompeted the same plasmid lacking the TA over hundreds of generations (Cooper & Heinemann 2000). Alternatively, the toxin's action within the cell may permit entry into a reversible bacteriostatic state in which energy conservation is achieved for persisting through unfavourable environments and/or harsh conditions (Miallau et al. 2009; Rotem et al. 2010).

1.2.5.2 Programmed Cell Death

This theory suggests toxin-induced programmed suicide of a sub-population of cells. The proposal is contested by many, as it is deemed unlikely from an evolutionary stand-point that a single-celled organism would encode a multitude of suicide elements as observed in organisms such as *M. tb* (Pandey & Gerdes 2005). Nonetheless, Mutschler et al (2011) suggest a mechanism of programmed cell death employed by zeta toxins, members of the epsilon/zeta TA family, in *E. coli*. The mechanism involves interference with MurA, an enzyme involved in peptidoglycan synthesis for bacterial cell wall production. The interaction was confirmed on solving the crystal structure of the zeta toxin bound to MurA (Mutschler et al. 2011).

1.2.5.3 Cellular Stress Response

Multiple studies have concluded that TA systems function in response to stressors to induce the NRP state when environmental conditions are unfavourable (Gerdes

2000; Gerdes et al. 2005). In response to stress signals, bacterial cells may trigger toxin activity in order to manipulate expression within the cell, tailoring the transcriptome to include only transcripts required for the current stress. This metabolic reprogramming would result in more efficient energy expenditure (Gerdes et al. 2005; Condon 2006; Ramage et al. 2009). A number of studies probing the importance of TA systems in *E. coli* have provided support for this hypothesis. Two TA systems in strains of extraintestinal pathogenic *E. coli* (*yefM-yoeB* and *ybaJ-hha*) have been associated with promoting colonization of the host bladder, while the *pasTI* TA system was found to be essential for survival within the host kidneys. Low-level expression of the PasT toxin in such strains significantly increased survival in response to nitrosative and oxidative stresses as well as during nutrient limitation (Norton & Mulvey 2012). RNA cleaving toxins such as SymE in *E. coli* have been proposed to be activated in response to the cellular stress response, playing a role in recycling damaged mRNAs that have been produced as a result (Kawano 2012).

Taken together, the observations made in *E. coli* along with recent research on *M. tb* (Arcus et al. 2005; Ramage et al. 2009; McKenzie 2011) have led to the hypothesis that the numerous TA proteins encoded within the *M. tb* genome are involved in the transitions between active and persistent/dormant phases of *M. tb* growth. The working model is that under adverse conditions, specific proteases are activated within the *M. tb* cell that break-down the antitoxin at a greater rate than the toxin, allowing the toxin to be released and exhibit its toxic effect to curtail bacterial replication. Cellular processes may manipulate the balance of toxin to antitoxin produced in order to fine-tune the onset and duration of the persistent state until more favourable conditions for growth are encountered (Rotem et al. 2010).

1.2.6 Distribution of Toxin-Antitoxin Systems in Mycobacteria

TA systems are ubiquitous in the genomes of free-living prokaryotes, notably so in many bacteria with pathogenic properties (Gerdes et al. 2005). The *M. tb* genome has been found to encode 88 putative TA loci (Ramage et al. 2009; Frampton et al. 2012; Min et al. 2012) and many interpret this massive expansion as an indication of their involvement in TB pathogenesis (Ramage et al. 2009). Of the identified TA systems encoded in the *M. tb* genome, 37% have been

mapped to regions noted as sites of horizontal gene transfer (HGT), suggesting that the mycobacteria may have originally acquired these TA genes from other species (Arcus et al. 2011). The majority of *M. tb* TA systems are found to be conserved only in mycobacteria with pathogenic properties, indicating that attaining these systems was an important step in evolution of *M. tb* (Ramage et al. 2009). Why *M. tb* has so many TA systems and the exact physiological role that they play in the cell is unclear. Expression profiling of *M. tb* TA systems under a number of environmental conditions, revealed four that were selectively upregulated during either hypoxic stress (*Rv2009-2010* and *Rv1955-1956*) or phagocytosis into macrophages (*Rv1560-1561* and *Rv0549c-0550c)* (Ramage et al. 2009). Interestingly, different TA systems were seen to be upregulated in response to different stressers, suggesting that specific subsets of TA systems are triggered in response to certain environmental changes as opposed to global/genome-wide TA system activation. Together these results hint at TA systems' involvement in *M. tb* cellular processes for overcoming stresses which are likely encountered *in vivo.*

1.3 The VapBC Toxin-Antitoxin Family

The *vapBC* (virulence associated proteins B and C) operons make up the largest of the nine identified TA families, however little is known about their physiological role. The association of many *vapBC* operons with repetitive elements, transposases and virulence factors implies a contribution to virulence (Arcus et al. 2005; Ramage et al. 2009). With the increasing technical feasibility and affordability of whole genome sequencing, a range of organisms encoding *vapBC* TA systems at varying abundance is being revealed. Noteworthy organisms displaying a large number include *Microcystis aeruginosa* (34 loci) *Pyrococcus kodakaraensis* (28 loci) and *Sulfolobus tokodaii* (24 loci) (Arcus et al. 2011). Among the mycobacteria, the expansion of *vapBC* systems is restricted to mammalian pathogenic species such as *M. tb* where at least 47 *vapBC* loci have been identified (Ramage et al. 2009; Ahidjo et al. 2011). In contrast, non-pathogenic *Mycobacterium smegmatis* and the pathogen *Mycobacterium leprae* reportedly encode one and none *vapBC* loci respectively. A mechanism of HGT among the gene pool may explain observations of homologous TA elements occurring between phylogenetically distant organisms,

while showing little conservation between isolates belonging to the same bacterial species (Arcus et al. 2005; Van Melderen & Saavedra De Bast 2009).

The VapBC TA family is characterised by the toxic component (VapC) belonging to the PilT N-terminal domain (PIN domain) family of proteins, while the anti-toxin component (VapB) falls into the protein family of transcription factors.

1.3.1 PIN-domain Proteins

The PIN-domains constitute an extensive protein family, with members identified in all three domains of life. Despite there being poor sequence conservation across the family, the 11 PIN-domain structures in the Protein Data Bank exhibit high conservation in three-dimensional structure (Arcus et al. 2011). PIN-domains possess a triad of strictly conserved acidic residues along with a fourth less well conserved acidic residue (Figure 1.3).

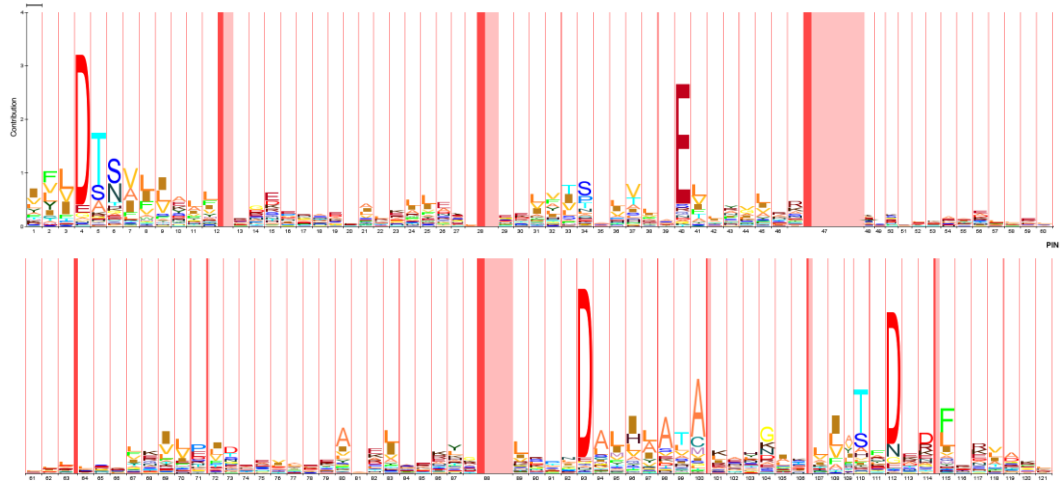


Figure 1.3 The Hidden Markov Model which defines the PIN-domain family of proteins. The height of a letter represents the proportion of information it provides with respect to its position in the PIN-domain family. Column widths indicate the importance of this position in defining the family. Dark and light pink columns represent sequence regions containing insertions. Taken from PFam (Finn et al. 2010) using HMM Logo (Schuster-Bockler et al. 2004).

An additional feature of PIN-domain proteins is the positioning of a polar residue (Asn, Ser or Thr) directly following the first conserved Asp residue (Arcus et al. 2011). The conserved tertiary structure positions the quartet of conserved residues (three asparagines (D) and one glutamine (E)) and the polar residue in close spatial proximity within a negatively charged pocket constituting the active site (Figure 1.4). Positioned here, these residues play an important role in sequestering Mg^{2+} ions that are required for catalytic activity (Arcus et al. 2011).

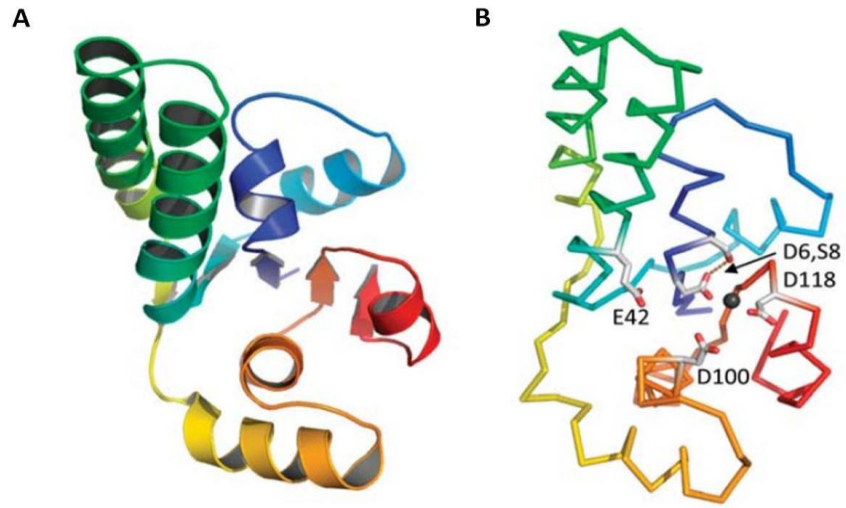


Figure 1.4 Structure of the PIN-domain. A, A cartoon schematic of the PIN-domain coloured blue to red from N- to C-terminus. **B**, The same structure as A, indicating the four conserved acidic residues, the conserved polar residue and a black Mg^{2+} ion. Figure taken from Arcus et al (2011).

The Pfam database currently lists 6253 annotated PIN-domain protein sequences occurring in 1341 species (Finn et al. 2010). The number of PIN-domain sequences mapped to the *M. tb* genome is notably higher than that found in any other studied organism. The PIN-domain motif is thought to be associated with ribonuclease function (Clissold & Ponting 2000; Ramage 2010). Accordingly, in mycobacteria, PIN-domain proteins exhibit their toxicity through endogenous ribonuclease activity, cleaving ssRNA in a sequence-specific and Mg^{2+}/Mn^{2+}-dependent manner (Robson et al. 2009).

1.3.2 Ribonuclease Activity of VapC Proteins

1.3.2.1 Proposed Catalytic Mechanism of VapC

The definitive catalytic mechanism of PIN domain proteins is currently unknown, however recently, two mechanisms for catalysis of ssRNA cleavage by VapC PIN-domain proteins were proposed. The data support a reaction proceeding via a one metal, or two metal catalysis mechanism (Figure 1.5). Crystal structures of $VapC_{PAE0151}$ from *Pyrobaculum aerophilum* reveal the first metal to be coordinated by the second and third conserved aspartate residues, and this metal is implicated in the activation of the water nucleophile. This cleavage mechanism is believed to leave ssRNA products with a 5'-terminal monophosphate group, as is indicated by matrix-assisted laser desorption ionization time of flight mass spectrometric (MALDI TOF MS) analysis of VapC RNase assay cleavage products (McKenzie 2011; Duyvestyn 2012).

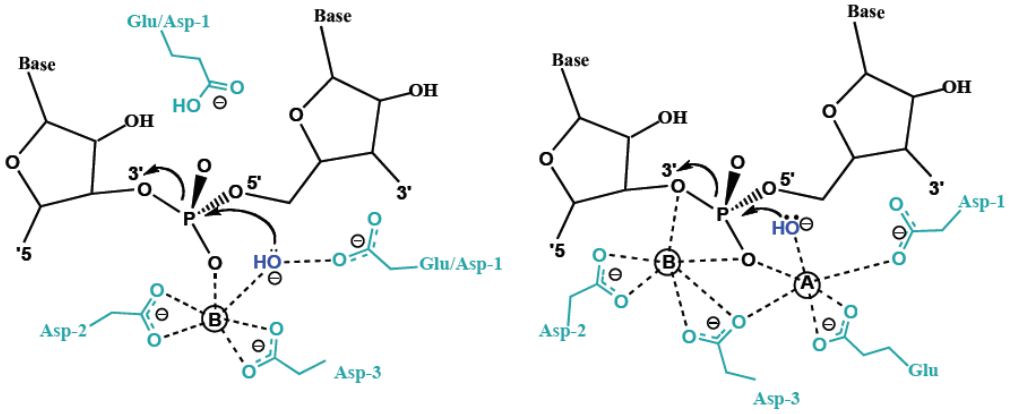


Figure 1.5 Proposed catalytic mechanism of PIN-domains. A, One-metal catalysis mechanism. **B**, Two-metal catalysis mechanism. Conserved acidic residues of the active sites are shown in cyan, the nucleophile in blue, the RNA substrate in black and the metal ion as a labelled black circle. Arrows represent electron movement. Figure taken from Duyvestyn (2012).

1.3.2.2 Screening for VapC Sequence-Specificity using Pentaprobes

It is theorised that the four conserved acidic residues in the PIN-domain active site of VapC enzymes interact with substrate ssRNA in a manner that allows specific bases to be recognised (Duyvestyn 2012). A sensitive and robust method for the determination of RNase sequence-specificity of a number of VapC proteins has recently been developed (McKenzie et al. 2012b). This method utilizes RNA Pentaprobes, and smaller RNA oligos of parent Pentaprobes, as substrates to probe the specificity of VapC. Together, twelve Pentaprobe RNA sequences encode every combination of five bases. Post-assay, VapC RNase reactions are subjected to MALDI TOF MS analysis in order to detect the masses of substrate cleavage products. MALDI TOF MS analysis is particularly suited to RNA molecules as they are less prone to the loss of bases during the ionization process which occurs with DNA (Nordhoff et al. 1993). MALDI TOF MS data can then be used to infer VapC cut sites in the substrate RNA sequence. An outline of the process from VapC purification to cut site determination is detailed in Figure 1.6.

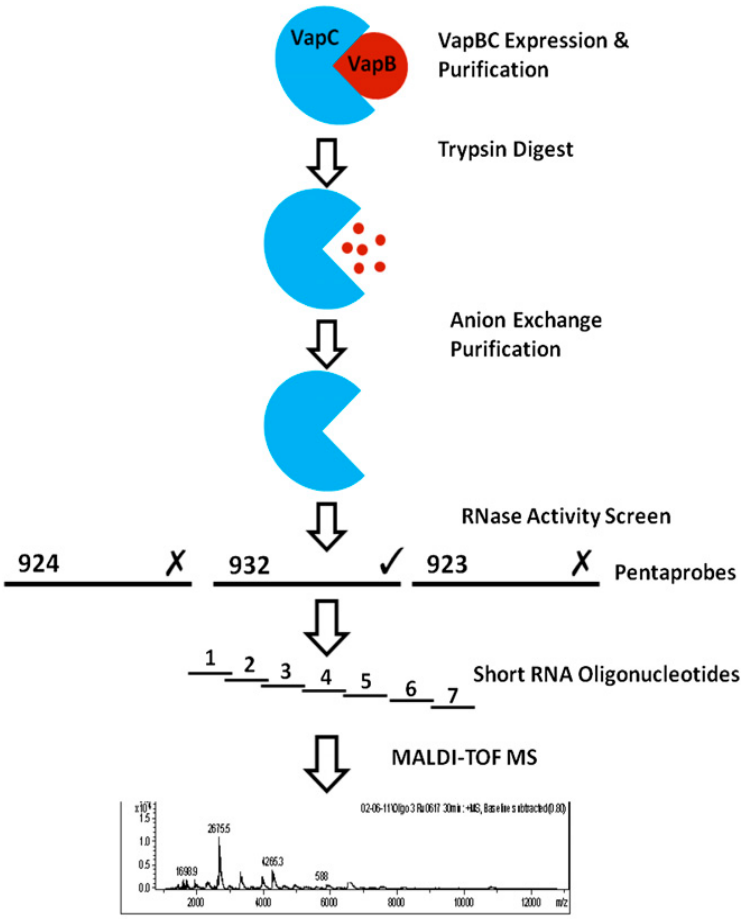


Figure 1.6 Schematic diagram of VapC purification and sequence specificity analysis. VapC toxicity is neutralised through co-expression with its cognate antitoxin VapB. VapB is removed through a trypsin mediated proteolytic degradation and the remaining VapC is purified by anion exchange chromatography. RNase sequence-specificity and activity is screened using Pentaprobe RNA and short Pentaprobe oligonucleotide RNA (ticks or crosses indicate cleavage or no cleavage respectively). Oligo RNA cleaved is examined by MALDI TOF MS and in-house software is used to determine VapC cut sites. Figure adapted from McKenzie et al (2012b).

1.3.2.3 The Impact of Secondary Structure on RNase-Specificity

It has recently been proposed that recognition of secondary structural motifs in the RNA substrate is an important contributor to the specificity of VapC-mediated mRNA degradation (McKenzie et al. 2012a; McKenzie et al. 2012b). Using the Pentaprobe system as described above, it was seen that $VapC_{MS1284}$ from *M. smegmatis* cleaved full Pentaprobe RNA far more efficiently than it cleaved its corresponding shorter oligonucleotides. This was attributed to differences in secondary structure at or around the VapC recognition sites in the shorter RNA fragments (McKenzie 2011; McKenzie et al. 2012a). Microarray identification of *in vivo* $VapC_{MS1284}$ RNA targets and their subsequent testing in RNase assays *in vitro* revealed that the three target transcripts experimentally characterised to be efficiently cut appeared to share the same basic predicted secondary structure: a stable hairpin immediately downstream of the target cut sequence. A sequence alignment and consensus secondary structure was created based on these three substrates (Figure 1.7).

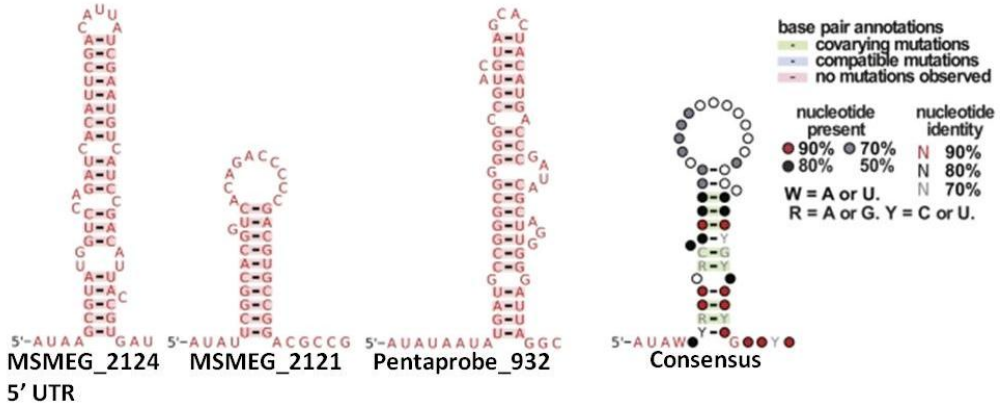


Figure 1.7 Secondary structure characterisation of VapC targets. The predicted RNA secondary structures for the three experimentally validated $VapC_{MS1284}$ targets found in the MSMEG_2124 5′ untranslated region (UTR), MSMEG_2121 mRNA, and the artificial Pentaprobe 932 sequence along with a consensus model of predicted $VapC_{MS1284}$ targets. Figure adapted from McKenzie, et al. (2012a).

Further investigation found that substrates containing the predicted optimal consensus structure were cleaved more efficiently than those with other secondary structures. It is currently theorised that the optimal VapC recognition motif is likely to be a combination of the correct target sequence along with a certain secondary structure (McKenzie et al. 2012a).

1.3.3 Characterised Autoregulation in *vapBC* Toxin-Antitoxin Systems

Transcriptional autoregulation of *vapBC* operons through binding of the VapBC protein complex to the *vapBC* promoter DNA has been noted in a number of systems from disparate taxa.

1.3.3.1 The fitAB *System of* Neisseria gonorrhoeae

The *fitAB* (fast intracellular trafficking) systems consist of a VapB antitoxin (FitA), and a VapC toxin (FitB). FitAB-mediated transcriptional autoregulation of the *fitAB* operon in the *Neisseria gonorrhoeae* genome has been investigated. In this system, four FitA antitoxins and four FitB toxins form a tetramer of heterodimers that binds to the *fitAB* promoter dsDNA. The N-terminus of the FitA antitoxin contains a ribbon-helix-helix (RHH) domain that was found through DNAse I footprint assays to bind and protect a 62 bp region within the promoter. This region includes an 8 bp IR, $^{5'}$TGCTATCA-N_{14}-TGATAGCA$^{3'}$, and covers the putative -10 promoter sequence for the operon (Wilbur et al. 2005). Figure 1.8 provides a structural representation of FitA and FitB bound to promoter DNA and indicates the positioning of the IR within this region.

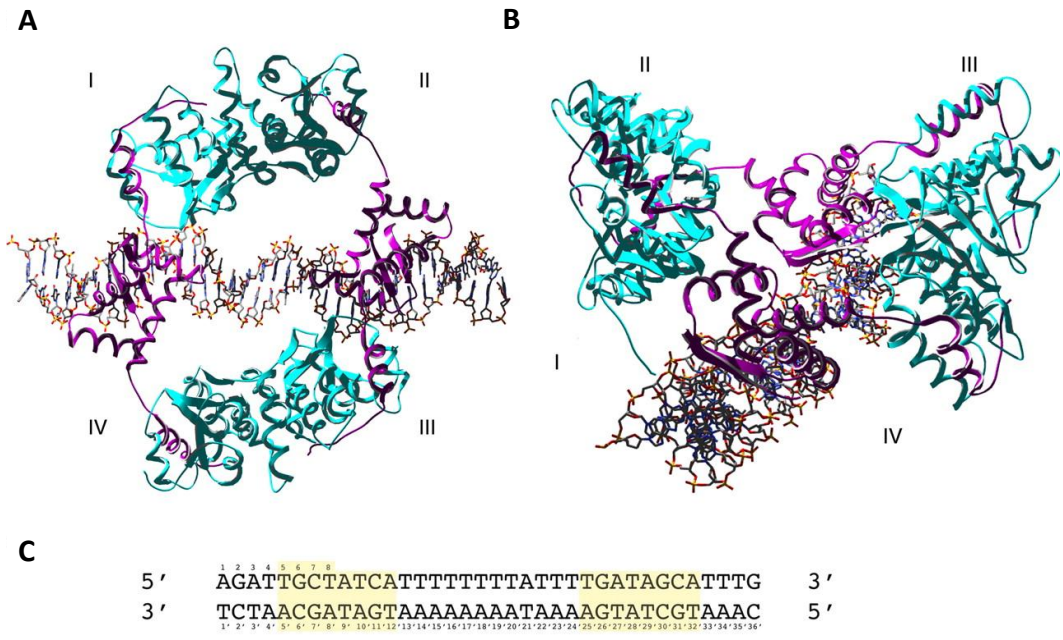


Figure 1.8 Structure of the FitAB-IR36 complex. A, a ribbon diagram of FitA (magenta) and FitB (cyan) bound to a 36 bp region of the *fitAB* promoter DNA. **B**, a rotated view of A, depicting two FitA β-sheets binding to the same DNA helix face. **C**. Sequence of the 36 bp promoter DNA region with the 8 bp IR emphasised in yellow. Figure taken from Mattison et al. (2006).

On evaluation of structural data for crystals of FitAB bound to a 36 bp region of the *fitAB* promoter it was seen that specific interactions between antitoxin residues and DNA bases were only being made at six of the eight nucleotide positions in each IR half, with interactions at the remaining two nucleotides being made non-specifically between amino acid residues and the DNA phosphate backbone (Mattison et al. 2006). This observation suggests that the antitoxin of VapBC systems may potentially recognise and bind to both perfect and near-perfect palindromic sequences in the cognate promoter. The following model has been proposed for how the FitAB-promoter interaction contributes to the pathogenic physiology of *N. gonorrhoeae* (Mattison et al. 2006): when in an extracellular environment, FitAB is bound to promoter DNA, repressing transcription of the *fitAB* operon. Upon epithelial cell invasion, FitAB releases the DNA, allowing *fitAB* operon expression and slowing *N. gonorrhoeae* replication.

1.3.3.2 The vapBC2 *System of* Rickettsia felis

Investigation of the *vapBC2* system from *Rickettsia felis* revealed that a VapBC2 octomer binds a 27 bp region of dsDNA in its own promoter. Within this region, a near-perfect palindrome with half-sequences of $^{5'}$TA(A/G)TATATA(C/T)TAA$^{3'}$ was suggested as being essential for VapBC2 binding (Mate et al. 2011). As in the

case of FitAB bound to DNA, the VapBC2 DNA binding complex consists of four toxins and four antitoxins forming a tetramer of heterodimers, arranged circularly via alternative antitoxin-antitoxin and toxin-toxin dimerisation interfaces (Figure 1.9).

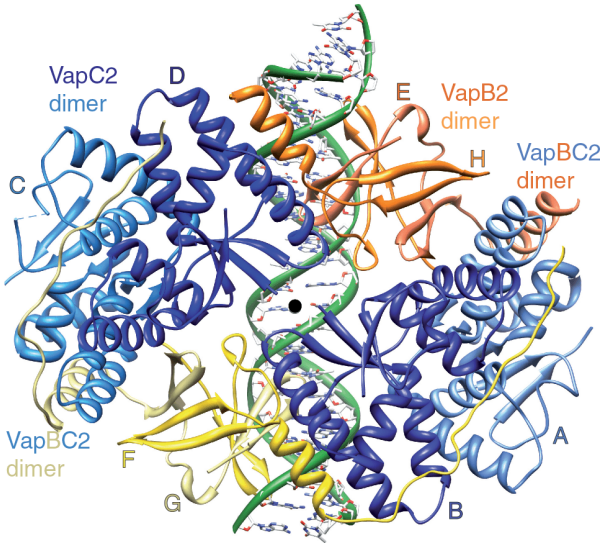


Figure 1.9 The VapBC2-DNA complex. VapB2 antitoxins are depicted in shades of blue, VapC2 toxins in shades of yellow and orange. Toxin and antitoxin dimers are labelled in relation to the colour of their counterparts. Monomers are labelled A – H. Figure adapted from Maté et al. (2011).

The antitoxins interact with the concave face of the curved DNA, where one VapB2 homodimer is seen to directly recognize one region within the near-perfect palindrome, while a second region is recognised by the other VapB2 homodimer. In contrast to the FitAB system, the DNA-binding domain of the VapB2 antitoxin does not display a RHH domain, rather VapB2 homodimers form a swapped-hairpin β–barrel fold as is observed for MazE (Mate et al. 2011).

1.3.3.3 The ntPR *System of* Sinorhizobium meliloti

The *ntrPR* operon of *Sinorhizobium meliloti* encodes a protein pair that forms a TA module belonging to the VapBC family. Transcription of this operon is found to be negatively autoregulated by binding of the NtrPR complex to a 141 bp region of the *ntrPR* promoter through DNA-binding domains at the N-terminal of NtrP. DNase I footprint experiments demonstrated that within the promoter, a 10 bp direct repeat sequence ($^{5'}$GGCATATACA-N_4-GGCATATACA$^{3'}$) was the specific binding site of NtrP. Electrophoretic mobility shift assay (EMSA) testing of DNA oligonucleotides containing either half of the direct repeat revealed that one half is sufficient to sustain VapBC binding. The direct-repeat overlaps the

annotated transcriptional start-site, suggesting VapBC repression through steric inhibition of RNA-polymerase (Bodogai et al. 2006).

1.3.3.4 *The* vapBC-1 *System of Nontypeable* Haemophilus influenza

Expression of the *vapBC-1* TA operon in this organism is similarly seen to be repressed by the VapBC complex. Unlike previously described systems, *vapBC-1* is unique in that both the VapBC-1 complex and VapC-1 alone are observed to exhibit DNA binding activity to the promoter dsDNA, recognizing a perfect IR ($^{5'}$GTATATACT-N_{12}-AGTATATAC$^{3'}$). Altering nucleotide sequence in one half of the IR substantially diminishes binding of VapBC-1, while alteration in both halves completely abolishes VapBC-1-promoter DNA interaction. The binding requirement for both halves of the IR indicates recognition by a dimer. Surprisingly, VapB-1 alone displays no DNA interaction with the promoter region even at high concentrations; therefore its proposed role is in the direction of VapC-1 to the promoter binding site (Cline et al. 2012).

1.4 The *vapBC* Toxin-Antitoxin Systems in Mycobacteria

1.4.1 The *vapBC* System of *Mycobacterium smegmatis*

M. smegmatis, a non-pathogenic relative of *M. tb*, has been used as a model system for studying VapBC proteins as it only possesses a single *vapBC* operon. This, and its close relatedness to *M. tb*, makes *M. smegmatis* an ideal system for gathering results that may be extrapolated to *M. tb*, leading to evidence as to the corresponding function for homologous TA elements in the *M. tb* genome. The *MSMEG_1283/4* operon of *M. smegmatis* encoding the *vapBC* system is found to be constitutively expressed at low levels through transcription as a leaderless mRNA throughout the growth cycle (McKenzie 2011).

1.4.1.1 VapC Toxicity

A number of studies have used the overexpression of the toxin, often in a heterologous host, to demonstrate toxicity of the toxin module. In such an approach, Robson et al (2009) investigated toxicity of the *M. smegmatis* chromosomally encoded $VapC_{MS1284}$ protein when conditionally expressed under the control of a tetracycline-inducible promoter. The effect of $VapC_{MS1284}$ overexpression was monitored in both an *M. smegmatis* mc^2155 wild-type strain and an *M. smegmatis* $mc^2155\Delta$*vapBC* knockout strain. The effect of

vector-derived $VapC_{MS1284}$ overexpression was seen to be more marked in the knockout *M. smegmatis* strain than in the wild-type. *M. smegmatis* mc^2155 expressing $VapC_{MS1284}$ exhibited a 10-fold reduction in colony forming units (CFU)/ml compared with strains expressing either the empty vector or VapBC complex, while the knockout strain exhibited a 100-fold difference when $VapC_{MS1284}$ expression was induced. These observations suggest an ability of both endogenous $VapB_{MS1283}$ in the wild-type strain and co-expressed vector-derived $VapB_{MS1283,}$ to offset the toxic effect of induced vector-derived $VapC_{MS1284}$. Strains expressing $VapC_{MS1284}$ only formed barely-visible, pinpoint-sized colonies. A VapC-induced alteration in cell morphology was also noted, as electron microscopy revealed that the length of single cells was slightly increased in $VapC_{MS1284}$-expressing strains compared to the lengths seen in strains expression $VapBC_{MS1283/4}$ or the empty vector (Robson 2010).

Further investigations supported a mechanism of $VapC_{MS1284}$-mediated growth inhibition and toxicity in *M. smegmatis* through inhibition of translation rather than transcription or DNA replication, a result consistent with the RNase activity of VapC enzymes (Robson et al. 2009). Observed toxicity in growth experiments was attributable to the presence and effects of the VapC protein rather than the mRNA transcript through identification of $VapC_{MS1284}$ protein *in vivo* via Western blot using polyclonal antibodies raised against purified $VapC_{MS1284}$. The $VapC_{MS1284}$ translational inhibition was classed as bacteriostatic, as removal of tetracycline reversed growth inhibition. Observation of this in the knockout strain indicated that $VapB_{MS1283}$ was not involved in this reversal process. An analogous phenotype of toxin-mediated reversible growth arrest in absence of the antitoxin has been observed for the *M. tb relBE* TA system when it is overexpressed in *M. smegmatis* strains (Korch et al. 2009). These findings allude to the presence of currently uncharacterised intracellular mechanisms for degradation of the toxin (Robson et al. 2009).

1.4.1.2 VapC RNase-Specificity

The toxic activity of $VapC_{MS1284}$ was found to be mediated by inhibition of translation through an endoribonuclease mechanism targeting AUA(U/A) sequences on ssRNA *in vitro* (McKenzie 2011). $VapC_{MS1284}$ has been demonstrated to both cleave and bind to RNA, where binding was indicated by a

smearing of the RNA in assay lanes on urea denaturing gels where $VapC_{MS1284}$ was present (McKenzie et al. 2012a). Microarray analysis of the transcriptome in response to $VapC_{MS1284}$ overexpression revealed that downregulated transcripts and their upstream regions were rich in the AUA(U/A) $VapC_{MS1284}$ recognition sequence. Among the downregulated transcripts were those with annotated roles in carbon metabolism and transport (Robson 2010). These transcripts play a fundamental role in the utilization of glycerol during aerobic respiration. Through regulating the mRNA involved in this process, $VapC_{MS1284}$ may be playing a role in balancing anabolic and catabolic processes within the cell.

1.4.1.3 Intracellular Roles of VapB and the VapBC Complex

In concordance with the classical TA system interactions model, an autoregulatory role has been observed for the $VapBC_{MS1283/4}$ complex of the *MSMEG_1283/4* operon. Three-fold higher expression levels were observed for a *promoter-vapB-lacZ* fusion construct in a Δ*vapBC* deletion strain than in the wild-type strain, indicating that the *MSMEG_1283/4* operon is subject to autoregulation, and that either one or both TA proteins are required in this system for repression (Robson et al. 2009).

Supporting this, EMSAs revealed that the purified $VapBC_{MS1283/4}$ complex specifically binds a 126 bp region of *MSMEG_1283/4* promoter DNA. Within this region, both halves of a 6 bp perfect IR ($^{5'}$TATAGA–N_{13}-TCTAT$^{3'}$) are essential for VapBC binding. The IR overlaps both -35 and -10 promoter elements (Robson et al. 2009).

In addition to comprising a component of the repressor complex in autoregulation, $VapB_{MS1283}$ is suggested as being a factor required for translation of $VapC_{MS1284}$ mRNA transcript. A lack of intracellular copies of $VapC_{MS1284}$ protein in *M. smegmatis* mc^2155 Δ*vapB* deletion strains has been noted, despite the occurrence of elevated *MSMEG_1284* transcript levels (Robson 2010). This highlights the importance of RNA processing and translational coupling in proper VapC expression from the *M. smegmatis MS1283/4* operon.

1.4.2 The *vapBC* Systems of *Mycobacterium tuberculosis*

After extensive investigation of the *M. tb* chromosomally encoded TA systems, it was found that of the 88 putative systems, at least 47 belong to the VapBC family

(Ramage et al. 2009; Ahidjo et al. 2011). This represents the highest number of *vapBC* systems seen among intracellular pathogens, yet despite this, their roles in *M. tb* physiology are unclear and the evolutionary processes that have led to this great expansion of *vapBC* elements in the *M. tb* genome remain enigmatic. Many of the *vapBC* operons in the *M. tb* genome are found to be linked to repetitive elements, transposases and virulence factors, suggesting their possible origin via HGT. While originating from mobile elements, it may be the case that through the course of evolution these systems have adopted important functional roles within the cell (Arcus et al. 2005).

There are conflicting reports of the occurrence of 'cross-talk' between *M. tb* TA system components of separate operons. In a study conducted by Zhu et al (2010), several non-cognate TA interactions were observed in *M. tb*, both within and between different TA families (e.g. MazF toxins and VapB antitoxins). If such a phenomenon is enlisted, this could add an additional layer of sophistication to the TA network and in turn to the potential metabolic regulatory control that the organism can exhibit. In contrast to these findings, examinations of *M. tb vapBC* systems by Ahidjo et al (2011) and Ramage et al (2009) found *M. tb* VapB antitoxins to be specific to their cognately encoded VapC toxins, whereby the cognate VapB was the only antitoxin able to rescue VapC-induced growth inhibition (Ramage et al. 2009).

As previously mentioned, a functional role for certain *M. tb* VapBC homologues (*Rv2009-Rv-2010, Rv1955-1956, Rv1560-1561* and *Rv0549-Rv0550*) in the stress response has been suggested given the observation of their increased activation under hypoxic conditions and macrophage infection (Ramage et al. 2009). Structures have been solved for three VapBC TA complexes from *M. tb* (Miallau et al. 2009; Min et al. 2012) which illustrate a tight binding between VapB and VapC as a tetramer of VapBC heterodimers, similar to the structures observed for the FitAB complex from *N. gonorrhoeae* (Mattison et al. 2006; Arcus et al. 2011) and the VapBC2 complex from *R. felis* (Mate et al. 2011).

1.4.2.1 VapC Toxicity

The physiological role of *M. tb vapBC* TA systems has been probed through observing the toxicity of VapC proteins on *M. smegmatis* and/or *M. tb* cell

cultures. A toxicity survey of ten *M. tb* VapC toxins has been conducted by Ahidjo et al (2011). In this work, putative toxin or putative TA pairs were conditionally expressed under the control of a tetracycline-inducible promoter and effects on *M. smegmatis* and *M. tb* growth were observed. The screen identified four VapC toxins (Rv0549c, Rv0595c, Rv2549c and Rv2829c) which displayed such toxicity. Toxicity of Rv3320c was also revealed when this VapC was constitutively expressed (Ahidjo et al. 2011). The effects of *M. tb* VapC expression on growth of *E. coli, M. smegmatis* and/or *M. tb* has also been reported by other groups (Gupta 2009; Ramage et al. 2009). While there is some concordance among results, for many VapCs there are conflicting reports of toxicity. This is thought to be due to differences in expression vectors, growth conditions, strains used or translation initiation signals. Reported lack of toxicity for certain *M. tb* VapC proteins has been suggested to be due to one or more of these factors leading to reduced protein expression and, in turn, insufficient VapC levels required for any evident effect. This has led to an incorrect classification as 'non-toxic'. Indeed, it was found that Rv2549c growth inhibition was only observed once the induced protein level exceeded a certain threshold (Ahidjo et al. 2011). As such, it is likely that the current number of VapC proteins described as 'toxic' is an under-representation.

1.4.2.2 The Rv0065a/c *and* Rv0617a/c vapBC *Systems*

The *Rv0065a/c* and *Rv0617a/c* operons of *M. tb* encode functional TA systems belonging to the VapBC family. $VapC_{Rv0065}$ and $VapC_{Rv0617}$ are Mg^{2+}-dependent, sequence-selective ribonucleases whose activity is selectively and completely inhibited by their cognate antitoxin (Ahidjo et al. 2011). $VapC_{Rv0065}$ shares 26% amino acid identity and 50% amino acid similarity with the toxic *M. tb* $VapC_{Rv0549c}$ protein, and $VapC_{Rv0617}$ shares 38% sequence identity and 55% sequence similarity with the toxic *M. tb* $VapC_{Rv3320c}$ protein (Ahidjo et al. 2011).

$VapC_{Rv0065}$ and $VapC_{Rv0617}$ proteins have been observed to target ssRNA rather than dsRNA, ssDNA or dsDNA (Ahidjo et al. 2011; McKenzie et al. 2012a). While it has been noted that VapC proteins of the *Shigella flexneri* 2a virulence plasmid pMYSH6000 and *Salmonella enteric* serovar Typhimurium LT2 *vapBC* operon are tRNases, cleaving initiator $tRNA^{fMet}$ between the anticodon stem and loop (Winther & Gerdes 2011), there has been no evidence of such an activity for

mycobacterial VapC proteins $VapC_{MS1284}$, $VapC_{Rv0065}$ or $VapC_{Rv0617}$ (McKenzie et al. 2012a).

A general method for the overexpression and purification of TB VapC proteins has been developed, and the Pentaprobe system (Section 1.3.2.2) applied to determine their RNase sequence-specificity (McKenzie et al. 2012b). H37Ra genomic DNA was used as template for the cloning of these operons into expression vectors as the *Rv0065a/c* and *Rv0617a/c* operons display 100% sequence conservation between H37Ra and H37Rv strains. A combination of RNase assays against the twelve Pentaprobes sequences and their shorter oligos and subsequent MALDI TOF MS analysis of cleavage products led to the conclusion that $VapC_{Rv0065}$ and $VapC_{Rv0617}$ have the same sequence-specificity, targeting GC-rich 4-mers in the mRNA. Table 1.2 provides a summary of the RNase properties determined for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ along with those for $VapC_{PAE0151}$ from *P. aerophilum.*

Table 1.2 Sequence-specificity of three characterised VapC proteins

VapC ORF	Organism	Pentaprobes cut	932 oligos* cut	Sequence-specificity
Rv0065	*M. tb*	922 923 924 925 926 927 932	1-7	(G/C)G(G/C)(G/C/A)
Rv0617	*M. tb*	922 923 924 925 926 927 932	1-7	(G/C)G(G/C)(G/C/A)
PAE0151	*P. aerophilum*	922 923 924 925 927 932	3-6	GG(U/G)G

*A schematic diagram of the 932 Pentaprobe and oligonucleotides is included in Appendix C5. Table adapted from McKenzie et al (2012b)

$VapC_{Rv0065}$ and $VapC_{Rv0617}$ produce near identical banding patterns and fragment masses when assay products for the same substrate are run on urea denaturing gels or analysed by MALDI TOF mass spectrometry respectively (McKenzie et al. 2012b). It has been noted that there is a level of redundancy in their RNAse specificity, whereby $VapC_{Rv0065}$ and $VapC_{Rv0617}$ may cleave a range of other sequences albeit with less efficiency and at a slower rate (McKenzie et al. 2012b). The apparent lack of individuality in RNase specificity between the *M. tb* VapC proteins, along with their GC-rich targeting within the GC-rich *M. tb* genome (GC content of 66%), is confounding, suggesting a potentially wide range of RNA transcripts in the organism containing the target sequence. (McKenzie et al. 2012a). This specificity contrasts that which was previously determined for $VapC_{MS1284}$ of *M. smegmatis*, and unlike $VapC_{MS1284,}VapC_{Rv0065}$ and $VapC_{Rv0617}$ exhibit no binding activity against ss or dsRNA (Ahidjo et al. 2011).

In concordance with conclusions drawn from the *M. smegmatis vapBC* system, it has been proposed that in addition to recognition of a primary sequence, $VapC_{Rv0065}$ and $VapC_{Rv0617}$ require an additional layer of target recognition residing in the mRNA secondary structure (McKenzie et al. 2012b). Whether these two VapC target the same or unique RNA secondary structural motifs remains to be determined.

1.5 Research Objectives

The abundance of *vapBC* operons in the *M. tb* genome has led us to focus on the biological roles of the proteins that they encode. Elucidating such roles for the array of approximately 47 *vapBC* TA systems encoded in the *M. tb* genome presents a challenging task. As preliminary work on the RNase specificity of VapC proteins encoded from the *Rv0065a/c* and *Rv0617a/c vapBC* operons had been conducted, these two systems were chosen to be investigated further for this thesis research.

The aim of this thesis was to determine the phenotypic effect of these two *vapBC* systems and investigate their underlying biochemistry. The main objectives to fulfil this aim were:

1) To determine the *in vivo* phenotypic effect of $VapC_{Rv0065}$ and $VapC_{Rv0617}$ conditional expression in *M. smegmatis* strains.
2) To investigate the specificity and activity of the $VapC_{Rv0065}$ and $VapC_{Rv0617}$ RNases.
3) To investigate the DNA-binding activity and thus autoregulatory role of the VapBC complex in the *Rv0065a/c* and *Rv0617a/c* TA systems.

Chapter Two: Materials & Methods

Tables of bacterial strains, plasmids, buffers and media used in this study can be found in Appendix A. Sequencing results can be found in Appendix B. Protein and RNA information can be found in Appendix C.

2.1 General Materials and Methods

2.1.1 DNA Manipulations

2.1.1.1 Mycobacterium tuberculosis *Genomic DNA*

M. tb H37Ra strain genomic DNA was obtained from the Proteins & Microbes Lab, University of Waikato.

2.1.1.2 Escherichia coli *Plasmid DNA Extraction*

E. coli plasmid DNA was extracted from 3 ml of an overnight LB culture using the QIAprep Spin Miniprep Kit (Qiagen, Netherlands) according to the manufacturer's instructions. A 30 µl elution buffer volume was used.

2.1.1.3 Agarose Gel Electrophoresis

Separation of DNA fragments was achieved via agarose gel electrophoresis. DNA fragments >1000 bp were run on 0.8 % (*w/v*) gels, fragments 200-400 bp on 1.5 % (*w/v*) gels, and fragments <200 bp on 2% (*w/v*) gels. A 1 Kb Plus DNA Ladder (Invitrogen, USA) was included as a standard of band sizes. Gels were prepared containing 1x SYBERSafe^{TM} DNA gel stain (Invitrogen, USA). Samples were mixed with 2.5 µl 5 x DNA Loading Dye prior to their loading onto the gel. All gels were run in 1x TAE buffer (40 mM Tris-acetate, 2 mM EDTA) at 100 V for 45 – 60 minutes. Gels were visualised on a blue light box (Invitrogen, USA) and images were captured by a camera.

2.1.1.4 DNA Extraction from an Agarose Gel

A scalpel blade was used to excise the desired band from the agarose gel. DNA was extracted from the gel using a QIAquick Gel Extraction Kit (Qiagen, Netherlands) according to manufacturer's instructions. A 30 µl elution buffer volume was used.

2.1.1.5 DNA Purification from Solution

A QIAquick PCR Product Purification Kit (Qiagen, Netherlands) was used to purify DNA from 25 µl polymerase chain reaction (PCR) samples according to the manufacturer's instructions. A 30 µl elution buffer volume was used.

2.1.1.6 DNA Quantification

DNA was quantified using a Nanodrop 2000 spectrophotometer (Nanodrop Technologies, USA) at a measurement wavelength of 260 nm. The purity of DNA samples was assessed by calculating 260/280 and 260/230 ratios, which indicate protein and carbohydrate contamination respectively.

2.1.1.7 Restriction Enzyme Digest

Restriction enzyme digests were performed as per the manufacturer's instructions. Buffers were chosen as recommended by the manufacturer. A double digest was performed if a buffer compatible with both restriction enzymes was available. Reactions included ≤30 µl of DNA, 2 µl of enzyme (Invitrogen (USA) or Roche Applied Science (Switzerland)), appropriate 1x buffer and MilliQ H_2O up to 100 µl. Incubations were performed at 37 °C for three hours.

2.1.1.8 DNA Ligation

Ligation reactions were carried out as per the manufacturer's recommendations using 1 U of T4 DNA Ligase (Invitrogen, USA), 1x Ligase Buffer (Invitrogen, USA), a ratio of 1:3 vector:insert and MilliQ H_2O up to a total volume of 20 µl. Reactions were incubated for four hours at room temperature or at 16 °C overnight.

2.1.1.9 Annealing of DNA Oligonucleotides

Single-stranded complementary DNA oligonucleotides were designed and analysed using Geneious Pro (Version 5.02) (BioMatters Ltd, NZ). Oligonucleotides were ordered through Invitrogen (USA) and made up to a final concentration of 100 µM in 1x TE buffer. Equal molar ratios of forward and reverse oligonucleotides were added to 10 µl Binding Buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl). Samples were heated at 95 °C for 5 minutes and cooled to room temperature over 45 minutes. The single-stranded oligonucleotides and annealed product were run alongside each other on a 2 % agarose gel (Section 2.1.1.3) to assess the success of the annealing reaction.

2.1.1.10 Two Halves Mutagenesis

The two halves mutagenesis technique was used to mutate bases in perfect IR halves of a 130 bp region of the *Rv0617a/c* promoter to adenines (A). This method, outlined in Figure 2.1, involves introduction of the mutation into overlapping halves of the target region of DNA using PCR and mutagenic primers designed using the QuikChange Primer Design software (www.genomics.agilent.com). Joining of the halves through PCR is then possible due to the shared, mutated sequence region present at one end of each overlapping half. The result is a full length region of DNA containing the desired mutation.

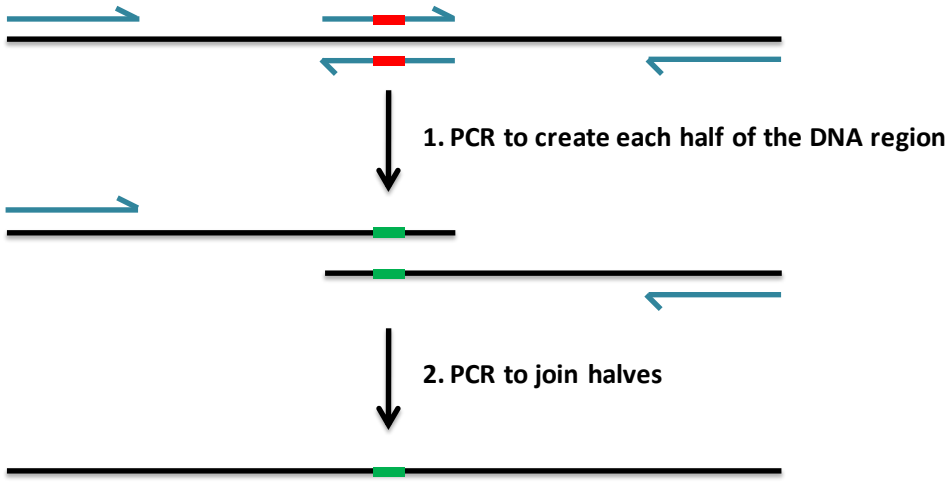


Figure 2.1 Schematic diagram of two halves mutagenesis. The desired mutation in the DNA region and mutagenic primers are shown in red.

2.1.1.11 Preparation of Electrocompetent Escherichia coli

A glycerol stock of the appropriate *E. coli* strain was streaked onto an LB agar plate and incubated overnight at 37 °C. A single colony was used to inoculate 10 ml of LB medium which was then incubated overnight at 37 °C, shaking at 200 rpm. The overnight culture was used to inoculate 1 L of LB media. The 1 L culture was incubated at 37 °C, 200 rpm until an OD_{600} of 0.5-0.7 was reached. The culture was chilled on ice for 45 minutes before being centrifuged at 4000 g for 15 minutes at 4 °C. The cell pellet was resuspended in an equal volume of chilled 10 % *(v/v)* glycerol and centrifuged at 4000 g for 15 minutes at 4 °C. The supernatant was discarded, the pellet resuspended in 500 ml of chilled 10 % *(v/v)* glycerol, and the cells were centrifuged at 4000 g for 15 minutes at 4 °C. The supernatant was discarded and the pellet was resuspended in 20 ml of chilled 10 % *(v/v)* glycerol. The entire volume was transferred to a pre-chilled 50 ml

Falcon tube and centrifuged at 4000g for 15 minutes at 4 °C. The supernatant was tipped off and the pellet was resuspended in 2 or 3 ml of chilled 10 % *(v/v)* glycerol, depending on whether the OD_{600} of the 1 L culture was 0.5 or 0.7 respectively. Aliquots (40 µl) of cells suspension were dispensed into 1.5 ml microfuge tubes, flash frozen in liquid nitrogen and stored at -80 °C.

2.1.1.12 Electroporation of Escherichia coli

The following protocol was used for transformation of both DH5α and TOP10 electrocompetent *E. coli* strains. Plasmid DNA or ligation reaction (1 µl) was added to a 1.5 ml microfuge tube containing 40 µl of electrocompetent *E. coli* cells thawed on ice. The total volume was transferred to a pre-chilled 0.2 cm electroporation cuvette (BioRad Laboratories, USA) on ice. The cuvette was tapped to settle the solution and the sample was electroporated with a Bio-Rad Gene Pulser™ (Bio-Rad Laboratories, USA) at 2.5 kV, 25 µF capacitance and 200 Ω resistance. For recovery, 1 ml of SOC medium was immediately added to the electroporated cells, and the entire volume was transferred from the cuvette into a 1.5 ml microfuge tube. Recovering cells were incubated for 1 hour at 37 °C, 200 rpm. Aliquots of 50 and 100 µl were plated onto LB agar plates supplemented with the appropriate antibiotic. Colonies from these plates were used to seed 5 ml LB broth cultures containing the appropriate antibiotic. These cultures were incubated at 37 °C, 200 rpm overnight, allowing plasmid DNA isolation the following day according to the method outlined in Section 2.1.1.2.

2.1.1.13 Preparation of Electrocompetent Mycobacterium smegmatis

A glycerol stock of the appropriate *M. smegmatis* strain was streaked onto an LB agar plate supplemented with 0.05 % Tween-80 (LBT) and incubated overnight at 37 °C. A single colony was used to inoculate 5 ml of LBT media which was then incubated overnight at 37 °C, shaking at 200 rpm, until an OD_{600} of ~0.7 was reached. A 1 ml volume of the overnight culture was used to inoculate 100 ml of 7H9/ADC + 0.05 % Tween-80 media. The 100 ml culture was incubated at 37 °C, 200 rpm until an OD_{600} of 0.5-0.7 was reached. The culture was chilled on ice for 90 minutes before being centrifuged at 4000 g for 20 minutes at 4 °C. The cell pellet was resuspended in 100 ml of chilled 10 % *(v/v)* glycerol and centrifuged again at 4000 g for 20 minutes at 4 °C. The majority of supernatant was discarded,

leaving a small amount for resuspension of the pellet. The volume was transferred to a pre-chilled 50 ml Falcon tube and centrifuged at 4000 g for 20 minutes at 4 °C. The supernatant was tipped off and the pellet was resuspended in 0.2 ml chilled 10 % *(v/v)* glycerol. Aliquots (40 µl) of cells suspension were dispensed into 1.5 ml microfuge tubes, flash frozen in liquid nitrogen and stored at -80 °C.

2.1.1.14 Electroporation of Mycobacterium smegmatis

The following protocol was used for transformation of the *M. smegmatis* strains $mc^2$4517, $mc^2$155 and $mc^2$155Δ*vapBC*. Plasmid DNA or ligation reaction (1 µl) was added to a 1.5 ml microfuge tube containing 40 µl of electrocompetent *M. smegmatis* cells thawed on ice. Chilled glycerol (260 µl of 10% *(v/v)*) was added and the total volume (300 µl) was transferred to a 0.2 cm electroporation cuvette (BioRad Laboratories, USA) chilled on ice. The cuvette was tapped to settle the solution and electroporated with a Bio-Rad Gene Pulser™ (BioRad Laboratories, USA) at 2.5 kV, 25 µF capacitance and 1000 Ω resistance. For recovery, 1 ml of 7H9/ADC media was immediately added to the electroporated cells and the entire volume was transferred from the cuvette into a 15 ml Falcon tube. Cells were incubated for 3 hours at 37 °C, 200 rpm. Aliquots of 10, 50 and 150 µl were plated onto 7H10/ADC + 0.05% *(v/v)* Tween-80 agar plates supplemented with the appropriate antibiotic.

2.1.2 RNA Manipulation

2.1.2.1 *Total RNA Extraction from* Mycobacterium smegmatis

Method developed by Ali Ruthe, Proteins & Microbes Lab, University of Waikato.

A glycerol stock of wild-type *M. smegmatis* $mc^2$155 was streaked on a low salt LBT agar plate in the absence of antibiotics. The plate was incubated in a plastic container at 37 °C for three to four days. A saturated paper towel was placed in the container to prevent plate desiccation. One colony was selected to inoculate 10 ml of LBT medium and the culture was incubated for approximately three days at 37 °C. The 10 ml of culture was immediately added to 40 ml of 5 M GITC pH 7.0. Cells were pelleted by centrifugation (4600 g for 15 minutes at 4 °C) and supernatant was discarded. Fresh 5 M GITC (0.5 ml) was used to resuspend the cell pellet, after which the entire volume was transferred to a 2 ml tube containing 0.3 g of 0.1 mm and 2.5 mm zirconia beads. A Fastprep FP120 bead beater

(Thermosavant, USA) was used to mix the beads for 20, 25, 30 and 35 second time periods, with one minute cooling intervals in between. Following this, brief centrifugation removed foam. One tenth of the volume (50 µl) of 2 M sodium acetate pH 4.0 was added and the solution was gently mixed. An equal volume (500 µl) of water-saturated phenol was added, the solution was mixed vigorously, and the tube was left to rotate for 10 minutes. 1-bromo, 3-chloro propanate (100 µl) was added, followed by agitation of the solution for one minute and incubation on ice for five minutes. The solution was centrifuged and the top aqueous phase removed. Steps from the addition of 2 M sodium acetate pH 4.0 were then repeated, with the volumes adjusted accordingly. After the final centrifugation step, the top layer was removed, added to an equal volume of isopropanol, and chilled on ice for 30 minutes to precipitate the RNA. The sample was centrifuged (13000 g, 4 °C, 30 minutes), 1 ml of 70 % ethanol was added, centrifuged (13000g, 4 °C, 15 minutes), the supernatant was discarded, 1 ml of 100 % ethanol was added, centrifuged (13000g, 4 °C, 15 minutes) and the supernatant was discarded. A 25 µl volume of 10 mM Tris-HCl pH 7.0, 0.5 mM $MnCl_2$ was used to resuspend the pellet. DNA was degraded through the addition of 1 µl DNase (Promega, USA) per µg RNA, followed by incubation for 30 minutes at 37 °C, 600 rpm. DNase activity was quenched through the addition of 1.1 µl EGTA stop solution per 1 µl of DNase used, followed by heat treatment for 10 minutes at 65 °C. The concentration of RNA was determined using the Nanodrop 2000 spectrophotometer (Section 2.1.2.4) and RNA was stored at -80 °C.

2.1.2.2 RNA Precipitation

RNA samples were precipitated to eliminate salts and contaminants by addition of 2 M NaOAc pH 4 (10 % the volume of RNA) followed by 2.5 x the volume of 100 % ethanol. The RNA was left to precipitate at -20 °C for 30 - 60 minutes. Samples were centrifuged at 13000 rpm for 30 minutes at 4 °C and the supernatant was discarded. The pellet was washed with 1 ml chilled 70 % ethanol, centrifuged at 13000 rpm for 10 minutes at 4 °C, washed with 1 ml 100 % chilled ethanol and centrifuged once more at 13000 rpm for 10 minutes at 4 °C to re-pellet the RNA. The supernatant was discarded and the tube was left at room temperature for 5 minutes to allow the pellet to air-dry. The pellet was

resuspended in 25 µl of water or Tris/$MnCl_2$. For long term storage the pellet was resuspended in 50 µl ethanol and frozen at -80 °C. Concentration of RNA was determined using the Nanodrop 2000 spectrophotometer (Section 2.1.2.4).

2.1.2.3 Purification of 16S and 23S rRNA from Total RNA

Aliquots (20 µl) of isolated total RNA (Section 2.1.2.1) were loaded on a 1% TAE agarose gel containing Sybersafe along with 2 µl of 5 x DNA Loading Dye. The gel was run in 1x TAE buffer at 100 V for 1 hour. Bands were visualised on a blue light box and 16S and 23S bands (at 2.5kb and 1.5kb respectively) were excised using a sterile scalpel. Thick walled non-autoclaved microfuge tubes (0.6 ml) were pierced at the bottom using a flamed needle and a small sterile glass bead was placed in the tubes to block the hole. Two to three excised gel bands were added per tube, and tubes were frozen at -80 °C for 30 minutes. The 0.6 ml tubes were placed into larger 1.5 ml non-autoclaved microfuge tubes. Tubes were centrifuged at 13000 rpm for 6 minutes at 4 °C to elute nucleic acids from the gel bands. The volume of liquid eluted from gel bands into the larger microfuge tube was measured and RNA was then precipitated from this (Section 2.1.2.2).

2.1.2.4 RNA Quantification

RNA concentration was measured using a Nanodrop 2000 spectrophotometer (Nanodrop Technologies, USA). Absorbance was measured at 260 nm and the 260/280 ratio was used to indicate sample purity.

2.1.3 Polymerase Chain Reaction

2.1.3.1 Primers

Primers were designed and analysed using Geneious Pro (Version 5.02) (BioMatters Ltd, NZ). This program was used to check melting temperatures and possible secondary structures such as primer dimers or hairpins. Primers were ordered through Integrated DNA Technologies (IDT, USA) and made up to a final concentration of 100 µM in 1x TE buffer. Primers were stored at - 20 °C. For use in ligation cloning, primers were designed to have a restriction site flanked by four bases.

2.1.3.2 PCR for Amplification of Genomic DNA and Plasmid Inserts

For amplifying genomic DNA or plasmid inserts, primers were used at 10 pmol per reaction. For each set of primers, a temperature gradient PCR was initially

conducted to determine the optimum primer annealing temperature. Reactions were carried out in a 25 µl volume. Single reactions included either 10 – 100 ng of genomic DNA from the *M. tb* H37Ra strain, or 1 ng of plasmid DNA. One unit of DNA polymerase was added per reaction. In all cases, annealing time was dependent on amplicon length (1 kb = 1 minute). The components of reactions and the PCR cycling regimes applied were in accordance with the polymerase enzyme used, as outlined in the following sections.

2.1.3.3 Platinum Pfx *PCR*

PCR samples set up with Platinum *Pfx* DNA polymerase (Invitrogen) contained 1x *Pfx* Amplification Buffer, 1 mM $MgSO_4$, 1 U Platinum *Pfx* DNA polymerase, 0.3 mM deoxynucleotide mix (dATP, cCTP, dTTP & dGTP), 0.2 µM of each primer and 10 to 100 ng template DNA.

PCR Cycling Conditions:

94 °C	2:00		
94 °C	0:15	x 29	min:sec
Tm	0:30		
68 °C	0:30		
68 °C	7:00		

2.1.3.4 Taq *PCR*

PCR samples set up with *Taq* DNA polymerase (Roche) contained 1x *Taq* PCR buffer, 1.5 mM $MgCl_2$, 2.5 U *Taq* DNA polymerase, 0.2 mM deoxynucleotide mix (dATP, cCTP, dTTP & dGTP), 0.2 µM of each primer and 10 to 100 ng template DNA.

PCR Cycling Conditions:

94 °C	2:00		
94 °C	0:15	x 29	min:sec
Tm	0: 30		
72 °C	0:30		
72 °C	7:00		

2.1.3.5 iProof PCR

PCR samples set up with iProof™ DNA polymerase (Bio-Rad) contained 1x iProof GC Buffer, 1 mM $MgCl_2$, 0.5 U iProof DNA polymerase, 0.2 mM deoxynucleotide mix (dATP, cCTP, dTTP & dGTP), 0.2 µM each primer and 10 to 100 ng template DNA.

PCR Cycling Conditions:

98 °C	0:30		
98 °C	0:10	x 25	min:sec
Tm	0: 30		
72 °C	0:15		
72 °C	7:00		

2.1.3.6 Touchdown PCR

The touchdown PCR protocol was used with iProof™ DNA polymerase.

PCR Cycling Conditions:

98 °C	0:30		
98 °C	0:10	x 9	
70 °C	0:30		
-1 °C per cycle			
72 °C	0:15		min:sec
98 °C	0:10	x 20	
60 °C	0:30		
72 °C	0:15		
Go to step 6, 20x			
72 °C	7:00		

2.1.3.7 Colony PCR

Colony PCR was used to screen for the presence of an insert in a particular plasmid. A single colony was picked off a plate and resuspended in 10 µl of 0.2 mM ampicillin in a 1.5 ml microfuge tube by pipetting up and down multiple times. One µl of this solution was used as the template in a 25 µl PCR with *Taq* polymerase (Section 2.1.3.4)

2.2 Methods relating to *Mycobacterium smegmatis* Growth Experiments

2.2.1 Construction of Tetracycline Inducible Constructs

For tetracycline inducible expression of VapC and VapBC proteins, the 667 bp *Rv0065a/c* operon, the 437 bp *Rv0065Cc* gene, the 662 bp *Rv0617a/c* operon and the 438 bp *Rv0617c* gene were amplified from H37Ra *M. tb* genomic DNA (Proteins & Microbes lab, University of Waikato) using the primers listed in Table 2.1 and Platinum *Pfx* polymerase (Section 2.1.3.3). Primers introduced a synthetic ribosome-binding consensus site (GGAGG) at the 5'-end of PCR products and flanking *Bam*HI and *Spe*I restriction enzyme sites. PCR products were digested (Section 2.1.1.7) and ligated (Section 2.1.1.8) into *Bam*HI/*Spe*I sites of the pMind vector (Proteins & Microbes lab, University of Waikato), placing them under control of the transcriptional regulator *tetR* and the intervening operator region *tetO*. Plasmids were transformed into TOP10 electrocompetent *E. coli* (Section 2.1.1.12). Colony PCR (Section 2.1.3.7) along with plasmid purification (Section 2.1.1.2) and sequencing based on the Applied Biosystems

3130xl Genetic Analyser (WDSF, University of Waikato) confirmed the presence of each insert in the correct orientation to the *tetRO* promoter region.

2.2.2 Conditional VapC and VapBC Protein Expression

Plasmids were electroporated into *M. smegmatis* mc^2155 (wild-type) and *M. smegmatis* Δ*vapBC* strains (Section 2.1.1.14). In the Δ*vapBC* strain, the chromosomally encoded *MSMEG_1283/4 vapBC* operon had been replaced with the kanamycin resistance cassette *aphA-3* through allelic exchange mutagenesis (Robson 2010).

M. smegmatis strains harbouring plasmids 0617vapBC_pMind, 0617vapC_pMind and pMind were grown in LBT medium for 38 hours to an OD_{600} between 0.2 and 0.4. An aliquot of each LBT starter culture (100 µl) was used to seed a second starter culture in Hartman's-de Bont minimal medium (HdB) supplemented with 0.2 % *(v/v)* glycerol, 50 mM MES and 0.05 % *(v/v)* Tween-80. Cultures were grown for 21 hours and then diluted to an OD_{600} of 0.0025 in 200 ml HdB medium in a 500 ml flask. OD_{600} values were monitored until they reached 0.1 – 0.15 (19 hours), at which point protein expression was induced by addition of 20 ng/ml tetracycline (Tc). This concentration was previously reported as sufficient for protein expression while remaining sub-inhibitory (Blokpoel et al. 2005; Robson 2010). Samples of culture were taken every six hours for the next 102 hours in order to monitor cell growth (OD_{600} values) and viability (colony forming unit (CFU) values). Cell viability was examined through serially diluting samples 100-fold from 10^{-2} to 10^{-6} in 1x phosphate buffered saline (PBS) (0.02 M phosphate (0.0038 M NaH_2PO_4, 0.15 M NaCl, pH 7.4)). Three 5 µl aliquots of each dilution were spotted onto plates containing both hygromycin B and tetracycline, and hygromycin B only plates.

All media were supplemented with 50 µg/ml hygromcyin B for plasmid maintenance and media for Δ*vapBC* strains were additionally supplemented with 50 µg /ml kanamycin.

Table 2.1 Primers used for cloning into pMind vector

Primer		Primer Sequence RE + RBS*	RE
F	0065vapC_pMind	AAATTTGGATCCGGAGGAATAATGGTGG ATGAATGTGTA	*Bam*HI
	0065vapBC_ pMind	AAATTTGGATCCGGAGGAATAATGGCTA CCATTCAAGT	*Bam*HI
	0617vapC_pMind	AAATTTGGATCCGGAGGAATAATGGTGA CGGTGCTGCT	*Bam*HI
	0617vapBC_pMind	AAATTTGGATCCGGAGGAATAATGGTGC GCACTACCATC	*Bam*HI
R	0065vapC_pMind	AAATTTACTAGTTCATTGGTGGTCGTTGG AATGAGTACC	*Spe*I
	0065vapB_ pMind	AAATTTACTAGTTCATTGGTGGTCGTTGG AATGAGTACC	*Spe*I
	0617vapC_pMind	AAATTTACTAGTTCACCGAACGAGTTTGA TTTCGCACG	*Spe*I
	0617vapBC_pMind	AAATTTACTAGTTCACCGAACGAGTTTGA TTTCGCACG	*Spe*I

*RE, restriction enzyme site; RBS, ribosome binding site

2.3 Methods Relating to Protein Expression and Purification

2.3.1 Small Scale Protein Expression in *Mycobacterium smegmatis*

Small scale protein expression trials were used to screen for VapB and VapC protein expression after transforming pYUB28b plasmids containing the *Rv0065a/c* and *Rv0617a/c* operons into electrocompetent *M. smegmatis* $mc^2$4517 cells.

2.3.1.1 Small Scale Protein Expression

Transformed cells were streaked onto antibiotic selective plates (low salt LBT plates supplemented with kanamycin (50 µg/ml) and hygromycin (50 µg/ml)). Plates were incubated in a plastic container at 37 ^{0}C for 3 – 4 days. A saturated paper towel was placed in the container to prevent plate desiccation. A single transformed colony was used to inoculate 10 ml of PA-0.5G/Tween-80 medium. The culture was incubated for 48 hours at 37 °C, 200 rpm. A 100µl aliquot of the PA-0.5G/Tween-80 seeder culture was used to inoculate 10 ml of ZYP-5052/Tween-80 medium (1:100 dilution) and this culture was incubated at 37 °C for 96 hours, 200 rpm. Cells were pelleted by centrifugation at 4500 g for

20 minutes at 4 °C. The supernatant was discarded, and cell pellets were resuspended in l ml of lysis buffer (50 mM sodium phosphate buffer pH 7.4, 200 mM NaCl, 20 mM imidazole). A whole cell sample was run on a sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel for analysis, consisting of 5 µl of resuspended cell pellet, 10 µl of lysis buffer and 5 µl of 4 x SDS loading buffer. The resuspended cells were lysed on ice using a Misonic Sonicator (Qsonica, USA). Sonication involved 3 x 15 second, and 2 x 12 second bursts at setting 3.5. Cell lysate was centrifuged for 20 minutes at 13000 g, 4 °C to pellet the insoluble fraction. The supernatant was decanted into a separate microfuge tube. The pellet was resuspended in 200 µl of lysis buffer. Soluble and insoluble fraction samples were taken to be run on an SDS-PAGE gel for analysis, consisting of 15 µl of either fraction added to 5 µl of 4 x SDS loading buffer.

2.3.1.2 His-tag Binding Tests

The soluble fraction of cell lysate was subjected to His-tag binding tests. A 50 µl aliquot of Ni Sepharose (GE Healthcare, UK) was added to a 1.5 ml microfuge tube. This was washed with 1 ml of lysis buffer and centrifuged at 2000 g for 30 seconds to pellet the Ni sepharose. The supernatant was discarded, and 200 µl of the soluble fraction of cell lysate was added. The lysate/Ni sepharose mixture was incubated for 15 minutes at room temperature whilst shaking at 1000 rpm. Centrifugation at 2000 g for 30 seconds was conducted to pellet the Ni sepharose. A 10 µl sample of the supernatant was added to 5 µl of 4 x SDS loading buffer for analysis on an SDS-PAGE gel. The remaining supernatant was discarded and the Ni sepharose was once again washed by addition of 1 ml lysis buffer followed by centrifugation at 2000 g for 30 seconds. The washing process was repeated two times, after which 25 µl of 4 x SDS loading buffer was added to the Ni sepharose for analysis by SDS-PAGE. SDS-PAGE gels were run and stained as outlined in Sections 2.3.5.4 and 2.3.5.5 in order to detect the presence of VapB and VapC protein in samples.

2.3.2 Large Scale Protein Expression in *Mycobacterium smegmatis*

2.3.2.1 Streaking Plates

Glycerol stocks of $mc^2$4517 strains containing the pYUB28b plasmid containing either the *Rv0065a/c* or *Rv0617a*/*c* operons were streaked onto antibiotic selective

plates (low salt LBT plates supplemented with kanamycin (50 µg/ml) and hygromycin (50 µg /ml)). Plates were incubated in a plastic container at 37 ^{0}C for 3 – 4 days. A saturated paper towel was placed in the container to prevent plate desiccation.

2.3.2.2 Seeder Cultures

Expression of VapBC protein was achieved using auto-induction medium supplemented with Tween-80 (0.05 %), kanamycin (50 µg /ml) and hygromycin (50 µg/ml). A sterile loop was used to pick one colony of the desired *M. smegmatis* strain. The colony was then used to inoculate 10 ml of PA-0.5G/Tween-80 medium, which was incubated at 37 ºC, 200 rpm for 48 hours and positioned horizontally to achieve maximum aeration.

2.3.2.3 Expression Cultures

A 10 ml seeder culture was used to inoculate 1 L of ZYP-5052/Tween-80 medium supplemented with kanamycin (50 µg/ml) and hygromycin (50 µg/ml). The culture was grown in a baffled flask at 37 ºC, 200 rpm for 96 hours.

2.3.2.4 Glycerol Stocks

For long term storage of bacterial cultures, glycerol stocks were prepared by combining 0.5 ml of 50% *(v/v)* glycerol and 0.5 ml of fresh liquid bacterial culture. Stocks were stored at -80 ºC.

2.3.2.5 Harvesting Cells from Expression Cultures

Cells from large scale expression cultures were pelleted by centrifugation at 4600 rpm, 4 ºC for 20 minutes. The supernatant was discarded and pellets were stored at -80 ºC.

2.3.3 Purification of the VapBC complex

2.3.3.1 Fast Performance Liquid Chromatography (FPLC)

An ÄKTA Basic^{TM} or ÄKTA Prime^{TM} FPLC system (GE Healthcare, UK) was used to perform fast performance liquid chromatography. UV absorbance was monitored at a wavelength of 280 nm in order to detect the elution of proteins.

2.3.3.2 Purification of His-tagged Proteins through IMAC

Immobilised metal affinity chromatography (IMAC) was used to purify proteins which had been expressed with an N-terminal His-tag. The IMAC column used

was a 5 ml HiTrap™ Chelating HP column (GE Healthcare, UK). Prior to use, the column was primed by washing through 5 ml of 100 mM $NiCl_2$, followed by 20 ml of water and 20 ml of the lysis buffer to be used during the run.

Expression culture cell pellets were resuspended in 30 ml lysis buffer and an EDTA-free protease inhibitor tablet (Roche, Applied Science, Switzerland) was added. The cells were sonicated on ice using a Misonix Sonicator (Qsonica, U.S.A) at setting 7.5 in 30 second bursts until cell lysis occurred (approximately 3-5 minutes total sonication time). Insoluble material was pelleted by centrifugation at 13000 rpm for 20 minutes at 4 °C. Supernatant was filtered through 1.2 and 0.45 µM Minisart filters (Sartorius AG, Germany). The filtered supernatant was loaded dropwise onto a 5 ml HiTrap™ Chelating HP column (GE Healthcare, UK). The column was attached to an FPLC system and 20 ml of lysis buffer was run through at a rate of 1 ml/min to elute unbound proteins. To elute bound proteins, a 75 ml gradient of 0-50 % elution buffer (lysis buffer + 1M imidazole) was run at a rate of 1 ml/min while collecting 2 ml fractions. Fractions that corresponded to the UV peak were analysed on an SDS-PAGE gel (Section 2.3.5.4). After protein purification was complete, the column was washed with 20 ml of water and stored in ethanol. Between purifications of different proteins, the column was washed with 10 ml 100 mM EDTA pH 8.0 to strip Ni^{2+} ions and then re-primed with 5 ml of 100 mM $NiCl_2$.

2.3.3.3 Size Exclusion Chromatography of the VapBC Complex

To further purify protein, size exclusion chromatography (SEC) was performed using a HiLoad™ 16/60 Superdex™ column (GE Healthcare, UK). VapBC protein fractions from the IMAC purification were pooled and concentrated down to 5 ml (Section 2.3.5.2). The column was pre-equilibrated in SEC Buffer (50 mM sodium phosphate, pH 7.0, 200 mM NaCl) overnight. The protein solution was loaded onto the equilibrated size exclusion column through a 0.2 µM filter (Pall, USA). Fractions were analysed on an SDS-PAGE gel (Section 2.3.5.4).

2.3.3.4 Dialysis of the VapBC Complex

Dialysis was conducted to reduce the salt content of the protein sample to a level suitable for anion-exchange chromatography. SEC fractions spanning the protein peak were pooled into 6-8 kDa molecular weight cut off Spectra Por® dialysis

tubing (Spectrum Laboratories, UDA) cut to the appropriate length based on the volume of liquid to dialyse (tubing held 3.3 ml/cm). Dialysis was conducted in a 1 L beaker containing 800 ml of dialysis buffer (50 mM sodium phosphate, pH 7.0, 100 mM NaCl) with gentle stirring at 4 °C. After 2 – 3 hours incubation the dialysis buffer was replaced with 800 ml of fresh buffer. Dialysis was continued overnight at 4 °C with gentle stirring.

2.3.4 Tryptic Digest of the VapBC Complex

A ratio of 0.1 mg trypsin : 0.7 mg VapBC protein was used to preferentially degrade the more labile VapB protein. Digests were incubated for 1 hour at room temperature, and stopped by addition of an equal amount of trypsin inhibitor (Sigma Aldrich, USA) to the amount trypsin that was initially added. Samples were then incubated at room temperature for 15 minutes with shaking (500 rpm).

2.3.5 Purification of VapC

2.3.5.1 Anion Exchange Chromatography

Anion exchange chromatography was performed using a 5 ml HiTrap™ Q Anion Exchange column (GE Healthcare, UK). The column was washed with 20 ml of water and pre-equilibrated with 40 ml of dialysis buffer. Samples were filtered through a 0.2 µM Minisart filer (Sartorius AG, Germany) before being loaded onto the column. The column was attached to an ÄKTA Basic™ or ÄKTA Prime™ FPLC system. A 50 ml gradient of 0-100% anion elution buffer (50 mM sodium phosphate buffer pH 7.4, 1 M NaCl) was used to elute bound proteins. Fractions were analysed on an SDS-PAGE gel (Section 2.3.5.4).

2.3.5.2 Concentration of Protein

Protein samples were concentrated using 20 ml or 2 ml Vivaspin concentrators (Sartorius AG, Germany) with 5 kDa molecular weight cut-offs. Concentrators were spun at 3600 g, 4 °C until the desired concentration or volume was achieved.

2.3.5.3 Measurement of Protein Concentration

The concentration of protein samples was measured using a Nanodrop 2000 spectrophotometer (Thermo Scientific, USA). This measures absorbance of the sample at 280 nm and calculates protein concentration using the Beer-Lambert equation, $A = \varepsilon c l$, where A is the absorbance at 280 nm, ε is the theoretical molar extinction coefficient ($M^{-1}cm^{-1}$), c is the concentration (M) and l is the path length

in cm. The online tool ProtParam was used to calculate theoretical molar extinction coefficients for proteins/protein complexes and these values were used in the following formula to calculate the actual concentration values:

$$\frac{\text{Nanodrop protein concentration value}}{\text{Molar extinction coefficient}} = \text{True protein concentration}$$

2.3.5.4 SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE gels were prepared in a SE275 Hoeffer gel casting apparatus (Hoeffer® Inc, USA). Gels consisted of a pH 8.8 16.5% acrylamide resolving gel layer beneath a pH 6.8 5% acrylamide stacking gel layer. Acrylamide was added from a 30% acrylamide/bis stock solution 37.5:1 (BioRad Laboratories, USA). Protein samples were prepared for analysis by mixing with 4 x SDS loading buffer at a 3:1 ratio. Samples were heated at 95 °C for 5 minutes before being loaded onto gels. In order to estimate protein band sizes, samples were run alongside a Precision Plus Protein™ Unstained Standard protein ladder (Biorad Laboratories, USA). Gels were run in 1 x SDS-PAGE running buffer at 70 V through the stacking layer and voltage was increased to 150 V once the resolving layer was reached.

2.3.5.5 Coomasie Blue Stain for SDS-PAGE

The quick coomassie blue staining method was used to stain SDS-PAGE gels (Wong et al. 2000). The gel was placed in a microwaveable container, submerged in Fairbanks stain A, microwaved for 30 seconds then cooled to room temperature for approximately 20 minutes with gentle shaking. The stain was then tipped off, the gel was rinsed with water and the above process was repeated using Fairbanks stains B, C and D.

2.4 Methods relating to RNase Assays and their Analysis

2.4.1 932 Pentaprobe RNA Fragments

2.4.1.1 Design of 932 Pentaprobe Fragments

The 932 Pentaprobe was used as a scaffold from which to design overlapping RNA fragments of 18-100 bp in length. Fragments were labelled based on their position along the 932 Pentaprobe. The two smallest RNA fragments (33 bp and 18 bp) were ordered through Integrated DNA Technologies (USA) and had the following sequences:

932 94-126: $^{5'}$GGGGUGAUAUAAUAUGAUGCCGGGCGGGGGCCG^{3}

932 109-126: $^{5'}$GAUGCCGGGCGGGGGCCG$^{3'}$

The remaining fragments were generated through PCR amplification using 932 Pentaprobe DNA as template, followed by *in vitro* transcription, as outlined below.

2.4.1.2 PCR Amplification of 932 Pentaprobe Fragments

932 Pentaprobe RNA and DNA was acquired from Jo McKenzie (Proteins and Microbes lab, University of Waikato). The 932 fragment DNA was amplified from 932 Pentaprobe DNA through PCR with *Taq* or Platinum *Pfx* polymerase (Sections 2.1.3.3 and 2.1.3.4). An additional round of PCR was conducted to add a T7 promoter sequence onto the 5' end of each DNA fragment. Primers used for each round of PCR are detailed in Table 2.2.

Table 2.2 Primers used for amplification of Pentaprobe 932 fragments

Primer		Primer Sequence
F	932 90-170	CACTGGGGTGATATAATATGATGCCG
	932 90-170 T7	ACGCTAATACGACTCACTATAGGCACTGGGGTGAT
	932 70-150	TCCAACCTCACACCACATTACACTGGGGTG
	932 70-150 T7	TAATACGACTCACTATAGGGAGATCCAACCTCACACC
	932 50-150	GTGCTGGAATTCTGCAGATCTCCA
	932 50-150 T7	TAATACGACTCACTATAGGGAGAGTGCTGGAATTCTG
R	932 90-170	TGCCTAATCCCACCTAGCGTATCG
	932 70-150	TCGGGTCATGTAGTGCTACGTTACGGCCC
	932 50-150	GTAACGTAGCACTACATGACCCGAT

2.4.1.3 **In Vitro** *Transcription of 932 Pentaprobe Fragments*

The T7 MEGAscript® kit (Ambion, USA) was used for transcription of PCR products to RNA. Reactions consisted of 1 µg PCR product template, 2 µl of each rNTP (75 mM stock), 2 µl T7 polymerase enzyme mix, 2 µl 10 x enzyme buffer and diethyl pyruvate carbonate (DEPC)-treated water to a final volume of 20 µl. Reactions were incubated at 37 °C for 2 hours. TURBO DNase (1 µl) was added and reactions were incubated at 37 °C for 15 min at 600 rpm. RNA was purified by precipitation as outlined in Section 2.1.2.2.

2.4.2 RNA Secondary Structure Predictions

RNA secondary structures for this thesis were predicted using RNAfold (Hofacker et al. 1994). This program predicts secondary structure through three kinds of dynamic programming algorithms: the minimum free energy (MFE) algorithm, which produces a single optimal structure (Zuker & Stiegler 1981), the partition function algorithm, which calculates base pairing probabilities based on thermodynamic principles (McCaskill 1990) and the suboptimal folding algorithm, which computes every suboptimal structure that is within a given energy range of the optimal energy (Wutchy et al. 1999).

2.4.3 General VapC Ribonuclease Assay Method

VapC ribonuclease assays were conducted against a range of RNA substrates over a 60 minute incubation time course. Assays were used to determine whether or not cleavage was occurring and, if so, where these cut sites were located on the substrate RNA. Prior to beginning assay work, the workspace and pipets to be used were cleaned with RNase Away® (Sigma-Aldrich, USA) to reduce the risk of RNase contamination. Individual assay reactions were set up for each time point to additionally reduce the possibility of RNase contamination. Assays were prepared on ice in 0.2 ml microfuge tubes and all incubations occurred at 37 ° C. Reactions to be analysed by urea denaturing gels were stopped by the addition of 10 µl 2x formamide loading buffer and heated at 70 °C for 5 minutes. Reactions to be analysed by mass spectrometry were heat inactivated at 95 °C for 5 minutes, after which salt and ethanol were used to precipitate the RNA (Section 2.4.6.1).

2.4.3.1 Pyrobaculum aerophilum *VapC Ribonuclease Assay*

Time Course Assay Reactions

Reactions for each time point consisted of the following: 1 µg purified RNA, 1 µg purified VapC protein, 6 µl Assay Buffer (20 mM Tris-HCl pH 9.2, 20 mM NaCl, 10 mM $MgCl_2$) and 2 µl of nuclease free H_2O. Assays were incubated for 5, 15, 30 and 60 minute periods.

RNA Only Negative Controls

RNA only negative control reactions were set up as above but excluding the VapC protein. One control was stopped immediately at 0 hr incubation while the other was left to incubate for the duration of the assay.

EDTA Control

A reaction was set up as per the time course samples however 40 mM EDTA was added to inhibit RNase activity. The reaction was incubated for the duration of the assay.

VapBC Negative Control

An assay reaction was set up as per the time course samples however the VapC was substituted for 1 µg of VapBC protein complex.

2.4.3.2 Mycobacterium tuberculosis *VapC Ribonuclease Assay*

Time Course Assay Reactions

Reactions for each time point consisted of the following: 1 µg purified RNA, 1 µg purified VapC protein, 6 µl Assay Buffer (20 mM sodium phosphate buffer pH 7.4, 20 mM NaCl, 10 mM $MgCl_2$) and 2 µl of nuclease free H_2O. Assays to be analysed by MALDI TOF-MS were carried out in an assay buffer consisting of 20 mM ammonium phosphate pH 7.4, 20 mM NaCl and 10 mM $MgCl_2$. Reactions were incubated for 5, 15, 30 and 60 minute periods.

Negative Controls

Negative controls for *M. smegmatis* VapC ribonuclease assays were the same as those for *P. aerophilum* assays except for the assay buffer which instead was that listed above (20 mM phosphate buffer pH 7.4, 20 mM NaCl, 10 mM $MgCl_2$).

2.4.4 Urea Denaturing Polyacrylamide Gel Electrophoresis RNA Analysis (Urea Denaturing-PAGE)

Urea denaturing-PAGE gels were prepared in a SE275 Hoeffer gel casting apparatus (Hoeffer® Inc, USA). Gels consisted of a resolving layer that varied from 5-20% acrylamide. The amount of acrylamide used was dependent on the size of the RNA samples being analysed: RNA molecules <150 bases were run on 20% gels, molecules 150-300 bases were run on 10% gels and molecules >1000 bases were run on 5% gels. Gels consisted of 6 M urea, 0.05% APS *(w/v)* and 0.05% TEMED *(v/v)*. Acrylamide was added from a 30% acrylamide/bis stock solution 37.5:1 (BioRad Laboratories, USA).

The gels were pre-run in 1x TBE buffer at 70 V for 20 minutes after which wells were flushed with 1 x TBE. Samples were loaded and the gel was run at 150 V until the xylene cyanol dye front neared the bottom of the gel.

2.4.5 Visualisation of RNA

SYBR™ Green II RNA Stain (Invitrogen, USA) was used to visualise RNA that had been run on polyacrylamide or agarose gels. Gels were incubated in the stain for 40 minutes with agitation and then visualised under UV light or on a blue light box (Invitrogen, USA). A low range ssRNA ladder (New England Biolabs, USA) was run alongside samples in order to determine the sizes of bands.

2.4.6 MALDI TOF Mass Spectrometry of RNA Oligonucleotides

2.4.6.1 Sample Purification and Preparation

RNA samples were purified through ammonium acetate and ethanol precipitation as this desalting technique was proven to result in maximum sample recovery (McKenzie et al. 2012b). To the RNA sample was added 5 M ammonium acetate (to a final concentration of 2 M) and three times the volume of ethanol as RNA sample. Precipitation steps that followed were as outlined in Section 2.1.2.2. The MALDI TOF MS matrix was prepared fresh daily. The matrix consisted of 10 µl 2.5 M diammonium citrate, 5 mg 3-hydroxypicolinic acid (3-HPA), 125 µl acetonitrile (ACN) and 365 µl of DEPC-treated water. The solution was vortexed well, and centrifuged at 13000 g for 5 minutes. Aliquots of matrix solution (1 µl) were spotted onto an Anchorchip™ target plate (Bruker Daltonics, USA) and left to air-dry. Oligonucleotide calibration standard (Bruker Daltonics, USA) (1 µl), or sample (1 µl) was overlayed on the dried matrix solution spots and left to air-dry.

2.4.6.2 MALDI TOF Set Up and Spectra Analysis

Samples were analysed on an Autoflex™ II MALDI TOF MS (Bruker Daltonics, USA) in linear mode. The laser power was typically set at 60 % however values varied depending on the ionisation requirements of the samples. The mass range selector was set to ‘low-range’, gain to 2500 V, acceleration voltage to 20 kV and at a pulsed ion extraction of 150 ns. A mass range of 2-12 kDa was collected. Spectra were firstly collected for the oligonucleotide calibration standard (Bruker Daltonics, USA) and these were used to calibrate the spectrometer through an

automatic polynomial correction. FlexAnalysis™ software (Bruker Daltonics, USA) was used to identify and label peaks of the spectra.

2.4.7 Developed Protocol for Identification of *In Vivo* VapC Targets

2.4.7.1 $VapC_{PAE0151}$ Ribonuclease Assay

An RNase assay with $VapC_{PAE0151}$was conducted against the 924 Pentaprobe RNA (Proteins & Microbes Lab, University of Waikato) as outlined in Section 2.4.3.1. Reactions were run on a 10% urea denaturing gel to judge ribonuclease activity (Section 2.4.4). The assay was repeated using an increased amount of 924 Pentaprobe and VapC (4 µg of each) and an incubation period of 30 minutes. RNA from reactions was precipitated as outlined in Section 2.1.2.2. RNA from the precipitated $VapC_{PAE0151}$ assay reactions (4 µg) along with 4 µg of purified Pentaprobe 924 RNA (control) was heated at 65 ° C for 5 minutes for denaturation before being placed on ice.

2.4.7.2 Poly(A)-tailing of RNA

Poly(A)-tailing of RNA samples was achieved through use of an mRNA-ONLY™ Prokaryotic mRNA Isolation Kit with Poly(A) Tailing (Epicentre, USA). Whilst on ice, the following was added to sterile 0.2 ml tubes: 2 µl Poly(A) Polymerase 10x Reaction Buffer, 2 µl 10 mM ATP, 0.5 µl Riboguard RNase Inhibitor, 4 µg RNA, 1 µl (4 units) Poly(A) polymerase and DEPC H_2O up to 20 µl. Tubes were incubated for 20 – 30 minutes at 37 °C after which the tailing reaction was immediately stopped by storage of samples at -20 °C. RNA was precipitated as outlined in Section 2.1.2.2 using 10 µl of RNase-free H_2O for resuspension.

2.4.7.3 Ligation of Acceptor Oligo

T4 RNA Ligase (New England Biolabs, UK), a ligase documented by the manufacturer to have specific activity toward RNA fragments with a 5'mono-phosphate as opposed to 5'-triphosphate, was used to ligate a synthesised RNA 'Acceptor Oligo' to samples with such a 5'-moiety (those cleaved by VapC). The RNA Acceptor Oligo designed for this protocol had the sequence 5'-GAGGAUCCCUGCAGGAAA-3'. This sequence was based on that of the RNA tag used in the ExactSTART™ Full-Length cDNA Library Cloning Kit (Epicentre, USA), which employs the same principle of using T4 RNA Ligase to

selectively tag the 5'mono-phosphate of RNA with a known RNA sequence. The RNA Acceptor Oligo sequence includes recognition sites for both *Bam*HI (GGATCC) and *Pst*I (CTGCAG). Ligation reactions consisted of the following: 10 µl of poly(A)-tailed RNA, 5 µl of Acceptor Oligo RNA at 10pmol/ul, 4 µl of 10x T4 RNA Ligase Buffer, 1 µl of T4 RNA Ligase, 4 µl of 10 mM ATP and 16 µl of 50% PEG8000. Incubation was at 37 °C for 30 minutes, followed immediately by 1st strand cDNA synthesis.

2.4.7.4 cDNA synthesis

The following reagents were added to the ligation reactions for 1st strand cDNA synthesis: 2 µl 100 µM oligo d(T)+*Pst*I fwd primer (Table 2.3), 4 µl 100 mM dNTPs and 6 µl DEPC H_2O. Samples were heated at 65 °C for 5 minutes, incubated for at least 1 minute on ice and centrifuged briefly. Superscript III RT (4v), 16 µl 5x FS Buffer and 4 µl 0.1 M DTT were added, the mixture was gently mixed, incubated at 50 °C for 50 minutes, and then inactivated by heating at 70 °C for 15 minutes. Reactions were nanodropped to assess the yield of 1st strand cDNA. 2nd strand cDNA was produced through Platinum *Pfx* PCR cycling (Section 2.1.3.3) using primers cDNA_PCR Fwd and cDNA_PCR Rev (Table 2.3). A low primer annealing temperature of 56 °C was chosen to promote primer annealing. Table 2.3 provides details of primers used for this protocol.

Table 2.3 Primers used in VapC target identification protocol

Primer		Primer Sequence*
F	oligo d(T)+*Pst*I	CTGGCACTGCAGTTTTTTTTTTTTTTA
	cDNA_PCR	GAGGATCCCTGCAGGAAA
R	cDNA_PCR	GACCGTGACGTCAAAAAAAAAA

*Restriction enzyme sites (*Pst*I and *Bam*HI) recognition sites are underlined

2.5 Methods relating to VapBC DNA Binding Experiments

Purified VapBC protein complex was used in EMSAs in order to examine its ability to bind to its cognate *vapBC* promoter region. EMSA experiments were performed using a DIG gel shift kit, 2nd Generation (Roche, USA). VapBC binding activity was tested against DNA fragments that were 3'-end labelled with the immunohistochemical marker digoxigenin-11-ddUTP. All primers used are listed in Table 2.4.

2.5.1 Amplification of *Rv0065a/c* and *Rv0617a/c* Promoter DNA

Primers Rv0617promoterF and Rv0617promoterR were designed using Geneious Pro (Version 5.02) (BioMatters Ltd, NZ) to amplify a region 130 bp upstream of the GTG translational start site of the *Rv0617a/c* operon. Primers Rv0065promoterF and Rv0065promoterR were used to amplify a region 101 bp upstream of the ATG translational start site of the *Rv0065a/c* operon. Primers Rv0065promoter dstrF and Rv0065promoter dstrR were used to amplify a 150 bp region spanning 38 bp upstream and 112 bp downstram of the ATG translational start site of the *Rv0065a/c* operon. Genomic H37Ra DNA (Proteins & Microbes lab, University of Waikato) was used as the template for amplifications. A temperature gradient PCR found the optimum primer annealing temperature for all primers to be 63.5 °C; this temperature was used in all subsequent PCRs using the Platinum *Pfx* enzyme (Section 2.1.3.3). Platinum *Pfx* was chosen as the polymerase due to its higher fidelity, 26 x higher than that of the *Taq* enzyme. Reactions were run on 2% agarose gels and the correct sized bands excised and purified using a QIAGEN gel extraction kit (Section 2.1.1.3 and Section 2.1.1.4).

2.5.2 Amplification of the *mazEF* Promoter

To assess the specificity of VapBC binding to the promoter of the operon from which it was encoded, a non-specific competition control was required in EMSA assays. Non-specific competitor DNA consisted of a non-target region of DNA from elsewhere in the *M. tb* genome. The promoter of an operon encoding proteins from the MazEF TA family, *Rv1494/5*, was chosen for this. *Rv1494* is a 303 bp gene annotated at coordinates 1686269-1686571 on the *M. tb* H37Ra chromosome. Based on this, a region 600 bp upstream of coordinate 1686269 was amplified from H37Ra genomic DNA (Proteins & Microbes lab, University of Waikato) using primers Rv1494promoterF and Rv1494promoterR. The DNA was then purified from solution (Section 2.1.1.5).

Table 2.4 Primers used for investigation of VapBC-promoter binding activity

Primer		Primer Sequence
F	Rv0065promoter	CTGGCCGATCTGCGCGCGGT
	Rv0065promoter dstr	GTGGATAAGCCGGCATTCTTAGCCG
	Rv0617 Bind2	GCCACCCAACACGGCTCAAAA
	Rv0617 Bind3	GCCGGCTCGTCTGCCAAAAGG
	Rv0617promoter	CCGCCACCCAACACGGCTCGT
	Rv1494promoter	CGAGATCTCCGAAGCGCTGGAGAAG
R	Rv0065promoter	AGCGGAGTTCACCGGCTAAGAATGCCG
	Rv0065promoter dstr	TCGATGAGCTTGGTGCGCATATACG
	Rv0617 Bind2	GAGCCGGCCGAGACAGCTTTT
	Rv0617 Bind3	CTGATCATGCTGGCCTTTTGG
	Rv0617promoter	GCCTCATCCTAACATCGCTGCATCGTGATGC
	Rv1494promoter	CGCGGGCCTCTCGCTAATCCAGCGTG

Primers in this table include those used for amplification of promoter regions (promoters of *Rv0065a/c*, *Rv0617a/c* and *Rv1494/5* operons) and two-halves mutagenesis.

2.5.3 Creation of Perfect IR Mutations

Two halves mutagenesis (Section 2.1.1.10) was used to create versions of the 130 bp *Rv0617a/c* promoter region where either the 5'-half, the 3'-half, or both halves of the 6 bp perfect IR were mutated to contain adenine (A) at the four central base positions. This method utilised mutagenic primers along with primers complementary to the 5' and 3'-ends of the *Rv0617a/c* 130bp promoter region. 'Bind 2' represents a version of the 130 bp promoter region where the upstream perfect IR half was mutated, while 'Bind 3' represents the same for the downstream perfect IR half. Primer pairs Rv0617 Bind2F and Rv0617 promoterR, and Rv0617 promoterF and Rv0617 Bind2R were used in iProof PCRs according to the manufacturer's instructions (Section 2.1.3.5) to create each half of the Bind2 strand. Halves were joined in a Platinum *Pfx* PCR (Section 2.1.3.3) using primers Rv0617promoterF and Rv0617promoterR. The same process was used to create Bind 3, however the touchdown PCR protocol (Section 2.1.3.6) was instead used for creation of the two halves. Bind 4, where both halves of the perfect IR were mutated, was created in the same manner as Bind 3 using Rv0617 Bind3 primers, and the Bind 2 sequence as the template.

2.5.4 Design of Promoter Oligonucleotides

The 130 bp *Rv0617a/c* promoter region was divided into five overlapping oligos of 40 or 50 bases in length for use in EMSAs in order to narrow down candidate near-perfect inverted repeats (NPIRs) to which $VapBC_{Rv0617}$ bound. The design

was such that both halves of each candidate IR were included in at least one of the oligos. Table 2.5 outlines the properties of these five promoter oligos. Oligos were ordered from Integrated DNA Technologies (IDT, USA) as sense and antisense ssDNA (Table 2.6). These strands were annealed as outlined in Section 2.1.1.9.

Table 2.5 Properties of designed *Rv0617a/c* promoter oligonucleotides

Oligo	Length (bp)	Covers both halves of:	Covers one half of:
1	40	$IR_{A,}$ IR_{B1}	IR_{B2}
2	40	IR_{B2}	$IR_{A,}$ $IR_{B1,}$ $IR_{C,}$ IR_{D}
3	40	$IR_{C,}$ IR_{E}	$IR_{D,}$ $IR_{F,}$ IR_{G}
4	50	$IR_{D,}$ $IR_{E,}$ $IR_{F,}$ IR_{G}	$IR_{B2,}$ IR_{H}
5	50	$IR_{H,}$ $IR_{I,}$ IR_{J}	IR_{D}, IR_{F}

2.5.5 Electrophoretic Mobility Shift Assays (EMSAs)

2.5.5.1 Labelling Target DNA

Double-stranded DNA (100 ng) and sterile MilliQ water were added to a microfuge tube to a final volume of 10 µl. Reactions were set up on ice, and to the oligonucleotide was added 4 µl 5 x Labelling Buffer (Roche), 4 µl 25 mM $CoCl_2$ solution, 1 µl 1 mM DIG-ddUTP solution (Roche) and 1 µl 400 U/µl terminal transferase (Roche). Reactions were mixed and spun briefly, then incubated at 37 °C for 15 minutes. Reactions were stopped by addition of 2 µl of 0.2 M EDTA pH 8.0 and returned to ice. Sterile MilliQ water (3 µl) was added to bring the final reaction volume to 25 µl and concentration to 4 ng/ul (0.155 pmol/µl) of labelled DNA. The DIG-labelled DNA was diluted to 0.4 ng/ul (4.80 fmol) with TEN Buffer (10 mM Tris, 1 mM EDTA, 0.1 M NaCl, pH 8.0) for use in EMSAs.

2.5.5.2 Preparation of VapBC for EMSA

VapBC molarity calculations were based on the tetrameric VapBC complex sizes of 98.04 kDa (98040 g/mol) and 94.8 kDa (94800 g/mol) for $VapBC_{Rv0065a/c}$ and $VapBC_{Rv0617a/c}$ respectively. These values were calculated in ProtParam (au.expasy.org/tools/protparam.html). Purified VapBC protein was diluted to the correct molarity for binding assays in 50 mM potassium phosphate buffer pH 7.3 containing 200 mM NaCl and 0.05 mg/ml bovine serum albumin (BSA).

Table 2.6 Primers used in creation of *Rv0617a/c* promoter oligos.

ssDNA Oligo		ssDNA Oligo sequence
F	Oligo 1	CCGCCACCCAACACGGCTCGTGCGCTGTCTCGGCCGGC TC
	Oligo 2	CTCGTGCGCTGTCTCGGCCGGCTCGTCTGCCGCACGGC CA
	Oligo 3	GCTCGTCTGCCGCACGGCCAGCATGATCAGTCCCGTTG GA
	Oligo 4	CACGGCCAGCATGATCAGTCCCGTTGGAATACCGGTGA GCGTCGGCGCGC
	Oligo 5	GGTGAGCGTCGGCGCGCGCATCACGATGCAGCGATGT TAGGATGAGGCGGTG
	Oligo 5 mut1	GGAAAAAATCGGCGCGCGCATCACGATGCAGCGATGT TAGGATGAGGCGG
	Oligo 5 mut2	GGTGAGCGTCGGCGCGAAAAAAACGATGCAGCGATGT TAGGATGAGGCGG
	Oligo 5 mut3	GGTGAGCGTCGGCGCGCGCATCACGATGCAGCGATGT TAGGAAAAAAAAG
	Oligo5 mut4	GGTGAGCGTCGGCGCGCGCATCACGATGCAAAAAAAT TAGGATGAGGCGG
R	Oligo 1	GAGCCGGCCGAGACAGCGCACGAGCCGTGTTGGGTGG CGG
	Oligo 2	TGGCCGTGCGGCAGACGAGCCGGCCGAGACAGCGCAC GAG
	Oligo 3	TCCAACGGGACTGATCATGCTGGCCGTGCGGCAGACG AGC
	Oligo 4	GCGCGCCGACGCTCACCGGTATTCCAACGGGACTGATC ATGCTGGCCGTG
	Oligo 5	CCGCCTCATCCTAACATCGCTGCATCGTGATGCGCGCG CCGACGCTCACC
	Oligo 5 mut1	CCGCCTCATCCTAACATCGCTGCATCGTGATGCGCGCG CCGATTTTTTCC
	Oligo 5 mut2	CCGCCTCATCCTAACATCGCTGCATCGTTTTTTTCGCGC CGACGCTCACC
	Oligo 5 mut3	CTTTTTTTTCCTAACATCGCTGCATCGTGATGCGCGCGC CGACGCTCACC
	Oligo 5 mut4	CCGCCTCATCCTAATTTTTTTGCATCGTGATGCGCGCGC CGACGCTCACC

2.5.5.3 EMSA Reactions

EMSA reactions were set up on ice to a final volume of 6.5 µl. Into each reaction was added the appropriate volume of purified VapBC protein, 0.4 ng DIG labelled DNA, 2 µl 5 x Binding Buffer (Roche), 0.5 µl 1 mg/ml poly dI-dC (Roche) and MilliQ water. Binding reactions were incubated for 20 minutes at room

temperature. Reactions were placed back on ice and 2.5 µl of loading buffer with bromophenol blue (Roche) was added.

2.5.5.4 EMSA Controls

Specific competition

A reaction was set up as outlined in Section 2.5.5.3 with the addition of 125 x excess (50 ng) of un-labelled target DNA (*Rv0617a/c* promoter for $VapBC_{Rv0617a/c}$ assays, and *Rv0065a/c* promoter for $VapBC_{Rv0065a/c}$ assays).

Non-specific competition

A reaction was set up as outlined in Section 2.5.5.3 with the addition of 125 x excess (50 ng) of un-labelled non-target DNA (600 bp region of the *M. tb mazEF* promoter).

2.5.5.5 Native TBE Polyacrylamide Gel Electrophoresis

Tris-borate-EDTA (TBE) native acrylamide gels (6%) were prepared the day before use to ensure complete polymerisation. Gels were pre-run in 1 x TBE Buffer at 60 V for 20 minutes. Samples were loaded onto the gel and run at 150 V until the dye-front had migrated approximately two-thirds of the way down the gel.

2.5.5.6 Contact-Blotting and Crosslinking of Oligonucleotides

One plate was removed from the 6% TBE gel. A dry Hybond-N^+ membrane (GE Healthcare, USA) trimmed to size was placed onto the gel, followed by three layers of dry Whatman 3 MM blotting paper, a glass plate, and a load of 100 g. Transfer was complete after 30 minutes. The membrane was transferred DNA side up onto a sheet of Whatman 3 MM blotting paper soaked in 2 x SSC (300 mM NaCl, 30 mM $Na_3C_6H_5O_7$, pH 7.0) and placed onto a sheet of tinfoil. Crosslinking was achieved using a Bio-Link BLX-254 BRL UV Crosslinker Irradiation System (Life Technologies, USA) which irradiated the membrane with 120 mJ at UV 254 nm for 20 seconds. The membrane was either used immediately, or stored air-dried between two sheets of Whatman 3 MM blotting paper at 4 °C.

2.5.5.7 Chemiluminescent Detection

All incubations were performed at 15 – 25 °C with agitation in a small plastic container. The membrane was washed for 2 minutes in 40 ml washing buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5, 0.3 % *(v/v)* Tween-20), and then

incubated for 30 minutes in 40 ml 1x blocking solution which was prepared by diluting a 10x blocking solution stock (Roche) in maleic acid buffer (0.1 M Maleic acid, 0.15 M NaCl, pH 7.5). An antibody solution was prepared by centrifuging Anti-Digoxigenin-AP (Roche) for 5 min at 10,000 rpm, and diluting the antibody 1 : 10,000 (75 mU/ml) in 1 x blocking solution. The membrane was incubated in the antibody solution for 30 minutes. Two 15 minute wash steps were conducted in 40 ml washing buffer, followed by equilibration for 2-5 minutes in 40 ml detection buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5). The membrane was placed between two plastic sheets and incubated with 1 ml of CSPD solution (Roche) (diluted 1:100 in detection buffer) for 5 minutes at room temperature. The membrane was sealed in the plastic and incubated at 37 °C for 10 minutes to enhance the luminescence reaction. The membrane was exposed to film (Amersham Hyperfilm^{TM} MP) for 20 – 25 minutes. In a dark room, films were developed for at least 2 minutes in Kodak Developer and replenisher (Sigma Aldrich) and fixed for at least 30 seconds in Kodak Fixer and replenisher (Sigma Aldrich). Films were visualised on a light box and photographed.

Chapter Three: Results

3.1 The Effects of Conditional Expression of $VapC_{Rv0617}$ on *Mycobacterium smegmatis* Growth and Morphology

Studies conducted on a number of the 47 *vapBC* TA systems encoded within the *M. tb* genome have revealed that some, but not all of the *M. tb* VapC proteins have a growth inhibitory effect when overexpressed in mycobacterial species (Gupta 2009; Ramage et al. 2009; Ahidjo et al. 2011). To determine whether or not the $VapC_{Rv0617}$ protein from *M. tb* displays toxicity, observations were made of its effect on the growth and viability of *M. smegmatis* cells when conditionally overexpressed. In addition, the toxicity was classified as either bacteriostatic or bactericidal depending on whether or not *M. smegmatis* growth could be rescued when expression of the VapC protein was halted. Ideally experiments would be conducted in the native organism *M. tb;* however due to hazard restrictions the non-pathogenic relative *M. smegmatis* was chosen as an appropriate model.

3.1.1 The Effect of $VapC_{Rv0617}$ on Cell Growth and Viability

For conditional expression experiments, four inducible constructs were created using pMind vectors, containing either an entire *vapBC* operon or a single *vapC* gene under the control of a tetracycline inducible promoter (Section 2.2.1). Constructs were designated 0065vapBC_pMind, 0065vapC_pMind, 0617vapBC_pMind and 0617vapC_pMind. All constructs were designed to encode a synthetic ribosomal binding site (RBS) at the 5' end in order to ensure translation *in vivo*. The whole-operon constructs were included in the experiment as a means to confirm that any toxicity observed was a result of the VapC protein alone and not of the VapBC complex. These constructs also served as controls to ensure that any growth inhibition was not an artefact of the general overexpression of a protein. Each construct was transformed into two electrocompetent strains of *M. smegmatis*; mc^2155 (wild-type strain) and $mc^2155\Delta$*vapBC* (deletion strain where the single *vapBC* operon of *M. smegmatis* was deleted from the chromosome) (Section 2.1.1.14). The deletion strain was prepared by Jennifer Robson (University of Otago).

M. smegmatis mc^2155 wild-type and Δ*vapBC* deletion strains containing 0617vapBC_pMind, 0617vapC_pMind and an empty pMind vector were grown up for large-scale conditional expression studies as outlined in Section 2.2.2. After 19 hours of incubation, the 200 ml cultures grown in HdB medium reached OD_{600} values between 0.1 - 0.15 and thus were simultaneously induced with tetracycline (20 ng/ml). Blokpoel et al (2005) previously reported that greatest expression using the pMind vector occurred 24 hrs after 20 ng/ml tetracycline induction. Optical density at 600 nm was therefore monitored 6 hourly over the following 121 hours, and 11 measurements of viability (plating of cells for CFU determinations) were taken over a 96 h period post-induction.

Analysis of changes in the OD_{600} readings over the 121 hour period (Figure 3.1) indicated that both *M. smegmatis* deletion strains and wild-type strains entered the stationary phase after an incubation time of approximately 37 hours. Deletion strains expressing $VapBC_{Rv0617a/c}$ and the empty vector reached an OD_{600} maximum of ~3 and maintained OD_{600} readings within the range of 2.57 - 3.04 through the remainder of the 121 hr sampling period. The deletion strain expressing $VapC_{Rv0617}$ exhibited slight inhibition of growth, where a reduced OD_{600} maximum of ~2.3 was reached after 37 hours. The OD_{600} of the toxin-expressing deletion strain fell to a final reading of ~1.4 over the sampling period.

$VapC_{Rv0617}$-mediated growth inhibition was similar in the deletion strain and the wild-type strain expressing the $VapC_{Rv0617}$ toxin. The wild-type strain expressing $VapC_{Rv0617}$ reached a maximum OD_{600} of 2.2 after 49 hrs incubation. This was in comparison to the OD_{600} maximum of ~3 that was reached at the 37 hour incubation point in the wild-type strains expressing $VapBC_{Rv0617a/c}$ or empty vector. Final OD_{600} readings at 96 hours post-induction were 3.03, 2.85 and 1.48 for wild-type strains expressing empty vector, $VapBC_{Rv0617a/c}$ and $VapC_{Rv0617}$ respectively, while these values were 2.8, 2.65 and 1.4 for deletions strains expressing empty pMind vector, $VapBC_{Rv0617a/c}$ and$VapC_{Rv0617}$ respectively.

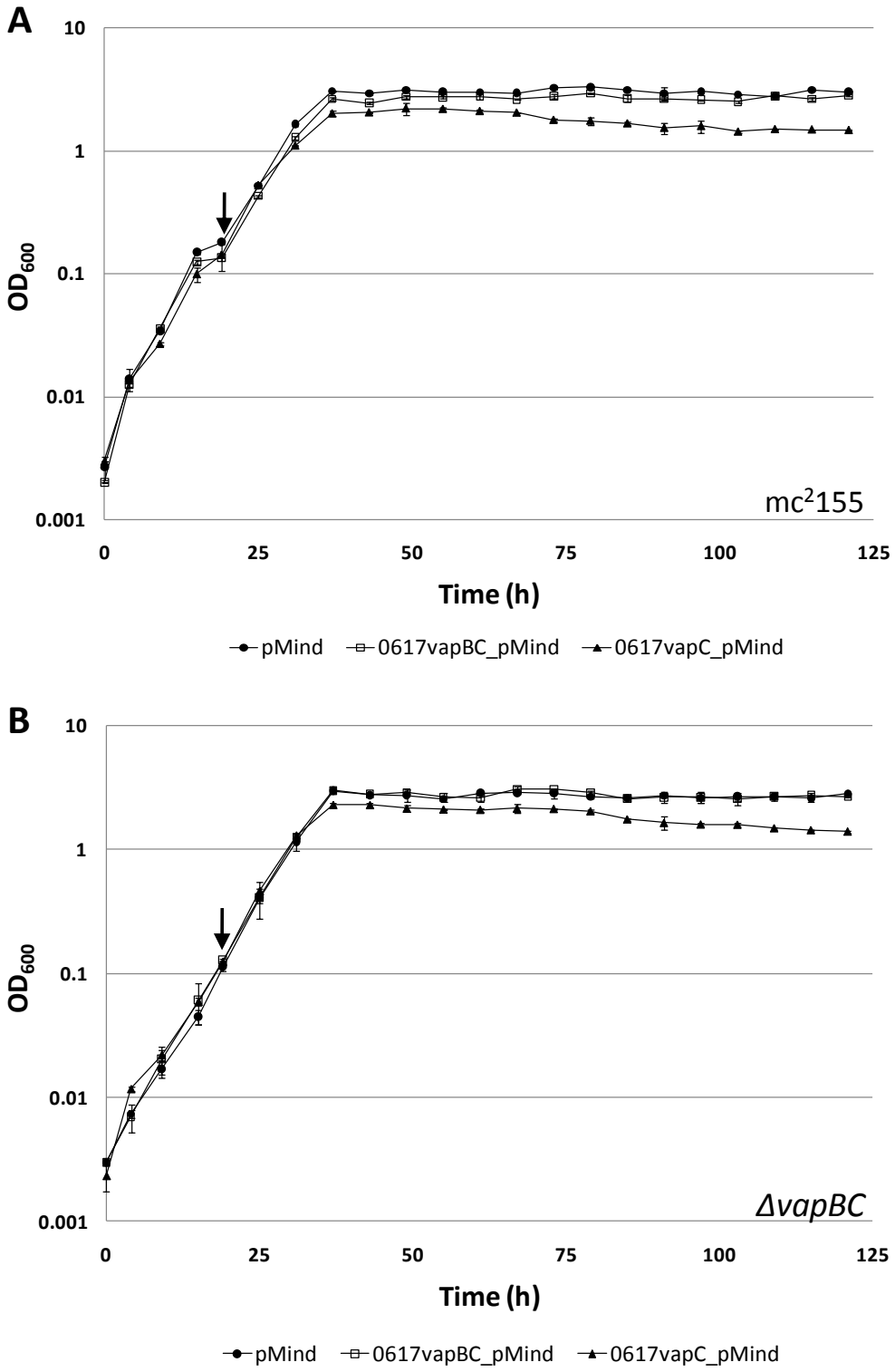


Figure 3.1. The effect of $VapC_{Rv0617}$ overexpression on the growth of *M. smegmatis* mc^2155 and Δ*vapBC* strains. OD_{600} was monitored for *M. smegmatis* mc^2155 (**A**) and Δ*vapBC* (**B**) strains harbouring plasmids 0617vapBC_pMind (expressing VapBC; open squares), 0617vapC_pMind (expressing VapC; solid triangles) or pMind (expressing empty vector; solid circles). Protein expression was induced during early exponential growth (OD_{600} of 0.1-0.2) by the addition of 20 ng/ml tetracycline (indicated by black arrow). Results shown indicate the mean ± from one independent experiment representative of three biological replicates.

In order to further investigate the effect of $VapC_{Rv0617}$ on cell viability and to determine whether this effect was of a bacteriostatic or bactericidal nature, samples of the culture were taken at 11 time points after tetracycline induction (6 hourly for 48 hours, followed by 24 hourly for 48 hours) and plated onto LBT agar plates containing both hygromycin (50 µg/ml) and tetracycline (20 ng/ml), or hygromycin (50 µg/ml) alone (Section 2.2.2). All plates for the *M. smegmatis* deletion strain were additionally supplemented with kanamycin at 50 µg/ml The inclusion of tetracycline in the plates served to maintain overexpression of the pMind construct ($VapBC_{Rv0617a/c}$, $VapC_{Rv0617}$ or empty vector). At each time point, a 1 ml sample of culture was taken and three 5 µl serial dilutions of this were spotted onto the plate. Plates were incubated for 3 days at 37 °C, after which colonies were counted for determination of CFUs and any differences in morphology between strains were noted. An example of colony growth for a dilution series of two wild-type *M. smegmatis* mc^2155 cultures carrying an empty pMind vector 24 hours after tetracycline induction is shown in Figure 3.2.

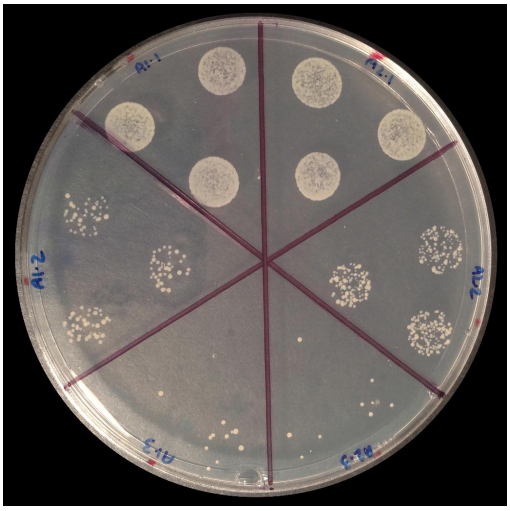


Figure 3.2 Colonies resulting from 5 µl spots of culture dilution series of 10^{-2}, 10^{-4} and 10^{-6} after three days incubation at 37 °C. Dilutions for two replicates of the *M. smegmatis* mc^2155-pMind strain are pictured, starting both clockwise and anticlockwise downward from the top two segments.

The degree of $VapC_{Rv0617}$ toxicity became more apparent when comparing measurements of CFUs for both wild-type and deletion strains expressing the $VapC_{Rv0617}$ toxin, VapBC complex or empty vector (Figure 3.3). Again, the relative effects of $VapC_{Rv0617}$, $VapBC_{Rv0617a/c}$ or empty vector expression between the wild-type and deletion strain were similar. By 96 hours post-induction, the wild-type strain expressing $VapC_{Rv0617}$ exhibited an 87-fold reduction in cell

viability compared to that of the strain expressing the empty vector, and an 86-fold reduction compared to the strain expressing the $VapBC_{Rv0617a/c}$ complex (Figure 3.3A). Similarly, induction of VapC expression in deletion strains led to 86-fold and 82-fold reductions in cell viability when compared to deletion strains expressing empty vector and $VapBC_{Rv0617}$ respectively (Figure 3.3B).

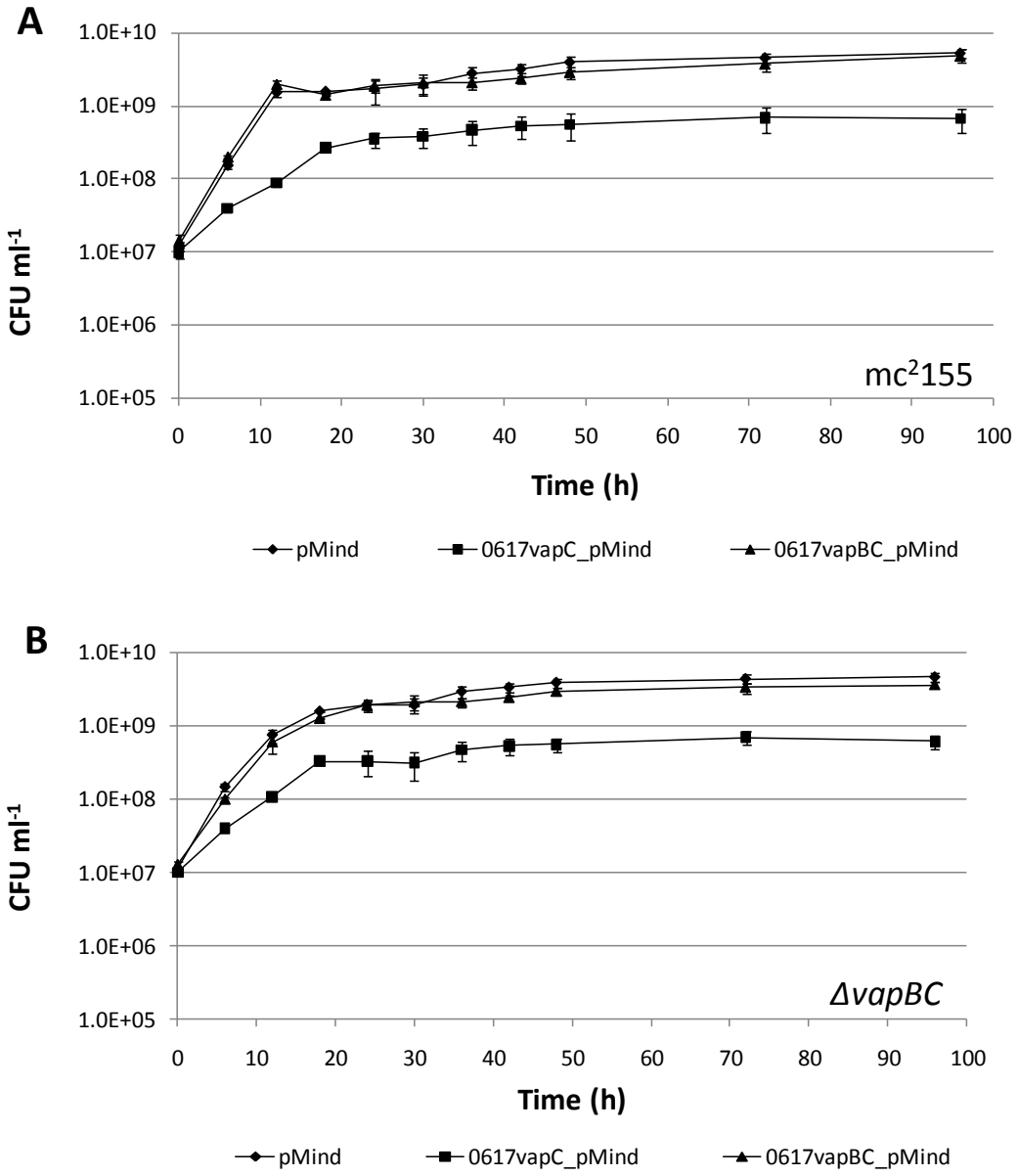


Figure 3.3 The effect of $VapC_{Rv0617}$ overexpression on the viability of *M. smegmatis* mc²155 and Δ*vapBC* strains. The effect of conditionally expressing $VapC_{Rv0617}$, $VapBC_{Rv0617a/c}$ or empty pMind vector on cell viability was examined in *M. smegmatis* mc²155 (**A**) and a Δ*vapBC* deletion strain (**B**). Strains expressing empty vector are denoted by solid circles, strains expressing $VapC_{Rv0617}$ by solid squares and strains expressing $VapBC_{Rv0617a/c}$ by open squares. CFU viability measurements were taken in the 96 h period following induction of expression with 20 ng/ml tetracycline. Graphs shown are from data acquired from one independent experiment representative of three biological replicates. Results shown indicate the mean ± SD of three technical replicates of each of three biological replicates.

3.1.2 The Effect of $VapC_{Rv0617}$ on Cell Morphology

At all sampling time points, both *M. smegmatis* mc^2155 and *M. smegmatis* Δ*vapBC* strains expressing either $VapBC_{Rv0617a/c}$ or the empty pMind vector exhibited normal colony growth on the agar plates in both the presence and absence of tetracycline. In contrast, when grown on tetracycline-supplemented plates, both the wild-type and deletion strains expressing $VapC_{Rv0617}$ formed fewer colonies, many of which were smaller than average (Figure 3.4). Colony growth was normal when $VapC_{Rv0617}$ expressing strains were grown on agar plates in absence of tetracycline, indicating the inhibitory effect was of a bacteriostatic nature.

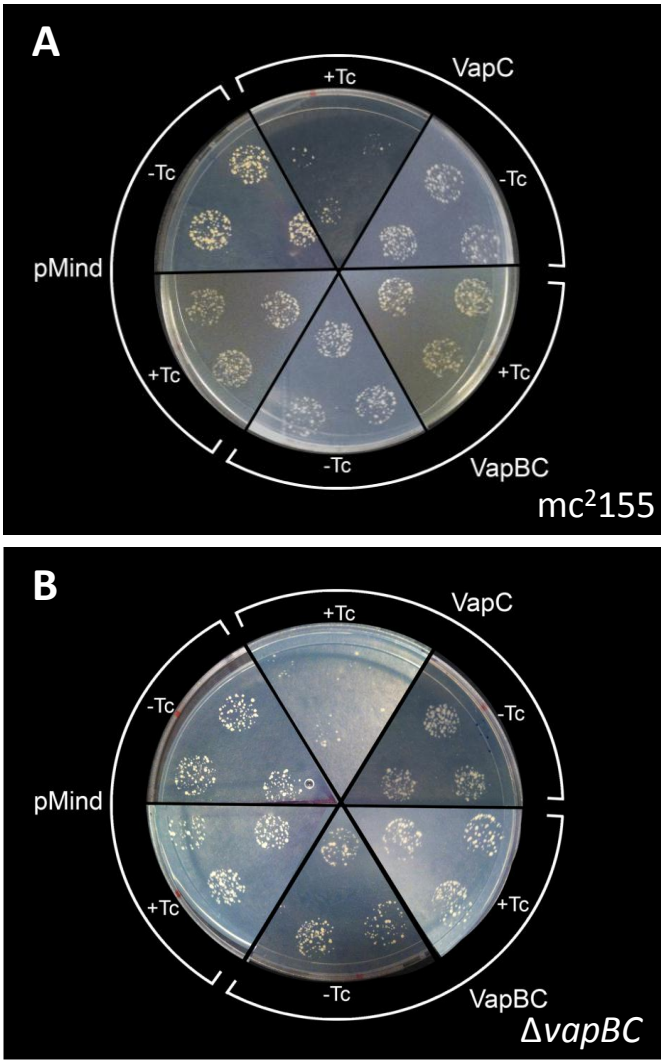


Figure 3.4 The effect of conditional expression of VapBC, VapC, and empty pMind vector on the growth of *M. smegmatis* mc^2155 and Δ*vapBC* strains. A mosaic of agar plates showing colony growth of the *M. smegmatis* mc^2155 strain (**A**) and the Δ*vapBC* deletion strain (**B**) containing pMind, 0617vapBC_pMind or 0617vapC_pMind in the presence (+Tc) and absence (-Tc) of tetracycline. Plates represent cultures 48 hours post-induction. Each segment includes three technical replicates of *M. smegmatis* culture growth at 48 hours post-induction diluted 10,000-fold.

3.2 Protein Expression and Purification

For activity assays, the non-toxic $VapBC_{Rv0065a/c}$ and $VapBC_{Rv0617a/c}$ protein complexes and toxic $VapC_{Rv0065}$ and $VapC_{Rv0617}$ proteins of *M. tb* were successfully expressed and purified as outlined in this section.

3.2.1 Protein Expression in *Mycobacterium smegmatis*

Expression and purification of the toxic VapC proteins alone has previously represented a significant hurdle in their biochemical characterisation (Mattison et al. 2006; Ramage et al. 2009). The expression protocol followed in this thesis for *M. tb* VapBC and VapC purification was based on that developed for expression of the *M. smegmatis* protein $VapC_{MS1284}$.

Expression of $VapC_{MS1284}$ was initially attempted in *M. smegmatis*; however the toxicity of this protein posed a problem. In an attempt to escape the toxicity, plasmids containing this construct were driven to mutation, resulting in a nonsensical stop-codon upstream of the true *vapC* stop codon. Subsequent expression attempts of $VapC_{MS1284}$ as both a His-tagged fusion protein and as a maltose-binding protein (MBP) fusion protein were unsuccessful, resulting in insolubility and the formation of large soluble VapC aggregates respectively. The large soluble VapC aggregates appeared to protect the MBP from being proteolytically degraded which hampered downstream purification. In order to overcome these issues, an alternate approach was taken that involved initial expression and purification of the entire $VapBC_{MS1283/4}$ complex. Given its lack of toxicity, the complex is more amenable to overexpression *in vivo* (Robson et al. 2009). Expression of the VapBC complex in *E. coli* was ruled out after it was determined that the *E. coli* transcriptional machinery does not recognise the overlap in bases occurring between the *vapB* and *vapC* genes in the *vapBC* operon, resulting solely in expression of insoluble VapB (Summers 2007; McKenzie 2011). Expression of the VapBC complex in *M. smegmatis* was much more successful. Greater success in this organism may be due to its larger degree of similarity to *M. tb*, meaning it has the correct mycobacterial molecular systems in place required for the production of *M. tb* proteins.

The $VapBC_{Rv0065a/c,}$ $VapBC_{Rv0617a/c,}$ $VapC_{Rv0065}$ and $VapC_{Rv0617}$ proteins were expressed in this manner, and IMAC and size exclusion chromatography were

carried out to purify either VapBC complex. Functional VapC protein could subsequently be acquired through a limited trypsin digest to remove the labile VapB antitoxin which is reported as being more readily affected by proteolytic degradation (Gerdes et al. 2005). Anion exchange chromatography was then required to separate the VapC toxin from the trypsin and degraded VapB.

3.2.2 Cloning of *Rv0065a/c* and *Rv0617a/c* Operons into pYUB28b

E. coli-mycobacterial shuttle plasmids, pYUB28b, containing either the *Rv0065a/c* TA operon or the *Rv0617a/c* TA operon (amplified from H37Ra strain genomic DNA) were obtained from the Proteins & Microbes Lab, University of Waikato. The pYUB28b plasmid places the introduced construct under the control of a T7 promoter and enables purification with the addition of a C-terminal His-tag. Plasmids were initially transformed into electrocompetent *E. coli* TOP10 cells (Sections 2.1.1.11 and 2.1.1.12) due to the ease of plasmid purification from this strain. Purified plasmids were sequenced to confirm the presence and accuracy of the insert sequence before being transformed into the *M. smegmatis* expression strain $mc^2$4517 (Sections 2.1.1.13 and 2.1.1.14).

3.2.3 Small Scale Protein Expression Trials

Following transformation, expression of VapB and VapC protein from the pYUB38b vector was confirmed through small scale protein expression trials (Section 2.3.1). For both *Rv0065a/c* and *Rv0617a/c* protein expression, trials revealed the presence of VapB and VapC protein in the Ni sepharose fraction, indicating that both proteins were successfully being expressed and that the VapC protein possessed the C-terminal His-tag. VapB and VapC proteins were absent in the insoluble fraction and in wash fractions (data not shown).

3.2.4 IMAC Purification of VapBC

Large scale expression cultures were grown up from glycerol stocks as outlined in Section 2.3.2. Cell pellets of these cultures were lysed by sonication, centrifuged, and the supernatant was filtered through 1.2 and 0.45 µM filters. The *Rv0065a/c* and *Rv0617a/c* constructs within the pYUB28b plasmid had been designed so as to add a C-terminal His-tag. This tag enabled isolation of the VapBC complex through IMAC purification with a HiTrap™ Chelating HP column (GE Healthcare, UK), (Section 2.3.3.2.) The chromatogram and corresponding

SDS-PAGE gel in Figure 3.5 depict the purification, column load and column flow-through fractions of a $VapBC_{Rv0065a/c}$ sample. Non-specific proteins bound loosely to the column are eluted off early in the void peak. $VapBC_{Rv0065a/c}$ elutes later in the gradient at approximately ~310 – 405 mM imidazole. Figure 3.6 depicts the chromatogram and corresponding SDS-PAGE gel for a $VapBC_{Rv0617a/c}$ IMAC purification. Here, $VapBC_{Rv0617a/c}$ elutes off the column at approximately 375 – 460 mM imidazole. Peak A in both spectra represents the void peak and contains only non-target proteins. The peaks labelled B and C in both spectra represent points of VapBC elution.

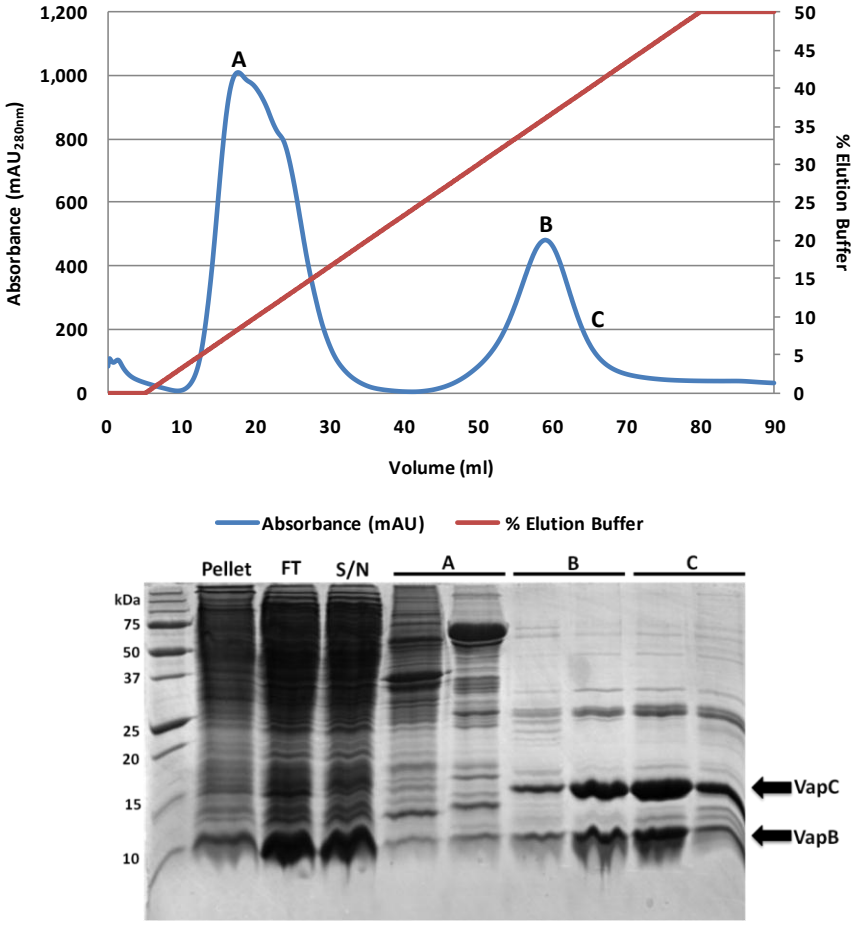


Figure 3.5 $VapBC_{Rv0065a/c}$ IMAC purification and corresponding 16.5% SDS-PAGE gel. The chromatogram depicts the UV absorbance and elution profile of $VapBC_{Rv0065a/c}$. Lanes in the SDS-PAGE gel denote insoluble protein (Pellet), column flow through (FT) and soluble protein (supernatant, S/N). Fractions and their corresponding elution position on the spectrum are labelled A-C. Positions of VapB and VapC proteins are indicated on the gel.

In both gels the pellet, supernatant and flow through display an array of protein species, with enrichment seen for the VapC and VapB bands. Both SDS-PAGE gels for IMAC purifications displayed VapB and VapC protein bands in the later eluted fractions along with a band at approximately 25 kDa. This band may

represent the toxin dimer which, due to the strength of the interactions involved, had not been denatured by the reducing agent or the heat treatment prior to loading the gel, or by the denaturing reagents in the gel itself.

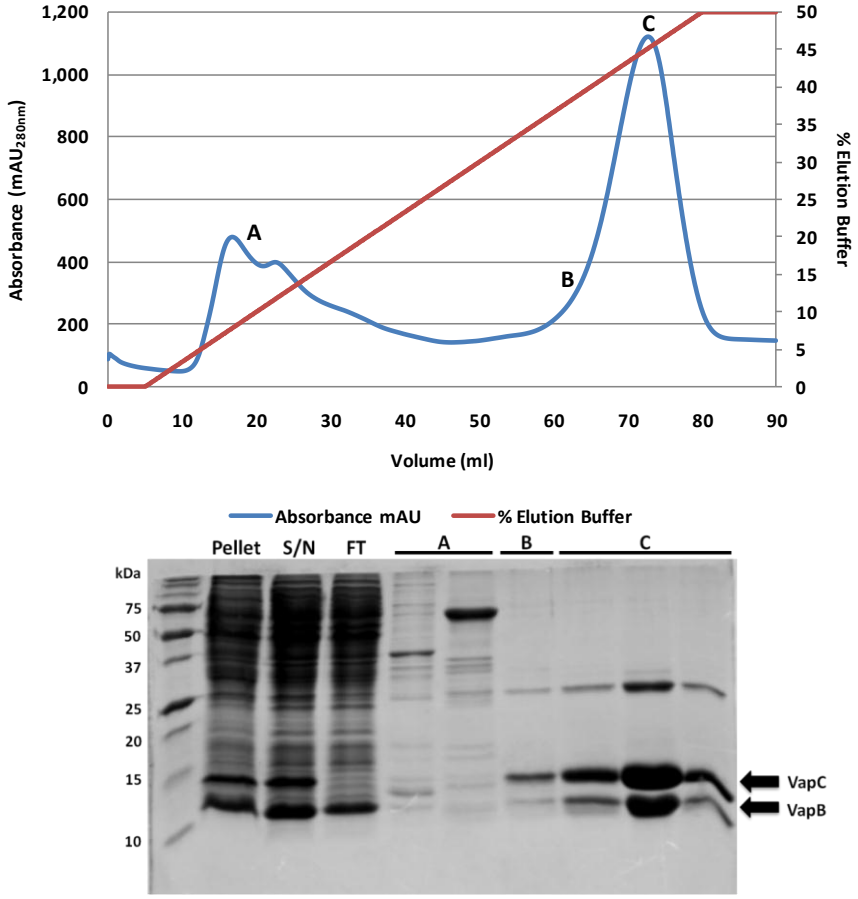


Figure 3.6 VapBC$_{Rv0617a/c}$ IMAC purification and corresponding 16.5% SDS-PAGE gel. The chromatogram depicts the UV absorbance and elution profile of VapBC$_{Rv0617a/c}$. Lanes in the SDS-PAGE gel denote insoluble protein (Pellet), column flow through (FT) and soluble protein (supernatant, S/N). Fractions and their corresponding elution position on the spectrum are labelled A-C. Positions of VapB and VapC proteins are indicated on the gel.

3.2.5 Size Exclusion Purification of VapBC

Fractions containing the VapBC protein complex were pooled and concentrated to 5 ml (Section 2.3.5.2) before being run through a HiLoad^{TM} 16/60 Superdex^{TM} size exclusion column (SEC) (GE Healthcare, UK) (Section 2.3.3.3). Figure 3.7 and Figure 3.8 depict the chromatograms and corresponding SDS-PAGE gels for VapBC$_{Rv0065a/c}$ and VapBC$_{Rv0617a/c}$ size exclusion purifications respectively. It can be seen in both figures that fractions in the region of points B and C contained eluted VapBC protein complex. The presence of a weak higher molecular weight band on gels at ~25 kDa may again represent VapC dimers which had not been fully denatured.

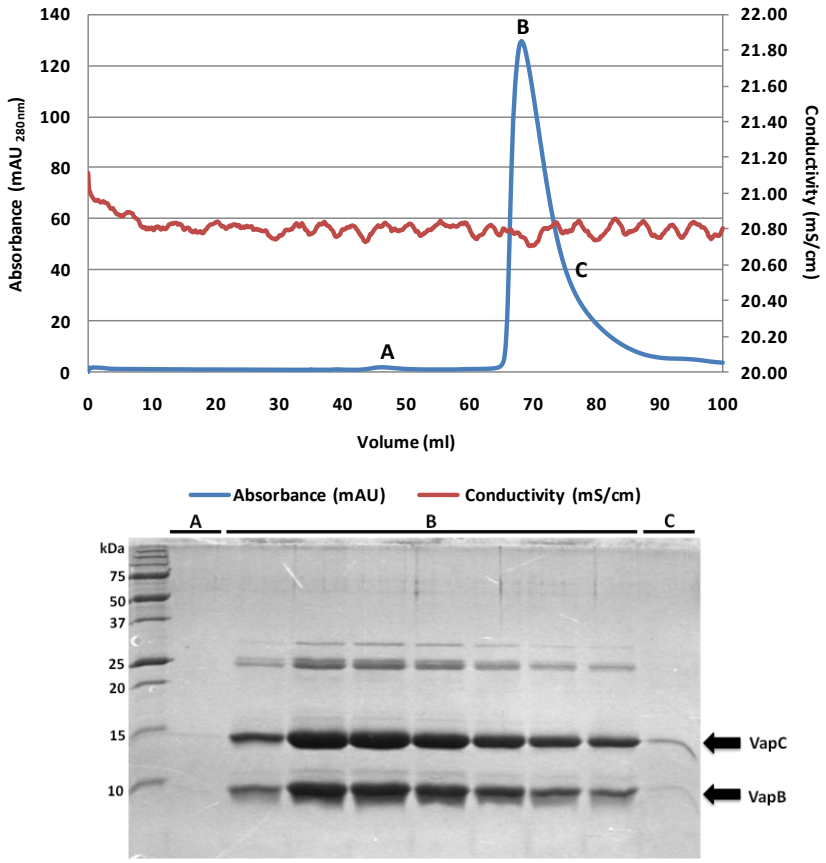


Figure 3.7 $VapBC_{Rv0065a/c}$ size exclusion column purification and corresponding 16.5% SDS-PAGE gel. The chromatographm depicts the UV absorbance and conductivity profile of $VapBC_{Rv0065a/c}$. The SDS-PAGE gel denotes fractions from elution positions A, B and C on the spectrum. Positions of VapB and VapC proteins are indicated on the gel.

Calibration of the S200 16/60 size exclusion column was carried out by Marisa Till (Proteins & Microbes lab, University of Waikato). This allowed calculation of K_{av} (gel phase distribution coefficient) values, and subsequent construction of a calibration curve. The equation from the calibration curve was $K_{av} = -0.159$ (MW) + 2.1739 which is rearranged to form the equation $MW = e^{((Kav-2.1739)/-0.159)}$. Taking into account the calculated K_{av} values of VapBC complex peaks from the *Rv0065a/c* and *Rv0617a/c* SEC purifications (0.277 and 0.173 respectively), this led to calculated MWs of 151.71 kDa ($VapBC_{Rv0065a/c}$) and 292.86 kDa ($VapBC_{Rv0617a/c}$). These values were consistent with those observed in previous purifications (Jo McKenzie, unpublished results). While the observed MW value for $VapBC_{Rv0065a/c}$ is similar to that seen for $VapBC_{MS1282/4,}$ which purifies as a tetramer of VapBC heterodimers bound to a 36 bp segment of DNA (152 kDa) (McKenzie 2011), the observed $VapBC_{Rv0617a/c}$ MW value suggests purification of a larger complex.

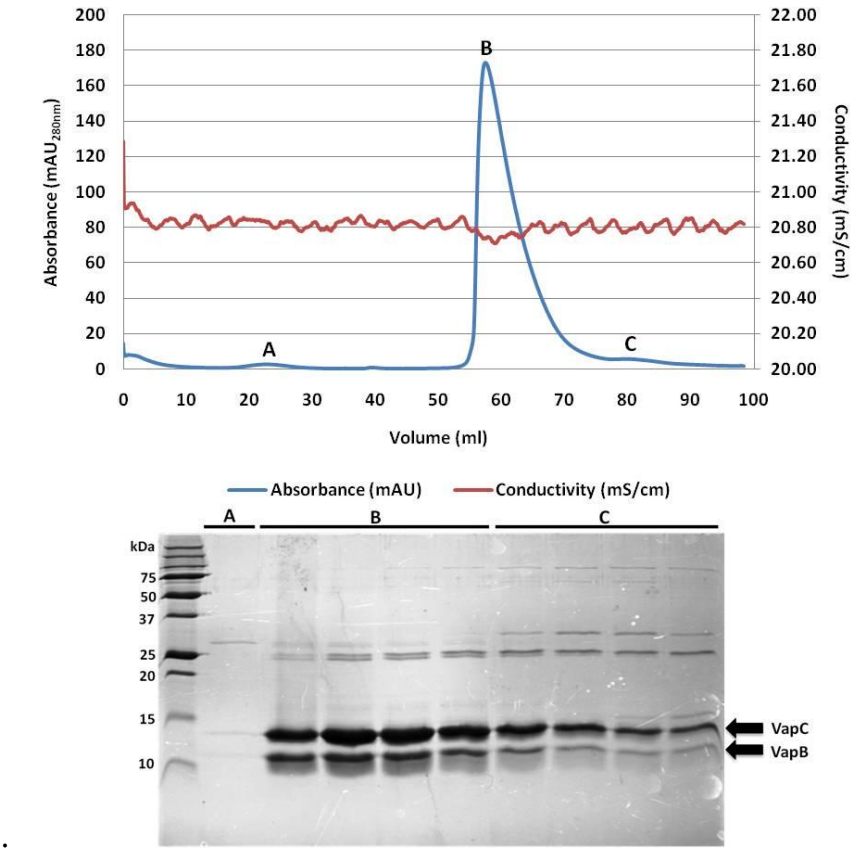


Figure 3.8 $VapBC_{Rv0617a/c}$ size exclusion column purification and corresponding 16.5% SDS-PAGE gel. The chromatogram depicts the UV absorbance and conductivity profile of $VapBC_{Rv0617a/c}$. The SDS-PAGE gel denotes fractions from elution positions A, B and C on the spectrum. Positions of VapB and VapC proteins are indicated on the gel.

Fractions containing eluted VapBC protein were pooled and dialysed into a lower salt buffer overnight (Section 2.3.3.4). Dialysis was required to reduce the salt content of the sample, making it more amenable to subsequent anion exchange purification. The concentration of the VapBC sample was measured at 280 nm using the Nanodrop 2000 spectrophotometer (Nanodrop Technologies, USA) as outlined in Section 2.3.5.3, taking into account the molar extinction coefficient of the VapBC complex (0.649 for $VapBC_{Rv0065a/c}$ and 0.534 for $VapBC_{Rv0617a/c}$) in order to obtain a more accurate concentration value.

3.2.6 Isolation of VapC

A common characteristic of TA systems is greater lability of the antitoxin compared with the toxin (McKenzie 2011). Instability indices of $VapB_{Rv0064}$ (84.53), $VapC_{Rv0065}$ (43.18), $VapB_{Rv0616}$ (47.12) and $VapC_{Rv0617}$ (21.81) are in agreement with this, where larger numbers for both VapB indicate greater instability. Instability indices were calculated using ProtParam (Gasteiger et al. 2005). Based on this property, a limited trypsin digest was used to selectively

degrade VapB whilst leaving functionally active VapC. One hour incubation at room temperature using a ratio of 0.1 mg trypsin: 0.7 mg protein was found to result in optimal degradation of VapB, after which the reaction was stopped by addition of trypsin inhibitor (Section 2.3.4). For functional studies, the VapC protein must be purified away from the degraded VapB, trypsin and the inhibitor. Anion exchange chromatography was suitable for achieving this due to the differing isoelectric points of trypsin (10.5) and VapC (6.34 and 6.5 for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ respectively) (values calculated using ProtParam (Gasteiger et al. 2005)). Trypsin digested protein sample was loaded onto a HiTrap™ Q anion exchange column (GE Healthcare, UK) through a 0.2 µM filter, and anion exchange chromatography was performed to purify the VapC toxin (Section 2.3.5.1). Inactivated trypsin in the sample was eluted off the anion exchange column early on while the VapC toxin had a longer retention time on the column, eluting off at ~200-300 mM NaCl. Figure 3.9 and Figure 3.10 depict the chromatograms and corresponding SDS-PAGE gels for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ anion exchange purifications respectively.

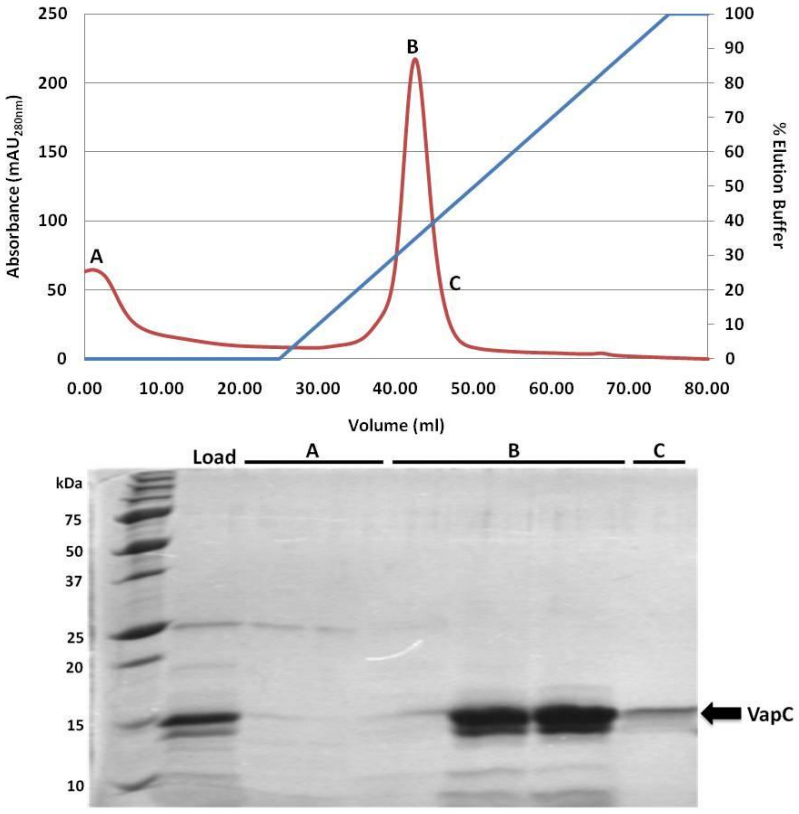


Figure 3.9 Anion exchange purification of a trypsin digest of $VapBC_{Rv0065a/c}$ and corresponding 16.5% SDS-PAGE gel. Trypsin is eluted off the column early on (A) while $VapC_{Rv0065}$ exhibits a greater retention time, remaining bound until early in the NaCl gradient (B & C).

Both gels depict presence of VapC bands and an absence of the previously noted VapB bands. The two smaller sized bands on the gel in Figure 3.9 differing in size to VapB and instead may represent small amounts of VapC degraded during the trypsin digest. The $VapC_{Rv0065}$ and $VapC_{Rv0617}$ bands ran at a slightly smaller size than seen in previous purification gels, attributable to cleavage of the ~1.5 kDa His-tag during the trypsin incubation. Such trypsin-mediated His-tag cleavage was noted during purification of the $VapC_{MS1284}$ protein (McKenzie 2011).

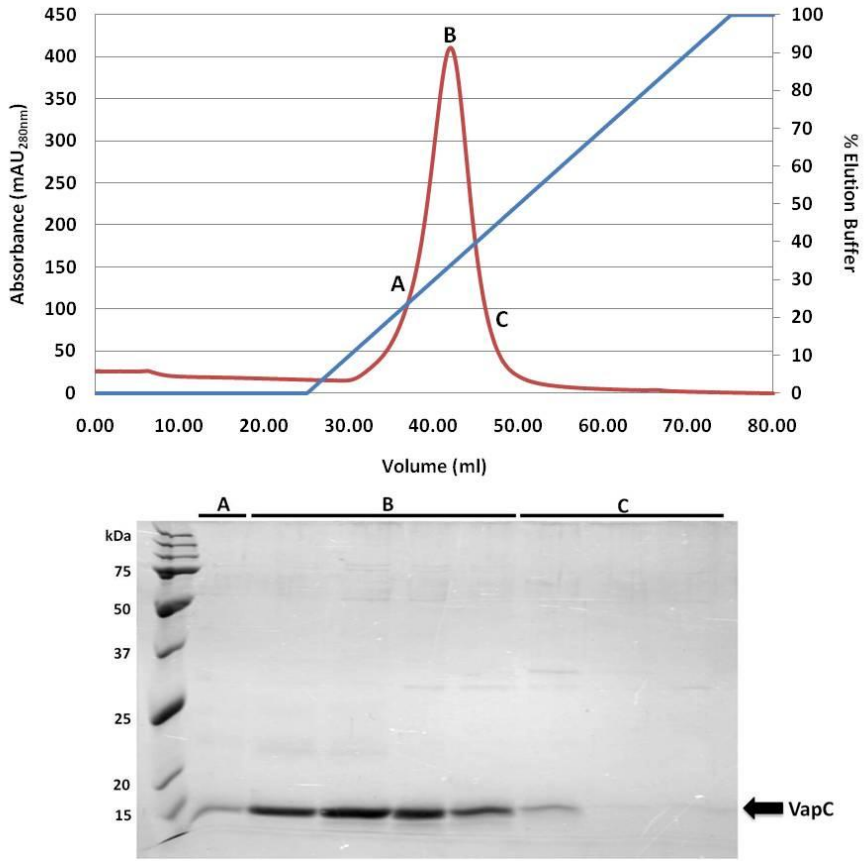


Figure 3.10 Anion exchange purification of a trypsin digest of $VapBC_{Rv0617a/c}$ and corresponding 16.5% SDS-PAGE gel. Trypsin is eluted off the column early on (A) while $VapC_{Rv0617}$ exhibits a greater retention time, remaining bound until early in the NaCl gradient (B & C).

Following anion-purification, fractions containing eluted VapC protein were pooled and concentrated, taking into account the molar extinction coefficients of $VapC_{Rv0065}$ (0.824) and $VapC_{Rv0617}$ (0.816) (values calculated using ProtParam (Gasteiger et al. 2005)).

3.3 VapC Ribonuclease Activity

3.3.1 The Importance of RNA Secondary Structure in VapC Ribonuclease Activity

To date, the impact of RNA secondary structure on VapC ribonuclease activity has yet to be investigated thoroughly. To examine this, ribonuclease assays were conducted *in vitro* with $VapC_{Rv0065}$ and $VapC_{Rv0617}$ against a number of synthesised RNA substrates. These substrates possessed varying secondary structures while retaining GC-rich 4-mers; sites previously determined to be targeted for cleavage by either VapC (Section 1.4.2.2). In these assays, negative controls included either EDTA (as PIN-domain proteins are Mg^{2+}/Mn^{2+}-dependent ribonucleases), VapBC (as cognate VapB inactivates VapC activity through formation of a benign complex), or no VapC (to ensure an absence of contaminating ribonuclease activity).

3.3.1.1 932 Pentaprobe RNA VapC Ribonuclease Activity Assays

Preliminary experiments conducted against all twelve Pentaprobe RNA substrates showed that $VapC_{Rv0065}$ and $VapC_{Rv0617}$ exhibited the most activity against the 932 Pentaprobe (Section 1.4.2.2). To further confirm this, $VapC_{Rv0065}$ and $VapC_{Rv0617}$ were tested in RNase assays against single-stranded 932 Pentaprobe RNA acquired from Joanna McKenzie (Proteins & Microbes lab, University of Waikato). Results were in agreement with previous findings, as both VapC proteins displayed Mg^{2+}-dependent, sequence-specific ribonuclease activity against the 932 Pentaprobe, shown by degradation products on the gel (Figure 3.11).

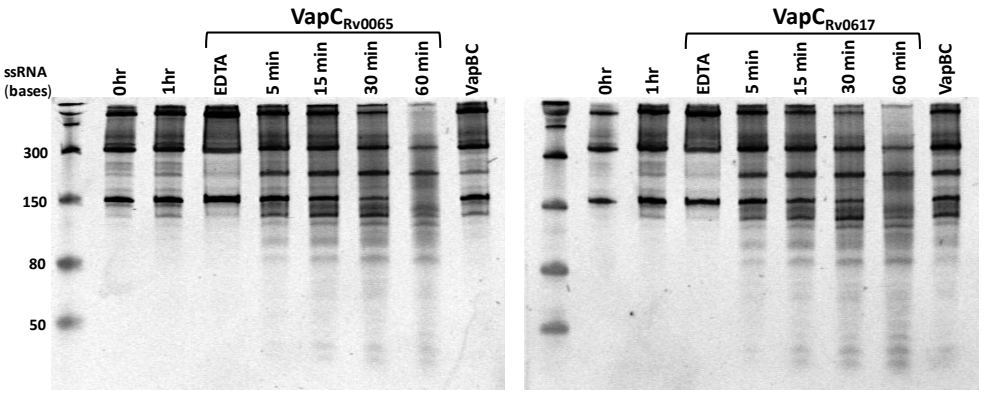


Figure 3.11 $VapC_{Rv0065}$ and $VapC_{Rv0617}$ ribonuclease assays against 932 Pentaprobe RNA. VapC displays sequence-specific RNase activity against the 932 Pentaprobe. Labels refer to RNA only negative controls (0 hr, 1 hr), EDTA negative control (EDTA), assay time points (5, 15, 30, 60 min) and activity of the $VapBC_{0065a/c}/VapBC_{0617a/c}$ complex (VapBC). A low range ssRNA ladder is shown down the side of the 10% urea denaturing gel.

Although urea denaturing polyacrylamide gels were used in these experiments, the large amount of secondary structure in the 932 Pentaprobe was observed as multiple high molecular-weight bands at the top of the gel. RNA only 0 hr and 1 hr samples showed no RNA degradation, proving that there were no contaminating RNases present in the assay buffer. No degradation in the EDTA control samples indicated an absence of metal-independent RNases, such as Ribonuclease A, in the protein preparation. If metal-independent RNases were present here, degradation of RNA would still be seen upon inhibition of VapC with EDTA. Comparing RNA degradation between the VapC samples to those with either $VapBC_{Rv0065a/c}$ or $VapBC_{Rv0617a/c}$ revealed that VapB inhibited ribonuclease activity of its cognate VapC. A shared specificity between these VapC ribonucleases was indicated by the near identical fragmentation patterns that the two enzymes produced. Under these assay conditions, $VapC_{Rv0065}$ and $VapC_{Rv0617}$ had comparable activity. In both assays, a signification proportion of the 932 Pentaprobe was degraded by the 5 minute incubation time point. With increasing incubation, the number of smaller sized degradation products increased. This suggests that an optimal recognition site was being targeted quickly, after which other sub-optimal sites were also recognised for cleavage over a longer time-scale.

3.3.1.2 932 Pentaprobe RNA Fragments VapC Ribonuclease Activity Assays

Due to the activity exhibited against this substrate, the 932 Pentaprobe was chosen as the scaffold from which to design shorter RNA fragments. To determine the role of secondary structure in the specificity of these VapC proteins, the overlapping RNA fragments were designed to have differing secondary structures, while all encompassing a 15 bp GC-rich region of the 932 Pentaprobe (Section 2.4.1). The structures of 932 fragments were predicted by RNAfold (Vienna RNA webservers, TBI) (Section 2.4.2) and these are included in Appendix C4. A goal of this work was to create a short substrate which would include the minimal, optimal secondary structure required for cleavage. A schematic of the positioning of the designed fragments in relation to the longer 932 Pentaprobe is provided in Figure 3.12.

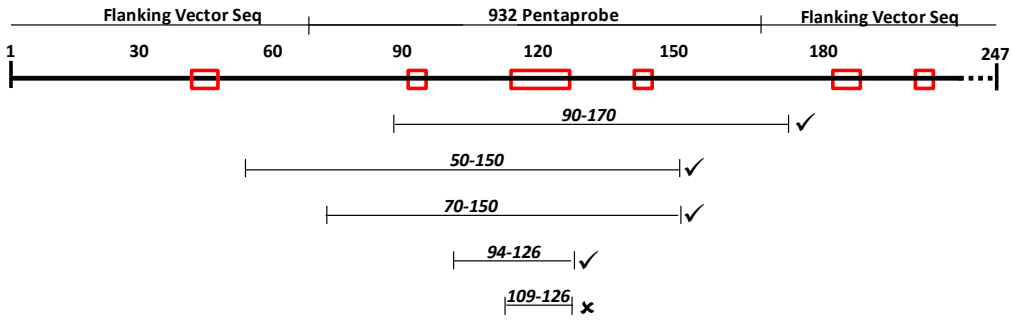


Figure 3.12 The 932 Pentaprobe and subset fragments. The 932 Pentaprobe RNA (1-247) consists of the Pentaprobe sequence flanked by regions of vector. Positioning of Pentaprobe fragments 90-170, 50-150, 70-150, 94-126 and 109-126 in relation to the full 932 Pentaprobe is shown. Numbers indicate nucleotide position along the 932 Pentaprobe RNA. Red boxes represent GC-rich regions of the RNA. Ticks/crosses indicate that both/neither VapC protein displayed activity against the fragment. The 15 bp GC-rich centre is located at position 112-126 on the 932 Pentaprobe RNA.

Each 932 RNA fragment was incubated *in vitro* with purified $VapC_{Rv0065}$ or $VapC_{Rv0617}$ at 37 °C in assay buffer (Section 2.4.3.2). Negative controls included were as previously described. RNase activity against the fragments was analysed by running assays on 20 % urea denaturing gels (Section 2.4.4). Results for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ assays are shown in Figure 3.13 and Figure 3.14 respectively.

$VapC_{Rv0065}$ and $VapC_{Rv0617}$ degraded four of the five 932 Pentaprobe fragments; 90-170, 50-150, 70-150 and 94-126. In all assays conducted, RNA only 0 hr and 1 hr samples exhibited no RNA degradation indicating that there were no contaminating RNases present in the assay buffer. It had been previously established that $VapC_{Rv0065}$ and $VapC_{Rv0617}$ efficiently cleave the 932 94-126 fragment (denoted 'Oligo 3' in previous experiments (Section 1.4.2.2)), and these assays confirmed this. While the gels suggested that there was a degree degradation in the initial 932 94-126 RNA preparation (seen as smearing in all lanes), there was definite VapC-mediated RNase activity occurring against this substrate as indicated by the presence of degradation bands in the VapC 60 minute sample that were absent in the controls.

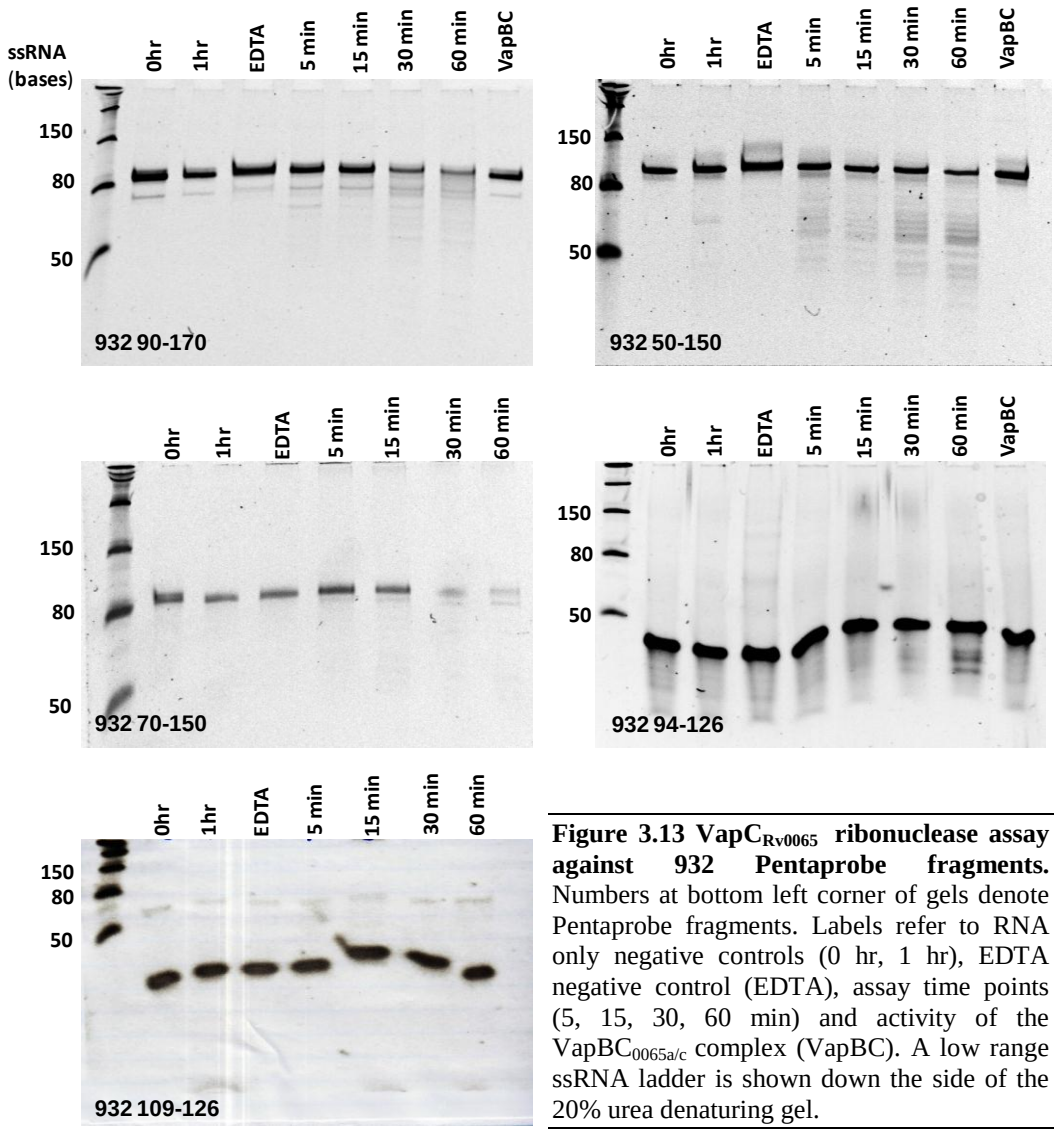


Figure 3.13 $VapC_{Rv0065}$ ribonuclease assay against 932 Pentaprobe fragments. Numbers at bottom left corner of gels denote Pentaprobe fragments. Labels refer to RNA only negative controls (0 hr, 1 hr), EDTA negative control (EDTA), assay time points (5, 15, 30, 60 min) and activity of the $VapBC_{0065a/c}$ complex (VapBC). A low range ssRNA ladder is shown down the side of the 20% urea denaturing gel.

$VapC_{Rv0065}$ and $VapC_{Rv0617}$ did not appear to bind the Pentaprobe fragment RNA molecules that they cleaved, as no smearing was seen that was specific to lanes in which VapC protein had been added. This is in contrast to observations made for $VapC_{MS1284}$ from *M. smegmatis* (Section 1.4.1). Substrate fragmentation patterns were virtually identical when comparing $VapC_{Rv0065}$ and $VapC_{Rv0617}$ mediated cleavage. By analysing the sizes of major cleavage products and correlating these with the positioning of bases along the substrate strands, suggestions can be made as to what cut sites mediate production of the degradation products. The cleavage product bands of sizes ~6 bases smaller than the 932 70-150 and 932 90-170 substrates are in agreement with cleavage at the GC-rich regions present at the termini of these substrates (refer back to Figure 3.12). Degradation bands running at ~60 and ~40 bases in the 932 50-150 assays may correlate to a single cleavage

event at this substrate's GC-rich centre. Interestingly, there are only very weak degradation bands corresponding to a single cleavage event at the GC-rich centre for the 932 90-170 and 932 70-150 fragments. In these two fragments, the GC-rich region is predicted to reside in a stem region of the RNA. In contrast, this region forms a hairpin loop in the 932 50-150 fragment. This motif has previously been associated with VapC activity; it was suggested that a hairpin loop immediately downstream of the cut site is preferable for cleavage by $VapC_{MS1284}$ (Section 1.3.2.3). No activity was observed for either VapC against the small 932 109-126 Pentaprobe RNA fragment.

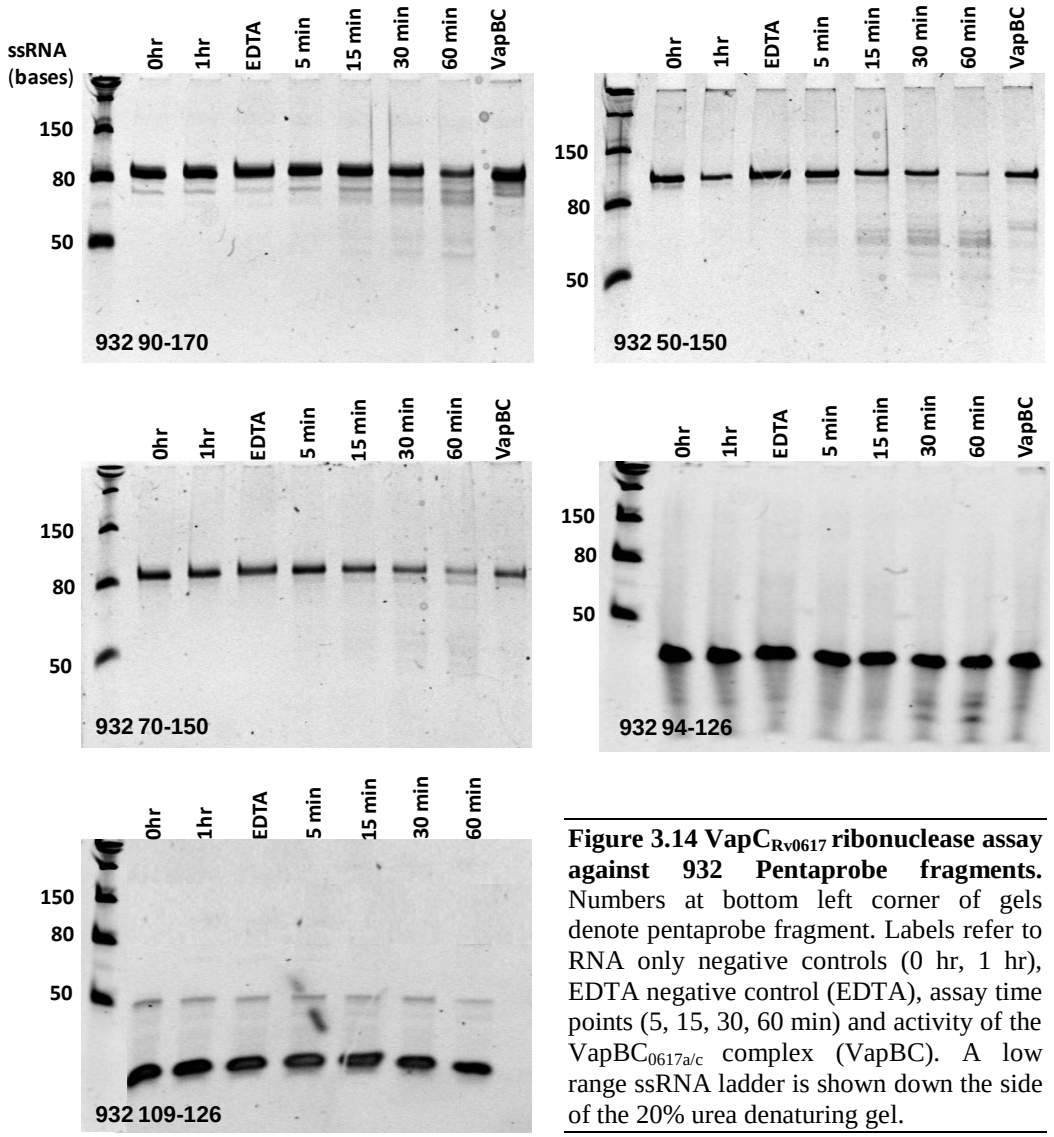


Figure 3.14 $VapC_{Rv0617}$ ribonuclease assay against 932 Pentaprobe fragments. Numbers at bottom left corner of gels denote pentaprobe fragment. Labels refer to RNA only negative controls (0 hr, 1 hr), EDTA negative control (EDTA), assay time points (5, 15, 30, 60 min) and activity of the $VapBC_{0617a/c}$ complex (VapBC). A low range ssRNA ladder is shown down the side of the 20% urea denaturing gel.

Collectively, results of the assays in Figure 3.13 and Figure 3.14 corroborate the previously drawn conclusion that $VapC_{Rv0065}$ and $VapC_{Rv0617}$ target the same recognition sequence of GC-rich 4-mer regions in ssRNA.

In order to strengthen evidence for the negative result seen against the 932 109-126 fragment for either VapC protein, assays against this substrate were repeated and analysed by MALDI TOF MS (Figure 3.15).

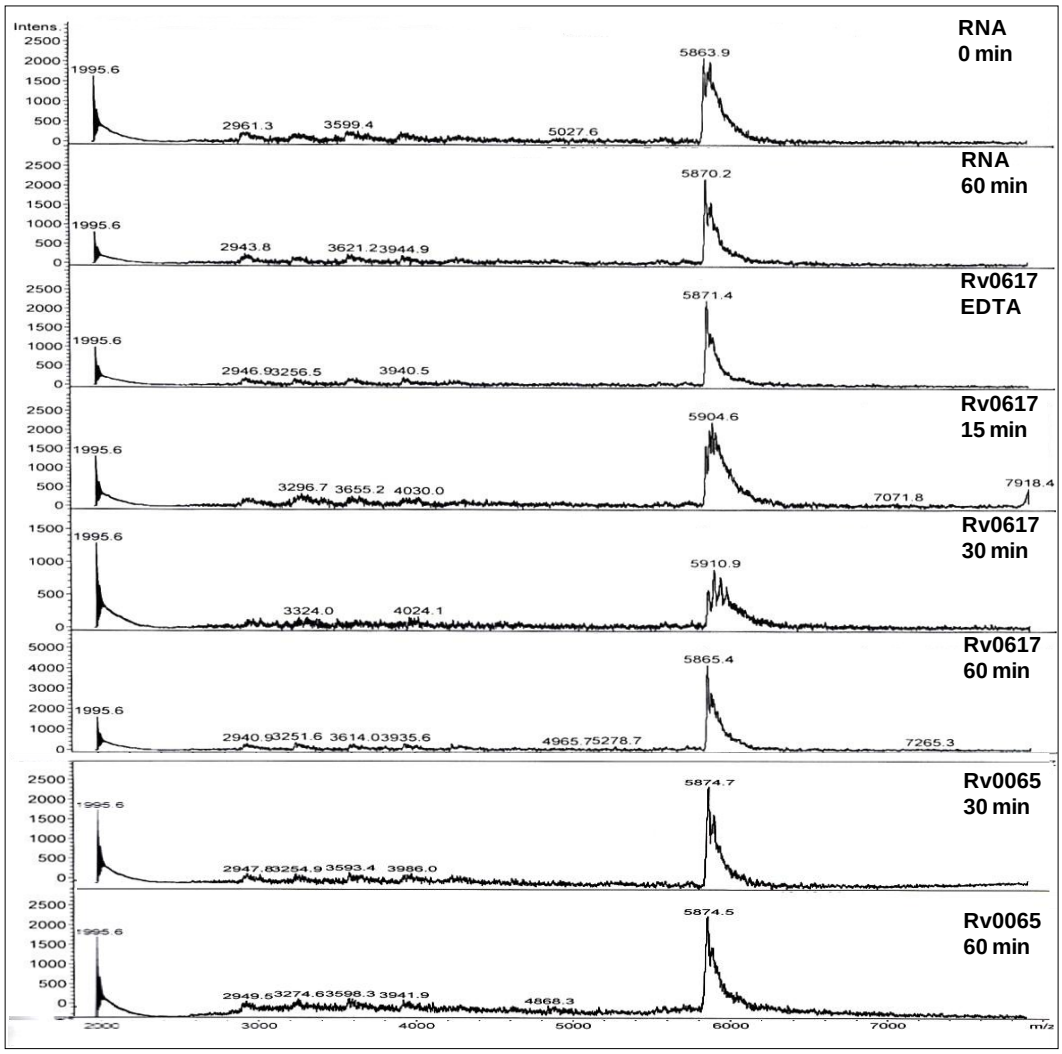


Figure 3.15 932 109-126 RNA, VapC$_{Rv0617}$ and VapC$_{Rv0065}$ MALDI TOF MS results. Labels refer to RNA only negative controls (RNA 0, 60 min), EDTA negative control (Rv0617 EDTA), VapC$_{Rv-617}$ assay time points (Rv0617 15, 30, 60 min) and VapC$_{Rv0065}$ assay time points (Rv0065 30, 60 min). Data analysed by Bruker Data Analysis and baseline subtracted.

The mass spectra showed no degradation of the 932 109-126 fragment in negative controls nor in reactions including VapC$_{Rv0617}$ or VapC$_{Rv0065}$. Signal intensity of the molecular ion peak was slightly reduced in the Rv0617 30 min assay time point sample; however is restored in the 60 min time point sample. This reduced intensity and absence of degradation product peaks suggests that initially there was less RNA added to this reaction tube. Multiple peaks in the Rv0617 30 min molecular ion peak may be due to the formation of RNA adducts (i.e. $M^{+}Na$). The MFE predicted secondary structure of the 932 109-126 fragment was compared

with that of the 94-126 fragment using RNAfold prediction software (Figure 3.16) in an attempt to rationalize why one was cleaved preferentially over the other.

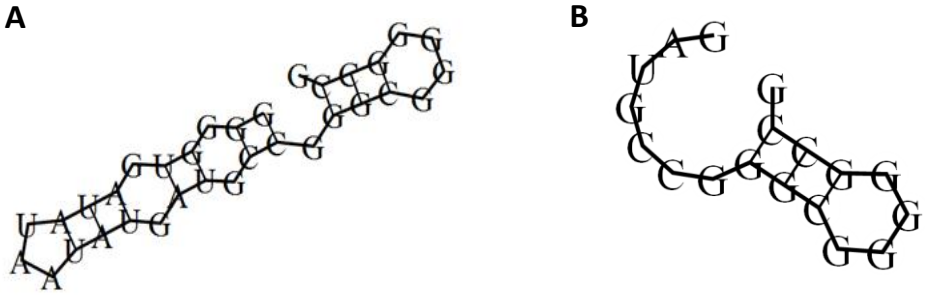


Figure 3.16 MFE secondary structure predictions for the 932 94-126 and 109-126 RNA. A, Predicted secondary structure of the 932 94-126 fragment. **B**, Predicted secondary structure of the 932 109-126 fragment. Predictions were made using RNAfold (Vienna RNA webservers, TBI).

The 932 109-126 RNA is predicted to have a large terminal loop region that is absent in the 932 94-126 structure. Both structures retain the 15 bp GC-rich region in a hairpin loop portion; however the 932 109-126 fragment has lost the longer stem region extending down from this. This comparison, along with analysis of the structures and assay results of the longer 932 fragments, could suggest that the presence of both a GC-rich cut site in a hairpin loop and a long stem region proximal to this are required for efficient VapC cleavage.

The free energy of the predicted secondary structure for each of the RNA substrates tested is given in Table 3.1. Out of the four longest substrates (932 1-247, 932 90-179, 932 50-150 and 932 70-150), both VapC proteins displayed the least activity against 932 70-150. This RNA substrate has the most positive free energy value, and therefore the least predicted secondary structure of the four.

Table 3.1 Minimum free energy values at 37 °C for each RNA substrate as calculated by RNAfold WebServer

RNA Substrate	Free Energy kcal/mol (at 37 °C)
932 1-247	-88.8
932 90-170	-17.7
932 50-150	-22.4
932 70-150	-14.9
932 94-126	-9.5
932 109-126	-4.9

The more negative the free energy value, the larger the amount of secondary structure in the RNA molecule and the higher the likelihood that this structure will form.

While a definitive secondary structural motif required for VapC cleavage has not been reported, these results support the growing evidence of such a phenomenon.

3.3.2 VapC Ribonuclease Activity Against *In Vivo* Substrates

In *M. smegmatis*, research conducted by Robson (2010) resulted in a number of postulated *in vivo* targets of the singular VapC protein, including transcripts involved in sugar metabolism and transport (Section 1.4.1). To date, no target transcripts have been identified *in vivo* for *M. tb* proteins $VapC_{Rv0065}$ and $VapC_{Rv0617.}$ Elucidation of *in vivo* targets will be instrumental in shedding further light on the biochemical role that the *M. tb* encoded VapC proteins play in the bacterium's metabolism.

The 16S and 23S rRNA constitute indispensible components of the prokaryotic ribosome and are essential for cell viability. The 16S and 23S rRNA bands can be easily identified when total RNA is separated on a 1% agarose gel, running at sizes of approximately 1.5 kb and 2.9 kb respectively. Due to extensive research into the ribosome and its many components, the secondary structure of both 16S and 23S rRNA is well documented. As such, identification of cut sites along these rRNA molecules may elucidate secondary structure pre-requisites for VapC cleavage. These factors led to the investigation of whether the 16S and 23S rRNA were targets for $VapC_{Rv0065}$ and $VapC_{Rv0617}$ –mediated ribonuclease activity.

A number of approaches were considered for isolation of rRNA. These included i) isolation of ribosomes from *M. smegmatis* culture using strong anion-exchange chromatography (Trauner et al. 2011) followed by total RNA extraction of the ribosomes; ii) an *in vitro* approach via PCR amplification of rRNA genes directly from *M. tb* H37Ra genomic DNA followed by *in vitro* transcription and purification (this approach would have the disadvantage in that the resulting synthesised transcripts would be lacking in post-transcriptional modifications that potentially take place *in vivo* to produce the mature rRNA molecules) and iii) isolation of total RNA followed by specific separation of rRNA from the total RNA pool. This last approach seemed the most feasible. A suggested method for determining the location of cut sites along the rRNA involved use of post-assay reverse transcription followed by real-time PCR analysis with primers situated at intervals along the 16S rRNA transcript. This was deemed undesirable as primers

used in the reverse transcription step would be subject to bias, which could skew results. Instead, assays against the rRNA were run on low percentage urea denaturing gels in an attempt to gain high resolution between the resulting fragmentation pattern bands.

3.3.2.1 Mycobacterial Total RNA Ribonuclease Activity Assays

Initial experiments aimed to roughly gauge the activity of $VapC_{Rv0065}$ and $VapC_{Rv0617}$ against mycobacterial rRNA. Due to the unavailability of *M. tb* total RNA, the activity of VapC against total RNA from both *M. smegmatis* mc^2155 and *Mycobacterium bovis* was tested.

Preliminary assays investigating the ribonuclease potential of the VapC proteins indicated that $VapC_{Rv0065}$ and $VapC_{Rv0617}$ cleave *M. smegmatis* total RNA (McKenzie, unpublished results). These assays were repeated, specifically focusing on activity against the 16S and 23S rRNA bands. *M. smegmatis* total RNA was isolated from cultures using the GITC/phenol-chloroform extraction protocol as outlined in Section 2.1.2.1. RNase assays against each VapC protein were set up as described in Section 2.4.3.2 which included 15, 30 and 60 min time points. Controls included 0 hr and 1 hr RNA only samples, an EDTA control and a VapBC control ($VapBC_{Rv0065a/c}$ or $VapBC_{Rv0617a/c}$). Assays were run on 1% TAE agarose gels for analysis (Section 2.1.1.3) (Figure 3.17).

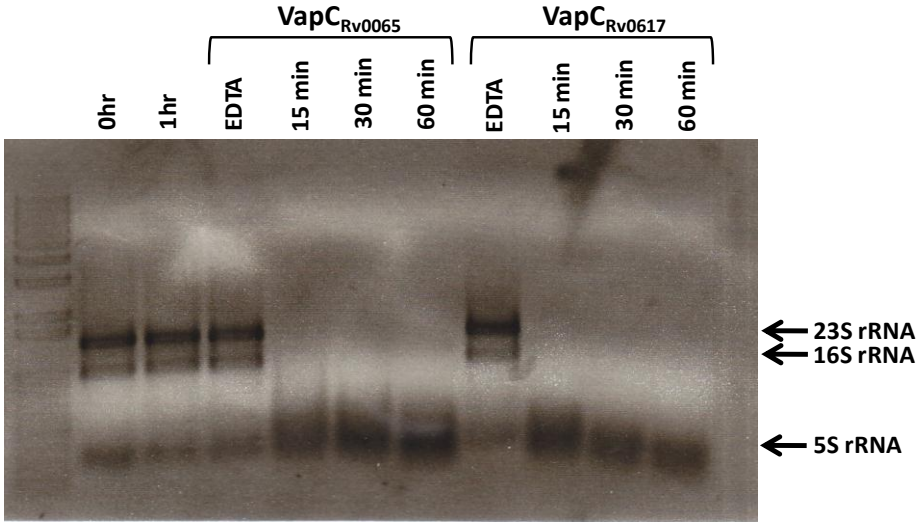


Figure 3.17 Ribonuclease activity of $VapC_{Rv0065}$ and $VapC_{Rv0617}$ from *M. tb* using *M. smegmatis* total RNA as a substrate. VapC proteins display Mg^{2+}-dependent activity as addition of EDTA abolishes activity. RNA degradation in the presence of both VapC is seen over a period of 15 – 60 minutes. Negative controls are shown by 0 hr and 1 hr RNA only samples along with EDTA controls. A $1Kb^+$ molecular weight marker shows molecular masses of dsDNA.

$VapC_{Rv0065}$ and $VapC_{Rv0617}$ displayed Mg^{+}-dependent ribonuclease activity against the *M. smegmatis* total RNA, as inclusion of EDTA in the assay buffer abolished ribonuclease activity. The disappearance of 23S and 16S rRNA bands and accumulation of smaller RNA fragments with increasing VapC incubation suggests both VapC proteins were active against the rRNA. Comparing this assay to that conducted with the 932 Pentaprobe RNA, it can be seen that $VapC_{Rv0065}$ and $VapC_{Rv0617}$ cleaved the rRNA in this total RNA sample at a much faster rate.

M. bovis total RNA was acquired from the Proteins & Microbes lab, University of Waikato. Due to the limited amount, only 15 and 30 minute time points were selected for assays. These assays revealed activity for both $VapC_{Rv0065}$ and $VapC_{Rv0617}$ against the *M. bovis* 16S and 23S rRNA (Figure 3.18) (assay with $VapC_{Rv0065}$ not shown).

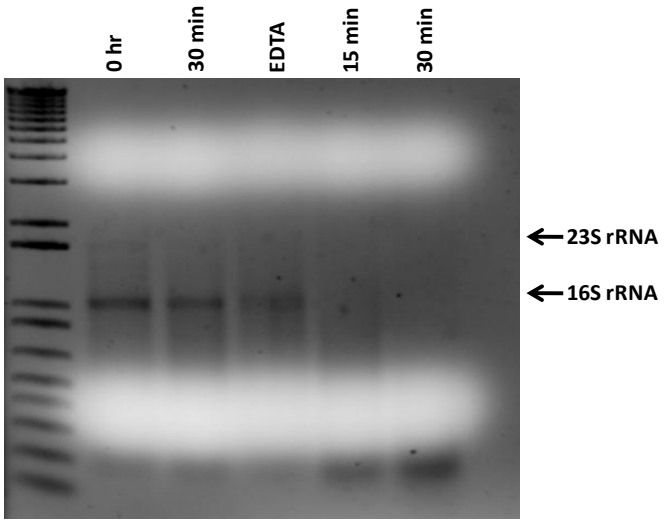


Figure 3.18 Ribonuclease activity of $VapC_{Rv0617}$ from *M. tb* using *M. bovis* total RNA as a substrate. $VapC_{Rv0617}$ displays Mg^{2+}-dependent activity as addition of EDTA abolishes activity. RNA degradation in the presence of $VapC_{Rv0617}$ is seen over a period of 15 – 30 minutes. Negative controls are shown by 0 hr and 1 hr RNA only samples along with EDTA controls. A $1Kb^{+}$ molecular weight marker shows molecular masses of dsDNA.

3.3.2.2 Mycobacterium smegmatis *rRNA Ribonuclease Activity Assays*

In order to further investigate VapC–mediated cleavage of the rRNA, 16S and 23S rRNA was purified from *M. smegmatis* total RNA as outlined in Section 2.1.2.3. Ideally the native *M. tb* rRNA would be purified; however due to unavailability and technical restrictions *M. smegmatis* was chosen as a model. Genomic sequences for *M. smegmatis* $mc^{2}155$ and *M. tb* H37Rv 16S rRNA genes were identified as having a nucleotide conservation score as 95% on the National

Centre for Biotechnology Information (NCBI) website. Figure 3.19 depicts a total RNA preparation run on a 1% TAE agarose gel (Section 2.1.1.3).

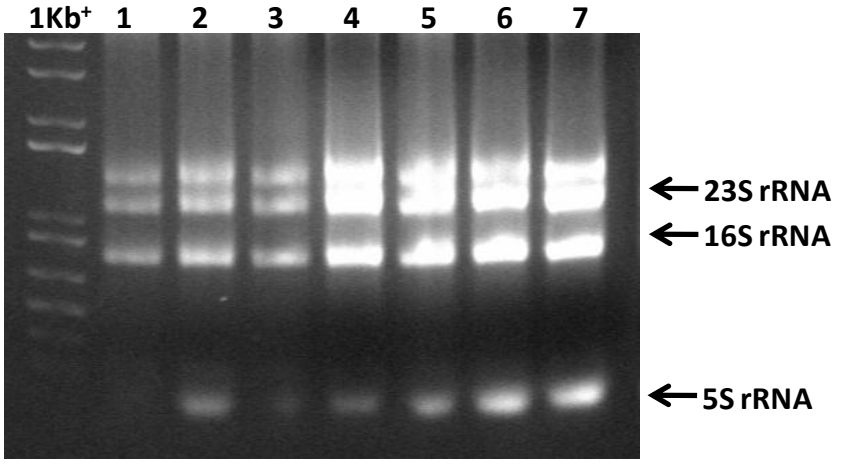


Figure 3.19 Isolation of total RNA from *M. smegmatis.* Lanes 1-7 represent RNA samples that were isolated using a GITC/phenol-chloroform protocol for RNA extraction. Bands for 23S, 16S and 5S rRNA are indicated. Samples are run alongside a $1Kb^{+}$ ladder (Invitrogen) for size estimation.

It was consistently noted that two bands were present at the 23S rRNA region. These two bands may reflect differing secondary structure in the 23S rRNA molecule. Purified 16S and 23S rRNA was used in RNase assays with $VapC_{Rv0065}$ and $VapC_{Rv0617}$ and the resulting fragmentation patterns were observed in an attempt to gauge their specificities against these substrates.

$VapC_{Rv0065}$ and $VapC_{Rv0617}$ ribonuclease assays against the purified 16S rRNA were conducted as outlined in Section 2.4.3.2. Controls included 0 hr and 1hr RNA only samples, an EDTA control and a cognate VapBC control. Figure 3.20 depicts these assay samples run on 5 % urea denaturing gels prepared as outlined in Section 2.4.4.

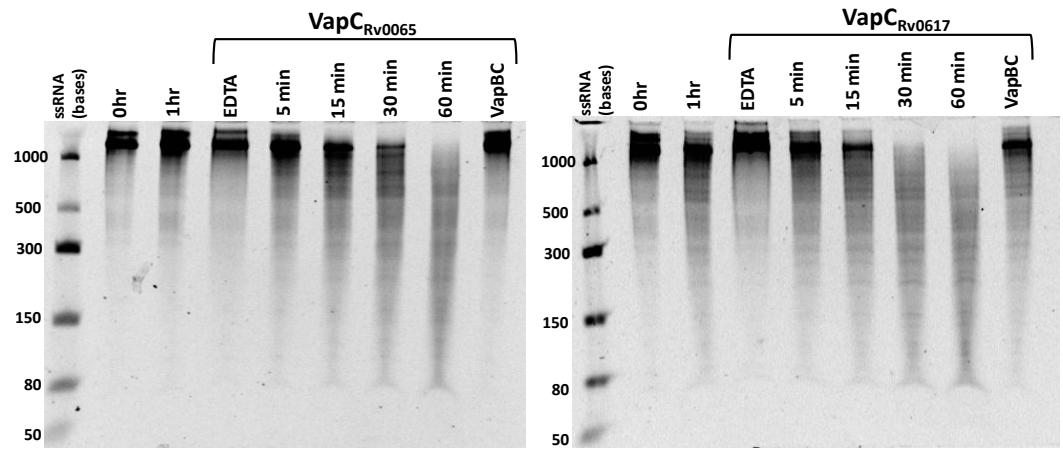


Figure 3.20 Ribonuclease activity of VapC$_{Rv0065}$ and VapC$_{Rv0617}$ from *M. tb* against isolated *M. smegmatis* 16S rRNA. Labels refer to RNA only negative controls (0 hr, 1 hr), EDTA negative control (EDTA), assay time points (5, 15, 30, 60 min) and activity of the VapBC$_{Rv0065a/c}$ or VapBC$_{Rv0617a/c}$ complex (VapBC). A low range ssRNA ladder is shown down the side of the 5% urea denaturing gel.

As indicated by the presence of degradation products in lanes including either VapC protein, VapC$_{Rv0065}$ and VapC$_{Rv0617}$ both displayed Mg^{2+}-dependent ribonuclease activity against isolated and purified 16S rRNA from *M. smegmatis*. VapBC TA complexes displayed no ribonuclease activity (Figure 3.20).

Ribonuclease activity of VapC$_{Rv0065}$ and VapC$_{Rv0617}$ against isolated and purified 23S rRNA was tested following the same protocol as for the 16S rRNA assays. Controls in each assay included 0 hr and 1 hr RNA only samples, an EDTA control and a cognate VapBC control. Figure 3.21 depicts assays against both VapC proteins run on 5% urea denaturing gels prepared as outlined in Section 2.4.4.

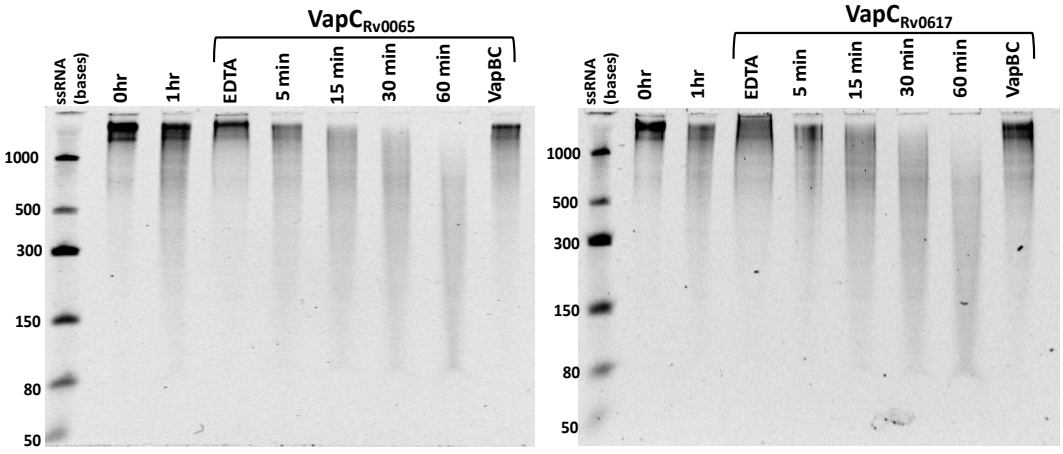


Figure 3.21 Ribonuclease Activity of VapC$_{Rv0065}$ and VapC$_{Rv0617}$ from *M. tb* against isolated *M. smegmatis* 23S rRNA. Labels refer to RNA only negative controls (0 hr, 1 hr), EDTA negative control (EDTA), assay time points (5, 15, 30, 60 min) and activity of the VapBC$_{Rv0065a/c}$ or VapBC$_{Rv0617a/c}$ complex (VapBC). A low range ssRNA ladder is shown down the side of the 5% urea denaturing gel.

As indicated by the presence of degradation products in lanes including either VapC protein, $VapC_{Rv0065}$ and $VapC_{Rv0617}$ both displayed Mg^{2+}-dependent ribonuclease activity against isolated and purified 23S rRNA from *M. smegmatis*. VapBC TA complexes displayed no ribonuclease activity.

Both $VapC_{Rv0065}$ and $VapC_{Rv0617}$ appear to have greater activity against the 23S rRNA, since the majority of the full-length substrate was degraded after 15 minutes incubation, whereas against the 16S rRNA an equivalent degree of degradation required 60 minutes. Unfortunately in both the 16S and 23S rRNA assays, no clear fragmentation pattern could be discerned. This hindered the comparison of RNase specificity of either VapC for these substrates, and prevented approximation of the location of cut sites along the rRNA based on the sizes of fragments produced. The ratio of protein to RNA in subsequent RNase assays was decreased in an attempt to see VapC acting only upon primary cut sites, thus simplifying the degradation pattern. These assays still displayed a multitude of differently sized fragments appearing as a smear on the gel (data not shown).

3.3.3 Proposed Protocol for the Identification of *In Vivo* VapC Targets

Identification of the *in vivo* targets of the *M. tb* proteins $VapC_{Rv0065}$ and $VapC_{Rv0617}$ will shed light on their physiological roles and potential importance in *M. tb* virulence. A protocol was developed to specifically identify mRNA targets of VapC through selective RNA tagging, reverse transcription and subsequent PCR amplification. Bacterial mRNA within the cell is believed to natively occur with a 5'-triphosphate group on the first transcribed nucleotide (Schoenberg 2007). The developed protocol takes advantage of the singular 5'-monophosphate left on the 3' cleavage product by VapC (McKenzie 2011; Duyvestyn 2012) and uses this modification to selectively tag cleaved mRNAs with an RNA oligo of known sequence. Non-specific 3' polyA-tailing allows the tagged mRNA molecules to be reverse transcribed into cDNA using primers complementary to the RNA oligo tag and the polyA-tail. A round of PCR on the cDNA strand amplifies these tagged molecules to a quantity sufficient for cloning and subsequent sequencing. A flowchart of this process is detailed in Figure 3.22.

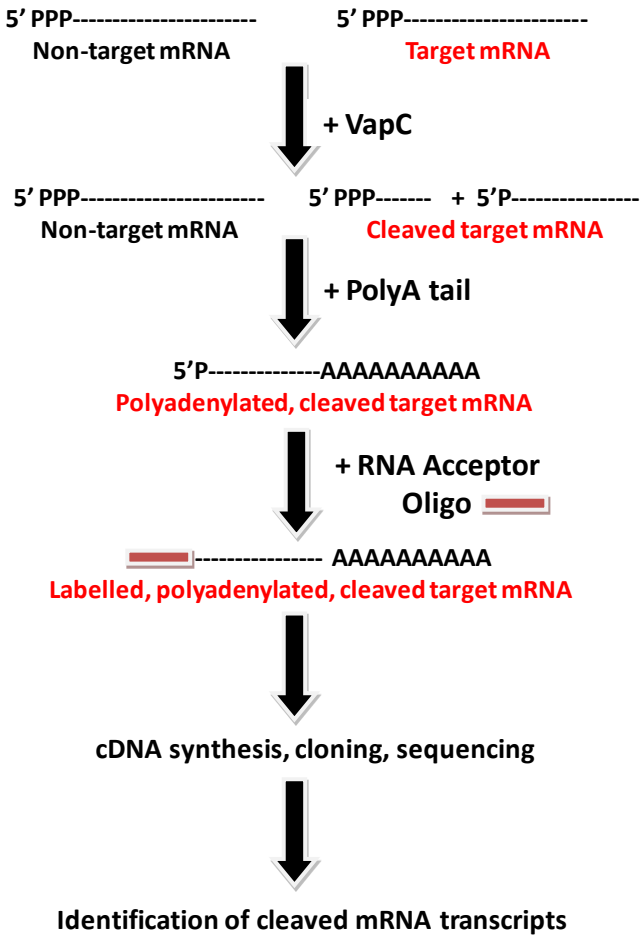


Figure 3.22 Proposed protocol for identification of *in vivo M. tb* VapC mRNA targets.

3.3.3.1 Development of Protocol with VapC$_{PAE0151}$ and the 924 Pentaprobe

The *E.coli* VapC$_{PAE0151}$ protein was used in the development of this protocol. This protein was chosen over mycobacterial VapC species as it was readily available in a purified form, and preliminary VapC$_{PAE0151}$ RNase assays assessed on urea denaturing gels revealed distinct fragmentation patterns against a number of the Pentaprobe RNA substrates (McKenzie et al. 2012b). An assay that produced a small number of cleavage products was desired as this would simplify their subsequent labelling and identification. IMAC nickel-purified VapC$_{PAE0151}$ was acquired from the Proteins & Microbes lab, University of Waikato. In order to test the protocol, the 924 Pentaprobe was chosen as a substrate. When RNase assays against this substrate were run on a urea denaturing gel (Sections 2.4.3.1 and 2.4.4), VapC$_{PAE0151}$ cleaved the Pentaprobe into three major degradation products of approximately 120 bases, 80 bases and 65 bases (Figure 3.23). The principles

of this protocol dictate that these three products would specifically be labelled and amplified.

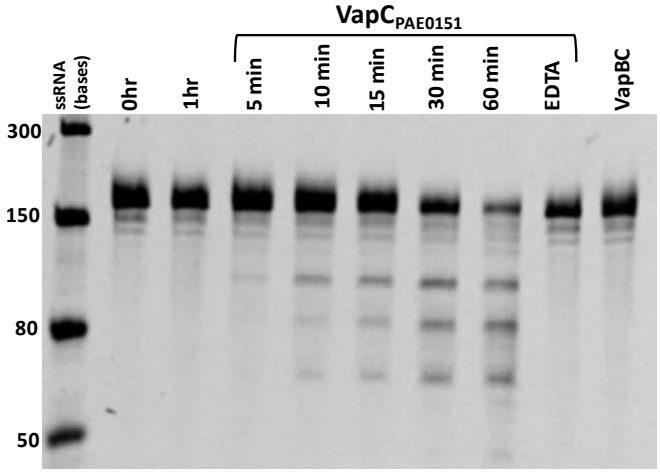


Figure 3.23 Ribonuclease activity of VapC$_{PAE0151}$ from *Pyrobaculum aerophilum* against the 924 Pentaprobe. Ladder = low range ssRNA marker. Lanes from left to right represent two RNA negative controls, assay time points at which samples were taken, an EDTA control and a VapBC control. Figure adapted from (McKenzie et al. 2012b).

The protocol was followed through and stages were completed as outlined in Section 2.4.7. After cDNA synthesis, rather than following on to cloning and sequencing, the cDNA was instead run on a 2% TAE agarose gel in order to quickly and roughly determine whether bands corresponding to the three primary 924 degradation products had been selectively tagged and amplified. The protocol was independently followed through three times. Figure 3.24 depicts a gel representative of all results observed.

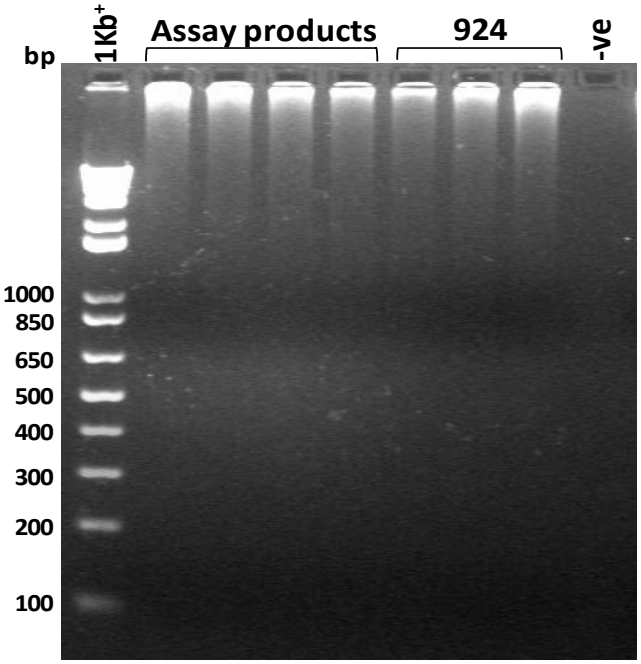


Figure 3.24 2% TAE agarose gel of PCR amplified and tagged RNase assay products/Pentaprobe 924 RNA. The resulting 2% gel after following the proposed protocol for identification of VapC target transcripts. 'Assay product' lanes indicate samples where RNA originated from VapC$_{PAE0151}$-924 assay products. '924' lanes indicate samples where RNA originated from purified 924 Pentaprobe. A negative control where template was excluded from the PCR mixture is run in right-most lane. A 1Kb$^+$ molecular weight marker shows molecular masses of dsDNA.

No bands corresponding to either the full length 924 Pentaprobe, or VapC-produced degradation fragments of the 924 Pentaprobe, were present on gels. The gels displayed a significant amount of unknown high molecular mass nucleic acid material which remained near the wells. Due to difficulties encountered in pinpointing problematic step(s) in the protocol, it was decided at this point that no further work was to be done to remedy the problems encountered. The protocol shows promise for the development of a tool which could selectively identify VapC targets from a sample of mixed mRNA, and further optimisation will be required to achieve this.

3.4 VapBC Transcriptional Autoregulation through DNA Binding

TA systems, including those belonging to the VapBC family, are well documented to have an autoregulatory function whereby the TA complex binds back to the promoter DNA to sterically repress transcription (Section 1.2.4). The following experiments examined the binding abilities of $VapBC_{Rv0065a/c}$ and $VapBC_{Rv0617a/c}$ against promoter DNA of their corresponding operons, which would suggest autoregulatory roles for these complexes in their TA system functioning. If promoter-VapBC binding was observed, it was then determined what sequence motif in the promoter region was required for binding. As previously discussed, studies characterising a number *vapBC* systems have reported IR sequences in the promoter whose presence is essential for maintaining the VapBC-DNA interaction.

3.4.1 Identification of Promoter Regions

The *Rv0065* and *Rv0617* genes, encoding $VapC_{Rv0065}$ and $VapC_{Rv0617}$ respectively, have previously been found to be leaderless (McKenzie, unpublished work) indicating that their transcriptional start points (TSPs) coincide with the start point of their translation. The genomic coordinates and nucleotide sequence of both *Rv0065a/c* and *Rv0617a/c* operons were identified using the website http://cmr.jcvi.org and searching within the *Mycobacterium tuberculosis* H37Ra genome. The *Rv0617a/c* operon was identified as starting at position 710780 on the chromosome and therefore the region 710650-710780 (130 bp) was taken as the putative promoter. The *Rv0065a/c* operon started at position 71587, and the region 71486-71587 (101 bp) was taken as the putative promoter. Both putative promoter regions contained a 6 bp perfect IR. The regions were analysed for distinguishing promoter features.

A characteristic element of prokaryotic promoters is the hexameric sequence located around the region -10 nucleotides upstream of the TSP, referred to as the Pribnow box. Mycobacterial Pribnow boxes are known to be disparate to those characteristic of *E. coli* (Bashyam et al. 1996). A study by Bashyam et al (1996) surveyed the characteristics of Pribnow box-like sequences in mycobacterial promoters and identified conserved nucleotides at positions in the -10 hexameric sequences of *M. tb* samples; 80% T, 90% A, 60% Y, 40% g, 60% A and 100% T,

where the letter Y represents a pyrimadine base and capital letters indicate bases exhibiting conservation of 50% or more (Bashyam et al. 1996). The most important bases in the hexameric sequence were denoted as being at the first, second and sixth positions. These positions are functionally more important than in *E.coli* and are highly conserved among mycobacterial promoter sequences (Bashyam et al. 1996). The remaining positions in mycobacteria are thought to have high vulnerability to GC intrusions due to the higher GC content of mycobacterial DNA (Bashyam et al. 1996).

Within the 130 bp putative *Rv0617a/c* promoter region, nucleotides 5'-GATGTT-3' occurring at positions -16 to -11 relative to the TSP may be a -10 recognition element. This sequence was annotated as a -10 element in the *M. tb* promoter T26 (Bashyam et al. 1996). Nucleotides TAGGAT at positions -12 to -7 relative to the TSP may also be a putative -10 element as this sequence fulfils the requirements of a sigma factor A recognition site (the sigA consensus sequence is identified as 5'-TA(G/T)(A/G)AT-3' (Gomez & Smith 2000; Zaunbrecher et al. 2009)). The SigA protein has been identified as an expression modulator of genes contributing to virulence (Wu et al. 2004), a category that the *vapBC* genes fall into. No previously annotated -10 recognition elements were found in the 101 bp *Rv0065a/c* promoter region. The -35 regions in the *Rv0617a/c* and *Rv0065a/c* promoters bear no resemblance to the 5'-TTGACA-3' motif seen in *E. coli* and several other bacteria (Bashyam et al. 1996). This is not surprising as a number of studies have reported that no single -35 sequence is conserved among mycobacterial promoters (Bashyam et al. 1996). Each promoter region was found to contain a 6 bp perfect IR. Figure 3.25 outlines hypothetical -10 and -35 elements along with perfect IRs in the putative *Rv0065a/c* and *Rv0617a/c* promoter regions. Details regarding the preparation of putative promoter DNA used in EMSAs can be found in Section 2.5.1.

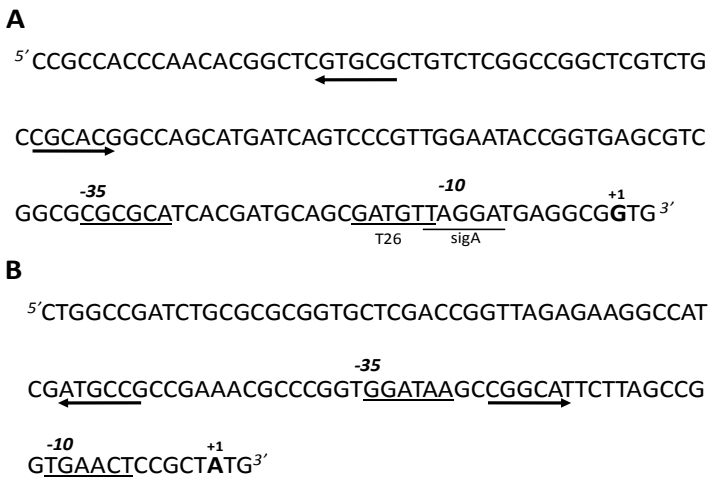


Figure 3.25 Putative *Rv0065a/c* and *Rv0617a/c* promoter elements. Underlined nucleotides represent putative promoter elements. TSP is indicated (+1, bold). Perfect IR are indicated by divergent arrows. **A**, 130 bp *Rv0617a/c* promoter region. Reported *M. tb* T26 promoter and sigA binding site are labelled. **B**, 104 bp *Rv0065a/c* promoter region.

3.4.2 VapBC$_{Rv0065a/c}$ EMSAs with Putative *Rv0065a/c* Promoter DNA

The 101 bp sequence of putative *Rv0065a/c* promoter DNA was DIG-labelled and used in an EMSA with purified VapBC$_{Rv0065a/c}$ protein as described in Section 2.5.5. VapBC was used in EMSA assays in the range 0 nM - 470 nM. Labelled DNA (4 ng) was used in each reaction. No shift was observed at any VapBC concentration (Figure 3.26).

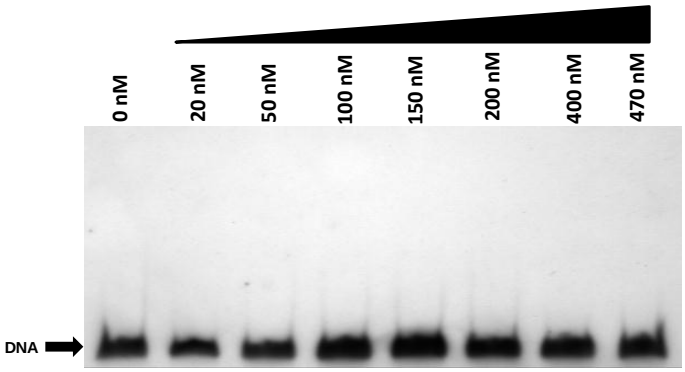


Figure 3.26 VapBC$_{Rv0065a/c}$ does not bind to a 101 bp region of the *Rv0065a/c* promoter. EMSA experiments were performed with a 101 bp DIG-labelled region of promoter DNA (0.4 ng) encompassing a perfect 6 bp IR. Concentrations of VapBC are shown above each lane

It was theorised that the recognition site for VapBC$_{Rv0065a/c}$ may be downstream of the TSP. The region of DNA being tested was altered to span 38 bp upstream and 112 bp downstream of the TSP. The 112 bp region downstream of the TSP contained three additional perfect IR. The 150 bp region was amplified from *M. tb* H37Ra genomic DNA (Proteins & Microbes lab, University of Waikato), DIG-labelled and tested with VapBC$_{Rv0065a/c}$ in EMSAs. No shift was observed in the VapBC concentration range of 0 nM to 450 nM (Figure 3.27). To ascertain whether the lack of VapBC$_{Rv0065a/c}$ DNA-binding activity was due to the complex

co-purifying with nucleic acids, rendering it unable to bind additional DNA, $VapBC_{Rv0065a/c}$ purified through IMAC nickel, size-exclusion and anion exchange chromatography was run on a 1% TAE agarose gel and stained with SYBERSafe^{TM} DNA gel stain (Invitrogen, USA). The gel showed no presence of VapBC-associated DNA (data not shown).

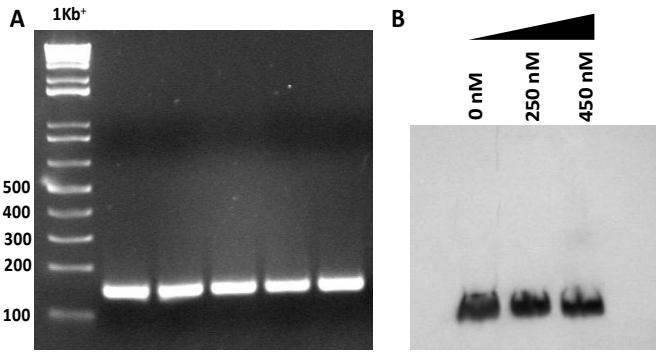


Figure 3.27 $VapBC_{Rv0065a/c}$ does not bind to a 150 bp region of the *Rv0065a/c* promoter. **A**, PCR amplification of a 150 bp DNA fragment proximal to the *Rv0065a/c* TSP. **B**, EMSAs were performed with the 150 bp DNA region (0.4 ng) encompassing a perfect 6-bp IR sequence. Concentrations of VapBC (nM) are shown above each lane.

3.4.3 Cross-talk between VapBC Complexes and Promoter DNA

EMSAs were used to investigate whether the $VapBC_{Rv0065a/c}$ complex would bind the *Rv0617a/c* promoter region, and vice versa with the $VapBC_{Rv0617a/c}$ complex. VapBC protein was tested at a concentration of 400 nM, and in both cases, no DNA-binding activity was observed with VapBC against its non-cognate promoter region (data not shown).

3.4.4 $VapBC_{Rv0617a/c}$ EMSAs with Putative *Rv0617a/c* Promoter DNA

The 130 bp putative *Rv0617a/c* promoter region DNA was DIG-labelled and used in EMSAs with purified $VapBC_{Rv0617a/c}$ protein as described in Section 2.5.5. Assays included 4 ng of DNA and VapBC in the range 0 nM - 350 nM (Figure 3.28). A full shift indicating DNA binding was observed at 150 nM VapBC. Introduction of 125 x excess non-labelled DNA to indicate specific competition (130 bp *Rv0617a/c* promoter, 'specific' lane on gel) out-competed binding of the labelled DNA, seen in the near abolishment of the shift. A non-labelled, non-specific competition control (*mazEF* promoter DNA) was included to assess the specificity of $VapBC_{Rv0617a/c}$ binding toward the cognate promoter. The gene *Rv1495* in the *M. tb* genome is identified as a MazF toxin homolog (Huang & He 2010) and the *Rv1494* gene a MazE antitoxin homolog. Therefore, a 600 bp region upstream of the *Rv1494* start site was amplified (Section 2.5.2) and taken to be the *mazEF* promoter. Addition of 125x excess of *mazEF* promoter DNA

('non-specific' lane on gel) did not have an effect on the shift, indicating no competition for binding with the labelled DNA.

3.4.4.1 EMSAs with Perfect Inverted Repeat Mutants

The 130 bp *Rv0617a/c* promoter region contains a 6 bp perfect IR (5'-CGTGCG-N_{22}-CGCACG-3') with halves positioned at nucleotides -112 and -84 relative to the transcriptional and translational start site of the *Rv0617a/c* operon. To assess whether the perfect IR was required for the VapBC binding that was exhibited with this region, either half or both halves were mutated as outlined in Section 2.5.3. Mutations consisted of changing the four central bases of an IR half to adenine. Adenine substitution was thought to cause the most disruption to protein-DNA interactions given the high GC content of the *M. tb* genome. DNA sequencing based on the Applied Biosystems 3130xl Genetic Analyser (WDSF, University of Waikato) confirmed that mutations had been achieved. Sequencing results are included in Appendix B1.

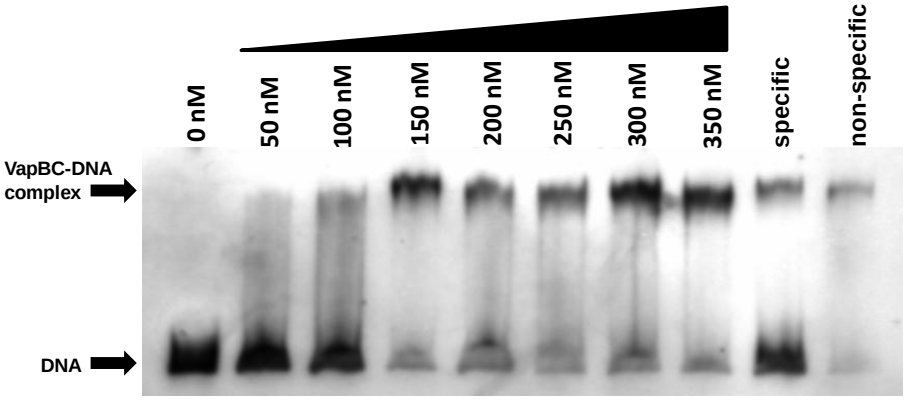


Figure 3.28 VapBC$_{Rv0617a/c}$ binds specifically to a 130 bp region of the *Rv0617a/c* promoter. EMSA experiments were performed with a DIG-labelled 130 bp promoter fragment (0.4 ng) encompassing a perfect 6-bp IR sequence. Concentrations of VapBC are shown above each lane. Non-specific and specific binding is proven in lanes 9 and 10 using 125-fold excess (50 ng) of unlabelled 130-bp promoter and 600-bp *mazEF* promoter DNA respectively.

Mutations in either half or both halves of the perfect IR did not affect DNA-binding as, in either case, EMSAs displayed no difference in shift compared to that seen with the wild-type sequence (Figure 3.29). These results indicate that the perfect IR in the promoter region of the *Rv0617a/c* gene is not essential for the VapBC$_{Rv0617a/c}$-DNA interaction.

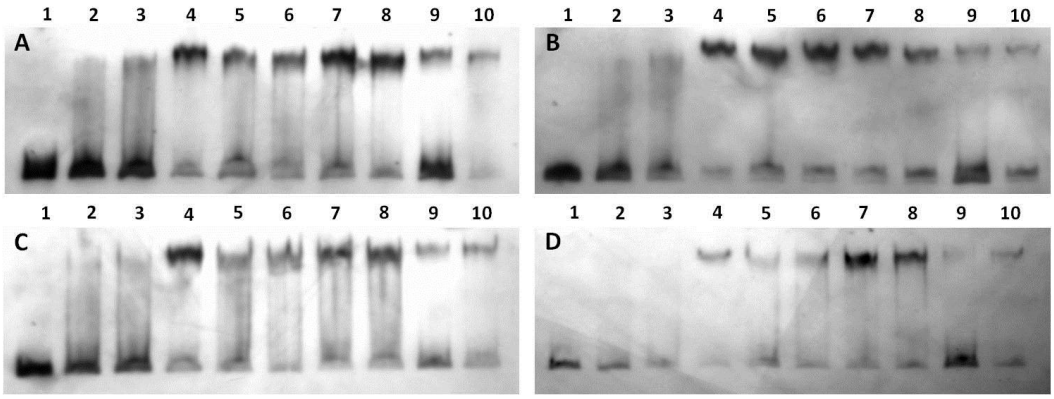


Figure 3.29 VapBC$_{Rv0617a/c}$ does not bind to the perfect IR in the *Rv0617a/c* 130 bp promoter region. EMSAs with VapBC$_{Rv0617a/c}$ were run against native and mutated versions of the 130 bp promoter region. **A**, native 130 bp promoter region. **B**, upstream half of the perfect IR mutated. **C**, downstream half of the perfect IR mutated. **D**, both halves of the perfect IR mutated. Lanes 1-8 of each assay include VapBC at concentrations of 0, 50, 100, 150, 200, 150, 300 and 350 nM. Non-specific and specific binding is proven in lanes 9 and 10 using 125-fold excess (50 ng) of unlabelled 130-bp promoter and 600-bp *mazEF* promoter DNA respectively.

3.4.4.2 Identification of Near-Perfect Inverted Repeats within the Rv0617a/c *Promoter*

An investigation of the importance of NPIRs was sparked after analysis of the crystal structure of the FitAB TA complex bound to DNA (Mattison et al. 2006). This structure revealed that specific protein residue – DNA base interactions were only occurring at 6 of the 8 nucleotide positions in each perfect IR half. At positions 4 and 5 interactions were occurring non-specifically between the protein residues and the phosphate backbone of the DNA (Section 1.3.3). Taking this into consideration, the 130 bp *Rv0617a/c* putative promoter region was examined for NPIRs that agreed with the following parameters; at least five bp in length; no more than two base mismatches between halves; a spacer of at least five bp between halves. This screen resulted in eleven NPIRs identified in the *Rv0617a/c* 130 bp putative promoter region, any of which VapBC$_{Rv0617a/c}$ could potentially be recognising and requiring for binding. The NPIRs were labelled A through J, based on their proximity to the 5'-end of the promoter sense strand. Table 3.2 provides a summary of identified NPIRs. Their positioning along the 130 bp promoter region is indicated in Figure 3.30.

Table 3.2 Properties of NPIRs in the 130 bp *Rv0617a/c* putative promoter region

IR Code (5'-3')	Genomic position from TSP*	Mismatches (bp)	IR Length (bp)	Spacer Length (bp)	Full IR covered in Oligo***:
A	-98, -118	2	6	14	1
B1**	-103, -114	1	5	6	1
B2**	-77, -103	1	5	21	2
C	-60, -80	2	6	14	2
D	-36, -75	2	6	33	3
E	-56, -73	1	5	12	3, 4
F	-45, -64	3	8	11	4
G	-48, -58	1	5	5	4
H	-30, -42	1	6	6	5
I	-27, -1	1	7	19	5
J	-14, -26	1	6	6	5

*Genomic positions were measured by counting from TSP (GTG) to the 3'-most base of each IR half

**B1 and B2 share one half of their IR sequence

***Oligo design is outlined in Section 2.5.4

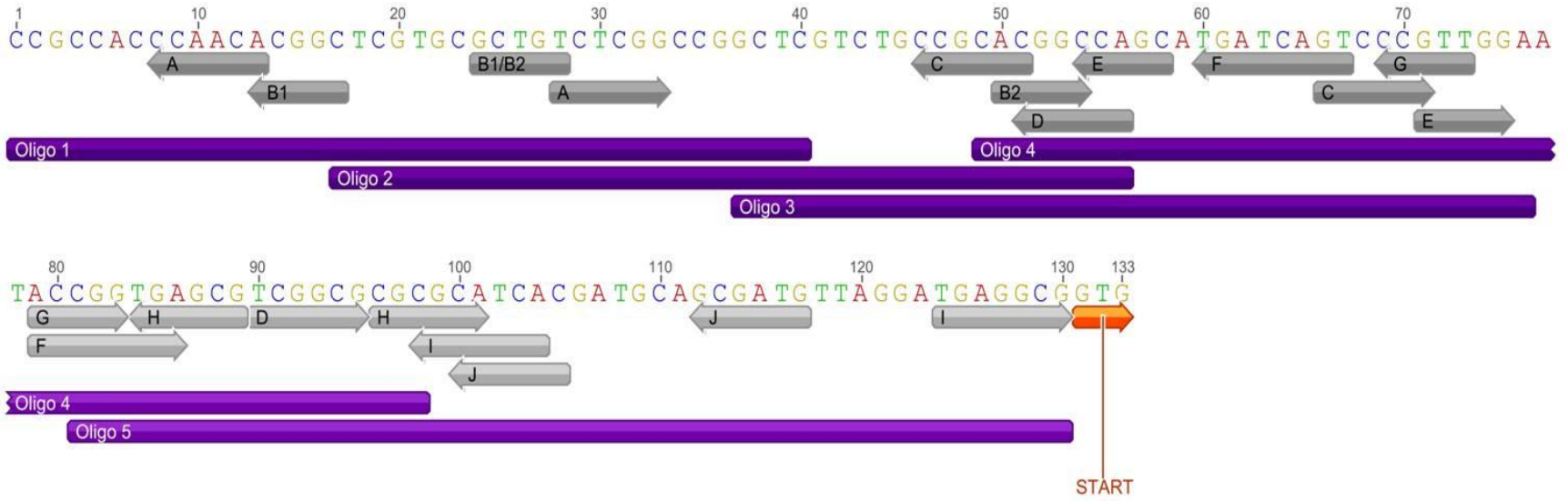


Figure 3.30 Putative *Rv0617a/c* promoter region indicating NPIRs and designed oligo positioning. NPIR (divergent grey arrows) with a maximum of three mismatches, minimum length of 5 bp and minimum spacer length of 5 bp were identified in the *Rv0617a/c* 130 bp region of DNA directly preceding the TSP (indicated in orange). NPIRs are labelled A – J. Five oligos (purple) were designed to span the promoter region such that both halves of each IR were included in at least one of the oligos.

3.4.4.3 EMSAs with Rv0617a/c *Promoter Oligos and Oligo Mutants*

In order to narrow down the region in the promoter to which the $VapBC_{Rv0617a/c}$ complex binds, and in turn to reduce the number of potential NPIR candidates required for binding, five short, overlapping 40 – 50 bp DNA oligos were designed to span the 130 bp *Rv0617a/c* putative promoter region (Section 2.5.4) (Figure 3.30). These five oligos were used in EMSAs with $VaBC_{Rv0065a/c}$ at concentrations of 0 – 350 nM protein (Figure 3.31). A shift indicating $VapBC_{Rv0065a/c}$ -DNA binding was only observed against oligo 5, which covers the region immediately upstream of the *Rv0617a/c* TSP.

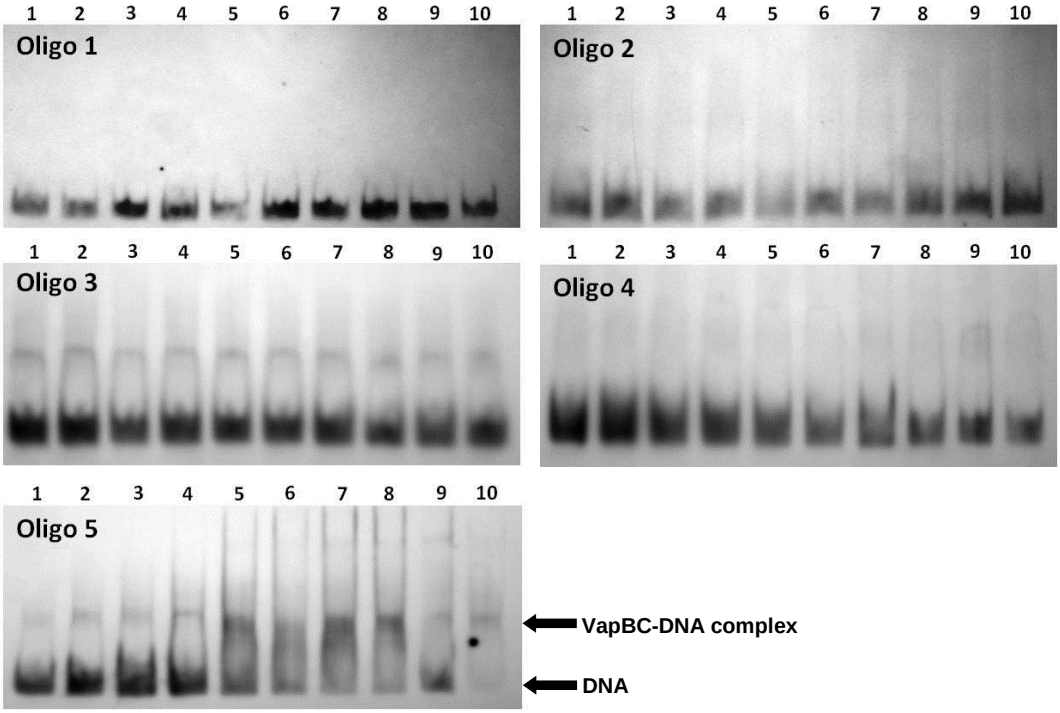


Figure 3.31 $VapBC_{Rv0617a/c}$ binds specifically to the 50 bp region of DNA directly preceding the TSP. EMSA experiments were performed to test $VapBC_{Rv0617a/c}$ binding against five oligos that span the 130 bp *Rv0617a/c* promoter region. Lanes 1 to 8 represent VapBC at varying concentrations (nM) of 5, 10, 100, 150, 200, 250, 300 and 350. Non-specific and specific binding are demonstrated in lanes 9 and 10 using 125 x excess (50 ng) of unlabelled 130 bp promoter and 600 bp *mazEF* promoter DNA respectively.

This narrowed down potential target NPIRs to those with both halves in oligo 5 (NPIRs H, I and J). Variants of oligo 5 were created where regions of NPIRs H, I and J were mutated to ascertain which of the three was essential for binding (Section 2.5.4). Assays with DNA variants that had mutations in H and I NPIRs displayed unaffected shifts (Bind 2 and 4), whereas when either half of NPIR J was mutated (Bind 3 and 5) the shifts were significantly impaired (Figure 3.32).

This led to the conclusion that NPIR J is the recognition sequence in the promoter required for $VapBC_{Rv0617a/c}$ binding.

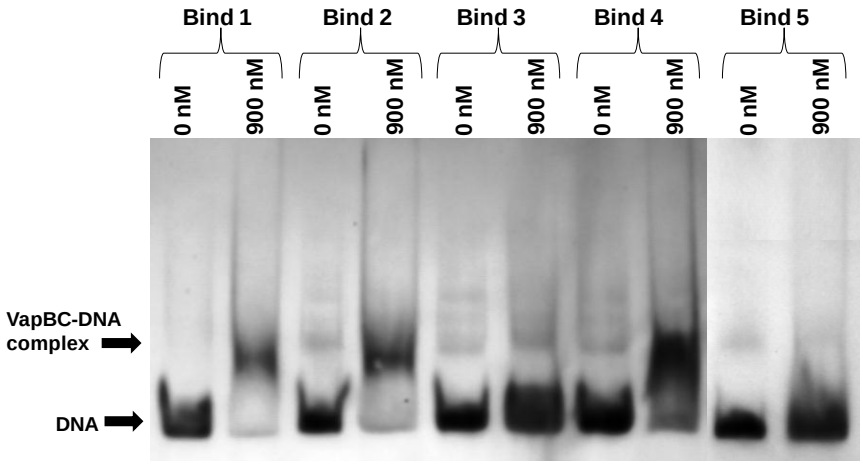


Figure 3.32 Testing $VapBC_{Rv0617a/c}$ binding against four oligo 5 variants. EMSA experiments were performed to test $VapBC_{Rv0617a/c}$ binding against four mutated promoter region oligos where IR halves had been selectively mutated out. Bind 1 (native oligo 5), Bind 2 (5' half of NPIR H mutated), Bind 3 (5' halves of NPIRs H, I and J mutated), Bind 4 (3' half of NPIR I mutated) and Bind 5 (3' half of NPIR J mutated).

EMSAs against Binds 1, 3 and 5 were repeated with a greater $VapBC_{Rv0617a/c}$ concentration range to validate results (Figure 3.33). These assays corroborated the previous findings, as $VapBC_{Rv0617a/c}$ binding activity against oligo 5 DNA was again impaired when either half of NPIR J was mutated. This was visualised as a complete abolishment of shift when the upstream NPIR half was mutated, and a near-complete abolishment when the downstream NPIR half was mutated.

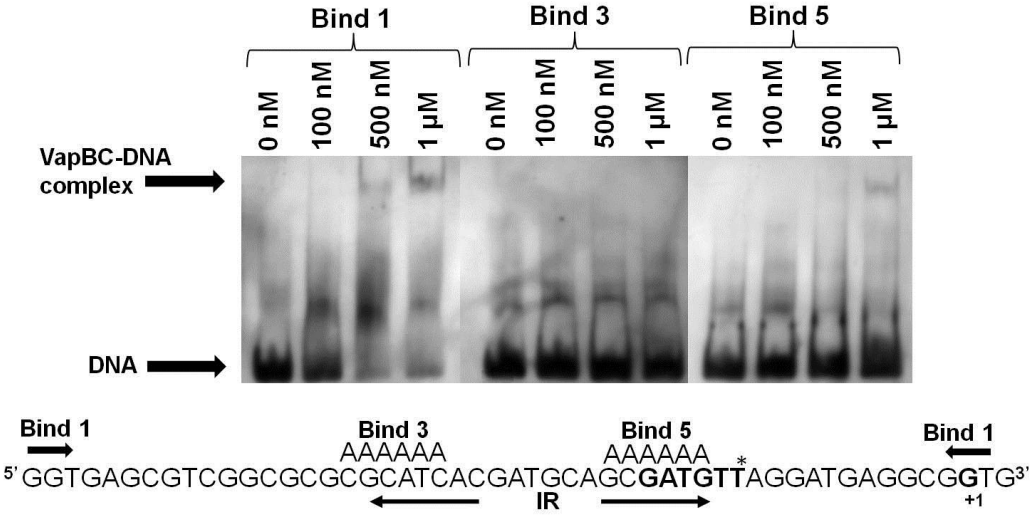


Figure 3.33 $VapBC_{Rv0617a/c}$ requires NPIR J for binding to the *Rv0617a/c* promoter region. EMSA experiments were performed to test $VapBC_{Rv0617a/c}$ binding against two mutated promoter region oligos where IR halves had been selectively mutated out. Bind 1 (native oligo 5), Bind 3 (5' halves of NPIRs H, I and J mutated) and Bind 5 (3' half of NPIR J mutated).

VapBC-binding competition assays were performed using $VapBC_{Rv0617a/c}$ protein complex (500 nM) and unlabelled oligo 5 DNA fragments with a native sequence (Bind 1), non-target NPIR H mutated (Bind 2) or either half of target NPIR J mutated (Binds 3 and 5) at 25 x and 50 x that of the DIG-labelled Bind 1 (0.4 ng) (Figure 3.34). A lack of shift indicated that the unlabelled DNA provided in excess was bound by the $VapBC_{Rv0617a/c}$ complex, and as such, 'out-competed' the DIG-labelled Bind 1 DNA.

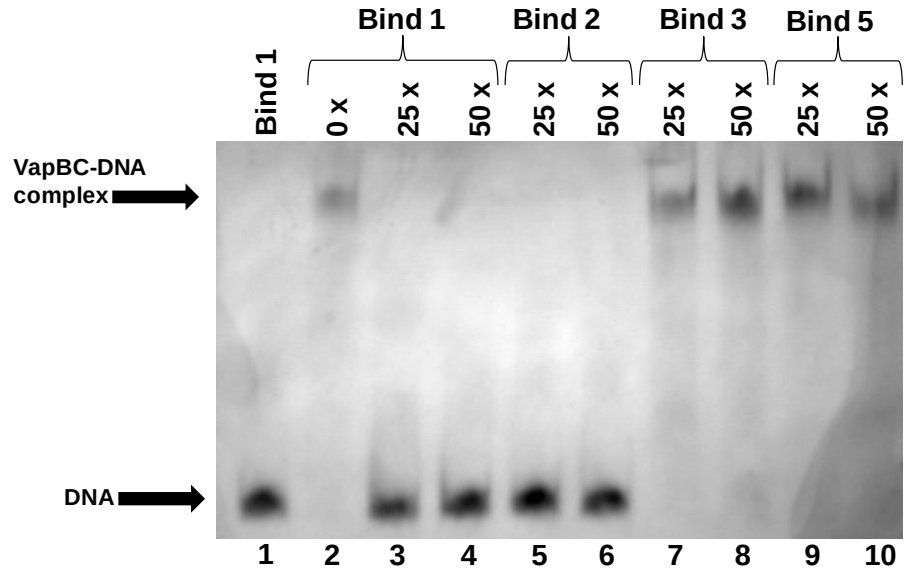


Figure 3.34 $VapBC_{Rv0617a/c}$ gel shift competition experiments. Competition experiments were performed with VapBC complex (500 nM) and DIG-labelled Oligo 5 'Bind 1' DNA. Increasing amounts of unlabelled competitor DNA were used at 25x and 50x that of DIG-labelled Bind 1. Concentration factors and varieties of competitor DNA used for each shift are shown above each lane.

Lanes 3 and 4 indicate that 25 x and 50 x excess of the unlabelled, native Bind 1 were sufficient to outcompete the DIG-labelled Bind 1 DNA, visualised as a complete loss of shift compared with that seen for Bind 1 in the absence of competition (Lane 2). A complete retardation of shift was also seen with addition of unlabelled Bind 2 (Lanes 5 and 6). This further confirms that the NPIR H is not required for $VapBC_{Rv0617a/c}$ binding to this region. In contrast, excess levels of unlabelled Bind 3 and Bind 5 were unable to out-compete labelled Bind 1, indicating that $VapBC_{Rv0617a/c}$ had lost affinity to these oligo 5 variants due to the mutations introduced into NPIR J.

Chapter Four: Discussion & Future Research

4.1 Discussion

4.1.1 General Background

Since their discovery more than 20 years ago (Gerdes et al. 1986), a multitude of studies on TA systems have resulted in a number of postulated biological roles for these ubiquitous prokaryotic genes, including involvement in the stress response, long term dormancy and formation of MDT persister cells (Frampton et al. 2012). The *vapBC* TA systems are widely spread among prokaryotes and display dramatic expansions in a number of pathogenic organisms. Due to their relatively recent identification, the biological roles of *vapBC* systems remain mainly uncharacterised. Database mining efforts are constantly identifying new *vapBC* systems, and the count for those encoded in the *M. tb* genome has reached at least 47 (Ramage et al. 2009; Arcus et al. 2011). To investigate the function of *vapBC* operons in *M. tb*, I have examined the biology and biochemistry of VapC ribonucleases and VapBC protein complexes of two of the 47 *vapBC* operons encoded on the *M. tb* chromosome. Characterisation of these proteins aimed to provide insight into their physiological roles within the *M. tb* cell and into how these roles may contribute to the virulence, persistence and drug resistance of this pathogen.

In this section the results of this thesis are discussed, focusing on VapC toxicity, VapC RNase specificity, transcriptional autoregulation mediated by VapBC, and the wider implications that the functioning of these systems may have on the pathogenicity of *M. tb*.

4.1.2 $VapC_{Rv0617}$ Overexpression is Toxic and Bacteriostatic to *Mycobacterium smegmatis* Strains

In order to assess the toxicity of $VapC_{Rv0617}$, overexpression constructs were designed to conditionally express $VapC_{Rv0617}$, $VapBC_{Rv0617a/c}$ or empty vector in both *M. smegmatis* mc^2155 wild-type and Δ*vapBC* deletion strains as a function of tetracycline inducer level. Similar experiments conducted for a number of *M. tb* VapC proteins have revealed a toxic role for the majority (Gupta 2009; Ramage et al. 2009; Ahidjo et al. 2011). In both the wild-type and deletion strains,

conditional expression of $VapC_{Rv0617}$ alone resulted in a substantial reduction in OD_{600} readings and bacterial CFUs over time. This is suggestive of induced cell stasis mediated by action of VapC. The inhibition of growth is presumably due to the effect of $VapC_{Rv0617}$ on the rate of translation and protein synthesis, given its ribonuclease activity and the fact that this effect was confirmed for the $VapC_{MS1284}$ protein when expressed in *M. smegmatis* (McKenzie 2011). Co-expression of the cognate *vapB* on the operon with *vapC,* and in turn, production of the entire $VapBC_{Rv0617a/c}$ protein complex, abrogated toxicity. This result is in agreement with the model of classical VapBC system interactions whereby toxicity of VapC can be tempered by expression of its cognate VapB (Arcus et al. 2011).

A greater $VapC_{Rv0617}$-mediated inhibitory effect was seen in CFU observations as opposed to OD_{600} readings. This may be attributed to the fact that the absorbance readings are influenced by the presence of both dead and live cells. In a parallel study of the toxicity of $VapC_{MS1284}$ in *M. smegmatis* (Robson 2010) it was observed that $VapC_{MS1284}$ overexpression resulted in altered cell morphology, namely in the lengthening of cells. This, along with cellular swelling, which is a commonly encountered specific cell change associated with exposure to toxins (Koshy et al. 1996; Kumar et al. 2012), may have contributed to increase the OD_{600} readings in the $VapC_{Rv0617}$ growth experiment.

Cultures expressing $VapC_{Rv0617}$ were seen to form fewer and slightly smaller colonies on tetracycline-supplemented agar plates; however normal growth resumed when these cultures were plated in the absence of tetracycline. This indicates that the toxic effect of $VapC_{Rv0617}$ is bacteriostatic rather than bactericidal, which is concordant with the mechanism of toxicity reported for a number of other toxins both inside (Pedersen et al. 2002; Gupta 2009; Robson et al. 2009; Ahidjo et al. 2011; Winther & Gerdes 2011; Cline et al. 2012) and outside (Pedersen et al. 2002) the VapBC family. The reversion back to normal growth upon removal of the tetracycline inducer also suggests an inactivation or break down of existing $VapC_{Rv0617}$ protein. The cellular factors leading to this instability of $VapC_{Rv0617}$ in the absence of tetracycline are not known. Stress-induced ATP-dependent serine proteases have been implicated in the degradation of the labile antitoxin in many TA systems (Yamaguchi & Inouye

2011; Yamaguchi et al. 2011). ClpXP has been theorised to play a role in influencing antitoxin stability in *M. tb* (Zhu et al. 2010) and ClpAP has been attributed to MazE antitoxin degradation (Aizenman et al. 1996). *M. tb* encodes two ClpP homologs, which form a mixed complex with an annotated role in the degradation of abnormally produced proteins (Raju et al. 2012). It may be speculated that such proteases also play a role in VapC degradation in absence of VapB.

It is important to note that the observed $VapC_{Rv0617}$ toxicity is only predicted to be due to the presence of $VapC_{Rv0617}$ protein. To confirm that toxicity is due to protein expression rather than to the effect of the *vapC* transcript, a western blot analysis for detection of $VapC_{Rv0617}$ would be required. In extrapolating the level of toxicity of $VapC_{Rv0617}$ seen in *M. smegmatis* strains to what would be expected *in vivo* in *M. tb* strains, it must be kept in mind that, on a proteomic level, genomic comparisons indicate that approximately 1218 of the 4034 (30.2 %) protein-coding genes of *M. tb* H37Ra do not have orthologues in *M. smegmatis* (Altaf et al. 2010).

4.1.3 *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* TA Elements Exhibit No Cross-talk

The theory of cross-talk occurring between TA elements encoded on separate operons is highly contested. A recent investigation by Wang et al (2013) revealed that the toxin MsqR of the *msqR/msqA* TA system enriches toxin *ghoT* mRNA both *in vitro and in vivo* through degrading the antitoxin *ghoS* mRNA. Similarly, a study conducted by Zhu et al (2010), identified several non-cognate TA interactions in *M. tb*, both within and between different TA families (e.g. MazF toxins and VapB antitoxins). In contrast to these findings, examinations of *M. tb vapBC* systems by Ahidjo et al (2011) and Ramage et al (2009) found *M. tb* VapB antitoxins to be specific to their cognate VapC toxins, whereby the cognate VapB was the only antitoxin able to rescue VapC-induced growth inhibition (Ramage et al. 2009). In this thesis research, $VapC_{Rv0617}$-mediated inhibitory effects were comparable between the *M. smegmatis* wild-type and deletion strains, a result which argues against functionally relevant cross-talk interactions occurring between toxin and antitoxin components of the *vapBC* systems in *M. smegmatis* and *M. tb*. If cross-talk were occurring, it would be expected that endogenous

$VapB_{MS1283}$ produced from the wild-type *M. smegmatis* mc^2155 strain may reduce $VapC_{Rv0617}$ activity, resulting in a lessened degree of growth-inhibition in the wild-type than in the knockout where no endogenous $VapB_{MS1283}$ is produced. Since no such difference in growth-inhibition was observed, it can be concluded that there is no interaction between $VapB_{MS1283}$ and $VapC_{Rv0617}$.

Results from EMSAs revealed that $VapBC_{Rv0617a/c}$ did not bind to a 600 bp promoter region upstream of the *M. tb* encoded *mazEF* TA system operon, nor did it bind to the fellow *Rv0065a/c vapBC* TA system promoter. This provides evidence against any degree of VapBC-mediated autoregulatory cross-talk occurring within the multiple *vapBC* systems or among different TA systems in *M. tb.*

4.1.4 $VapC_{Rv0065}$ and $VapC_{Rv0617}$ are Ribonucleases Targeting a Combination of RNA Sequence and Structure

It remains unclear how PIN-domain proteins achieve specificity in their ribonuclease activities. $VapC_{Rv0065}$ and $VapC_{Rv0617}$ RNase assays included in this thesis support the previously drawn conclusion that both these VapC proteins target GC-rich 4-mers in the RNA. Due to the high GC content of the *M. tb* genome, this points toward a role of $VapC_{Rv0065}$ and $VapC_{Rv0617}$ as global gene down-regulators of mRNA transcripts when conditions allow for VapC to be free from its cognate VapB inhibitor. Sequence specificity directing toxins to a broad range of targets within the cell is not uncommon in TA systems. For example, the MazF toxin from *E. coli* targets ACA sequences, which, due to its high prevalence across the genome, makes MazF an extremely potent toxin (Zhang et al. 2003).

The shared recognition sequence for both $VapC_{Rv0065}$ and $VapC_{Rv0617}$ is similar to the case seen in *P. aerophilum*, where two VapC proteins separately encoded in the genome, $VapC_{PAE0151}$ and $VapC_{PAE2754}$, were found to target the same RNA sequence (GGUG) (McKenzie 2011). These *P. aerophilum* VapC proteins differed from each in terms of catalytic rate, as $VapC_{PAE2754}$ displayed faster ribonuclease activity compared to $VapC_{PAE0151}$ (McKenzie 2011). In contrast, $VapC_{Rv0065}$ and $VapC_{Rv0617}$ appeared to have equivalent ribonuclease activity. If these proteins do in fact have the same activity, it could be speculated that activation of a single one of these two *M. tb* TA systems could be employed for a

lower level of transcriptome regulation, while the additive effect of activation of both would result in more rapid alterations to the transcriptome when required. RNA recognition sequences for the remaining ~45 VapC proteins encoded in the *M. tb* genome are yet to be determined, and it will be interesting to see whether they too target GC-rich 4-mers in the RNA.

A higher degree of specificity in the RNase activity of *M. tb* VapC proteins was suggested by results obtained from the $VapC_{Rv0617}$ conditional expression experiment conducted in *M. smegmatis.* The fact that $VapC_{Rv0617}$ displayed only slight toxicity suggested that a small subset of mRNA transcripts was being targeted for degradation. If the toxin was non-specifically degrading the majority of the transcriptome, or cleaving at every GC-rich 4-mer site of the mRNA, a much larger growth inhibitory effect would be expected.

Accordingly, results of $VapC_{Rv0617}$ and $VapC_{Rv0065}$ RNase assays against the 932 109-126 RNA fragment suggested that this additional activity determinant resides in the secondary structure of the RNA. If these VapC proteins solely recognised the GC-rich 4-mer cut sites, then the above fragment, given the activity observed against its slightly extended counterpart (932 94-126) and its significant GC content, would be predicted to be cleaved very efficiently. Instead, the complete loss of activity when the 932 94-126 fragment was truncated by just 12 bases while retaining all predicted cut sites is a strong indicator of the influence that secondary structure imparts on VapC RNase activity. As both $VapC_{Rv0065}$ and $VapC_{Rv0617}$ lacked activity against the 932 109-126 fragment, this suggests that they may exhibit specificity for the same RNA secondary structural element. The efficient cleavage by both VapC of substrates with a higher degree of secondary structure, (the 932 Pentaprobe and the longer 932 Pentaprobe fragments) further supports the theory that VapC requires the presence of a particular structural motif for RNase activity to be exhibited. The 932 fragments that were cleaved the most efficiently at their GC-rich centre had this sequence located in a hairpin loop which was flanked by a long stem region. This may be a motif required for efficient VapC cleavage, however it is also possible that the secondary structure motif recognised is positioned further from the cut site.

The optimal combination of sequence and secondary structural motifs associated with VapC RNase activity may be present in the rRNA component of *M. smegmatis* total RNA samples. The results of RNase assays with *M. smegmatis* total RNA revealed that the majority of rRNA in the sample was degraded within 15 minutes of incubation with either VapC. In contrast, RNase mediated degradation of the 932 Pentaprobe RNA occurred over a longer timescale, perhaps due to such VapC recognition elements being less abundant or lacking.

Narrowing down the site and identity of the secondary structural motif(s) required for VapC cleavage will be hampered by methods of structure determination. Current physical methods involved in RNA structure determination such as x-ray diffraction and NMR spectroscopy are time-consuming and costly, so methods involving the computational prediction of structure, such as use of the RNAfold software, are necessary (Mohsen et al. 2009). This approach, however, is hampered by its reliance on algorithmic predictions, meaning that results are associated with a degree of uncertainty.

VapC-mediated cleavage at certain secondary structure elements in the mRNA may serve to feed redundant mRNA transcripts into more generalised cell RNA degradation pathways. While the general system for mRNA degradation in mycobacteria is not well characterised, it is known that secondary structural elements such as those at the 5' and 3' UTRs can increase stability of the mRNA and provide a barrier against degradation by nucleases (Steege 2000). Results indicated that both $VapC_{Rv0065}$ and $VapC_{Rv0617}$ are able to cleave *in vitro* synthesised RNA with a high degree of predicted secondary structure. Therefore, *in vivo*, when the correct secondary structural element is present, VapC-mediated cleavage may remove this barrier, increasing the susceptibility of mRNA transcripts to enter general RNA degradation pathways.

An additional factor that may contribute to the specificity of VapC ribonuclease activity which has not yet been investigated is the potential influence of RNA modifications such as methylation. While not commonly associated with bacterial RNA, Yi & Pan (2011) suggest there is increasing evidence for RNA modifications playing a regulatory role in prokaryotic cells, especially during the stress response. The methylated nucleotides m^6A, m^1A, C_m, m^5C, G_m, m^1G, m^7G,

m^5U and U_m are common in both eukaryotes and prokaryotes (Motorin & Helm 2011). Bacterial rRNA methylation at several 2'-O sites along rRNA is reported to confer resistance to certain antibiotics such as aminoglycosides, macrolides and lincosamides through altering the affinity of these drugs for the ribosome (Motorin & Helm 2011). The 2'-O-methylation of tRNAs was shown to prevent cleavage by ribotoxins (Yi & Pan 2011). VapC proteins such as $VapC_{Rv0617}$ and $VapC_{Rv0065}$ that appear to target the same primary RNA sequence and secondary structure may in fact be differentiated by the RNA modifications that either one recognises.

The additional layers of VapC RNase specificity residing in the RNA secondary structure, and potentially in RNA modifications, would reduce the enormous number of target transcripts in the *M. tb* genome that would be cleaved based on sequence recognition alone. Different VapC proteins in the same organism may strategically target unique subsets of the transcriptome specific to different metabolic pathways. This would allow strict control over the genetic elements expressed at any one time, equipping the cell with a mechanism to 'reprogram' itself to quickly adapt to changing external factors and different stress conditions.

4.1.5 $VapC_{Rv0065}$ and $VapC_{Rv0617}$ Display Ribonuclease Activity Against 16S and 23S rRNA Isolated from *Mycobacterium smegmatis*

While both $VapC_{Rv0065}$ and $VapC_{Rv0617}$ enzymes displayed Mg^{2+}-dependent ribonuclease activity against purified *M. smegmatis* 16S and 23S rRNA, the RNase action was to such an extent that discrete degradation product bands could not be identified. Instead, the array of degradation products appears as a smear on the gel. This indicates that the VapC enzymes cleave at a multitude of sites along the rRNA sequences. Due to the lack of definite degradation product bands, it was decided that this approach for determining secondary structure requirements of the VapC enzymes was not to be pursued further. If the results of this experiment are taken to indicate $VapC_{Rv0617}$ and $VapC_{Rv0065}$ activities against rRNA *in vivo*, it must be kept in mind that ribosomal proteins associated with the rRNA *in vivo* may play a role in protecting the rRNA from such ribonuclease factors. For example, the MazF toxin from *E. coli* is seen to target ACA sequences and effectively cleaves both 23S and 16S rRNA *in vitro*, yet displays no such activity *in vivo* (Zhang et al. 2003).

4.1.6 $VapBC_{Rv0617a/c}$ Binds a NPIR in the *Rv0617a/c* Promoter

Analysis of a 150 bp region upstream of the *Rv0617a/c* transcriptional start point revealed two possible -10 elements: one which shared sequence with a -10 element annotated in another *M. tb* promoter ('T26') (Bashyam et al. 1996) and another which shared the consensus sequence for mycobacterial sigma factor A binding site (Gomez & Smith 2000; Zaunbrecher et al. 2009). No -35 element that showed similarity to other mycobacterial or *E. coli* -35 elements was found in this region; however this is unsurprising given that there is very little sequence conservation of this element among mycobacteria (Bashyam et al. 1996). Within the 150 bp region, one perfect IR was identified along with eleven NPIRs. EMSA experiments revealed that $VapBC_{Rv0617a/c}$ binds specifically to this 150 bp region of DNA. Mutations made to either half of the perfect IR in this region had no effect on gel shifts, implying that the perfect repeat was not essential for $VapBC_{Rv0617a/c}$ binding. Further EMSA experiments identified a 6 bp NPIR overlapping the -10 region, 'NPIR J' as being essential for $VapBC_{Rv0617a/c}$ binding, as complete or near-complete loss of shift was observed when either half was mutated. Loss of shift when either half-site alone is mutated suggests that $VapBC_{Rv0617a/c}$ must bind both sites of the NPIR to achieve repression.

While the majority of VapBC complexes characterised to date have displayed perfect IR recognition, the VapBC2 complex of *Rickettsia felis* similarly requires a near-perfect palindrome for binding (Mate et al. 2011). The fact that the NPIR that $VapBC_{Rv0617a/c}$ bound overlaps the putative -10 recognition site is consistent with a mechanism of autoregulation whereby binding of the complex sterically excludes RNA polymerase from the DNA. It is possible that the $VapBC_{Rv0617a/c}$ complex binds to the DNA in a similar manner to the VapBC complex ('FitAB') of *N. gonorrhoeae*; as a tetramer of heterodimers, where each VapB dimer binds to one half of the IR (Mattison et al. 2006). The two systems differ however in that the FitAB complex was shown to require only one half of the IR for binding.

Analysis of the 101 bp region upstream of the *Rv0065a/c* transcriptional start point revealed no similarity in the -10 and -35 regions to those seen in other mycobacterial promoters. Despite the presence of a IR sequence, EMSA experiments indicated no DNA binding activity of $VapBC_{Rv0065a/c}$ to this region, nor to a region extending downstream of the TSP. If $VapBC_{Rv0065a/c}$ does exhibit

autoregulatory activity through binding to its cognate promoter, this binding site must be relatively far from the TSP. Is has been noted in the *ccdAB* (Afif et al. 2001), *phd/doc* (Johnson et al. 1996) and *parDE* (Magnuson & Yarmonlinksy 1998) TA systems that an altered toxin:antitoxin ratio can result in the formation of TA complexes with varied stoichiometry that no longer exhibit DNA-binding activity. An alternate explanation for the absence of $VapBC_{Rv0065a/c}$ DNA-binding activity therefore may be that the ratio of toxin : antitoxin in the $VapBC_{Rv0065a/c}$ samples was such that it caused destabilisation of the repressor-DNA complex. This mechanism of alleviating repression when the toxin ratio is high is suggested to be a common mechanism in TA systems, and may function to promote production of the antitoxin to prevent harmful toxin activation when it is not required (Afif et al. 2001).

Conversely, the *Rv0065a/c* operon may not employ autoregulation through VapBC binding to the promoter region. Individualised transcriptional regulation mechanisms among the *vapBC* systems of *M. tb* may add an extra layer of complexity to the regulation of the *vapBC* TA system network, helping explain the persistence of such an expanded number of these operons in the genome.

4.1.7 VapBC Systems' Theorised Physiological Role in *Mycobacterium tuberculosis*

A hallmark of *M. tb* infections is the bacteria's ability to persist for months to years in an asymptomatic, dormant state. The mycobacterial $VapC_{MS1284}$ is proposed to maintain the balance of anabolic and catabolic rates in *M. smegmatis* through its ribonuclease activity (McKenzie 2011). Based on the RNase activity that $VapC_{Rv0065}$ and $VapC_{Rv0617}$ display *in vitro,* and the toxic, bacteriostatic effect that their unmasked expression has on mycobacterial cells *in vivo,* it is possible that they play a role in degrading redundant transcripts within the *M. tb* cell. This would serve to correct the metabolic imbalances that can result in the production of toxic intermediates (Russell & Cook 1995). Fine-tuning of catabolic and anabolic rates would allow conservation of energy, a factor that would be required for the *M. tb* persistence in the dormant state.

A theory that remains to be explored is the possibility that VapBC complexes and/or the VapC proteins are secreted/exported from the *M. tb* cell, allowing them

to act upon the mRNA of the host, i.e. that of the macrophage cell. Preliminary evidence for this theory includes recent studies which report the *Rv0617* ($VapC_{Rv0617}$) (Gomez et al. 2000) and *Rv0064* ($VapB_{Rv0064}$) (Malen et al. 2007) genes as encoding putative secreted *M. tb* proteins.

4.2 Future Resesarch

4.2.1 Determine the Toxicity of $VapC_{Rv0065}$ to *Mycobacterium smegmatis* Strains

The pMind expression constructs for investigating the effect of conditional overexpression of $VapBC_{Rv0065a/c}$ and $VapC_{Rv0065}$ were created and transformed into wild-type *M. smegmatis* mc^2155 and knockout *M. smegmatis* mc^2155 Δ*vapBC* strains. Conducting a growth experiment similar to that done for the *Rv0617a/c vapBC* TA system would allow more comparisons to be drawn between the two systems, and would provide greater insight into the role of this system *in vivo*.

4.2.2 Microarray Analysis of $VapC_{Rv0065}$ and $VapC_{Rv0617}$ Overexpression

In order to identify *in vivo* transcript targets, it may be of interest to compare the transcriptome of *M. smegmatis* cultures in the presence and absence of $VapC_{Rv0065}$ and $VapC_{Rv0617}$ expression through microarray analysis. As activity and specificity studies of these two VapC proteins *in vitro* has revealed a shared RNase specificity, it would be interesting to examine whether they would induce downregulation of the same sets of transcripts when acting *in vivo.* Studies conducted to investigate factors the leading to activation of either VapC may reveal whether these systems are triggered by the same or different stresses.

4.2.3 Further Examine the Specificity of *Mycobacterium tuberculosis* VapC Ribonucleases

It remains to be determined whether multiple VapC proteins from the same organism target shared or exclusive mRNA transcripts *in vivo*. Therefore, to better understand the collective role of VapC proteins in *M. tb* physiology, efforts should be made to express and purify the remaining ~45 VapC *M. tb* proteins. The Pentaprobe system combined with MALDI TOF MS should be applied to each to characterise VapC RNA sequence recognition sites. In addition, the influence of secondary structure and possibly RNA modifications should be probed further, and attempts made to determine what recognition motif(s) are required and where these reside in relation to the RNA cut site.

4.3 Conclusions

VapC proteins are present in half of all sequenced prokaryotes and notably so in pathogens such as *N. gonorrhoeae* (Wilbur et al. 2005; Mattison et al. 2006), *Haemophilus influenzae* (Daines et al. 2007; Cline et al. 2012), *Leptospira interrogans* (Zhang et al. 2004) and *M. tb* (Miallau et al. 2009; Ramage et al. 2009). This thesis has outlined the characterisation of proteins encoded by the *Rv0065a/c* and *Rv0617a/c vapBC* operons of the *M. tb* genome. Relating back to the initial research objectives, the following conclusions can be drawn:

1) $VapC_{Rv0617}$ is a toxic protein that exhibits an *in vivo* bacteriostatic effect when conditionally expressed in *M. smegmatis* strains.
2) $VapC_{Rv0065}$ and $VapC_{Rv0617}$ display the same sequence specificity and target GC-rich 4-mers in the RNA. Their activity appears to be additionally influenced by the secondary structure of the RNA substrate.
3) The $VapBC_{Rv0617a/c}$ complex binds specifically to a NPIR sequence in the putative *Rv0617a/c* promoter DNA, presumably to fulfil a transcriptional autoregulatory role. In contrast, the $VapBC_{Rv0065}$ complex displays no such DNA binding activity with putative *Rv0065a/c* promoter DNA.

Ultimately, gaining a better understanding of VapC and VapBC protein functioning within the *M. tb* cell may reveal strategies for hindering activity of the VapC toxins. Encumbering their RNase action would prevent the ability of cells to selectively alter their transcriptome by this means. This would in theory hamper the cell's ability to regulate its growth rate in order to adapt to environmental stresses. In turn, the formation of persister cells which lead to latent *M. tb* infections would be prevented, and the efficacy of both antibiotic treatments and immune-driven attack would be increased.

Appendices

Appendix A: Reagents

A1: Bacterial Strains and Plasmids

Table A1.1 Bacterial Strains

Strain	Description
E. coli	
DH5α	High plasmid copy number, transformation efficiency and DNA yield. *fhuA2* Δ*(argF-lacZ)U169 phoA glnV44* ϕ*80* Δ(*lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
DH10B (TOP10)	High plasmid copy number, transformation efficiency and DNA yield. *F-mcrA* Δ(*mrr-hsdRMS-mcrBC)*ϕ*80* d*lacZ* ΔM15 Δ*lacX74 deoR recA1 araD139* Δ *(ara leu)7697 galU galK rpsL endA1 nupG*
M. smegmatis	
mc^2155	Electrocompetent lab strain of *M. smegmatis*
mc^2155Δ*vapBC* (JR121)	mc^2155Δ*vapBC*::*aphaA-3*; Km^r acquired from Jennifer Robson, University of Otago (2010)
$mc^2$4517	*M. smegmatis* expression strain with T7 RNA polymerase, Km^r

Table A1.2 Plasmids

Plasmid	Description
pYUB28b	*Mycobacterium* expression vector; Km^r Hyg^r
pMind	Tetracycline inducible vector; Km^r Hyg^r
0065vapBC_pMind	Tetracycline inducible vector with 0065a/c operon insert; Km^r Hyg^r
0065vapC_pMind	Tetracycline inducible vector with 0065c gene insert; Km^r Hyg^r
0617vapBC_pMind	Tetracycline inducible vector with 0617a/c operon insert; Km^r Hyg^r
0617vapC_pMind	Tetracycline inducible vector with 0617c gene insert; Km^r Hyg^r

A2: Buffers, Solutions and Media

Table A2.1 Buffers and Solutions

Name	Composition
5x DNA Loading Dye	0.05 % (*w/v*) bromophenol blue 0.25% (*w/v*) xylene cyanol, 30% (*v/v*) glycerol
Fairbanks Staining Solution A	0.05% (*w/v*) coomasie blue R-250, 25% (*v/v*) isopropanol, 10% (*v/v*) acetic acid
Fairbanks Staining Solution B	0.005% (*w/v*) coomasie blue R-250, 10% (*v/v*) isopropanol, 10% (*v/v*) acetic acid
Fairbanks Staining Solution C	0.002% (*w/v*) coomasie blue R-250, 10% (*v/v*) isopropanol, 10% (*v/v*) acetic acid
Fairbanks Staining Solution D	10% (*v/v*) acetic acid
2 x Formamide Loading Buffer	80% (*v/v*) formamide, 5 mM EDTA, 0.1% (*w/v*) bromophenol blue, 0.1% (*w/v*) xylene cyanol FF
GITC	295.4 g guanidine thiocyanate, 2.5 g N-lauroyl sarcosin, 3.9 g tri-sodium citrate, 3.6 ml 2-mercaptoethanol, 280 ml DEPC H_2O (total volume 500 ml)
4 x SDS loading Buffer	200 mM Tris-HCl pH 6.8, 8% (*w/v*) SDS, 40% (*v/v*) glycerol, 0.4% (*w/v*) bromophenol blue, 400 mM β-mercaptoethanol
SDS Resolving Gel Layer	16.5% acrylamide*, 0.1 % (*w/v)* SDS, 0.05% *(w/v)* APS, 0.05% *(v/v)* TEMED, 375 mM Tris (pH 8.8)
SDS Stacking Gel Layer	5% acrylamide*, 0.1 % (*w/v)* SDS, 0.05% *(w/v*) APS, 0.05% *(v/v)* TEMED, 128 mM Tris (pH.6.8)
SDS running Buffer	25 mM Tris-HCl ph 6.8, 0.1% (*w/v*) SDS, 190 mM glycine
TBE Gels	6% acrylamide*, 0.05% (*w/v*) APS, 0.05 % (*v/v*) TEMED, 15 % (*v/v*) 10 x TBE.
10 x TBE	0.89 M Tris-HCl, 0.89 M boric acid, 20 mM EDTA
1 x TBE	100 ml 10 x TBE + 900 ml H_2O
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0

*Acrylamide was added from a 30% acrylamide/bis stock solution 37.5:1 (BioRad Laboratories, USA).

Table A2.2 Liquid Media

Name	Composition
7H9/ADC/T	Dissolve 0.47 g 7H9 powder (BD) and 0.2% glycerol in 90 ml H_2O, autoclave, then add 10 ml ADC enrichment and 0.05% Tween-80 when cool
Hartmans-de Bont	Add components in the following order: 1 x trace metal stock*, 15 mM $(NH_4)_2SO_4$, 0.05% (*v/v*) Tween-80, 0.2% (*v/v*) glycerol, 50 mM MOPS. Autoclave, then add 8.9 M K_2HPO_4 and 7.08 M NaH_2PO_4
LB	1% bactotryptone, 0.5% yeast extract, 1% NaCl pH 8.0
LBT	1% bactotryptone, 0.5% yeast extract, 1% NaCl, 0.05% Tween-80 pH 8.0
Low salt LB	1% bactotryptone, 0.5% yeast extract, 0.5% NaCl pH 8.0
PA-0.5G	50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 0.5% glucose, 0.1 x metals mix**, 200 µg/ml each of 17 amino acids (no Cys, Tyr or Met). Individual components are autoclaved or sterile filtered before adding to sterile H_2O.
SOC	2% bactotryptone or bactopeptone, 0.55% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose
ZYP-5052	1% bactotryptone, 0.5% yeast extract, 50 mM Na_2HPO_4, 25 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 1 x metals mix

*****The 100 x trace metals stock for HdB media consisted of: EDTA (0.1 g), $MgCl_2.gH_2O$ (1 g), $CaCl_2.2H_2O$ (10 mg), $NaMoO_4.2H_2O$ (2 mg), $CoCl_2.6H_2O$ (4 mg), $MnCl_2.2H_2O$ (10 mg), $ZnSO_4.7H_2O$ (20 mg), $FeSO_4.7H_2O$ (50 mg), $CuSO_4.5H_2O$ to 100 ml

**The 1000x metals mix for PA-0.5G media consisted of sterile components made up to the following concentrtaions: 50 µM $FeCl_3$ in 0.12 M HCl (filter sterilized), 20 µM $CaCl_2$, 10 µM $MnCl_2$, 10 µM $ZnSO_4$, 2 µM $CoCl_2$, 2 µM $CuCl_2$, 2 µM $NiCl_2$, 2 µM Na_2MoO_4, 2 µM Na_2SeO_3, 2 µM H_3BO_3

Table A2.3 Solid Media

Name	Composition
7H10/ADC/T-agar	Dissolve 1.9 g 7H10 powder (BD), 0.5% glycerol in 90 ml H_2O, autoclave, the add 10 ml ADC enrichment and 0.05% Tween-80 when cool
LB-agar	1% bactotrypton, 0.5% yeast extract, 1% NaCl, 15 g/L agar pH 8.0
LBT-agar	1% bactotrypton, 0.5% yeast extract, 1% NaCl, 15 g/L agar, 0.05% Tween-80 pH 8.0
Low salt LB-agar	1% bactotrypton, 0.5% yeast extract, 0.5% NaCl, 15 g/L agar pH 8.0

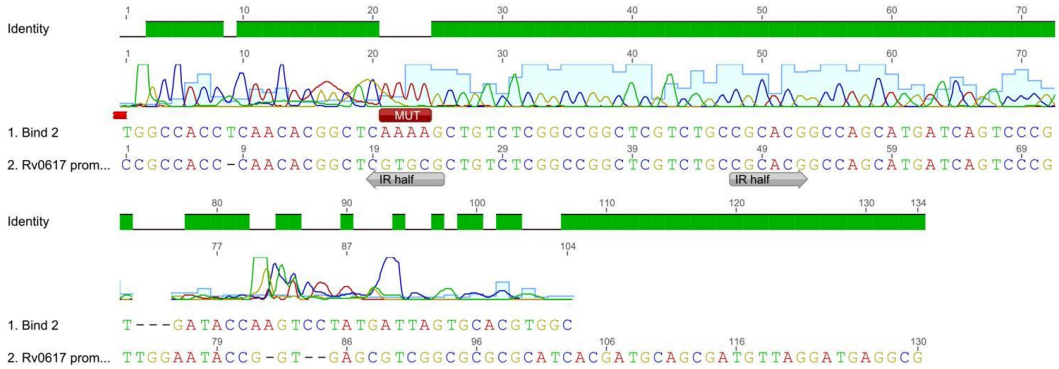
Appendix B: Sequencing Results

B1: Perfect IR Mutagenesis

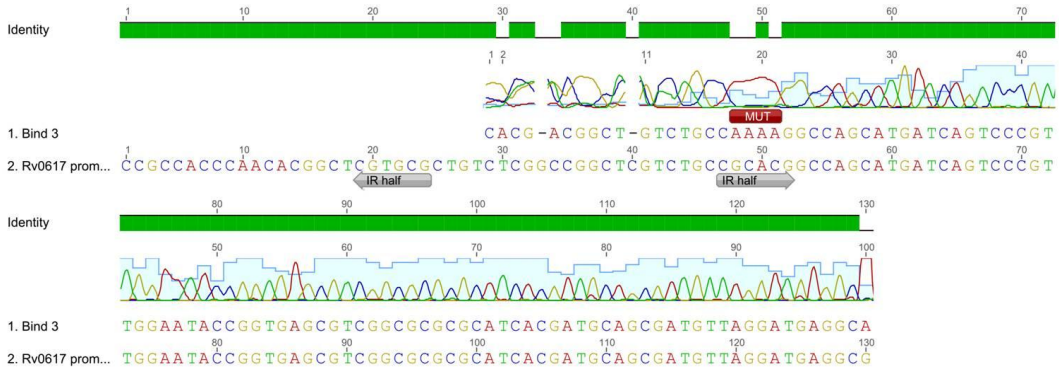
Sequencing results for two-halves mutagenesis of the perfect IR in the 130 bp *Rv0617a/c* promoter region. Bind 2 represents the upstream half mutated, Bind 3 the downstream half mutated, and Bind 4 both halves mutated. Shown for each mutant promoter sequence is the sequencing chromatogram results aligned with the wild-type *Rv0617a/c* 130 bp promoter region DNA sequence. IR; inverted repeat. MUT; mutated bases.

Mutant	Sequence

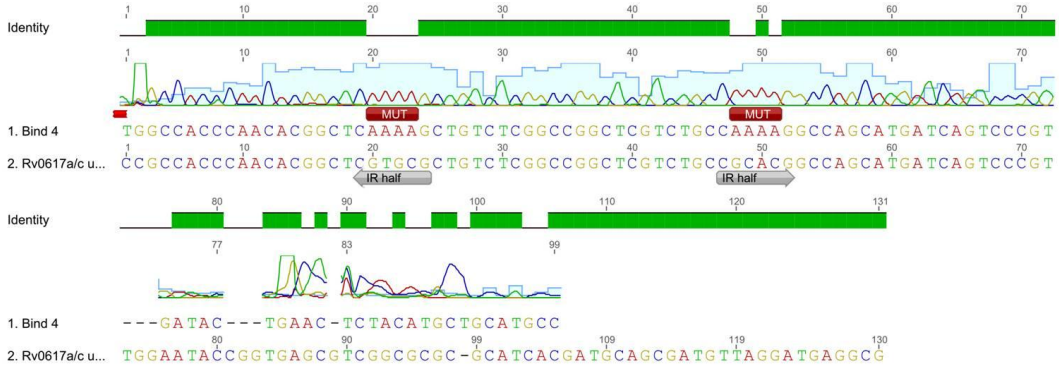
Bind 2



Bind 3



Bind 4



Appendix C: Protein and RNA Information

C1: *Mycobacterium* Protein Information

$VapB_{Rv0064}$ Protein Sequence
MATIQVRDLPEDVAETYRRRATAAGQSLQTYMRTKLIEGVRGRDKAEAI
EILEQALASTASPGISRETIEASRRELRGG*
80 amino acids
Molecular weight of monomer: 8.71
Theoretical pI: 9.12

$VapC_{Rv0065}$ Protein Sequence with C-terminal His tag
VDECVVDAAAVVDALAGKGASAIVLRGLLKESISNAPHLLDAEVGHALR
RAVLSDEISEEQARAALDALPYLIDNRYPHSPRLIEYTWQLRHNVTFYDAL
YVALATALDVPLLTGDSRLAAAPGLPCEIKLVRKLAAALEHHHHHH*
147 amino acids
Molecular weight of monomer: 15.8 kDa
Theoretical pI: 6.34

$VapBC_{Rv0065a/c}$ Protein Complex
Molecular weight of heterodimer: 24.51 kDa

$VapB_{Rv0616}$ Protein Sequence
VRTTIDLPQDLHKQALAIARDTHRTLSETVADLMRRGLAANRPTALSSDP
RTGLPLVSVGTVVTSEDVRSLEDEQ*
76 amino acids
Molecular weight of monomer: 8.2
Theoretical pI: 5.64

$VapC_{Rv0617}$ Protein Sequence with C-terminal His tag
VTVLLDANVLIALVVAEHVHHDAAADWLMASDTGFATCPMTQGSLVRF
LVRSGQSAAAARDVVSAVQCTSRHEFWPDALSFAGVEVAGVVGHRQVT
DAYLAQLARSHDGQLATLDSGLAHLHGDVAVLIPTTTKLAAALEHHHHH
H*
147 amino acids
Molecular weight of monomer: 15.5 kDa
Theoretical pI: 6.5

$VapBC_{Rv0617a/c}$ Protein Sequence
Molecular weight of heterodimer: 23.7 kDa

C2: *Pyrobaculum* Protein Information

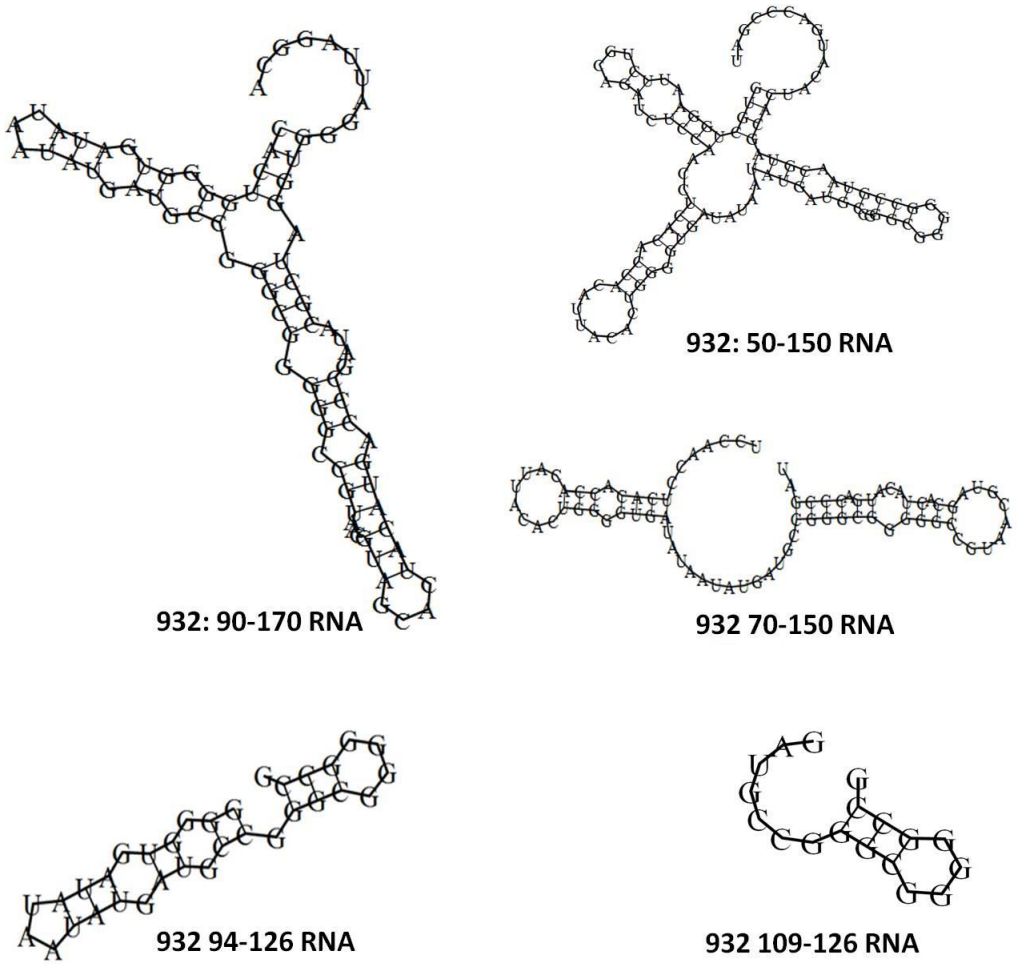
$VapC_{PAE0151}$ Protein Sequence with pProEX fusion tag
MSYYHHHHHHDYDIPTTENLYFQGAMKLVVDASAIAALYVPEERSEQAE
RAVSQAQELHTLDLAAYEVANDLWKHARRGLLREDEASNMLEELWEFF
KALKVHSYAEVLKDAFALALKHGVTVYDAAYVALAEKIGGKLLTLDRQ
LAEKFPALVTP*
156 amino acids
Molecular Weight: 17.7 kDa
Theoretical pI: 5.50

C3: Pentaprobe RNA and Flanking Sequences

A representation of the RNA sequence that is transcribed from the T7 promoter. Regions highlighted in yellow correspond to the Pentaprobe sequence (McKenzie 2011)

AGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAG
TGTGCTGGAATTCTGCAGATCTCCAACCTCACACCACATTACACTGGG
GTGATATAATATGATGCCGGGCGGGGGCCGTAACGTAGCACTACATGA
CCCGATACGCTAGGTGGGATTAGGCATCACACTGGCGGCCGCTCGAGC
ATGCATCTAGAGGGCCCTATTCTATAGTGTCACCTAAATGCTAGAGCT
CGCT

C4: Predicted Secondary Structure of 932 RNA Fragments



Minimum free energy structures predicted using RNAfold (Vienna RNA webservers, TBI).

C5: Predicted Secondary Structure of 932 RNA Oligonucleotides

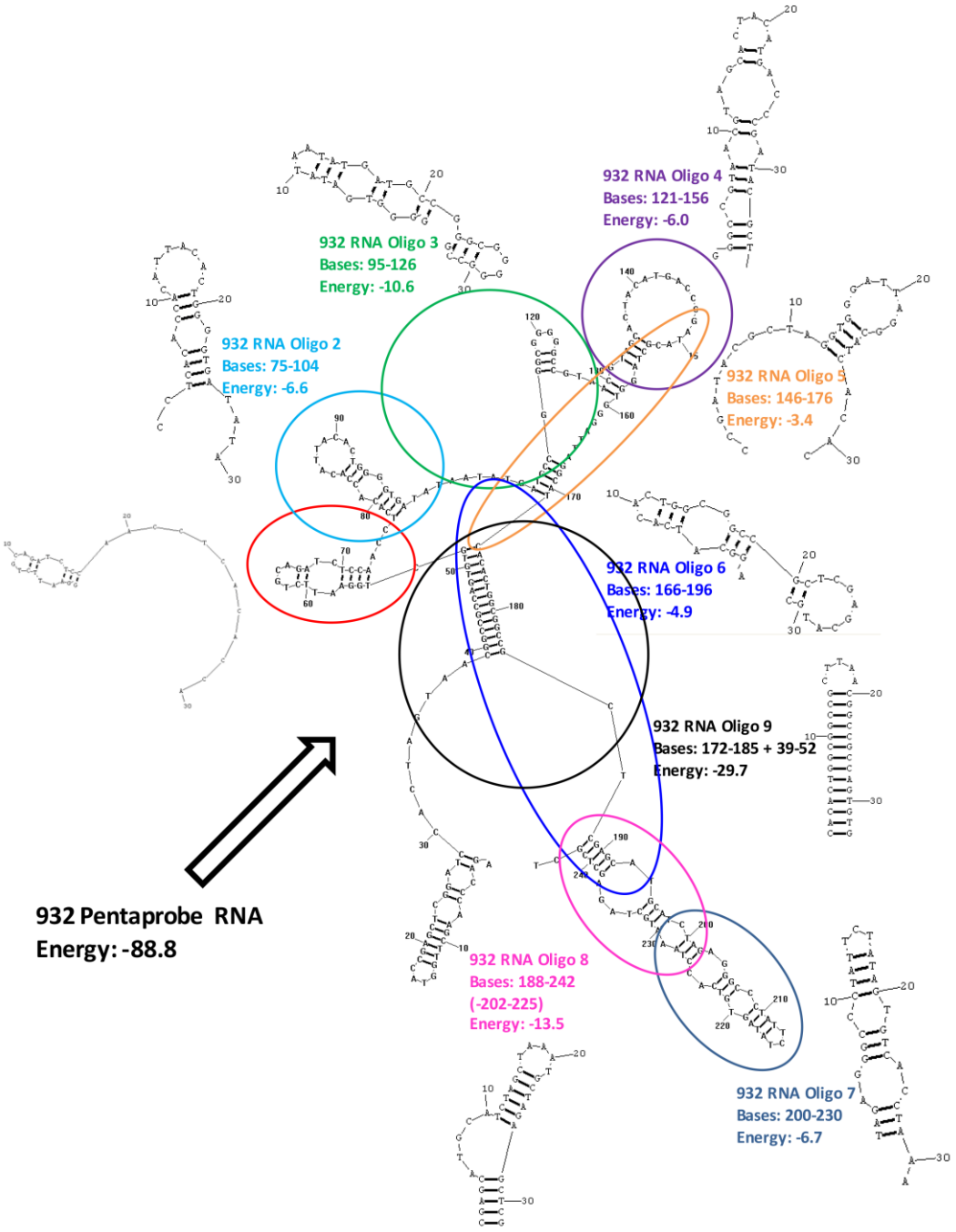


Figure taken from Mckenzie (2011)

References

Afif H, Allali N, Couturier M, Van Melderen L 2001. The ratio between CcdA and CcdB modulates the transcriptional repression of the ccd poison-antidote system. Molecular Microbiology 41(1): 73-82.

Ahidjo BA, Kuhnert D, McKenzie JL, Machowski EE, Gordhan BG, Arcus VL, Abrahams GL, Mizrahi V 2011. VapC toxins from *Mycobacterium tuberculosis* are ribonucleases that differentially inhibit growth and are neutralized by cognate VapB antitoxins. PLoS ONE 6(6): e21738.

Aizenman E, Engelberg-Kulka H, Glaser G 1996. An *Escherichia coli* chromosomal "addiction module" regulated by guanosine 3',5′-bispyrophosphate: a model for programmed bacterial cell death. Proceedings of the National Academy of Sciences of the United States of America 93: 6059-6063.

Altaf M, Miller CH, Bellows DS, O'Toole R 2010. Evaluation of the *Mycobacterium smegmatis* and BCG models for the discovery of *Mycobacterium tuberculosis* inhibitors. Tuberculosis (Edinb) 90(6): 333-7.

Amitai S, Yassin Y, Engelberg-Kulka H 2004. MazF-mediated cell death in *Escherichia coli*: a point of no return. Journal of Bacteriology 186(24): 8295-8300.

Anantharaman V, Aravind L 2003. New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. Genome Biology 4(12): 81.

Arcus VL, Rainey PB, Turner SJ 2005. The PIN-domain toxin-antitoxin array in mycobacteria. Trends in Microbiology 13(8): 360-365.

Arcus VL, McKenzie JL, Robson J, Cook GM 2011. The PIN-domain ribonucleases and the prokaryotic VapBC toxin-antitoxin array. Protein engineering, design & selection : Protein Engineering, Design & Selection 24(1-2): 33-40.

Babajan B, Chaitanya M, Rajsekhar C, Gowsia D, Madhusudhana P, Naveen M, Chitta SK, Anuradha CM 2011. Comprehensive structural and functional characterization of *Mycobacterium tuberculosis* UDP-NAG enolpyruvyl transferase (Mtb-MurA) and prediction of its accurate binding affinities with inhibitors. Interdisciplinary Scienes: Computational Life Sciences 3(3): 204-216.

Bahassi EM, O'Dea M, H., Allali N, Messens J, Gellert M, Couturier M 1999. Interactions of CcdB with DNA gyrase. Journal of Biological Chemistry 274(16): 10936.

Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S 2004. Bacterial persistence as a phenotypic switch. Science 305(5690): 1622-1625.

Bashyam MD, Kaushal D, Dasgupta SK, Tyagi AK 1996. A study of mycobacterial transcriptional apparatus: identification of novel features in promoter elements. Journal of Bacteriology 178(16): 4847-4853.

Blokpoel MCJ, Murphy HN, O'Toole R, Wiles S, Runn ESC, Stewart GR, Young DB, Robertson BD 2005. Tetracycline-inducible gene regulation in mycobacteria. Nucleic Acids Research 33(2): e22.

Blower TR, Short FL, Rao F, Mizuguchi K, Pei XY, Fineran PC, Luisi BF, Salmond GPC 2012. Identification and classification of bacterial type III toxin–antitoxin systems encoded in chromosomal and plasmid genomes. Nucleic Acids Research 40(13): 6158-6173.

Bodogai M, Ferenczi S, Bashtovyy D, Miclea P, Papp P, Dusha I 2006. The *ntrPR* operon of *Sinorhizobium meliloti* is organized and functions as a toxin-antitoxin module. Molecular Plant-Microbe Interactions 19(7): 811-822.

Brantl S 2012. Bacterial type I toxin-antitoxin systems. RNA Biology 9(12): 1-3.

Brown BL, Lord DM, Grigoriu S, Peti W, Page R 2013. The *Escherichia coli* toxin MqsR destabilizes the transcriptional repression complex formed between antitoxin MqsA and the *mqsRA* operon promoter. Journal of Biological Chemistry 288(2): 1286-1294.

Christensen-Dalsgaard M, Gerdes K 2006. Two *higBA* loci in the *Vibrio cholerae* superintegron encode mRNA cleaving enzymes and can stabilize plasmids. Molecular Microbiolgy 62(2): 397-411.

Christensen SK, Mikkelsen M, Pedersen K, Gerdes K 2001. RelE, A global inhibitor of translation, is activated during nutritional stress. Proceedings of the National Academy of Sciences of the United States of America 98(25): 14328-14333.

Christensen SK, Maenhaut-Michel G, Mine N, Gottesman S, Gerdes K, Van Melderen L 2004. Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the *yefM-yoeB* toxin-antitoxin system. Molecular Microbiology 51(6): 1705-1717.

Cline SD, Saleem S, Daines DA 2012. Regulation of the *vapBC-1* toxin-antitoxin locus in nontypeable *Haemophilus influenzae*. PloS one 7(3): e32199.

Clissold PM, Ponting CP 2000. PIN domains in nonsense-mediated mRNA decay and RNAi. Current Biology 10(24): R888-R890.

Condon C 2006. Shutdown decay of mRNA. Molecular Microbiology 61(3): 573-583.

Cooper TF, Heinemann JA 2000. Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. Proceedings of the National Academy of Sciences of the United States of America USA 97(23): 12643-12648.

Daines DA, Wu M, Yuan SY 2007. VapC-1 of nontypeable *Haemophilus influenzae* is a ribonuclease. Journal of Bacteriology 189(14): 5041-5048.

Duyvestyn JM 2012. Mechanism of RNA cleavage by VapC from *Pyrobaculum aerophilum*. Unpublished master's thesis, University of Waikato, Hamilton, New Zealand.

Engelberg-Kulka H, Glaser G 1999. Addiction modules and programmed cell death and antideath in bacterial cultures. Annual Review of Microbiology 53(1): 43-70.

Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC 2009. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proceedings of the National Academy of Sciences of the United States of America 106(3): 894-899.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K and others 2010. The Pfam protein families database. Nucleic Acids Research 38: D211-D222.

Frampton R, Aggio RB, Villas-Boas SG, Arcus VL, Cook GM 2012. Toxin-antitoxin systems of *Mycobacterium smegmatis* are essential for cell survival. Journal of Biological Chemistry 287(8): 5340-5356.

Garcia-Contreras R, Zhang X-S, Kim Y, Wood TK 2008. Protein translation and cell death: the role of rare tRNAs in biofilm formation and in activating dormant phage killer genes. PLoS ONE 3(6): e2394.

Garcia-Pino A, Balasubramanian S, Wyns L, Gazit E, De Greve H, Magnuson RD, Charlier D, van Nuland NAJ, Loris R 2010. Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell 142(1): 101-111.

Gasteiger E, Hoogland C, Gattiker A, Duvand S, Wilkins MR, Appel RD, Bairoch A ed. 2005. Protein identification and analysis on the ExPASy Server. Totowa, Humana Press.

Gerdes K 2000. Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. Journal of Bacteriology 182(3): 561-572.

Gerdes K, Rasmussen PB, Molin S 1986. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. Proceedings of the National Academy of Sciences of the United States of America 83(10): 3116-3120.

Gerdes K, Christensen SK, Løbner-Olesen A 2005. Prokaryotic toxin-antitoxin stress response loci. Nature Reviews Microbiology 3(5): 371-382.

Gill WP, Harik NS, Whiddon MR, Liao RP, Mittler JE, Sherman DR 2009. A replication clock for *Mycobacterium tuberculosis*. Nature Medicine 15(2): 211-214.

Gomez M, Smith I 2000. Determinants of mycobacterial gene expression. In: Harfull GF, Jacobs WR ed. Molecular genetics of mycobacteria. Washington DC, American Society for Microbiology Press.

Gomez M, Johnson S, Gennaro ML 2000. Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. Infection and Immunity 68(4): 2323-2327.

Gupta A 2009. Killing activity and rescue function of genome-wide toxin-antitoxin loci of *Mycobacterium tuberculosis*. FEMS Microbiology Letters 290(1): 45-53.

Han J, Lee J, Anandan T, Zeng M, Sripathi S, Jahng WJ, Lee SH, Suh J, Kang C 2010. Characterization of a chromosomal toxin-antitoxin, Rv1102c-

Rv1103c system in *Mycobacterium tuberculosis*. Biochemical and Biophysical Research Communications 400(3): 293-298.

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P 1994. Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie / Chemical Monthly 125(2): 167-188.

Huang F, He Z-G 2010. Characterization of an interplay between a *Mycobacterium tuberculosis* MazF homolog, *Rv1495*, and its sole DNA topoisomerase I. Nucleic Acids Research 38(22): 8219-8230.

Johnson EP, Strom AR, Helinski DR 1996. Plasmid RK2 toxin protein ParE: purification and interaction with the ParD antitoxin protein. Journal of Bacteriology 178: 1420-1429.

Kawano M 2012. Divergently overlapping cis-encoded antisense RNA regulating toxin-antitoxin systems from *E. coli*: *hok/sok, ldr/rdl, symE/symR*. RNA Biology 9(12): 1-8.

Keren I, Shah D, Spoering A, Kaldalu N, Lewis K 2004. Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. Journal of Bacteriology 186(24): 8172-8180.

Kong H, Lin L, Porter N, Stickel S, Byrd D, Posfai J, Roberts RJ 2000. Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 Genome. Nucleic Acids Research 28(17): 3216-3223.

Korch SB, Contreras H, Clark-Curtiss JE 2009. Three *Mycobacterium tuberculosis* Rel toxin-antitoxin modules inhibit mycobacterial growth and are expressed in infected human macrophages. Journal of Bacteriology 191(5): 1618-1630.

Koshy SS, Montrose MH, Sears CL 1996. Human intestinal epithelial cells swell and demonstrate actin rearrangement in response to the metalloprotease toxin of *Bacteroides fragilis*. Infection and Immunity 64(12): 5022-5028.

Kumar V, Abbas AK, Aster JC 2012. Cell Injury, Cell Death, and Adaptations. Robbins Basic Pathology. Philadelphia, Elsevier. Pp. 1-30.

Magnuson RD 2007. Hypothetical Functions of Toxin-Antitoxin Systems. Journal of Bacteriology 189(17): 6089-6092.

Magnuson RD, Yarmonlinksy MB 1998. Corepression of the P1 addiction operon by Phd and Doc. Journal of Bacteriology 180(23): 6342-6351.

Malen H, Berven FS, Fladmark KE, Wiker HG 2007. Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. Proteomics 7(10): 1702-1718.

Masuda H, Tan Q, Awano N, Wu KP, Inouye M 2012. YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*. Molecular Microbiology 84(5): 979-989.

Mate MJ, Vincentelli R, Foos N, Raoult D, Cambillau C, Ortiz-Lombardia M 2011. Crystal structure of the DNA-bound VapBC2 antitoxin/toxin pair from Rickettsia felis. Nucleic Acids Research 40(7): 3245-3258.

Mattison K, Wilbur JS, So M, Brennan RG 2006. Structure of FitAB from *Neisseria gonorrhoeae* Bound to DNA Reveals a Tetramer of Toxin-Antitoxin Heterodimers Containing Pin Domains and Ribbon-Helix-Helix Motifs. Journal of Biological Chemistry 281(49): 37942-37951.

McCaskill JS 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers 29(6-7): 1105 - 1119.

McKenzie J, Robson J, Berney M, Smith TC, Ruthe A, Gardner PP, Arcus VL, Cook GM 2012a. A VapBC toxin-antitoxin module is a posttranscriptional regulator of metabolic flux in mycobacteria. Journal of Bacteriology 194(9): 2189.

McKenzie JL 2011. The Biochemistry of VapBC Toxin-Antitoxins. Unpublished doctorol dissertation, University of Waikato, Hamilton, New Zealand.

McKenzie JL, Duyvestyn JM, Smith T, Bendak K, Mackay J, Cursons R, Cook GM, Arcus VL 2012b. Determination of ribonuclease sequence-specificity using Pentaprobes and mass spectrometry. RNA 18(6): 1267-1278.

Miallau L, Faller M, Chiang J, Arbing M, Guo F, Cascio D, Eisenberg D 2009. Structure and Proposed Activity of a Member of the VapBC Family of Toxin-Antitoxin systems. VapBC-5 from *Mycobacterium tuberculosis*. Journal of Biological Chemistry 284(1): 276-283.

Min A, B., Miallau L, Sawaya MR, Habel J, Cascio D, Eisenberg D 2012. The crystal structure of the Rv0301-Rv0300 VapBC-3 toxin—antitoxin complex from *M. tuberculosis* reveals a Mg2+ ion in the active site and a putative RNA-binding site. Protein Science 21(11): 1754-1767.

Mohsen AM, Khader AT, Ramachandram D 2009. Comparison of HSRNAFold and RNAFold Algorithms for RNA Secondary Structure Prediction. TENCON 2009 - 2009 IEEE Region 10 Conference. Singapore.

Motorin Y, Helm M 2011. RNA nucleotide methylation. Wiley Interdiscip Rev RNA 2(5): 611-631.

Murphy DJ, Brown JR 2007. Identification of gene targets against dormant phase *Mycobacterium tuberculosis* infections. BMC infectious diseases 7(1): 84-84.

Mutschler H, Gebhardt M, Schoeman RL, Meinhart A 2011. A Novel Mechanism of Programmed Cell Death in Bacteria by Toxin–Antitoxin Systems Corrupts Peptidoglycan Synthesis. PLoS Biol 9(3): e1001033.

Naito T, Kusano K, Kobayashi I 1995. Selfish Behaviour of Restriction-Modification Systems. Science 267(5199): 897-899.

Nordhoff E, Cramer R, Karas M, Hillenkamp F, Kirpekar F, Kristiansen K, Roepstorff P 1993. Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry. Nucleic Acids Research 21(15): 3347-3357.

Norton JP, Mulvey MA 2012. Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic *Escherichia coli*. PLoS Pathog 8(10): e1002954.

Ogura T, Hiraga S 1983. Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proceedings of the National Academy of Sciences of the United States of America 80(15): 4784-4788.

Pandey DP, Gerdes K 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Research 33(3): 966.

Pedersen K, Christensen SK, Gerdes K 2002. Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. Molecular Microbiology 45(2): 501-10.

Pieters J 2008. *Mycobacterium tuberculosis* and the Macrophage: Maintaining a Balance. Cell Host & Microbe 3(6): 399-407.

Rajni, Rao N, Meena LS 2011. Biosynthesis and Virulent Behavior of Lipids Produced by *Mycobacterium tuberculosis*: LAM and Cord Factor: An Overview. Biotechnology Research International 2011.

Raju RM, Unnikrishnan M, Rubin DHF, Krishnamoorthy V, Kandror O, Akopian TN, Goldberg AL, Rubin EJ 2012. *Mycobacterium tuberculosis* ClpP1 and ClpP2 function together in protein degradation and are required for viability *in vitro* and during infection. PLoS Pathog 8(2): e1002511.

Ramage HR 2010. The toxin-antitoxin systems of *Mycobacterium tuberculosis* and their role in persistence. Unpublished thesis, University of California, San Francisco.

Ramage HR, Connolly LE, Cox JS 2009. Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. PLoS Genetics 5(12): e1000767.

Rankin D, Turner L, Heinemann J, Brown S 2012. The coevolution of toxin and antitoxin genes drives the dynamics of bacterial addiction complexes and intragenomic conflict. Proceedings Biological Sciences 11: 11.

Robson J 2010. Characterization of the *vapBC* toxin-antitoxin operon of *Mycobacterium smegmatis*. Unpublished doctoral thesis, University of Otago, Dunedin, New Zealand.

Robson J, McKenzie JL, Cursons R, Cook GM, Arcus VL 2009. The *vapBC* operon from *Mycobacterium smegmatis* is an autoregulated toxin–antitoxin module that controls growth via inhibition of translation. Journal of Molecular Biology 390(3): 353-367.

Rotem E, Loinger A, Ronin I, Levin-Reisman I, Gabay C, Shoresh N, Biham O, Balaban NQ 2010. Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. Proceedings of the National Academy of Sciences of the United States of America 107(28): 12541-12546.

Ruiz-Echevarría MJ, de la Cueva G, Díaz-Orejas R 1995. Translational coupling and limited degradation of a polycistronic messenger modulate differential gene expression in the parD stability system of plasmid R1. Molecular and General Genetics 248(5): 599-609.

Russell J, Cook GM 1995. Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiological Reviews 59(1): 48-62.

Schoenberg DR 2007. The end defines the means in bacterial mRNA decay. Nature Chemical Biology 3(9): 535-536.

Schuster-Bockler B, Schultz J, Rahmann S 2004. HMM logos for visualization of protein families. BMC Bioinformatics 5(7).

Sharp JD, Cruz JW, Raman S, Inouye M, Husson RN, Woychik NA 2012. Growth and translation inhibition through sequence-specific RNA binding by *Mycobacterium tuberculosis* VapC toxin. The Journal of Biological Chemistry 287(16): 12835-12847.

Steege DA 2000. Emerging feature of mRNA decay in bacteria. RNA 6: 1078-1090.

Summers E 2007. Expression of VapBC from *Mycobacterium smegmatis* in *Escherichia coli*. University of Waikato.

Trauner A, Bennett MH, Williams HD 2011. Isolation of bacterial ribosomes with monolith chromatography. PLoS ONE 6(2): e16273.

Van Melderen L, Saavedra De Bast M 2009. Bacterial toxin–antitoxin systems: more than selfish entities? PLoS Genetics 5(3): e1000437.

Wagner E, Unoson C 2012. The toxin-antitoxin system tisB-istR1: Expression, regulation, and biological role in persister phenotypes. RNA Biology 9(12): 1-7.

Wang X, Lord DM, Hong SH, Peti W, Benedik MJ, Page R, Wood TK 2013. type II toxin/antitoxin MqsR/MqsA controls type V toxin/antitoxin GhoT/GhoS. Environmental Microbiology.

Wang X, Lord DM, Cheng H, Osbourne DO, Hong SH, Sanchez-Torres V, Quiroga C, Zheng K, Herrmann T, Peti W and others 2012. A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. Nature Chemical Biology 8: 855-861.

WHO 2012. Global Tuberculosis Report 2012.

Wilbur JS, Chivers PT, Mattison K, Potter L, Brennan RG, So M 2005. *Neisseria gonorrhoeae* FitA interacts with FitB to bind DNA through its ribbon−helix−helix motif. Biochemistry 44(37): 12515-12524.

Winther KS, Gerdes K 2011. Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. Proceedings of the National Academy of Sciences of the United States of America 108(18): 7403-7407.

Wong C, Sridhara S, Bardwell JC, Jakob U 2000. Heating greatly speeds Coomassie blue staining and destaining. Biotechniques 28(3): 426-428, 430, 432.

Wu S, Howard ST, Lakey DL, Kipnis A, Samten B, Safi H, Gruppo V, Wizel B, Shams H, Basaraba RJ and others 2004. The principal sigma factor sigA mediates enhanced growth of *Mycobacterium tuberculosis in vivo*. Molecular Microbiology 51(6): 1551-1562.

Wutchy S, Fontana W, Hofacker IL, Schuster P 1999. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. Biochemistry 37(42): 14719-14735.

Yamaguchi Y, Inouye M 2011. Regulation of growth and death in *Escherichia coli* by toxin–antitoxin systems. Nature Reviews Microbiology 9(11): 779-790.

Yamaguchi Y, Park J, Inouye M 2011. Toxin-antitoxin systems in bacteria and archaea. Annual Review of Genetics 45: 61-79.

Yi C, Pan T 2011. Cellular dynamics of RNA modification. Accounts of Chemical Research 44(12): 1380-1388.

Zaunbrecher MA, Sikes JRD, Metchock B, Shinnick TM, Posey JE 2009. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase *eis* confers kanamycin resistance in *Mycobacterium tuberculosis*. Proceedings of the National Academy of Sciences of the United States of America 106(47): 20004-20009.

Zhang Y, Zhang J, Hoeflich KP, Ikura M, Qing G, Inouye M 2003. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. Molecular Cell 12(4): 913-923.

Zhang YX, Guo XK, Wu C, Bi B, Ren SX, Wu CF, Zhao GP 2004. Characterization of a novel toxin-antitoxin module, VapBC, encoded by *Leptospira interrogans* chromosome. Cell Research 14(3): 208-216.

Zhu L, Sharp JD, Kobayashi H, Woychik NA, Inouye M 2010. Noncognate *Mycobacterium tuberculosis* toxin-antitoxins can physically and functionally interact. Journal of Biological Chemistry 285(51): 39732-39738.

Zuker M, Stiegler P 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Research 9(1): 133-148.